Research on the Data Mining Method for Design Knowledge of Industrial Robots Based on Association Rules

Weigang Li, Ph.D.
Professor in Industrial Engineering
Dean of Industrial Engineering
Harbin Institute of Technology
Harbin, China
weigangli@hit.edu.cn

Chi Wang, Bachelor
School of Mechatronics Engineering
Harbin Institute of Technology
Harbin, China
1530119205@qq.com

Jihong Yan*, Ph.D.
Professor in Industrial Engineering
Deputy Dean of School of Mechatronics Engineering
Head of intelligent Manufacturing Scientific Research Team
Harbin Institute of Technology
Harbin, China
jyan@hit.edu.cn

Abstract**—**In today's rapid industrial development, the application of industrial robots is becoming increasingly extensive, where the industrial robot needs diversified designs to adapt to rich application scenarios and different utilization conditions. How to design products quickly and reasonable according to the market demand has become an urgent problem in the development of industrial robots. The traditional design method of robots is mainly in accordance with experience, or functional analysis and module division of industrial robots which needs a lot of time cost and resources. In order to improve the design efficiency of industrial robots, this paper proposes a design method based on association rules through knowledge mining in which the robot structural feature parameters are accumulated and utilized. In this process, the K-means clustering method is used to discretize the data through the Euclidean metric between the feature parameters, and then the association relationships are mined by using the Apriori. Finally the association rules are summarized according to the physical meaning of the structural usefeature parameters. The method proposed in this paper provides a scientific basis for the rapid design of robots and improves the rationality of robot design.

Keywords**—**Industrial robots, Association rules, Data mining, Structural design parameters, Apriori

1. Introduction

With the rapid development of modern industry, many new modern technologies have emerged, and industrial robots are one of them (Fu et al 2020). The technologies which contains mechanization, digitization, electronics, automation, computers, and artificial intelligence. Because of the rich technology included, its practical application range is wide. For example, industrial robots have become an indispensable part of large-scale manufacturing. It has been widely used in many automated industries and flexible manufacturing systems (Mu 2020). Industrial robots have greatly liberated human resources which improved the efficiency of industrial production, and promote the process of industrial production automation. It vigorously transforms the mode of economic development and enhance the level of technological competition in my country's industry (Zhu 2018). Because of its programmability, stability, flexibility, and high precision, it is widely applied in the manufacturing industry. In order to adapt to the rich application scenarios, higher requirements are put forward for the diversity of the robot body structure.

How to design products quickly and accurately according to the market demand has become an urgent problem in the development of industrial robots. In order to solve the problems in the design and optimize the design process as much as possible, a research method of modular design is proposed. Xiao Chao et al. established a library of robot modules and realized automatic assembly between modules (Xiao et al 2015). Ge Weimin et al. proposed a locking mechanism and mechanism docking algorithm to realize the connection between robot modules (Ge et al 2016). Liu Shuang et al. completed the reconfigurable construction of the 50 kg load industrial robot body and conducted static analysis of important parts (Liu et al 2011). The analysis found that the modular industrial robot design has some shortcomings: (1) There are many types of modules and the structure is complex. (2) Modularity is not universal that it is limited to the modular design of certain joints or parts of the industrial robot (Feng et al 2019). In summary, modular robots are complex that have limitations.

With the rapid development of network information technology and the enhancement of massive data processing capabilities, we have entered the era of big data (Wang 2021). Wikipedia defines big data as the amount of data which is so large that it is difficult to obtain, manage, process, and organize information for the purpose of helping business decision-making within a reasonable time through the current mainstream software tools (Peng 2017). In practical applications, the significance of big data does not lie in its information content, and the most important thing is the information processing capabilities of big data. In the development of enterprises and industries, processing information quickly and accurately can help enterprises rapidly occupy the market. This is also the reason why big data has attracted the attention of enterprises (Chen 2021). In terminal artificial intelligence products, technicians can use data retrieval operations to directly retrieve application data which shorten the data transmission link, and help improve the efficiency of artificial intelligence. It also saves the resource cost and time cost of data transmission (Si et al 2020).

In the context of the big data era, data mining technology has gradually been widely used. Cao et al. analyzed the influencing factors of road damage conditions based on the Apriori through association rule mining (Cao.et.al 2021). Liu et al. established an association rule model for obtaining typical process sequences, and proposed the use of improved Apriori to obtain typical Algorithm of process sequence (Liu rt al 2006). Jing et al. proposed a data mining-based holographic virtual product conceptual design model by using data mining technology and methods, and established a data warehouse for product conceptual design scheme generation under the development of data mining tools (Jing et al 2003). Yu et al. analyzed the correlation of meteorological big data based on the Apriori (Yu et al 2020). By learning from other examples that apply data mining, this article proposes a design method based on association rules through knowledge mining in which the robot structural feature parameters are accumulated and utilized.

Data mining technology is used to realize the relationship between robot parameters and advances automatic of robot design by applying the rules to the design. Using data rationally from the robot structure parameter system makes the efficiency of robot design and modeling greatly improved, and at the same time look forward to the future development of the industry. In order to improve the design efficiency of industrial robots, this paper proposes a design knowledge mining method based on the application of robot structural feature parameters through association rules. The rest of this article is arranged as follows: the second section introduces the robot structure and data analysis and processing; the third section explains the selection of the algorithm. Finally, the association rules are summarized according to the physical meaning of the structural feature parameters. The method proposed in this paper provides a scientific basis for the rapid design of robots and improves the efficiency of robot design.

2. Structure Analysis and Data Collation

2.1 Structure analysis

As industrial robots are gradually applied to various fields, the design and modeling process of industrial robots is particularly significant. Therefore, this topic will be launched for the purpose of realizing automated modeling to improve the efficiency of robot design and modeling. Under the data default conditions, the robot modeling is completed with a high degree of completion by using the analysis of the strong association relationship between the data. The robot space work diagram which from the official website of the robot development company as shown in Figure 1 below. The robot space work diagram provides data such as the limit position of the robot movement and the length of the rod. According to these data, the position information of rotation center axis of each joint can be determined. On the basis of data, this paper analyzes the process of designing the robot through modeling on SolidWorks. The preliminary modeling is shown in Figure 2 below.

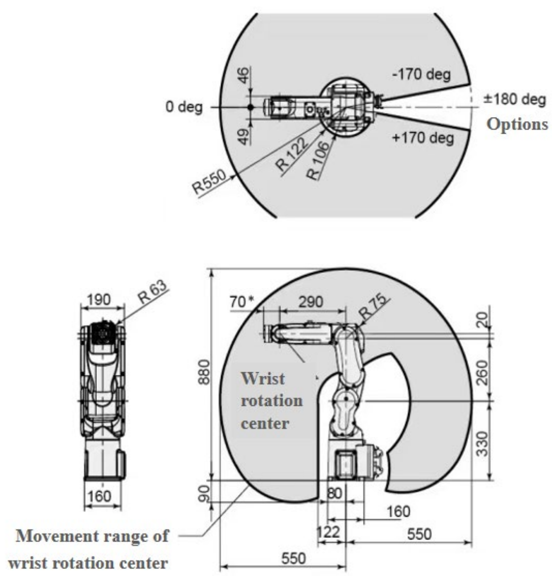

Fig.1. Robot space motion range diagram

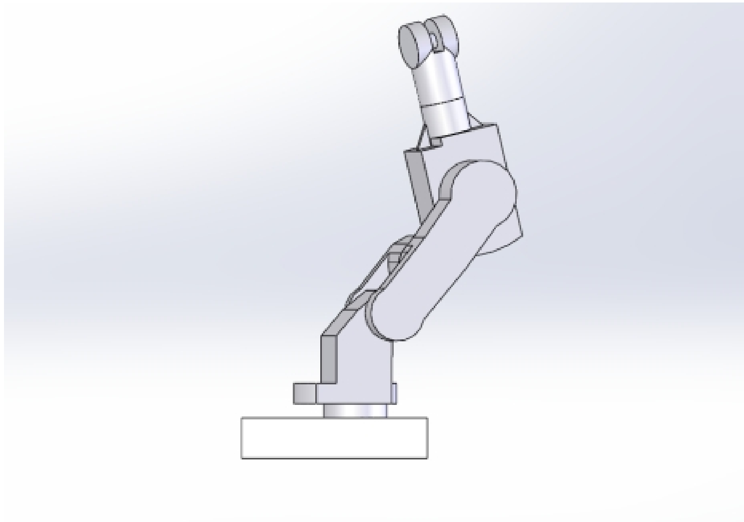

Fig.2. Preliminary modeling of robot based on SolidWorks

2.2 Data collation

After completing the robot modeling, the parameters of the key structural will be summed up, such as the extreme space position, the rod length and other information. The general shape and space size of the robot can be determined. The strong association relationship between the data can satisfy the known data to complete the modeling with the maximum degree of completion. However, the specific rod diameter and thickness of the robot arm with other related information cannot be determined.

Finally, 469 pieces of robot data were collected from 7 companies. Each piece of data contains 18 characteristic parameters such as range, payload, accuracy, and quality. The data source is shown in Table 1

TABLE 1. Robot data source

Company	Number
EFORT	26
ROKAE	14
FANUC	113
YASKAWA	96
KUKA	112
ABB	64
SIASUN	44

3. Parameter Discretization and Association Rule Analysis

3.1 Parameter discretization

In order to simplify the data structure, it is usually necessary to adopt the data discretization technology to discretize the large and complex data. Discretization of continuous data aims to divide the data range into several discrete intervals, and then get different symbols to represent the attribute values falling in each subinterval. The results obtained by using discretized data in data mining are more convenient, compact and accurate, and more convenient for us to use.

In this paper the K-means clustering algorithm aims to preprocess the data of the robot's structural feature parameters which divide the interval. When using the K-means to discretize data, the selection of the K is very important. In this paper, the contour coefficient method is used to complete the determination of the K. Taking the maximum range of motion of an industrial robot as an example, more than 400 kinds of data collected from seven companies are imported into an Excel table. The contour coefficient method is impressed on generating the two-dimensional coordinate system shown in Figure 3 below. According to the principle of the contour coefficient method, the larger the axis value, the better the clustering effect. As shown in the figure, the maximum K of the contour coefficient can be obtained which means that our optimal number of clusters is 2, but when the K is 2, SSE (sum of the squared errors, the sum of squares of errors) is also larger. Selecting the K of 4 means clustering the range into 4 categories. After selecting the appropriate K, the data is divided into four categories through the K-means algorithm to complete the discretization. In this paper, the Euclidean distance is selected to implement the K-means algorithm. The specific implementation process is shown in Figure 4.

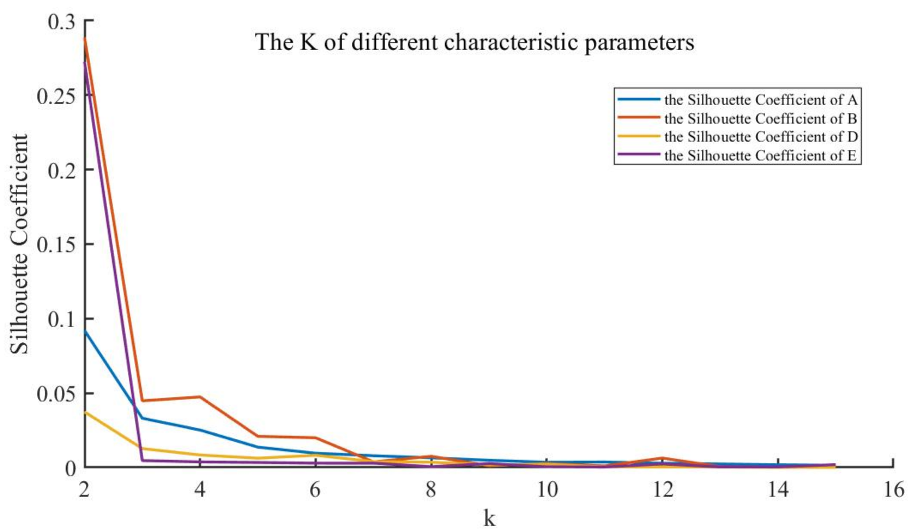

Fig.3. The K and the contour factor

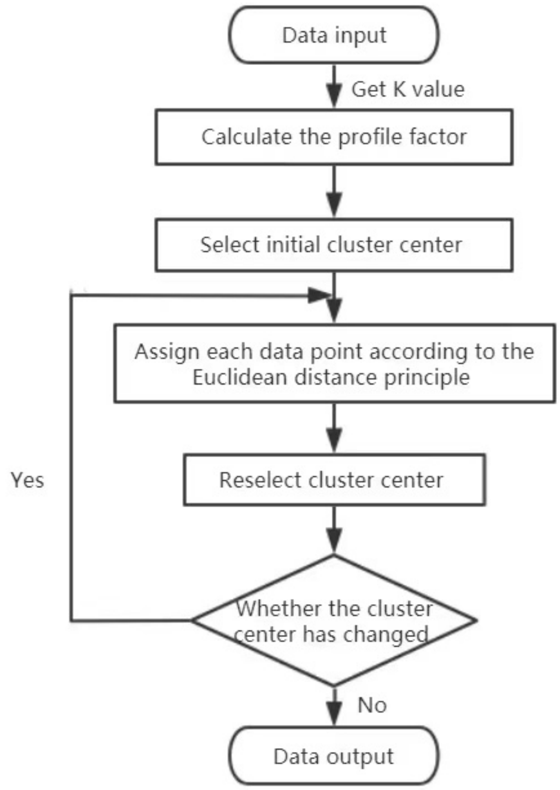

Fig.4. K-means algorithm discretization data flow

3.2 Association rule analysis

Association analysis is an important branch of data mining. Association rule analysis supports to find out the potential connections between data and conduct a full-scale analysis of data in different dimensions, so as to make more ideal decisions and predict development trend of things more accurately. Two algorithms which Apriori and FP-tree are commonly used in the association rule analysis. The core problem of researching the related transaction rules and data mining is that all association rules need to meet the support of data A item set in the database of a given transaction. At the same time, when the data A item set is retrieved, the rule data B item set also meets the confidence of being retrieved.

According to the content of the subject project, the main purpose is to finally realize the automated modeling of the robot which requires a certain degree of familiarity with the robot structure and a certain amount of robot data reserve. Develop strong associations between different types of data by using the appropriate algorithm, then complete automated modeling. Under the research of the robot structure and the understanding of the purpose of the project, the selection of algorithms is mainly concentrated on Apriori and FP-tree. By learning of the algorithm, the Apriori was finally selected to realize the mining of the relationship between the robot structural feature parameters.

The Apriori is a classic association relationship analysis algorithm whose main idea is to find the largest frequent item set in the transaction item set that use the largest frequent item set and the preset minimum confidence threshold to generate strong association rules. The computing power of all data attributes, data values, and data volumes traversed in association analysis is huge. The key principle of the Apriori is to split the itemset into the inclusion relationship of itemsets and subsets. When the subset meets the criteria of frequent itemsets, then its original itemsets must be frequent itemsets. The key principle of the Apriori simplifies the calculation amount of search traversal from the mathematical logic. The specific implementation steps are divided into three steps, as shown in Figure 5.

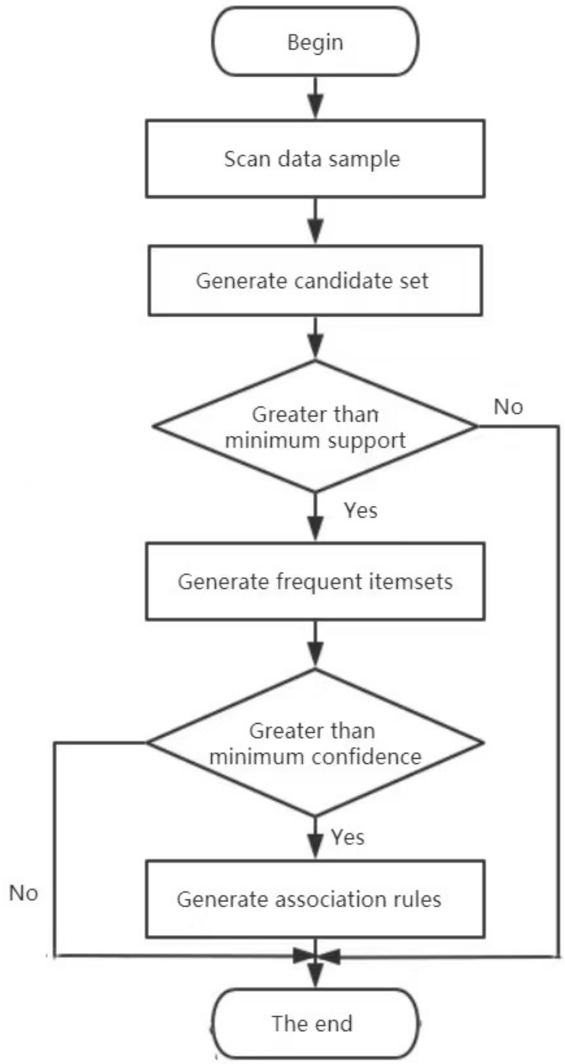

Fig.5. Apriori implementation steps

4. Experimental Verification and Conclusion

After learning of the algorithm, the association rule algorithm is selected through combining with the characteristics of common algorithms. According to the advantages and disadvantages of the FP-growth and Apriori, the algorithm of the project finally determines to use Apriori since it is mainly suitable for numerical or nominal data. Therefore, data needs to be preprocessed to achieve discretization by means of suitable algorithm. Discretize the collected data and perform renaming operations. For example, set the code name of the range to "A". According to the previous section, the K-means algorithm is used to divide the range into 4 intervals. The order is A1, A2, A3 and A4. The same method is used to discretize and rename data such as payload, accuracy, quality, protection level and height. Follow the above steps in sequence to get the data code shown in Table 2.

TABLE 2. Robot data code

Name	Codename	Name	Codename
Range	A	Forward range	J
Payload	B	Backward range	K
Protection level	C	Axis 1 speed	L
Precision	D	Axis 2 speed	M
Quality	E	Axis 3 speed	N
Application	T	Axis 4 speed	O
Number of axes	G	Axis 5 speed	P
Height	F	Axis 6 speed	Q
Up range	H	Axis 7 speed	R
Down range	I	Axis 8 speed	S
Controller	U		

After the data discretization process, the application of the Apriori is implemented by using Pycharm, and the Excel data is run after importing. The frequent 1 item set, frequent 2 item set, frequent 3 item set and frequent 4 item set are obtained respectively with a support degree of 0.1. Then get the relationship between the robot structural feature parameters. Take the relationship between A, B, D and E as an example, as shown in Table 3 below.

TABLE 3. Examples of the relationship between A、B、D and E

Serial number	A known1	A known2	Unknown1	Unknown2	Confidence
I	B2	\	D3	\	0.65986
II	B1	\	E1	\	0.64685
III	A1	\	E1	B1	0.91358
IV	E3	B2	D3	\	0.72358

Analyze the running results and take one of the rules as an example.

From the data discretization naming, it can be seen that: B2—payload is (340,900); D3—accuracy range is (0.04,0.08]. Then the operation result of Article I shows that when the payload is at (340,900], its accuracy range at (0.04, 0.08], and the confidence of this conclusion is about 0.65986.

Aiming at the final data to be substituted, arithmetic mining is carried out through Apriori. A series of association rules are obtained, and the rules are classified into ten categories shown in Table 4.

TABLE 4. Rule classification

Serial number	Classification
I	The relationship between A, B, D, E
II	The relationship between L, M, N, O, P, Q
III	The relationship between B, E, L, M, N, O, P, Q
IV	The relationship between B, L, M, N, O, P, Q
V	The relationship between E, L, M, N, O, P, Q
VI	The relationship between A, L, M, N
VII	The relationship between D, L, M, N, O, P, Q
VIII	The relationship between B, E, H, I, J, K
IX	The relationship between B, H, I, J, K
X	The relationship between E, H, I, J, K

5. Summary and Outlook

This paper collected and sorted out the robot structural feature parameters by understanding of the robot design process with modeling. According to the discretization preprocessing of K-means and the association relationship mining of the Apriori, ten types of rules were obtained that guide the process of the robot design. In the following research, we hope that the application and other structural feature parameters of the robot can be mined to obtain the corresponding association relationships.

Acknowledgements

This work was supported in part by the JCKY Project of China (#JCKY2019603C016).

References

Fu, Q. L, Lu, Q. D, and Jiang, S. Y. The development status and prospects of industrial robots in my country, *Modern Manufacturing Technology and Equipment*, vol. 56, no. 10, pp. 217-218, 2020.

Mu, Q, Analysis of the development status and future trends of industrial robots, *Southern Agricultural Machinery*, no. 28, pp. 89-90, 2020.

Zhu, J. Z, Research on the Application of Industrial Robot Technology, *Modern Industrial Economics and Informatization*, vol. 8, no.15, pp. 57-58+67, 2018.

Xiao, C, Zhou, Y. L, and Sheng, H. Y, et al. Modular Design of Mechanical Noumenon for Industrial Robots, *China Mechanical Engineering*, vol. 27, no. 8, pp. 1018 − 1025, 2016.

Ge, W. M, Zhang, Y, and Wang, X. F, et al. Locking Mechanism Design and Docking Research on Self-reconfigurable Robot, *Journal of Mechine Design*, vol. 32, no. 4, pp. 46 − 50, 2015.

Liu, S, Yin, G. F, and Li, X. Q, et al. Configuration of Industrial Robots Based on Reconfigurable Modularization and Its Statics Analysis, *Machinery Design & Manufacture*, no. 11, pp. 205 − 207, 2011.

Feng, H. J, Miao, Q, and Fan, S. Q, et al. Modular and Parametric Design for Mechanical Structure of Industrial Robots, Machine Tool & Hydraulics, vol.47, no. 21, pp. 60 − 63, 2019.

Wang, R. H. Analysis of scientific and technological information research under the background of big data era, *Technology and Market*, vol. 28, no. 02, pp.183+185, 2021.

Peng, J. Y. Analysis of the opportunities and challenges faced by scientific and technological information work in the era of big data, *Low Carbon World*, vol. 35, no. 28, pp. 289-290, 2017.

Chen, X. N. The development trend of mechanical automation under the background of big data, *Science and Technology Innovation and Productivity*, no.01, pp.73-75+78, 2021.

Si, K. P, Fan, T. C, and Fan, L.M. The effective application of artificial intelligence in computer network technology under the background of the big data era, *Computer Knowledge and Technology*, no. 33, pp. 176-177, 2020.

Cao, L, Xu, L, Yang, F, and Jia, P. F. Analysis of influencing factors of road damage based on association rule mining, *Computer System Applications*, vol.30, no. 01, pp. 186-193, 2021.

Liu, S, Tian, X.T, Zhang, Z. M, Xu, J. X, and Zhu, M. Q. A typical process sequence acquisition method based on Apriori algorithm, *Computer Integrated Manufacturing System*, no.08, pp.1279-1283, 2006.

Jing, X. W, Yi, H, and Zhao, L. C. Research on Product Conceptual Design Modeling Based on Data Mining, *Computer Integrated Manufacturing System-CIMS*, no.11, pp.950-954+1000, 2003.

Yu, L. M, Li, Z, Yang, S. Q, Zhang, X. B, and Fu, Q. Correlation analysis of meteorological big data based on Apriori algorithm, *Information and Computer (Theoretical Edition)*, vol. 32, no. 15, pp.75-77, 2020.

Biography

Jihong Yan is a professor (since 2005) in Industrial Engineering at Harbin Institute of Technology (HIT), she is also the deputy dean of School of Mechatronics Engineering and head of intelligent manufacturing scientific research team at HIT. She received her PhD from Harbin Institute of Technology. Then she joined Tsinghua University, the University of Wisconsin, and Pennsylvania State University as a postdoctoral researcher. Dr. Yan is the director of National High-end Equipment Manufacturing Virtual and Simulation Experiment Teaching Center, head of Research Oriented Teaching Innovation Team for High-end Equipment Manufacturing of the Ministry of Industry and Information Technology of China, vice chairman of Production System Special Committee of Chinese Mechanical Engineering Society, and chairman of Industrial Engineering Professional Committee of the Mechanical Engineering Society of Heilongjiang Province. Her main area of research is industrial big data, sustainable manufacturing, intelligent logistics and advanced maintenance of machinery. As a PI, Dr. Yan has worked on and accomplished 15 projects in intelligent manufacturing and sustainability related areas, funded by the NSF of China (NSFC), NSFNSFC joint-project funding, National key R&D plan project funding, National High-tech project funding, National "863" project funding, EU EPSRC project funding, High-tech funding from industries, and so on. She has authored and co-authored over 100 research papers and published 3 books, two papers were ranked ESI high cited articles. Currently there are 17 professors and engineers with her research team, the team dedicates to theoretical research and system implementation in the fields of intelligent operation optimization theory and methods of manufacturing systems, manufacturing IoT technologies and devices, and equipment health monitoring, etc.