Mining and analyzing process similarity of product module for DPIPP based on PLM database

W P Peng 1, T Zhang 1,2, Q H Zhang 1 and S W Ni 1

1School of Power and Mechanical Engineering, Wuhan University, 8 Donghunan Road, Wuchang District, Wuhan 430072, China

Abstract. A distributed parametrized intelligent product platform (DPIPP) on basis of product lifecycle management database has been proposed to reveal the interrelationships between product attributes (i.e., structure, function or process) in recent years. Product modules in DPIPP are classified according to composition and structure of products. Components belonging to different product modules may have an identical or similar process route and could be manufacture with the same type of equipment. Mining and analysing the process similarity of product modules can effectively reduce production cost and improve production efficiency. In this paper, a novel product data mining approach with an improved Euclidean distance formula is proposed to research the product module process similarity, which aims to illustrate the effect of the structural change on the product process, then provide theoretical support for improving production patterns and reducing production costs to adapt to mass customization. The feasibility and effectiveness of the method are verified via the analysing of the process similarities among three different types of valves.

1. Introduction

In order to design and product rapidly, the similarity and classification method of product module have been widely studied. Several intelligent methods such as the genetic algorithm, the hybrid fuzzy clustering algorithm and particle swarm optimization algorithm has been used in this field. Product process route is the core of enterprise production management, so it is essential to carry out similarity analysis of product module process route and merge similar process route, which could effectively reduce production cost. Lu et al [7, 8] constructed a new process model by taking advantage of an existing base model and process attribute information. Misaki et al. [9] established an effective method of searching the sheet metal parts based on bending process similarity. A similarity measure method of manufacturing process considering both unit similarity and sequence similarity on optimal matching of weighted bipartite graph was presented [10]. Li et al. [11] proposed a genetic algorithm, which could swiftly search for the optimal process plan for single manufacturing system as well as distributed manufacturing systems. Huang et al. [12] raised an algorithm for calculating process similarity in order to cluster process designs. Huang et al. [13] studied a novel numerical control machining process reuse approach by merging feature similarity assessment and data mining for computer-aided manufacturing models. Liu et al. [14] described a method of construction of product modules and analysis of process similarity. A rapid design technology of assembly process based on process similarity was studied [15].

In summary, most previous studies mainly concentrated on the similarity analysis of single attribute, but ignored the similarity analysis of multi-attribute. A distributed parametrized intelligent product
platform (DPIPP) consisted of product modules, association rules and data dictionary has been proposed to satisfy the requirements of mass customization in recently years [16]. The product module in DPIPP is a generalized module which consists of parts or components from different products and their information (BOM, process card, process dictionary, function description etc.), and is also the carrier of function, structure and process of products [17]. The large amount of product information in the platform provides a basis for mining the similarity of multi-attributes.

In this paper, a novel product data mining approach with an improved Euclidean distance formula is proposed to analysis the similarity of multi-attributes among process and structure, aims at further optimizing DPIPP.

2. Description of process similarity of product module

2.1. Product module in DPIPP

The components A1-2, A2-2 from product A1, product A2 with related information constitute the product module C1 as shown in figure 1. So the product module in DPIPP can be expressed as follows:

\[
g(M) = f(F_i, S_i, P_i)\]

(1)

where M is a certain generalized product module; g(M) represents the main information of M; \(F_i\) is the \(i^{th}\) function vector of M; \(S_i\) represents the \(i^{th}\) structure vector of M; \(P_i\) is the \(i^{th}\) process vector of M.

\[\text{Figure 1. Product modules derived from the PLM database in DPIPP.}\]

2.2. Process similarity of product module

Each product module in DPIPP contains the information of several parts or components. Therefore, there are several process routes in a product module. The process routes of the same product module are independent, but there may be some similarities among the process routes of different product modules. The process similarity of product modules is illustrated in figure 2, where \(M_1\) is product module 1; \(M_2\) is product module 2; The vector \(P_{11}\) represents one of the process routes of the \(i^{th}\) product module, in which the value is binary, 1 and 0 represent the process route contains some process technic or not. \(P_{11}\&P_{22}\) and \(P_{12}\&P_{22}\) are considered as the similar process routes through the comparison of all process information.

2
3. Process similarity analysis of product module

3.1. Process matrix constructing

Components in the same DPIPP product module have similar functions, thus the component structure becomes the primary factor affecting the process. To establish a structure matrix is the prerequisite for constructing a process matrix. For instance, $M_i$ is a product module in DPIPP, the process parameters of $M_i$ are affected by the following $k$ structural parameters ($S_1, S_2, \ldots, S_k, k \in N^+$). The values of the $k$ structural parameters can be represented as a vector. If there are $m$ ($m \in N^+ \rangle$ kinds of products with different structures in $M_1$, the number of structure vectors is $m$. The structural matrix of product module $M_i$ is expressed by the $m$ structure vectors as shown in equation (2):

$$
M^s_i = \begin{bmatrix}
S^s_{i1} & S^s_{i2} & \cdots & S^s_{ij} & \cdots & S^s_{ik} \\
S^s_{2i} & S^s_{22} & \cdots & S^s_{2j} & \cdots & S^s_{2k} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
S^s_{mi} & S^s_{m2} & \cdots & S^s_{mj} & \cdots & S^s_{mk}
\end{bmatrix}
$$

In above equation, $M^s_i$ is the structural matrix of product module $M_i$; $S^s_{ij}$ is a vector of the $j^{\text{th}}$ structure parameter of the product module $M_i$; $S^s_{ij}$ is the value of the $j^{\text{th}}$ structural parameter of the $i^{\text{th}}$ kind of product in product module $M_i$.

A structural vector of $M^s_i$ corresponds to a process route, but different structures may share the same process route. Hence, the process routes of the $m$ kinds of products in $M_i$ might be less than $m$. If there are $u$ ($u \leq m, u \in N^+$) different process routes in $M_i$, the process matrix can be expressed as follows:

$$
M^p_i = [P^1_i, P^2_i, \ldots, P^u_i]^{T}
$$

(3)

where $M^p_i$ represents the process matrix of product module $M_i$; $P^i$ represents the $i^{\text{th}}$ process route of $M_i$ ($i \in (1, 2, 3, \ldots, u)$), which is composed of several process technics.

However, the process information is usually composed of numbers and words, in order to make the process matrix analysable, all the process technics should be defined in the process dictionary. Part of
the process dictionary is shown in table 1. In the matrix $M_p$, 1 and 0 represent whether a certain process technic contained in some process route or not respectively.

### Table 1. Part of the process dictionary

| P_1  | P_2  | P_3  | P_4  | P_5  | ...... | P_r |
|------|------|------|------|------|-------|------|
| Casting | Forging | Rough-milling | Finish-milling | Rough-turn | ...... | Finish-grinding |

\[
M_p' = \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1j} & \cdots & P_{1r} \\
P_{21} & P_{22} & \cdots & P_{2j} & \cdots & P_{2r} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
P_{s1} & P_{s2} & \cdots & P_{sj} & \cdots & P_{sr} \\
\end{bmatrix} \quad \Rightarrow \quad M_p'' = \begin{bmatrix}
1 & 0 & \cdots & 1 & \cdots & 1 \\
1 & 1 & \cdots & 1 & \cdots & 1 \\
0 & 1 & \cdots & 0 & \cdots & 0 \\
1 & 0 & \cdots & 1 & \cdots & 1 \\
\end{bmatrix}
\]

### 3.2. Process similarity mining

In this subsection, the process route similarity degrees are calculated by the clustering algorithm [17] through the following two steps:

1. **Step 1.** Utilize an improved Euclidean formula to calculate the distances among different process routes;

2. **Step 2.** Convert the distances calculated in step 1 into the similarity degrees.

Suppose there are $u$ and $v$ process routes in $M_x$ and $M_y$ respectively ($u \in \mathbb{N}^+$, $v \in \mathbb{N}^+$). The process matrices are as follows:

\[
M_p'' = \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1j} & \cdots & P_{1r} \\
P_{21} & P_{22} & \cdots & P_{2j} & \cdots & P_{2r} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
P_{s1} & P_{s2} & \cdots & P_{sj} & \cdots & P_{sr} \\
\end{bmatrix} \quad \Rightarrow \quad M_p''' = \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1j} & \cdots & P_{1r} \\
P_{21} & P_{22} & \cdots & P_{2j} & \cdots & P_{2r} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
P_{s1} & P_{s2} & \cdots & P_{sj} & \cdots & P_{sr} \\
\end{bmatrix}
\]

where $\forall P_{ij}, P_{ij}' \in \{0,1\}$. A coefficient $r$ is introduced into the Euclidean distance formula, which is convenient for standardizing the results of the calculation is used to calculate distances between two process routes from different product module in equation (6).

\[
d_{ij} = d(P_i, P_j') = \left[ \frac{1}{r} \sum_{w=1}^{r} (P_{iw} - P_{iw}')^2 \right]^{1/2}
\]

where $r$ refers to the number of process technics in Table 1.

The standardized distance matrix $d(M_x, M_y)$ according to product modules $M_x$ and $M_y$ is expressed as follows:
where $d_{ij}$ represents the difference between the $i^{th}$ process route of product module $M_i$ and the $j^{th}$ process route of product module $M_j$.

Then the similarity degree $s_{ij}$ can be calculated in equation (8):

$$s_{ij} = 1 - d_{ij}$$

The similarity degree matrix of product modules $M_i$ and $M_j$ can be established as follow.

$$\text{Sim}(M_i, M_j) = \begin{bmatrix}
s_{11} & s_{12} & \cdots & s_{1v} \\
s_{21} & s_{22} & \cdots & s_{2v} \\
\vdots & \vdots & \ddots & \vdots \\
\end{bmatrix}$$

The complex relationships of the structural matrices, the process matrices and the similarity degree matrix of product module $M_i$ and $M_j$ are clearly presented in figure 3.

Figure 3. The mutual relationships of structural, process and similarity degree matrix.

Take the element $s_{12}$ in the similarity degree matrix for instance, $s_{12}$ represents the similarity of the $1^{st}$ process route in product module $M_i$ and the $2^{nd}$ process route in product module $M_j$. The similarity degree matrix connects the structure matrices and the process matrices. Process similarity is an important design resource, plays a significant role in the production line optimization and rapid design. The analysis results of the process similarity are screened according to a certain similarity threshold that is set with the enterprise requirements, stored into a specific table in DPIIPP.

4. Case
To verify the effectiveness of the process similarity analysis, an example of process similarity analysis of the product modules from a value company is presented. The typical structure of a valve is as follow.

Figure. 4 The structure and components of a gate valve.
1. Valve Body; 2. Disc; 3. Stem; 4. Bonnet; 5. Seal ring; 6. Seat ring; 7. Stuffing box; 8. Bracket; 9. Stem Nut; 10. Hand Wheel.

4.1. Product module screening
The cavity of a valve is mainly composed of valve body and valve deck. The valve body is a vital part of a valve, bears the connection, sealing and supporting functions. In this section, the process similarities between three valve body modules are studied to illustrate the analysis process.

The valve body modules of gate valve, cut-off valve and ball valve are analyzed, which are expressed with $M_1$, $M_2$ and $M_3$ respectively. The structures of these three valves and the structural parameters which can influence the process of valve body are all shown in figure 5.

Figure 5. The structures of the three valve bodies.

4.2. Process matrix constructing
The structural parameters of the three types of valves are extracted from the PLM database. These parameters will be stored in a unified format. Part of the structural parameters of the valve body for the gate valve is shown in table 2.
Table 2. Parts of the structural parameters of the gate valve’s body.

| No. | DN | L  | D  | D1 | D2 | b  | n-φ₄ |
|-----|----|----|----|----|----|----|------|
| Z-01| 50 | 250| 165| 125| 99 | 20 | 4-18 |
| Z-02| 50 | 250| 165| 125| 99 | 20 | 4-18 |
| Z-03| 65 | 265| 185| 145| 118| 20 | 4-18 |
| Z-04| 80 | 280| 200| 160| 132| 20 | 8-18 |
| Z-05| 100| 300| 220| 180| 156| 22 | 8-18 |
| Z-06| 125| 325| 250| 210| 184| 22 | 8-18 |
| Z-07| 150| 350| 285| 240| 211| 24 | 8-22 |

... ... ... ... ... ... ... ... ...

Table 3. Process dictionary of the valve body module.

| P₁ | P₂ | P₃ | P₄ | P₅ | P₆ |
|----|----|----|----|----|----|
| Bead weld | Turning | Chamfering | Rough-milling | Finish-milling | Drilling |

| P₇ | P₈ | P₉ | P₁₀ | P₁₁ | P₁₂ |
|----|----|----|-----|-----|-----|
| Rough-boring | Finish-boring | Reaming | Tapping | Ablation | Welding |

The processes of the three valve body modules are encoded according to the process dictionary. Parts of the gate valve’s process encodings are listed in table 4. There are some repeated process encodings as shown in table 4, such as the Z-03 and the Z-04. These repeated encodings will be merged in similarity analysis.

Table 4. Parts of the gate valve’s process encodings.

| No. | P₁ | P₂ | P₃ | P₄ | P₅ | P₆ | P₇ | P₈ | P₉ | P₁₀ | P₁₁ | P₁₂ |
|-----|----|----|----|----|----|----|----|----|----|-----|-----|-----|
| Z-01| 0  | 1  | 1  | 0  | 0  | 0  | 0  | 1  | 1  | 1   | 1   | 0   |
| Z-03| 0  | 1  | 1  | 0  | 1  | 1  | 0  | 1  | 1  | 1   | 1   | 0   |
| Z-04| 0  | 1  | 1  | 0  | 1  | 1  | 0  | 1  | 1  | 1   | 0   | 0   |
| Z-05| 1  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 1  | 1   | 1   | 0   |
| Z-06| 0  | 1  | 1  | 1  | 1  | 1  | 1  | 0  | 1  | 1   | 1   | 0   |
| Z-07| 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1  | 1   | 1   | 1   |

... ... ... ... ... ... ... ... ...

The process matrix is generated by the rest of process encodings. The process matrix of the gate valve can be expressed as follows.

\[
M_P^Z = \begin{bmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1
\end{bmatrix}
\]

4.1. Process similarity mining

The cavity of a valve is mainly composed of valve body and valve deck. The valve body is a vital part of a valve, bears the Connection, sealing and supporting functions. In this subsection, the process similarities between three valve body modules are studied to illustrate the analysis process. Following the construction of the three valves process matrices, the distances between these process matrices are calculated according to equation (6). Then these distances are transformed into the similarity degrees according to equation (8). The similarity degree matrices of the three product modules are as follows:
Finally, the similarity degrees of different product modules are all stored in DPIPP with the form of table 5, according to a certain similarity threshold $\mu$, the product module with same process route.

**Table 5.** Parts of the process similarity information between gate valve body and cut-off body.

| Similarity degrees | Two-dimensional arrays | Process codes $(1,0, \ldots, 1)$ | Structural parameters $(S_1, S_2, \ldots, S_{max})$ | Product numbers |
|--------------------|------------------------|----------------------------------|----------------------------------------------|-----------------|
| $S_{1,1}=1.00$     | $(z_1, j_1)$           | $0,1,1,0,0,0,0,0,1,1,1,0$        | $50,250,165,125,99,20,4,\Phi_{18}$           | x-1             |
|                    |                        |                                  | $50,250,165,125,99,20,4,\Phi_{18}$           | x-5             |
|                    |                        |                                  | $65,265,185,145,118,20,4,\Phi_{18}$           | x-7             |
|                    |                        |                                  | $80,280,200,160,132,20,8,\Phi_{18}$           | x-12            |
|                    |                        |                                  | $40,200,150,110,84,18,4,\Phi_{18}$           | y-1             |
|                    |                        |                                  | $50,230,165,125,99,20,4,\Phi_{18}$           | y-8             |
|                    |                        |                                  | $65,290,185,145,118,20,4,\Phi_{18}$           | y-10            |
| $S_{2,0}=0.71$     | $(z_2, j_2)$           | $0,1,1,0,0,0,1,0,0,1,1,1$        | $50,350,286,240,211,24,8,\Phi_{22}$           | x-3             |
|                    |                        |                                  | $200,400,340,295,266,24,8,\Phi_{22}$           | x-6             |
|                    |                        |                                  | $250,450,405,355,319,26,12,\Phi_{26}$           | x-10            |
|                    |                        |                                  | $300,500,460,410,370,28,12,\Phi_{26}$           | x-15            |
|                    |                        |                                  | $100,350,235,190,156,24,8,\Phi_{22}$           | y-6             |

4.2. Process similarity storing
5. Conclusions
In this paper, the product process similarity degree matrix related to structure is constructed on basis of the mining and analysing mutual relationship between process and structure of the product module in DPIPP. The product process information for string type is converted to binary by utilizing the process dictionary, which is convenient for compositing process matrix. The similarity degree between different process routes is calculated via an improved Euclidean distance formula, then the product module is clustered by a given threshold according to the practical requirement. The feasibility and the effectiveness of the approach has been verified though the valve case.

This research scientifically elaborates the influence of structure on process, and provide a theoretical support for further optimizing DPIPP through mining and analysing the product module process similarity. However, there are still many shortcomings that need to be improved in the following research, for instance, parameters of the product process and productive facilities should be considered because the same process may require different productive facilities due to the differences in process parameters. In addition, as the product data becomes more and more complex, intelligent product data mining methods ought to be apply to this field.

6. References

[1] Shi JY and Tao QB 2010 Research on a genetic algorithm for solving module partition problem 6th, Int. Conf. on Natural Computation (Yantai: China/ New York IEEE) pp 4215-18
[2] Du H 2010 Application of GN Algorithm in Product Module Partition International Journal of Digital Content Technology & its Applications 6(14)
[3] Xu ZW and Liu YX 2011 Study on Module Partition Method for CNC Lathe Design. In Advances in Computer Science, Intelligent System and Environment pp 749-753. (Berlin Heidelberg: Springer)
[4] Teng XY and Zhang TJ 2010 Module partition method based on hybrid fuzzy clustering. Int. Conf. on Mechatronics and Automation (Xi’an: China/ New York IEEE) pp 333-8
[5] Li H, Ji YJ, Gu XJ, et al 2012 Module partition process model and method of integrated service product. Computers in Industry; 63(4) pp 298-308
[6] Shan Q, Chen Y and Yan GR 2008 Product Module Identification Based on PSO. Int. Symp. on Information Science and Engineering (Shanghai: China/ New York IEEE) pp 231-235
[7] Lu JD, Yao K and Gao FR 2009 Process similarity and developing new process models through migration. AIChE journal 55(9) pp 2318-28.
[8] Lu JD and Gao FR 2008 Process modelling based on process similarity. Industrial & Engineering Chemistry Research 47(6) pp 1967-74.
[9] Misaki D and Aomura S 2003 Searching sheet metal parts based on bending process similarity. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 217(3) pp 427-38.
[10] Zhang YJ, Liu JS, Wang L 2015 Product Manufacturing Process Similarity Measure Based on Attributed Graph Matching. 3rd Int. Conf. on Mechatronics, Robotics and Automation (Shenzhen: China/ Paris: ATLANTIS) pp 1083-86
[11] Li L, Fuh J Y H and Zhang YF 2005 Application of genetic algorithm to computer-aided process planning in distributed manufacturing environments Robotics and Computer-Integrated Manufacturing 21(6) pp 568-78
[12] Huang K, Zhou ZT and Han YB 2004 An algorithm for calculating process similarity to cluster open source process designs. Grid and Cooperative Computing-GCC 2004 Workshops (Wuhan: China/ Berlin Springer) pp 107-14
[13] Huang R, Zhang SS, Bai XL et al 2015 An effective numerical control machining process reuse approach by merging feature similarity assessment and data mining for computer-aided
manufacturing models. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **229**(7) pp 1229-42

[14] Liu W, Liu ZY and Tan JR 2010 Construction of Product Modules Based on Process Similarity. *Journal of Computer-Aided Design & Computer Graphics* **22**(10) pp 1647-54.

[15] Song HW, Zuo DW, Xue SL et al 2009 Research on rapid design technology of assembly process based on process similarity. *China Manufacturing Information: Academic Edition* **38**(11) pp 29-31.

[16] Huang YT 2015 Research on distributed parameterized intelligent product platform and its database realization (Thesis: Wuhan University)

[17] Cai RC, Zhang ZJ and Tung A K H 2014 A general framework of hierarchical clustering and its applications *Information Sciences* **272** pp 29-48.

**Acknowledgments**

This study was supported by National Natural Science Foundation (No.51275362).