

Application Research of Data Mining in MES Quality Management

Jiahai Wang¹, Guozhao Xu^{1,*}

¹Tongji University, Shanghai 201804, China

*Corresponding author: Guozhao Xu

Abstract: With the increasing market competition, enterprises have continuously raised the requirements for product quality in order to gain a favorable position in the market competition. In industrial production, MES quality information management mainly involves the collection, statistical analysis, and utilization of information data that affect quality in daily production. In actual industrial production, the formation of product quality is the result of the interaction of numerous factors. Due to the many factors and scenarios involved, traditional data statistical analysis methods cannot accurately and effectively analyze the collected relevant data, and fail to fully mine the value of data. This paper proposes the application of data mining technology to MES quality information management systems and elaborates on the use of the K-means algorithm and Apriori association rule algorithm to analyze the related processing rules of parts in the production and processing process. The algorithm model is used to analyze the actual production and processing data of a certain enterprise, and finally, the value and application of parts association processing rules in actual enterprise production management are summarized.

Keywords: MES quality management system; data mining; association rules.

1. Introduction

Quality management is a control activity for the production and processing of products. Therefore, for general production activities, quality management functional modules should be defined based on the quality characteristics of the product, and reasonable quality control plans and process control standards should be established. Traditional quality management systems have not considered statistical analysis of the manufacturing process, but have only recorded quality data of the enterprise, paying little attention to the production process, which cannot meet the current requirements for quality and production efficiency in production management. MES quality management system realizes the sharing of production process data and quality management process data. By applying data mining technology to the MES quality management system, hidden association rules in the part machining process can be excavated from the data, which can be beneficial to the optimization of quality management in enterprise production and form a closed-loop quality control.

2. Data Mining Applied to MES Quality Management

2.1. Introduction to Data Mining and Main Step

As a cutting-edge discipline in computer science, data mining is a fusion of ideas from statistics, artificial intelligence, machine learning, signal processing, optimization, information retrieval, and pattern recognition, enabled by database technology, parallel computing, and distributed computing. Data mining refers to the process of discovering previously unknown useful patterns and valuable information in large data repositories. With the development of information technology such as the Internet, contemporary society has entered the era of big data, and data mining

technology has been widely applied in industrial production with broad application prospects.

Data mining is a cutting-edge discipline in computer science that integrates statistical analysis, artificial intelligence, machine learning, signal processing, optimization, information retrieval, and pattern recognition. It relies on the technological support of database, parallel computing, and distributed computing. The main tasks of data mining can be divided into predictive tasks and descriptive tasks. The primary steps in data mining are as follows:

(1) Data Preparation: Data preparation involves identifying data objects and data preprocessing. Firstly, the data mining object needs to be identified. On the basis of clarifying the data mining object and target, useful information can be obtained, and reasonable and feasible decisions can be made.

(2) Data Preprocessing: Due to the fact that the original data may be incomplete, contain anomalies, and duplicate data, it is necessary to preprocess the data before conducting data mining. The main methods of data preprocessing include data selection, data cleaning, data integration, data transformation, and data reduction. The main purpose of data preprocessing is to optimize the data analysis of data mining, reduce analysis costs, decrease time, and improve quality. After data preprocessing, data mining efficiency and quality can be effectively improved.

(3) Data Mining: After data preprocessing, the data has met the basic requirements of data mining. At this stage, it is necessary to clarify the task type based on the requirements and objectives of data mining and select appropriate data mining algorithms. By analyzing the preprocessed data, valuable information for actual production can be obtained, resulting in benefits for the enterprise.

(4) Validation and Application: After the data mining model is established, it is necessary to explain and evaluate the model. Actual production data can be used to verify the accuracy of the model, and the established model can be explained to help users understand the model.

2.2. Functional module design of MES quality management system integrated with data mining technology

The application of data mining techniques to MES (Manufacturing Execution System) quality management systems has become increasingly popular in recent years. In particular, the use of association rule mining based on the k-means algorithm has been shown to be an effective method for analyzing data and identifying related factors that

influence quality. By doing so, this approach can significantly improve the intelligence level of MES quality management systems, more effectively mine the value of quality data, and overcome the limitations of manual statistical data analysis. Moreover, integrating data mining technologies into the MES quality management system can lead to a reduction in labor costs and help enterprises to better manage production. The functional model of the MES quality management system with integrated data mining technology is illustrated in the accompanying figure.

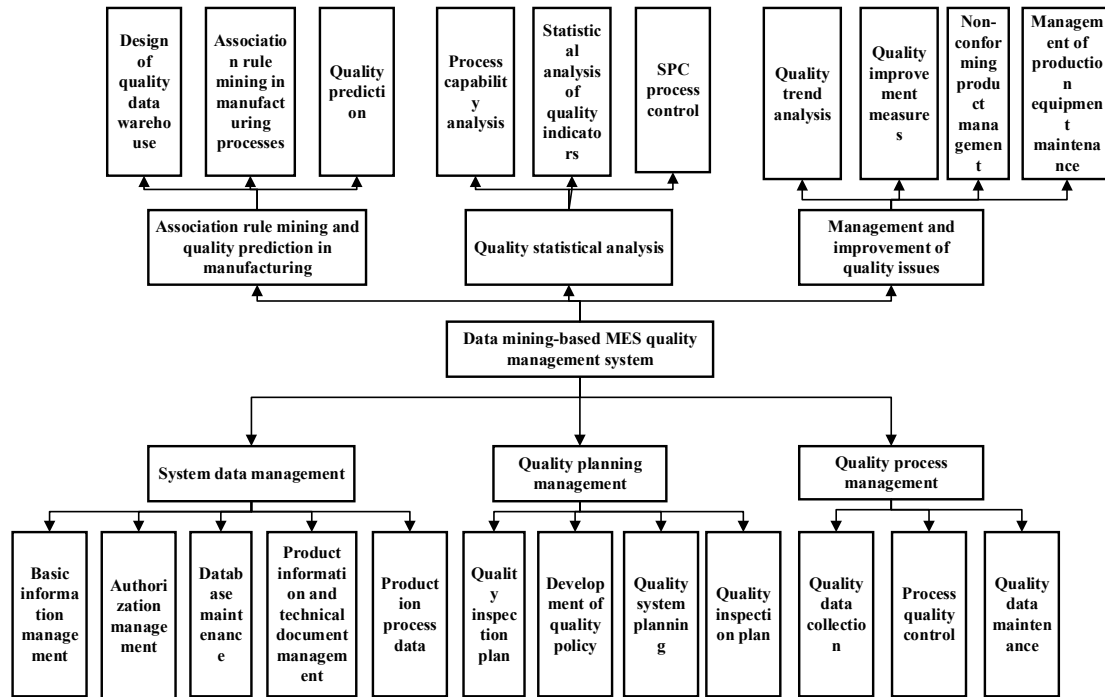


Figure 1. Functional model of MES quality management module

The functional model of the MES quality management system designed includes new modules for quality process rule mining and quality prediction. This paper primarily focuses on the analysis of the quality process rule mining module, which aims to discover the association rules in the part machining process and proposes a novel association rule mining algorithm model based on the K-means algorithm.

3. Application of Association Rules Based on K-means Algorithm in Obtaining Parts Processing Rules

This paper proposes a rule mining model for the machining process of parts, which is composed of clustering and association rule algorithms. The clustering algorithm is employed to analyze the historical quality-related data and calculate the clustering centers of qualified, defective, and scrapped products. By comparing the production status of the current product with each clustering center, the production status of the current product is classified into its corresponding cluster. Sampling in each cluster individually can ensure the consistency of the sample with the overall distribution. The association rule algorithm is used to learn and analyze the historical data to mine the hidden relationships among various quality influencing factors. The discovered rules will serve as new knowledge and rules to guide the manufacturing of new products and conduct quality

prediction, creating value for enterprises. The prediction results are obtained by combining the prediction results of clustering and association rule algorithms based on their historical prediction accuracy.

3.1. Research on Association Rules Algorithm Based on K-means Algorithm

3.1.1. Introduction to K-means Algorithm

As a commonly used algorithm in data mining, the K-means algorithm in clustering methods has the characteristics of simplicity, easy understanding, and efficient data processing. Clustering can divide data into meaningful clusters, which is the starting point for solving other problems, and is widely used in data mining. Due to the large amount of data in the production process, data integration and inconsistency, and the high real-time requirements for data mining, this paper proposes a K-means algorithm-based association rule algorithm model. Firstly, the K-means algorithm is used to calculate the clustering centers of qualified products, defective products, and substandard products[1]. Then, the association rule algorithm is applied to analyze the relevant data in the machining process of the parts.

The K-means algorithm, as a commonly used algorithm in data mining, is characterized by its simplicity, ease of understanding, and efficiency in handling data. Clustering data into meaningful clusters is the starting point for solving other problems, and K-means has been widely applied in data

mining. Given a dataset $Q = \{x_1, x_2, \dots, x_n\}$, the K-means algorithm minimizes the sum of squared errors by partitioning the clusters $C = \{C_1, C_2, \dots, C_k\}$.

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|_2^2 \quad (1)$$

Here, μ is the mean vector of a cluster, and the formula represents the closeness between the samples in a cluster and the mean vector. A smaller value of E indicates higher similarity between the samples in a cluster. To obtain the optimal solution of E , all possible clusters in the sample set D need to be partitioned. The K-means algorithm approximates the value of E through a greedy strategy and iterative optimization[2].

The K-means algorithm can be summarized as follows:

(1) Randomly select K suitable samples from the sample set Q as the initial mean vectors, also known as the initial cluster centers.

(2) For each data point in the sample set Q , a sphere neighborhood is defined with a positive number as the radius, and the number of samples falling within the spherical region is the density of that point. Choose the point with the highest density as the first initial cluster center.

(3) Compute the mean value of each set of cluster centers, and shift the associated cluster centers to the position of the mean value until the clusters or centroids no longer change.

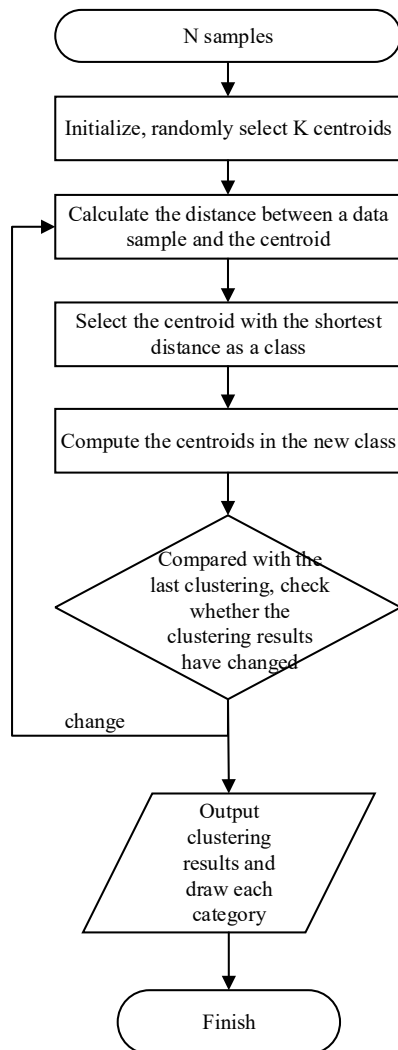


Figure 2. K-means algorithm flow chart

After applying the K-means algorithm, we can obtain the cluster centers of qualified products, defective products, and inferior products. By using the association rule algorithm, the interdependence and correlation between the data that affects the quality can be analyzed. It is possible to mine the value of each data item set by analyzing the association relationships between data item sets. The association rule algorithm is utilized to analyze historical production and processing data of parts and extract the correlation processing rules of the parts.

3.1.2. Introduction to Association Rules Algorithms

For a given dataset, an itemset is composed of all items in $D = \{d_1, d_2, d_3, \dots, d_m\}$. A single association rule can be represented as $A \rightarrow B$, where A and B are both subsets of itemset $M = \{t_1, t_2, t_3, \dots, t_n\}$, and A and B are disjoint. Association rule analysis algorithms rely on two important evaluation metrics, namely support and confidence[3].

Support is primarily used to indicate the proportion of transactions that contain the itemset in the entire transaction set. It can be denoted by the following equation:

$$Sup(A \rightarrow B) = \frac{|A \cup B|}{|D|} \quad (2)$$

Confidence is a key evaluation metric used in association rule analysis, which measures the proportion of transactions that contain both A and B , relative to the proportion of transactions that contain only A . It is represented as:

$$Conf(A \rightarrow B) = \frac{Sup(A \rightarrow B)}{Sup(A)} \quad (3)$$

In order to ensure the effectiveness of rules, traditional association rules mined from data must satisfy threshold requirements of support and confidence. However, the "support-confidence" framework may lead to the generation of some meaningless and valueless association rules. Therefore, some new evaluation metrics have been proposed to measure the quality of rules.

Lift is a measure of the increase in probability of occurrence of the consequent itemset B , given that the antecedent itemset A is known to be present in the transactions. It is calculated as the ratio of the joint probability of A and B to the product of the probabilities of A and B . The formula for calculating lift is:

$$Lift(A \rightarrow B) = \frac{Conf(A \rightarrow B)}{Sup(B)} \quad (4)$$

Lift is a measure of the degree of association between the antecedent and consequent of an association rule. A value greater than 1 indicates a positive correlation, while a value less than 1 indicates a negative correlation. A value of 1 indicates that the antecedent and consequent of the rule are independent of each other.

3.1.3. Removal of Redundant Association Rules

Due to the large amount of data in production processing, the number of association rules generated by association rule mining is usually quite large, which is not conducive to user understanding and analysis. Therefore, it is necessary to

remove redundant rules generated during association rule mining. In order to retain the rules that are of interest to users while removing duplicate and redundant rules, a clear definition of redundant rules is needed. According to the definition, if there are two association rules $R_1: X_1 \rightarrow Y_1(\text{sup}_1, \text{conf}_1)$ and $R_2: X_2 \rightarrow Y_2(\text{sup}_2, \text{conf}_2)$, where X_2 is a subset of X_1 and Y_2 is a subset of Y_1 , then R_2 is a subset of R_1 and can be deduced from R_1 . Therefore, R_2 is a redundant rule with respect to R_1 . Deleting redundant rules can effectively improve the efficiency of association rule mining and remove truly meaningless and valueless rules[4].

3.1.4. Association Rules Mining Model Based on K-means Algorithm

The association rule mining based on the K-means clustering algorithm involves dividing data into several clusters using the K-means algorithm, performing association

rule mining on each cluster using the association rule algorithm, and then combining the results. The specific steps of the association rule mining algorithm model based on the K-means clustering algorithm are as follows:

(1) Firstly, relevant factors that affect the machining quality of parts are identified from the MES database as the analysis data source for the K-means clustering algorithm.

(2) The K-means clustering algorithm is used to analyze the data, with qualified products, defective products, and inferior products serving as the clustering centers.

(3) By establishing the clustering centers, three clusters are formed, and the Apriori association rule analysis algorithm is used to analyze the data of each cluster, resulting in analysis results.

(4) All the results are then combined and analyzed to mine meaningful and valuable association rules for the production and processing of parts.

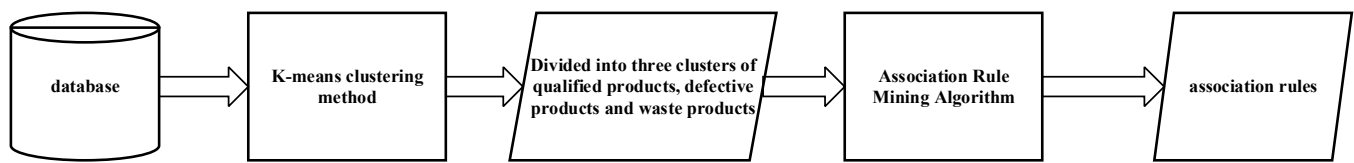


Figure 3. Flowchart of mining association rule mining model for parts processing

4. Model Test and Result Analysis

4.1. Data preparation

The data used in this study was collected from a certain automotive engine parts manufacturing company in Shanghai, with the aim of analyzing the correlation between human-machine-material-environment-measurement (HMME&M) factors and quality, and discovering the association rules of the production process of parts. A data warehouse with quality as the theme was established based on a large number of production and quality inspection records accumulated in the MES database of the company, which provided the data source for data mining. The K-means algorithm was

employed to perform cluster analysis on the data, and then association rule mining was conducted to discover the association rules among various influencing factors, thereby analyzing the correlation rules of the parts manufacturing process.

4.2. Data analysis

A portion of data was selected from the established part manufacturing quality data warehouse as sample data, mainly including basic information of machining personnel, tool information, machining process parameters, workshop environmental parameters, machining equipment information, required machining accuracy, etc. Some of the data are shown in Table 1 and Table 2.

Table 1. Partial sample data

serial number	part name	process name	...	tool name	cutting speed (m/min)	roughness value (μm)
1	crankshaft	Precision milling	...	Form milling cutter	60	3.20
2	gasoline engine piston	Precision milling	...	high-speed steel inserted-tooth milling cutter	60	2.18

Table 1. (continued table). Partial sample data

3	car concave cam	Precision turning	...	hard alloy turning tool	50	0.96
4	Automobile connecting rod	Precision milling	...	form milling cutter	60	1.96
...		

Table 2. Partial sample data

Operator ID	Proficiency level	Temperature (°C)	Metrology accuracy (mm)	Accuracy requirement (mm)	Result
7844	6	20	0.002	0.012	qualified
8945	5	23	0.002	0.012	qualified
3432	6	23	0.002	0.012	qualified
1483	5	19	0.002	0.012	unqualified
9135	4	22	0.002	0.012	qualified
3734	5	20	0.002	0.012	qualified
1748	6	22	0.002	0.012	qualified
...	

The historical machining records were analyzed using the K-means clustering algorithm with a setting of K=3. The Apriori algorithm was then applied with a minimum support threshold of 0.1 and 80% of the data used for training, while the remaining 20% was used for testing. To ensure the stability of the test results, each experiment was repeated eight times, and the average of the eight trials was taken as

the final test result. After the Apriori algorithm was applied, a total of 36 rules were mined from the data, with 34 of them being qualified rules and the remaining 2 being unqualified rules. The support, confidence, and lift of the mined rules were 0.68, 0.88, and 1.104, respectively. Some of the rules obtained for the correlation between the machining process and the quality of the parts are shown in Table 3.

Table 3. Association rules of some parts processing

serial number	rule guide	rule successor	confidence
1	Operator number: 7844, proficiency: 6, part name: gasoline engine piston, working step: Precision milling	qualified	0.88
2	Operator number: 7844, proficiency: 6, part name: automobile connecting rod, working step: Precision turning	unqualified	0.89

Table 3. (continued table). Association rules of some parts processing

3	Operation number: 8945, Proficiency: 5, Machine Tool Name: Ordinary Lathe C6132	unqualified	0.87
4	Machine name: vertical milling machine; machine number: X5647, step name: Precision turning	unqualified	0.90
5	Tool name: Carbide knife, workpiece material: No. 45 steel, cutting speed: 60 m/min	qualified	0.88
...

The association rule algorithm model based on K-means algorithm was used to analyze the various factors in the part processing process, and the processing association rules in Table 3 were obtained.

By comparing association rules 1 and 2, it is found that when the same operator processes different parts with different processing steps, the quality pass rate is not consistent. Specifically, for the employee with the operator ID of 7844 and proficiency level of 6, when processing the part "gasoline engine piston" with the operation step of "fine milling", the product quality has a higher pass rate.

Association rule 3 indicates that when the employee with the operator ID of 8945 and proficiency level of 5 operates the ordinary lathe C6132, the parts produced are mostly unqualified. The confidence level of this rule is 0.86, and it can be used to assess the performance of the employee and examine whether the employee is proficient in operating the ordinary lathe C6132.

Association rule 4 indicates that when the vertical milling machine with the machine number X5647 is performing fine machining, there is a high possibility of producing unqualified products. The confidence level of this rule is 0.90, which indicates that there is a fault in the vertical milling machine with the machine number X5647, and maintenance personnel need to be arranged to inspect and repair it.

Association rule 5 indicates that when the hard alloy tool is used to process the material 45# steel at a cutting speed of 60m/min, the produced parts are mostly qualified. The

confidence level of this rule is 0.88, which indicates that when processing material 45# steel with a cutting speed of 60m/min, a higher pass rate can be achieved. This rule can be used to guide workers in production and processing.

4.3. Evaluation of correctness of rules

In order to evaluate the obtained rule set, this study employs the score-based prediction method proposed by Li[5] to predict the test set samples analyzed. Given the rule set and test samples, subsets that correspond to them are identified. If the prediction results in the subset are consistent, the prediction result is assigned to the corresponding sample. Otherwise, the rule subsets are partitioned based on the prediction results, and the weighted scores of the combined data are calculated. The result corresponding to the group with the highest score is then assigned to the sample.

For a given sample set, we use overall classification accuracy and minority class classification rate (TPR).

Overall classification accuracy is a measure of the proportion of correctly classified samples out of the total number of samples. It is computed as the ratio of the number of correctly classified samples to the total number of samples:

$$\text{accuracy} = \frac{TN + TP}{NP} \quad (5)$$

Minority class classification rate, also known as True Positive Rate (TPR), is a measure of the proportion of

minority class samples that are correctly classified as belonging to the minority class out of the total number of minority class samples. It is computed as the ratio of the number of correctly classified minority class samples to the total number of minority class samples:

$$TPR = \frac{TP}{P} \quad (6)$$

In the evaluation of the performance of the K-means-based association rule algorithm, the true negatives (TN) represent the number of correctly classified qualified samples in the test set, true positives (TP) represent the number of correctly classified unqualified samples in the test set, total number of samples in the test set is represented by NP, and the number of unqualified samples in the test set is represented by P. After analyzing the obtained data, the accuracy and true positive rate (TPR) were calculated to be 94% and 92%, respectively. These results indicate that the K-means-based association rule algorithm has a certain degree of correctness.

5. Conclusion

In this paper, we propose to apply data mining techniques to the MES quality management system, using the K-means algorithm and association rule mining algorithm to establish a production process association rule mining model. Based on the results of association rule mining, there are many related rules in the historical processing records of the enterprise's

MES quality management system. The mining of the processing rules of parts in the production process can be used as process knowledge to assist in online adjustments during production, provide reference and decision-making support for the development of processing technology and the selection of raw materials, reduce the production cost of the enterprise, and enhance the core competitiveness of the enterprise in the market competition.

References

- [1] Wuest T. Identifying Product and Process State Drivers in Manufacturing Systems Using Supervised Machine Learning[M]. Springer International Publishing, 2015.
- [2] Maudes J, Bustillo A, Guerra A J, et al. Random Forest ensemble prediction of stent dimensions in microfabrication processes[J]. International Journal of Advanced Manufacturing Technology, 2017, 91(1-4):879-893.
- [3] Cheruku R, Edla D R, Kuppili V. SM-RuleMiner: Spider monkey based rule miner using novel fitness function for diabetes classification[J]. Computers in Biology & Medicine, 2016, 81:79-92.
- [4] Kashkoush M. Knowledge discovery models for product design, assembly planning and manufacturing system synthesis[Dissertation]. Canada: University of Windsor, 2015.
- [5] Wenmin Li, Jiawei Han and Jian Pei, "CMAR: accurate and efficient classification based on multiple class-association rules," Proceedings 2001 IEEE International Conference on Data Mining, San Jose, CA, USA, 2001, pp. 369-376.