

# Fuzzy association rules mining method for design knowledge of mechanical structure symmetry

Juncheng Yang<sup>1, a</sup>, Zhiyong Ma<sup>2, b</sup>

Faculty of Mechanical Engineering and Mechanics, Zhejiang Provincial Key Lab of Part Rolling Technology, Ningbo University

818 Fenghua Rd., Ningbo City, Zhejiang Province 315211, P. R. China

<sup>a</sup>email: nbuyjc@163.com, <sup>b</sup>email: mazhiyong@nbu.edu.cn

**Keywords:** Mechanical structure symmetry; fuzzy mining; association rules; design knowledge

**Abstract.** Structure symmetry is a common phenomenon in mechanical product structure, which plays important roles in realizing the design demands of mechanical product. This article using the method of multidimensional fuzzy association rules mining to mine design knowledge from the instance database, the mining results can guide the scientific application of structural symmetry in mechanical product structure design. First, the data to be excavated in the case database is processed by fuzzy method, and then association rules mining algorithm is used to mine the fuzzy association rules between structural symmetry and mechanical product function, performance and restriction, then the design knowledge of structural symmetry in mechanical product structure is established. The research result of this article is the theoretical basis of the further research on application rules of mechanical structure symmetry.

## 1. Introduction

Structure symmetry is a common phenomenon in mechanical product structure, the association rules between the mechanical structure symmetry and design requirements was proposed by using the data mining method in the instance database, establish the design knowledge of structural symmetry in mechanical product structure, which will guide the scientific application of structural symmetry design in mechanical product structure, and improve the technical, economic and social performance of mechanical products<sup>[1]</sup>. Structure symmetry design knowledge fuzzy association rule mining process includes the following steps in mechanical structure symmetry design instance database: data preparation, data mining, pattern evaluation and presentation, knowledge application, and knowledge maintenance. In all the steps, data mining is a core process<sup>[2]</sup>. We can summary and refine the association rules with high support and confidence in the structure symmetry instance database by the method of the frequent itemsets mining.

In this paper, the design knowledge mining is widely used to solve the mining method, we adopt the fuzzy association rules mining method to avoid the disadvantages of the boundary division, and adopt multidimensional association rules mining method to mine the rich and practical design knowledge. The article combines the advantages of fuzzy association rules and multidimensional association rule mining method, which takes the multi-angle and multi-level knowledge mining based on the symmetry of the instance database and the design requirement. The paper describes the general process of fuzzy knowledge mining, and illustrates the fuzzy mining method by standard instance database, which established before. We can summarize the general rules of fuzzy mining structure, establish the standardization system by the deep research on the design knowledge of design requirements, application effectiveness, application conditions and other elements, even we can summarize the application knowledge of different depth and different application range by abstracting different applications independently on different levels.

## 2. Fuzzy association rule mining

In the current study, this paper based on Dr. Ma's doctoral dissertation<sup>[3]</sup>, which clearly

described the general process of knowledge discovery. The mining process is shown in Figure 1. The research based on Agrawal et al. proposed in 1993 on association rule mining between itemsets in the database <sup>[4]</sup>. If  $I=\{i_1, i_2, \dots, i_m\}$  is an itemset, we take  $D$  as a collection of instance of  $T$  database, then  $T \subseteq I$ . Each instance transaction has a unique identifier TID. Usually, we use  $X \rightarrow Y$  to express association rules, in which  $X, Y$  represents disjoint itemsets, namely  $X \subset I, Y \subset I$ , and  $X \cap Y = \emptyset$ . Support and confidence threshold to show the correlation of association rules. Support is used to determine the frequency that rules can be used for a given set of data, reflecting the usefulness of the discovered rules; confidence is used to determine the frequency of the transaction of  $Y$  contains  $X$ , reflecting the certainty of the discovered rules<sup>[5]</sup>, namely  $\text{support}(X \rightarrow Y) = P(X \cup Y) = |\{T | T \in D \wedge (X \cup Y) \subseteq T\}| / |D|$ ,  $\text{confidence}(X \rightarrow Y) = P(Y | X) = |\{T | T \in D \wedge (X \cup Y) \subseteq T\}| / |\{T | T \in D \wedge X \subseteq T\}| = \text{support}(X \cup Y) / \text{support}(X)$ .

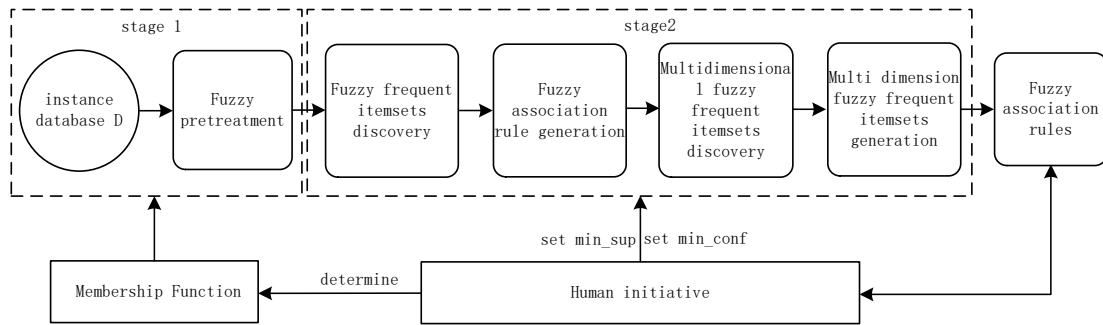


Fig.1. Fuzzy data mining model

## 2.1 Fuzzy association rules process

Ma and colleagues<sup>[3,6]</sup> have made standardization of collected instance, and established standard instance database. In this paper, we take the data mining of all mechanical structure symmetry design instance in the collected instance database, the association rules of structure symmetry  $\Rightarrow$  design requirement were mined, and the mapping process of the corresponding design requirements of the structure symmetry design instance is shown in Figure 2, each symmetric instance has corresponding functions, performances, and constraints.

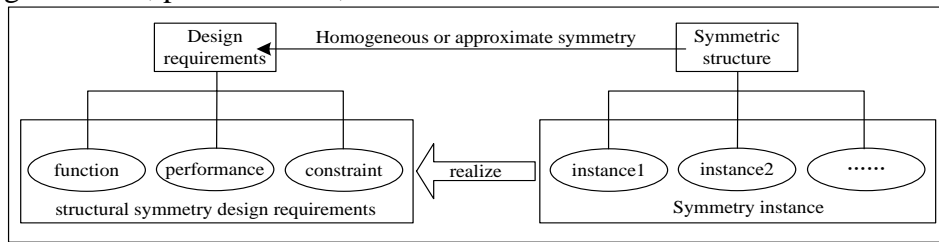


Fig.2. The relation between symmetry structure and design requirement

To simplify the design knowledge mining, the table represent the knowledge about the relationship between the structure symmetry and the design requirements in the structure symmetry instance database(horizontal represent symmetry structure, vertical represent design requirements), in the instance correlation analysis, support can describe the depth of the membership grade of each attribute by specific data, and show in the form of probability.

### (1) The discretization of quantitative attributes

Design knowledge mining is based on the probability of the mechanical structure symmetry design instance as the original database. However, the probability of an instance in the large database is too small, not easy to fuzzy data, so we need to standardize the original data. The object of analysis is used as samples, e.g.  $x_{i1}, x_{i2}, \dots, x_{ik}$ , we call  $X_{ik} = (x_{i1}, x_{i2}, \dots, x_{ik})$  as a sample set. The attributes of the instance database are quantified,  $x_{ik}$  represents the probability of  $i$  line of  $x_k$  to item  $k$  index in the instance database,  $k$  dimensional instance database can be shown as  $X_{ik} = (x_{i1}, x_{i2}, \dots, x_{ik}) (k=1, 2, \dots, m)$ , and separately calculate the average  $\mu_k$  and standard deviation  $\sigma_k$ .

of the original data, namely  $\mu_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$  ( $k=1,2,\dots,m$ ),  $\sigma_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \mu_k)^2}$  ( $k=1,2,\dots,m$ ), then

using  $x'_{ik} = \frac{x_{ik} - \mu_k}{\sigma_k}$  ( $k=1,2,\dots,m$ ) standardize the process of sample data. Due to the particularity

of the data in the instance database, the obtained data are not all in the range of  $[0,1]$ , which increase computational complexity, so it is necessary to deal with range formula  $x''_{ik}$ , namely

$$x''_{ik} = \left| \frac{x'_{ik} - \min\{x'_{ik}\}_{1 \leq k \leq n}}{\max\{x'_{ik}\}_{1 \leq k \leq n} - \min\{x'_{ik}\}_{1 \leq k \leq n}} \right| \quad (k=1,2,\dots,m) .$$

Zadeh proposed the relevant knowledge of fuzzy set theory<sup>[7]</sup>, how to mine the relationships about the attributes in the structure symmetry instance database? There is a good solution to introduce the concept of fuzzy association rules, the relationships about the instances attributes in the database can be represented by fuzzy concepts, it can effectively avoid the disadvantages of the boundary division, the method of fuzzy set is used to transform the probability of the instance, and obtain membership function of data in the structural symmetry instance database.

The determination of cluster centers and membership functions plays an important role in the analysis of fuzzy attributes, and the determination of the membership function is different, which depended on one's understanding of the fuzzy set, the practical experience and other factors, it is subjective. In this paper, we select the trapezoidal membership function, each quantitative attribute select 3 cluster centers, fuzzy set determined by membership function will map the data attributes in the structure symmetry instance database for  $z$  (low),  $\pi$  (middle),  $s$  (high) three attribute values, in which function  $z$  represents the low probability of an instance in the instance database, function  $\pi$  represents the middle probability of an instance in the instance database, function  $s$  represents the high probability of an instance in the instance database. Scanning the collected instance data source, the probability value of the corresponding attribute of each instance can be replaced by the membership grade. Membership function  $F$  as shown in Figure 3.

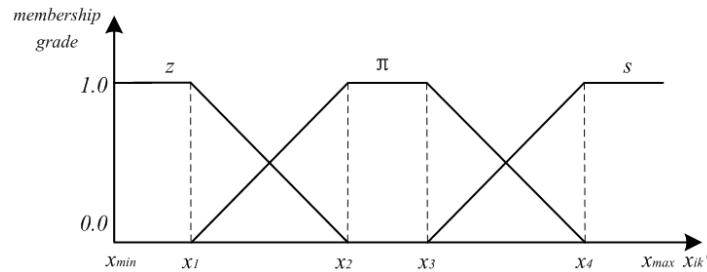


Fig.3. Membership function  $F$  of fuzzy sets

In Figure 3,  $x_1, x_2$  segment is determined by function  $z$  (low), function  $\pi$  (middle),  $x_3, x_4$  segment is determined by function  $\pi$  (middle), function  $s$  (high). In the process of fuzziness, there are 3 cluster centers in each attribute, which need to be normalized, and the non negative conditions should be

met, namely  $\sum_{i=1}^3 f_{ij} = 1, f_{ij} \geq 0$ .

$$z(x) = \begin{cases} 1 & x_{\min} \leq x'' < x_1 \\ \frac{x_2 - x}{x_2 - x_1} & x_1 \leq x'' < x_2 \\ 0 & x_2 \leq x \end{cases} \quad (1) \quad \pi(x) = \begin{cases} 0 & x'' < x_1 \\ \frac{x - x_1}{x_2 - x_1} & x_1 \leq x'' < x_2 \\ 1 & x_2 \leq x'' < x_3 \\ \frac{x_4 - x}{x_4 - x_3} & x_3 \leq x'' < x_4 \\ 0 & x_4 \leq x \end{cases} \quad (2) \quad s(x) = \begin{cases} 0 & x'' < x_1 \\ \frac{x - x_3}{x_4 - x_3} & x_3 \leq x'' < x_4 \\ 1 & x_4 \leq x'' \end{cases} \quad (3)$$

(2) Completing the data initialization by fuzzy membership function process the data in the instance database, then convert the original instance database  $X_{ik}$  into a new instance database  $X''_{ik}$ . The

fuzzy attribute is related to the dimension of the data initialization process, the new data in the database depends on the membership function defined above. We use Apriori algorithm to mine the regular form of interest, and combine the Apriori algorithm with the transaction containing fuzzy attributes. In the first iteration of the algorithm, the fuzzy probability of all the fuzzy 1-itemset is obtained, namely candidate fuzzy frequent 1-itemset  $C_1$ .

(3) Setting minimum support as  $\min\_sup$  and minimum confidence as  $\min\_conf$ , we call strong association rules that rules satisfy the threshold of support and confidence. Scanning the instance database, then delete items that do not meet  $\min\_sup$ , and get a fuzzy frequent 1-itemset  $L_1$ .

(4) Through self-joining  $L_1$ , namely  $L_1 \bowtie L_1$ , we can get candidate fuzzy frequent 2-itemset  $C_2$ , scan it, then delete items that do not meet  $\min\_sup$ , and get a fuzzy frequent 2-itemset  $L_2$ .

(5) Similarly, we can get candidate fuzzy frequent 3-itemset  $C_3$ , then delete items that do not meet  $\min\_sup$ , and get a fuzzy frequent 3-itemset  $L_3$ .

(6) In accordance with same method to perform until no candidate fuzzy frequent k-itemset is generated, then  $C_k = \phi$ , Apriori algorithm stop.

(7) Extracting meaningful association rules from the fuzzy frequent itemsets.

## 2.2 Multidimensional fuzzy association rules discovery

According to the data dimension involved in the actual database, the association rules are divided into single dimension and multidimensional association rules, the single dimensional association rules are only dealing with the relationship among different instances in a single attribute, e.g. structural symmetry realize the relationship about different functional properties. The multidimensional association rules relative the association rules in the single dimensional database, the method of mining rules in the multidimensional database<sup>[8]</sup>, namely dealing with some relationships between attributes, e.g. structural symmetry realize the relationship among functional interval, performance interval and constraint performance. multidimensional association rules not only need to consider the attributes of itemsets, but also need to consider the dimension of the itemsets. The specific mining method is based on the attribute property of the itemsets, and each dimension in the database needs to set the threshold of support and confidence.

In the whole process of extracting association rules, the threshold of support and confidence is very important. Of course, the threshold of support and confidence is not unique, we should revise repeatedly according to the final results of the mining results. To obtain more reasonable and more meaningful association rules in the structural symmetry instance database, and meet the requirements of the article, the paper combine the views of the expert system to carry out scientific value.

Structure symmetry design knowledge mining takes the multi-angle and multi-level knowledge mining mainly based on the structure symmetry, design requirements and so on, it also provides a reliable basis for the application of the concrete instance in the instance database.

The main mining process can be described as following:

(1) According to the frequent itemsets obtained above, and link each set of attributes, then add each original itemset to association rules.

(2) Using the Apriori algorithm to scan the extended data sets that generate candidate itemsets. If the set is larger than the predefined minimum support, then the set is frequent, otherwise not frequent. In the instance database, the design requirements of mechanical products can be divided into function, performance, and restriction; structure symmetry is mainly divided into symmetry, symmetry breaking and weak symmetry, in all structure symmetries, translation symmetry, rotation symmetry, and mirror symmetry are three basic symmetries, we can mine all the frequent itemsets by the level of the relationships.

(3) We should give a weighting on three interval respectively for different importance of function interval, performance interval and constraint interval. Each structure symmetry can realize the corresponding functions. Therefore, the function should be considered firstly, then constraints, and finally performance, namely function > constraints > performance. Getting the weight of collection  $F_i$  for the functional interval, the performance interval and the constraint interval based on the evaluation of the experts and the expert system,  $F_i = (\xi_1, \xi_2, \xi_3)$ .

(4)To get interesting association rules, we should cut off the boring association rules based on interest measure.

### 3. Instance analysis

The proposed design knowledge of fuzzy association rule mining method can guide the application of structure symmetry in mechanical design, the association rules of structure symmetry  $\Rightarrow$  design requirement were mined. The number of samples determines the reliability of rules. To get more convincing data, we collected more than 500 mechanical structure symmetry design instances from specialized books, academic articles, reference manuals, and design drawings<sup>[9,10,11]</sup>.

In this paper, to simplify the analysis process, we only analyze and compare the typical design requirements in the relationship between structure symmetry and design requirement in Figure 2, defined them respectively as A, B, C, D, E, F. We can analyze the relationships among 6 attributes by computing, here A indicates the function of clamping and pressing material, B indicates the function of bidirectional stop rotation, C indicates the constraint of material position parallel, D indicates the constraint of reducing manufacturing cost, E indicates the performance of expanding functional strength, F indicates the performance of expanding functional range. In this part, getting the weight of collection  $F_i$  for the function interval, the performance interval and the constraint interval based on the evaluation of the experts and the expert system, namely  $F_3=[(A,B)\times 0.5, (C,D)\times 0.3, (E, F)\times 0.2]$ .

Table.1. The relationship between structure symmetry and design requirements

design requirements structure		A	B	C	D	E	F
symmetry	translation symmetry	3.51%	6.97%	2.51%	1.92%	12.72%	16.23%
	rotation symmetry	1.68%	1.32%	2.12%	0.99%	12.98%	6.01%
	mirror symmetry	5.13%	3.55%	3.29%	1.38%	11.24%	4.54%
	.....						
symmetry breaking	.....						
weak symmetry	.....						
.....	.....						

Table 1 show that the support of the instance database is too low, we use the method described above to deal with the data in Table 1, each quantitative attribute select 3 cluster centers, namely z (low), $\pi$ (middle),s(high).In Figure3 membership function F of fuzzy set, we make  $x_1=0.1$ ,  $x_2=0.45$ ,  $x_3=0.55$ ,  $x_4=0.9$ ,  $x_{\max}=1$ , after converting the fuzzy instance database as shown in Table 2.

Table.2. Fuzzy database D

Design requirements structure		A			B			C			D			E			F		
symmetry	translation symmetry	1	0	0	0	0.32	0.68	0	0.36	0.64	0	0	1	0.9	0.10	0	0	0	1
	rotation symmetry	1	0	0	0.7	0.3	0	1	0	0	0.2	0.8	0	0	0	1	1	0	0
	mirror symmetry	0	0.09	0.91	1	0	0	0.77	0.23	0	0	0	1	0	0.3	0.7	0	1	0
	Count	2	0.09	0.91	1.7	0.62	0.68	1.77	0.59	0.64	0.2	0.8	2	0.9	0.4	1.7	1	1	1
	.....																		
symmetry breaking	.....																		
weak symmetry	.....																		
.....	.....																		

The generation of candidate fuzzy frequent itemsets and fuzzy frequent itemsets is shown in Figure 4, the example is analyzed to study the relationship among the function interval, the

performance interval and the constraint interval. Count the cluster centers in each attribute and set the minimum support count is 0.8, comparing the candidate fuzzy frequent itemsets generated by each scan with the minimum support count, this is a process of generating a fuzzy frequent itemsets. The last part of fuzzy frequent itemsets mining, self-joining  $L_4$ , we can get candidate fuzzy frequent itemset  $L_5$ ,  $L_5 = \{ \{A_{low}, C_{low}, D_{middle}, E_{high}, F_{low}\} \} \circ C_6 = \emptyset$ . End fuzzy frequent itemset mining.

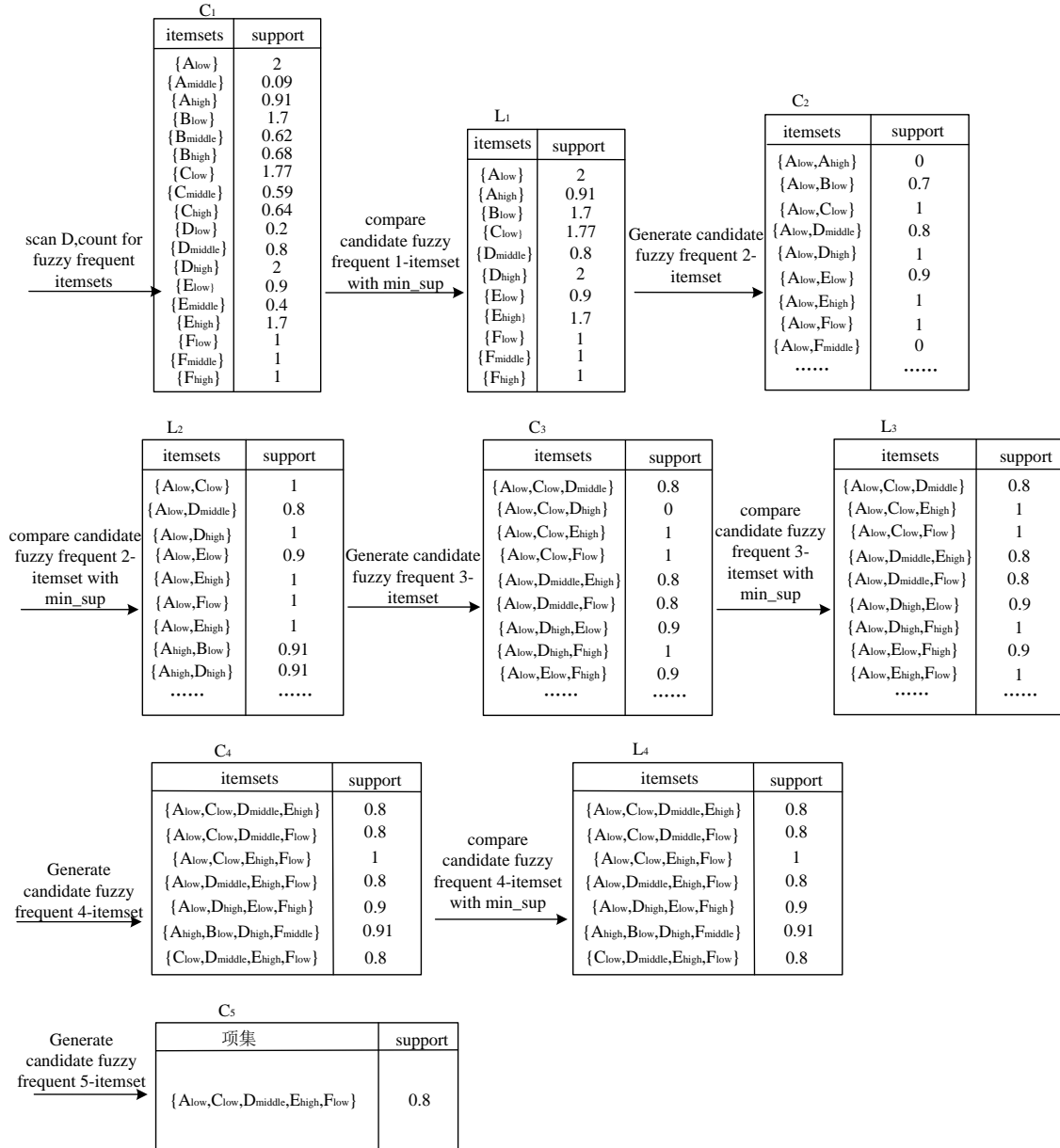


Fig.4. the generation of candidate fuzzy frequent itemsets and fuzzy frequent itemsets

Figure 4 show that all frequent fuzzy itemsets are obtained, we set  $min\_sup=0.044, min\_conf=0.8$ , and mining a number of frequent fuzzy itemsets, due to there are many association rules, the paper only lists some of the association rules, which meet the requirements, as shown in Table 3.

Table.3. Multidimensional association rules extraction in the instance database

Fuzzy association rules	Support	Confidence
$B_{low}, C_{low} \rightarrow E_{high}$	0.069	0.843
$B_{low}, E_{high} \rightarrow C_{low}$	0.069	0.885
$A_{low}, E_{high}, F_{low} \rightarrow D_{middle}$	0.044	0.8
$B_{low}, D_{high}, F_{middle} \rightarrow A_{high}$	0.051	0.91
$A_{low}, C_{low}, E_{high}, F_{low} \rightarrow D_{middle}$	0.044	0.8
.....	.....	.....

The table 3 show that some association rules extracted from table 3, if B=low and C=low, then E=high, with support=0.069, confidence=0.843; we can understand if a small amount of instances have the function of bidirectional stop rotation and the constraint of material position parallel, there will be the probability of 84.3% in which a large number of examples have the performance of expanding functional range. Similarly, we can get the other fuzzy association rules.

#### 4. Conclusion

In this paper, we use the method of multi dimension fuzzy data mining to find the association rules among the function interval, the performance interval and the constraint interval, the result of the knowledge mining guides the scientific application of structural symmetry design in mechanical product structure, and the mining method provides good reference for other fields of design knowledge mining. summarizes the application knowledge of different depth and different application range by different levels of abstraction of different applications. How to fully mine the structure symmetry design knowledge and optimize the mining results? They are the main work of the author's next step.

#### Acknowledgements

This work was supported by the National Science Foundation of China (No. 51305213) and the K.C.Wong Magna Fund in Ningbo University.

#### References

- [1] Feng P E, Zeng L B and Qiu Q Y. Research on mechanical function symmetry architecture and its application[J]. Journal of Mechanical Engineering, 2012, 48: 1-10.
- [2] Hand D J, Mannila H and Smyth P. Principles of data mining[M]. Cambridge: MIT Press, 2001.
- [3] Ma Z Y. Research on mechanical symmetry concept, function and acquisition of application knowledge[D]. Doctor Thesis of Zhejiang University, 2010.
- [4] Tan P N, Steinbach M, Kumar V. Introduction to data mining[M]. Boston: Addison-Wesley Longman Publishing, 2005.
- [5] Jiawei Han, Micheline Kamber, and Jian Pei. Data Mining: Concepts and Techniques(3rd Edition)[M]. Morgan Kaufmann, 2011.
- [6] Liu Fan, Ma Z Y, et al. Research on Concept System and Function of Mechanical Structure Static Symmetry[J]. Machine Design and Research. 2015(5):141-147.
- [7] L. A. Zadeh. Fuzzy sets[J]. Information and Control, 1965, 8(3): 338-353.
- [8] Hu Keyun, Tian Fengzhan, Huang Houkuan. Theory and application of data mining [M]. Tsinghua University Press, 2008.
- [9] Kozhevnikov H. Reference handbook of machine[M]. Beijing: China Machine Press, 1988.
- [10] Yan D K and Zhen L X. Lifting equipments[M]. Zhengzhou: Zhengzhou University Press, 2003.
- [11] Lin J R. Conformation of automobile (Engine volume) [M]. Beijing: Publishing House of Electronics Industry, 2004.