

Enterprise intelligent manufacturing data analysis technology based on big data analysis

Wenle Wang^{1,*}, Qilong Li¹, and Fuwen Zhu²

¹ School of Intelligent Manufacturing, Jiangsu Food & Pharmaceutical Science College, Huai'an 223001, China

² Huaigang Special Steel of Jiangsu Shagang Group, Huai'an 223001, China

Received: 11 December 2023 / Accepted: 18 March 2024

Abstract. The rise of big data has deeply influenced various industries, especially the intelligent manufacturing of enterprises. However, traditional data analysis methods are difficult to adapt to the storage and analysis of sea volume data in intelligent production. To address this issue, a method relying on big data analysis and cluster analysis is proposed to design data analysis techniques for enterprise intelligent manufacturing. The proposed improved algorithm is subjected to performance testing. The accuracy of this algorithm is 97%, which exceeds the comparison algorithm. The error is 6% and the running time is 5 s, both of which are below the comparison algorithm. The effectiveness of the enterprise intelligent manufacturing data analysis technology is tested. The experimental group completes orders in 4.1 weeks, 5.2 weeks, 3 weeks, 3.4 weeks, and 4.9 weeks, respectively, shorter than the control group. The product qualification rates for the experimental group are 92%, 93%, 95%, 92%, and 92%, respectively, which exceed the control group. In summary, the proposed enterprise intelligent manufacturing data analysis technology relying on big data and cluster analysis can better utilize data resources and information technology, improving the production efficiency and competitiveness of enterprises. It is hope that this research result can provide useful guidance and reference for the application and development of intelligent manufacturing data analysis technology in enterprises.

Keywords: Big data / intelligent manufacturing / data analysis / K-means / STK means

1 Introduction

Big data (BD) is a collection of data so large that it greatly exceeds the ability of traditional database software tools in terms of acquisition, storage, management, and analysis, and has four major characteristics: massive data scale, rapid data flow, diverse data types, and low value density. With the advances in technology and BD, intelligent manufacturing has become one of the important ways to upgrade the competitiveness and innovation ability [1]. In intelligent manufacturing, massive amounts of data are generated in various stages, including production, supply chain, sales, etc. [2]. These data contain valuable information and insights, but they also pose huge challenges for enterprises. How to extract valuable information from BD to support enterprise decision-making and optimize production processes need to be solved [3]. Data analysis can effectively mine the information contained in data. Data analysis refers to the use of appropriate statistical analysis methods to analyze a large amount of collected data, summarize, understand, and

digest them, in order to maximize the development of data functionality and leverage its potential. Data analysis is the process of conducting detailed research and summarizing data in order to extract useful information and form conclusions. Common data analysis methods include decision trees, support vector machines, and association rule mining. Among them, the data preparation of decision tree algorithms is relatively simple or even unnecessary, and decision trees can handle both data-driven and conventional attributes simultaneously. However, their processing of continuous data is difficult. Support Vector Machine is a generalized linear classifier that performs binary classification on data, capable of handling high-dimensional data and has strong generalization ability; However, there are drawbacks such as sensitivity to parameter and data scaling, as well as high computational complexity. Association rule mining can identify association rules between itemsets by analyzing the frequency and pattern of itemsets in the dataset, but it is not suitable for large-scale data. Traditional database technology has certain limitations in the ability to store, manage, and analyze data. Faced with the large amount of data generated in enterprise intelligent manufacturing, traditional databases are often unable to handle it effectively

* e-mail: wenle_wang123@outlook.com

[4,5]. However, influenced by BD technology, enterprises can gain efficient data analysis capabilities, thereby improving the management and decision-making abilities in quality control, process improvement, and service upgrading. To address this issue, big data analysis and clustering analysis are utilized to design data analysis techniques for enterprise intelligent manufacturing [6,7]. Data analysis and mining in enterprise intelligent manufacturing are carried out to achieve lean production and intelligent decision-making, and to provide scientific data support and decision-making reference for enterprises. The innovation of this study is to combine big data analysis and clustering analysis with other related technologies to study the analysis technology of enterprise intelligent manufacturing data, providing accurate and real-time decision support for in-depth analysis of enterprise production. The contribution of this study is to provide useful guidance and reference for the development and practice of intelligent manufacturing, promoting the widespread application of enterprise intelligent manufacturing in practical production. The first part introduces the application of BD analysis and the cluster analysis in recent years. The second part provides a detailed introduction to how to design enterprise intelligent manufacturing data analysis technology based on BD and clustering analysis. The third part analyzes the performance and verifies the practical application effect of enterprise intelligent manufacturing data analysis technology. The fourth part is a summary of the entire study.

2 Related work

With the advances in technology, BD is widely used. Big data analysis has become particularly important, covering various industries and fields, providing valuable insights and decision support for enterprises and organizations. In recent years, scholars have actively conducted research on the application of BD. To address whether a brand can sustain development in shopping malls, Du et al. proposed building a brand network map relying on consumption BD analysis. The results indicate that using big data can provide brand selection guidance for shopping malls [8]. Kulkarni et al. proposed a distributed data source based on fractional sparse fuzzy c-means algorithm and MapReduce framework to respond to the high computational complexity of traditional data analysis for BD analysis. The maximum accuracy is 90.6012%, which is better than the fractional sparse fuzzy c-means algorithm [9]. Chen et al. discussed a big data analysis relying on multiple linear regression models and artificial neural networks to address the impact of severe hydrological events on water quality and exacerbation of water pollution. This method takes into account the transport effects of compounds when conducting hazard analysis [10]. Thai et al. proposed a data collection and big data analysis based on the emergency medical service system process to build an operating system for the difficulty of reasonable allocation and

transportation of patients. This system can capture data for analysis and identification, and identify the most intelligent and cost-effective recommended hospitals [11]. The Sasikala team proposed a big data analysis based on crow search algorithm to optimize resource allocation in response to the complex structure of multimodal multimedia services in cloud platforms. The results show that the proposed algorithm can optimally allocate virtual machines to achieve the minimum response time [12].

In the process of data analysis, cluster analysis is often found. Through cluster analysis, the internal structure and patterns of data can be discovered, providing strong support for decision-making and promoting business optimization and development in various industries. Therefore, the research on clustering analysis has also received much attention. Agersted et al. proposed a towed instrument collection platform based on unsupervised clustering algorithm to address the difficulty in measuring biomass and abundance in the middle and upper layers of the ocean. This platform utilizes unsupervised clustering algorithm to solve the problem of low accuracy of traditional acoustic systems in detecting biomass and abundance in the middle and upper layers of the ocean. The results indicate that the platform can divide targets into different target groups to obtain correct backscatter information, thereby achieving more accurate biomass/abundance estimation [13]. Zhang et al. proposed a K-means clustering practical protocol model that integrates collaborative methods to address user privacy leakage during clustering. This model can ensure the security of user privacy information during clustering. Compared with traditional clustering models, the work efficiency of this model has not decreased [14]. Luo et al. proposed a contour density scanning clustering algorithm based on density clustering algorithm to address the low accuracy of fault identification in wind turbines. The results show that the algorithm can achieve automatic clustering. It exceeds the traditional algorithms in runtime and accuracy [15]. Cupak et al. designed a regression method based on non hierarchical clustering analysis to address the difficulty of low flow zoning in the upstream of the Vistula River basin. The results indicate that non hierarchical clustering analysis can be used for regional regression judgment. It is an effective tool for evaluating the low flow of rivers in southern Poland [16]. Tang et al. used Kalman filtering algorithm and K-means algorithm to cluster samples with low accuracy in predicting the estimated arrival time of aircraft taxiing. The results indicate that combining the cluster center sample trajectory sequence with the static path planned by the tower to predict aircraft taxiing has higher accuracy [17].

In summary, both big data analysis and cluster analysis have shown excellent performance in data analysis. However, there are still a few studies that combine the two. In response to the above issues, it is proposed to combine the two to study enterprise intelligent manufacturing data analysis technology, explore potential research value, and provide a new approach for the data analysis.

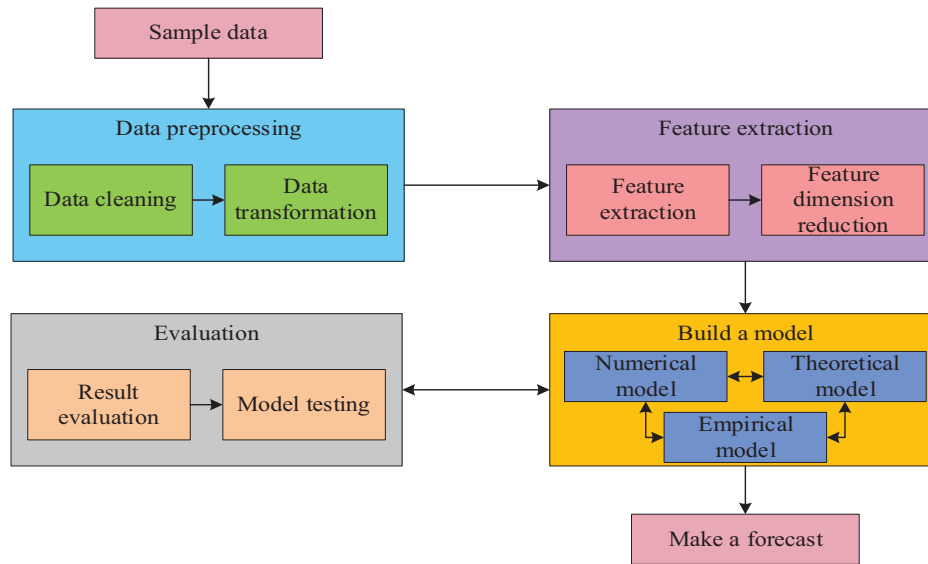


Fig. 1. Flow of data analysis.

3 Enterprise intelligent manufacturing data analysis based on BD analysis

With the advances in technology, manufacturing enterprises have entered the era of intelligent manufacturing. The traditional enterprise data analysis technology has limited storage and management capabilities for big data. Therefore, this chapter introduces the functional module of enterprise intelligent manufacturing data analysis relying on BD, as well as the improved enterprise intelligent manufacturing data analysis technology using clustering algorithms.

3.1 The design of analysis function module based on big data

In terms of production and manufacturing, the processes of mature enterprises have become stable after years of development. However, in the actual production process of manufacturing enterprises, there are often many uncertainties in parts, personnel, and tools that cannot be quantified or predicted, such as equipment failures caused by wear and decay of parts during processing, as well as quality changes caused by differences in parts provided by different batches and manufacturers. These uncertainties directly affect the judgments and decisions in production processes and scheduling. Therefore, data analysis is very important for manufacturing enterprises. The traditional data analysis process is shown in Figure 1 [18].

From Figure 1, in the analysis of enterprise manufacturing data, firstly, the sample data is preprocessed to eliminate abnormal data, error data, and abnormal data. A series of data transformations are carried out to obtain data that meets the requirements of feature extraction. Subsequently, in the feature extraction module, further data extraction is performed, laying the foundation for the data modeling. When the data feature is too large or there is redundancy, it is necessary to perform dimensionality

reduction processing to remove irrelevant features. In the modeling process, the model can be a numerical model obtained from characteristic calculations, an empirical model obtained from industry experts, or even a theoretical model obtained from theoretical calculations of product parameters. The purpose of evaluation is to verify the correctness of the model and add evaluation information to some data models that are numerically meaningless. To achieve intelligent manufacturing data analysis in enterprises, the BD analysis is introduced. BD analysis is a technology that mines hidden information and knowledge in large-scale data through multiple processes such as acquisition, storage, processing, analysis, and visualization, providing basis for various fields, optimizing operational efficiency, and improving intensification. The big data analysis cannot be separated from big data analysis platforms, which refer to infrastructure built to store, manage, and analyze data. It is based on high-performance data storage, processing, and computing capabilities, providing reliable and efficient data support by collecting, integrating, and processing large amounts of data. To achieve big data analysis for production and manufacturing, complete functions such as product quality evaluation, supplier part quality tracking, and enterprise equipment monitoring in the BD platform. A unified data information model needs to be established on the BD platform. The dataset needs to be constructed using cross database association retrieval, as shown in Figure 2.

In Figure 2, human computer interaction refers to the process in which a user sends instructions to a computer, which receives the instructions and executes the corresponding tasks. Through this approach, data can interact with the external environment. Real time decisions are made by the computer after the user issues instructions. In the enterprise intelligent manufacturing big data platform, the algorithm library is the core module of the platform. The efficiency of its algorithm directly affects the processing speed of big data platforms. The traditional serial algorithm in the algorithm library is changed to a

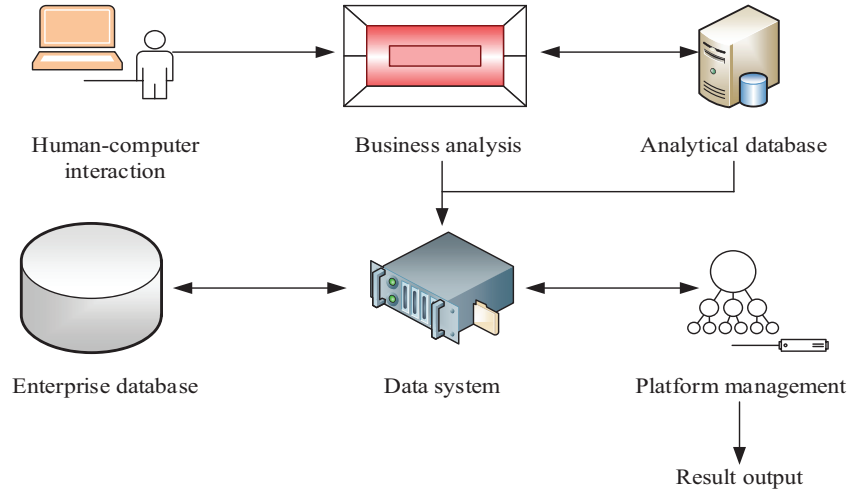


Fig. 2. Architecture diagram of enterprise intelligent manufacturing big data platform.

distributed parallel algorithm. K-Nearest Neighbor (KNN) is used to implement distributed parallel algorithms[19]. The core idea of KNN is as follows. If most of the K closest samples in the feature space belong to a certain category, then the sample also belongs to that category and has the characteristics of the samples in that category. The KNN is only relevant to very small adjacent samples when making category decisions. It is simple, easy to understand and implement, and does not require parameter estimation. Therefore, when it is used to build an enterprise intelligent manufacturing big data platform, the input data is first used to build a training dataset. The calculation is shown in equation (1).

$$T = (x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \quad (1)$$

In equation (1), $x_n \in X$ stands for the feature vector of the data. $y_n \in Y$ stands for the category of data. According to the rule of majority preference, the feature vector categories of the data are selected, as shown in equation (2).

$$y = \left(\arg \text{MAX}_{c_j} \sum_{x_i \in N_k(x)} I(y_i = c_j), i = 1, 2, \dots, N \right) \quad (2)$$

In equation (2), I represents the indicator function, which is a Boolean value. $N_k(x)$ represents the target point set. c_j represents the closest feature category. The distance calculation in KNN algorithm is shown in equation (3).

$$L_p(x_i, y_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^p \right)^{\frac{1}{p}} \quad (3)$$

In equation (3), L_p represents the distance to the target point $p(x_i, y_j)$. P stands for the adjacent points. l stands for the dimensionality of the data points. When $P=1$, it is the Manhattan distance. The calculation is shown in equation (4).

$$L_1(x_i, y_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}| \right) \quad (4)$$

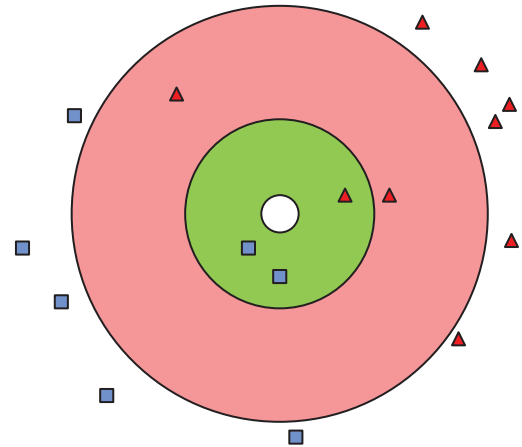


Fig. 3. Schematic representation of the KNN.

When $P=2$, it is the Euclidean distance. The calculation method is shown in equation (5).

$$L_2(x_i, y_j) = \left(\sum_{l=1}^n |x_i^{(l)} - x_j^{(l)}|^2 \right)^{\frac{1}{2}} \quad (5)$$

When $P=\infty$, the coordinate distance has the maximum. The calculation method is shown in equation (6).

$$L_\infty(x_i, y_j) = \text{MAX}_l |x_i^{(l)} - x_j^{(l)}|^2 \quad (6)$$

The calculation result of the distance formula is used as the nearest point sorting parameter. However, due to the error of the measurement data, in order to reduce the uncertainty, the study used the least squares method to process the data. By selecting and sorting the top K points, the categories of these points are statistically counted. The counting results of categories are sorted. The one to two categories with the highest count are the candidate results for classification, as shown in Figure 3.

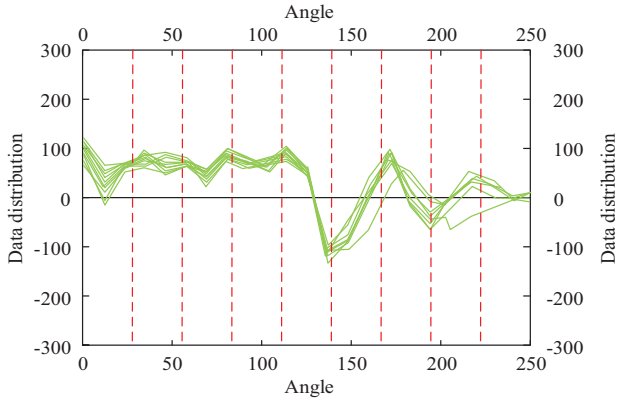


Fig. 4. Schematic diagram of the angle-window feature extraction method.

In [Figure 3](#), when K is 3, the three closest points to the center are selected as reference for classification. The nearest point includes 1 red triangle and 2 blue squares. At this point, the situation of points outside the black circle is not considered. Similarly, when K is 5, three red triangles and two blue squares are selected. The size of the K value is of great significance for KNN classification results. The selection of K is a crucial step in the enterprise intelligent manufacturing big data platform, greatly influenced the classification efficiency and accuracy.

3.2 Enterprise intelligent manufacturing data analysis technology based on STKmeans algorithm

When using enterprise intelligent manufacturing big data platforms for data analysis, there is a large-scale imbalance in enterprise data due to the individual needs or production processes of different enterprises. Conventional intelligent manufacturing big data analysis techniques are not accurate in analyzing such data. To this end, a feature extraction method using angle windowing and clustering analysis are proposed to optimize the data analysis technology of enterprise intelligent manufacturing. In each angle window, the mean in the angle window is calculated. The mean is used to represent the information of all points in the original window. The window number represents all angle information in the original window. This not only greatly simplifies the data, but also characterizes the feature information. Then the length of the data processed by the angle window also becomes consistent. Angle windowing feature extraction is shown in [Figure 4](#).

In [Figure 4](#), the data processed by the angle windowing feature extraction method includes extracting feature information from the data, which has a certain degree of rotation invariance. Combined with clustering analysis, the enterprise intelligent manufacturing data analysis technology is further optimized. Cluster analysis or clustering is the static classification method that divides similar objects into different groups or more subsets, so that member objects in the same subset have similar attributes. Cluster analysis itself is not a specific algorithm, but a general task that needs to be solved. It can be achieved through different algorithms, which have

significant differences in understanding the composition of clusters and how to effectively find them. The general concept of a cluster includes groups with smaller distances between cluster members, dense areas of data space, intervals, or specific statistical distributions. Therefore, clustering can be interpreted as a multi-objective optimization problem. Clustering can be achieved by calculating the similarity or distance between samples. There are five major categories in cluster analysis. The specific classification and representative algorithms are shown in [Figure 5](#).

In [Figure 5](#), K-means clustering algorithm (K-means), as one of the classic machine learning algorithms, is widely used in data clustering analysis. Therefore, it is used to improve enterprise intelligent manufacturing big data analysis technology [20]. During the classification process using the K-means algorithm, a similarity matrix is constructed based on data similarity. The expression is shown in [equation \(7\)](#).

$$A = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1j} \\ \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} \end{pmatrix} \quad (7)$$

In [equation \(7\)](#), i represents the i data objects. j represents that each data object has j features. When calculating the K-means, the clustering clusters K need to be set first. The cluster center calculation method is shown in [equation \(8\)](#).

$$u = \frac{1}{m} \sum x \quad (8)$$

In [equation \(8\)](#), u represents the cluster center vector. S represents a dataset. m stands for the quantity in S dataset. The K-means adopts Markov distance to calculate the distance between data objects. The mathematical expressions are shown in [equations \(9\), \(10\), \(11\), and \(12\)](#).

$$\mu = E\{X\} = X^{T_1} \left(\frac{1}{g}\right)_{g \times 1} \quad (9)$$

In [equation \(9\)](#), μ represents the mean. $\left(\frac{1}{g}\right)_{g \times 1}$ represents a g -dimensional column vector whose elements are all $\frac{1}{g}$. X represents the data sample matrix. T_1 represents transposition.

$$G = \frac{1}{g} X^{T_1} X \quad (10)$$

In [equation \(10\)](#), G represents the autocorrelation matrix.

$$\Sigma = E\{(X - \mu)^{T_1}\} = \frac{1}{g} X^{T_1} X - \mu \mu^{T_1} \quad (11)$$

In [equation \(11\)](#), Σ represents the covariance matrix.

$$D_4^2(X_i - X) = (x_i - \mu)^{T_1} \sum_i^{-1} (x_i - \mu) \quad (12)$$

In [equation \(12\)](#), D_4^2 represents the Markov distance from sample X_i to sample population X . Markov distance is calculated based on the population of data samples, which

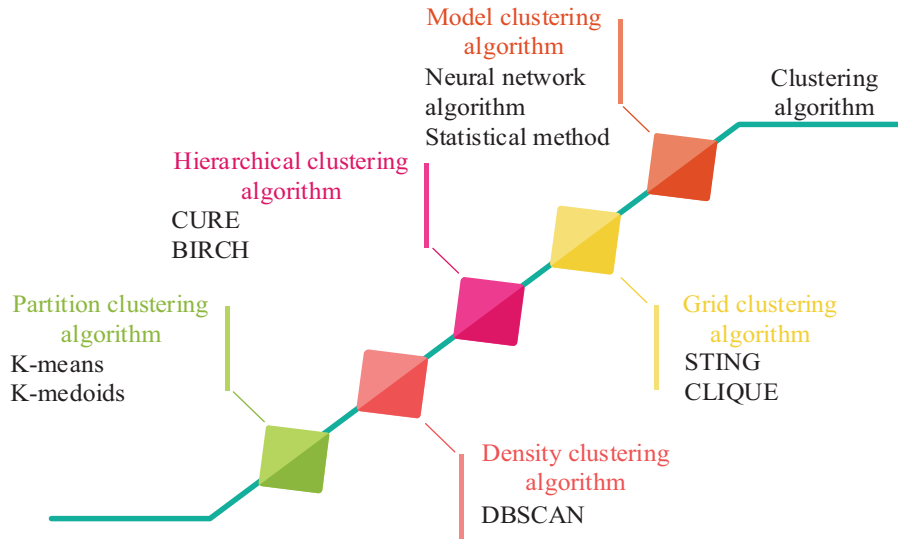


Fig. 5. Cluster algorithm classification and generation algorithm.

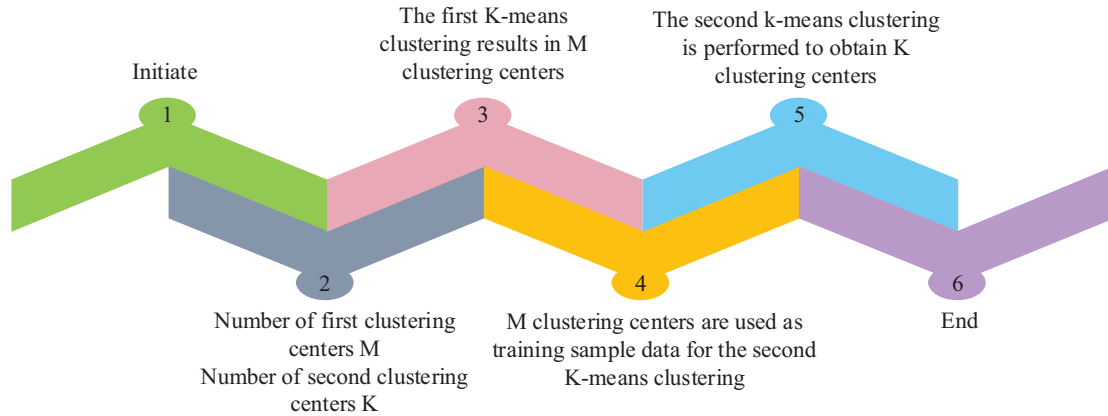


Fig. 6. Flow chart of the STKmeans.

can enhance the accuracy. The K-means is a commonly used clustering analysis method, but it also has some drawbacks. It is very sensitive to the selection of initial cluster centers. Different initial cluster centers have different results. This means that multiple experiments need to be conducted and the optimal clustering results need to be selected when applying the K-means. The clusters are determined by the user. Excessive human interference can cause differences in the results of each clustering. Moreover, users do not know which category the corresponding data is best classified into. There is no guarantee that the result after each clustering is the best clustering result. To address this issue, a second time K-means (STKmeans) is proposed to optimize the K-means algorithm in enterprise intelligent manufacturing data analysis technology. The STKmeans is a data analysis idea based on the K-means. It is a variant of the traditional K-means. The contour coefficient is used to determine the optimal clusters. The contour coefficient evaluates the quality of clustering by calculating the similarity between each sample and other samples within the cluster to which it belongs, as well as the similarity with the nearest samples from other clusters. The goal of the

STKmeans algorithm is to maximize the contour coefficients to obtain better clustering results. The workflow of this algorithm is shown in Figure 6.

In Figure 6, it can be applied to various datasets, including numerical and categorical data. Therefore, this algorithm can be used to improve the K-means algorithm, thereby optimizing the data analysis technology of enterprise intelligent manufacturing. In conventional enterprise intelligent manufacturing data analysis technology, the inaccurate results caused by data imbalance can be solved.

4 Comparison of improved algorithm performance and the application effect analysis of intelligent manufacturing data analysis technology

To prove the performance of the proposed STKmeans relying on the big data analysis and the effectiveness of the improved enterprise intelligent manufacturing data analysis technology, comparative experiments are conducted in

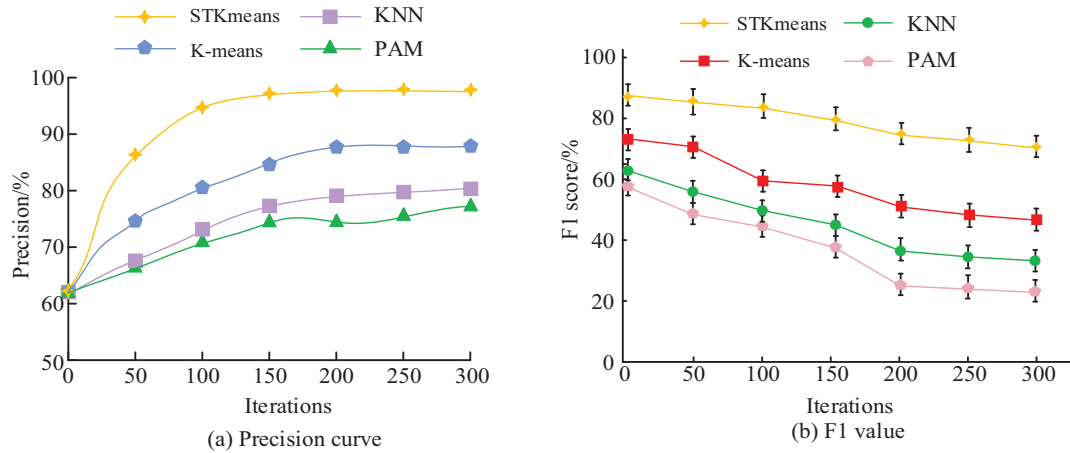


Fig. 7. Algorithm accuracy with F1 values.

the study. Commonly used clustering analysis algorithms are selected. Based on these algorithms combined with big data analysis, enterprise intelligent manufacturing data analysis technology is designed as the control group (CG) in the comparative experiment.

4.1 Performance comparison of proposed improved algorithm

To verify the performance, comparative experiments are designed. The KNN algorithm, K-means algorithm, and Partitioning Around Medoid (PAM) are used as the comparison algorithms. The Iris dataset is selected as the experimental dataset. The precision, F1 value, runtime, error, and accuracy of the algorithm are used as evaluation indicators. The precision and F1 value test results of KNN algorithm, K-means algorithm, PAM algorithm, and STKmeans algorithm are shown in Figure 7.

In Figure 7a, after 300 iterations, the accuracy of the STKmeans algorithm, K-means algorithm, KNN algorithm, and PAM algorithm are 97%, 88%, 80%, and 77%, respectively. The STKmeans algorithm has the highest accuracy. At around 100 iterations, the accuracy remains stable and the convergence speed is the fastest. In Figure 7b, after 300 iterations, the F1 values of the four algorithms are 70%, 45%, 32%, and 21%. The F1 of the STKmeans exceeds the comparison algorithm during the experimental process. The decrease in F1 value is below the comparison methods, indicating the optimal performance of the algorithm. The errors and runtime of STKmeans algorithm, K-means algorithm, KNN algorithm, and PAM algorithm are counted. The results are shown in Figure 8.

In Figure 8, the error and running time of the STKmeans are 6% and 5s, respectively. The K-means are 12% and 11s respectively. The KNN algorithm is 19% and 9s respectively. The PAM is 21% and 11s respectively. The performance of the STKmeans algorithm is superior to that of the comparison method. To verify the universality of the proposed method, comparative experiments are designed on three datasets, Wine, Mnist, and Iris. Table 1 displays the test results.

In Table 1, the Wine dataset is used for comparative experiments. When the training samples are 75 and 150, the accuracy of the STKmeans algorithm is 94.61% and 91.72%, respectively. Then the Mnist dataset is used for comparative experiments. When the training samples are 75 and 150, the accuracy of the STKmeans algorithm is 93.66% and 90.13%, respectively. Finally, the Iris dataset is used for comparative experiments. When the training samples are 75 and 150, the accuracy of the STKmeans algorithm is 95.58% and 91.75%, respectively. The STKmeans algorithm has the highest accuracy on these three datasets. As the training samples increase, the recognition rate is minimally affected, indicating that the performance exceeds the comparison algorithm.

4.2 The practical application effect of the proposed enterprise intelligent manufacturing data analysis technology

To verify the practical application effect of the enterprise intelligent manufacturing data analysis technology based on BD analysis and STKmeans algorithm proposed in the study, a certain engine manufacturing enterprise is selected. The production process is analyzed to improve the product qualification rate of the enterprise. Two groups are randomly selected for comparative experiments. A group uses the enterprise intelligent manufacturing data analysis technology proposed in the study, namely the experimental group (EG). The other one, without any changes, is called the CG. The experiment uses the Davies-Bouldin index (DB), product qualification rate, professional score, etc. as evaluation indicators. The experimental results of the DB index are shown in Figure 9.

From Figure 9, the data analysis results between the EG and the CG shows that the division of enterprise intelligent manufacturing data analysis technology proposed in the study is relatively clear. The compactness and dispersion of the data analysis structure are more reasonable. The DB index of the EG is below the CG. The smaller the DB index, the better the data analysis results. The proposed enterprise intelligent manufacturing

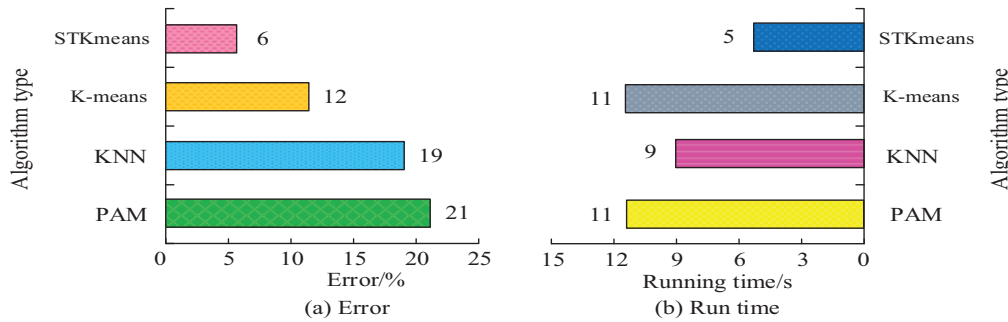


Fig. 8. Error and running time.

Table 1. Accuracy on different datasets.

Number of training samples	Dataset Type	PAM	KNN	K-means	STKmeans
75	Wine	66.74%	75.52%	88.49%	94.61%
	Mnist	64.83%	73.79%	85.84%	93.66%
	Iris	66.12%	74.95%	82.97%	95.58%
150	Wine	59.74%	68.61%	84.59%	91.72%
	Mnist	68.51%	67.42%	81.84%	90.13%
	Iris	69.71%	67.64%	81.04%	91.75%

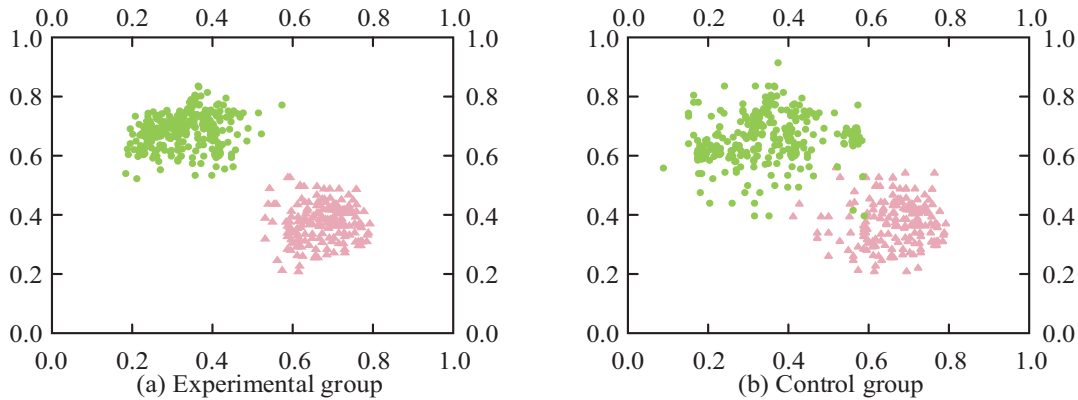


Fig. 9. The DB index of the data analysis technique.

data analysis technology is superior to traditional data analysis technology. The EG and CG are assigned 5 orders with the same quantity, respectively. The order completion time and product qualification rate results are shown in Figure 10.

In Figure 10, the completion time of the five order experimental groups is 4.1 weeks, 5.2 weeks, 3 weeks, 3.4 weeks, and 4.9 weeks, respectively. The completion time for the CG is 5 weeks, 6.5 weeks, 3.6 weeks, 4 weeks, and 6.4 weeks, respectively. The product qualification rates of the five order EG are 92%, 93%, 95%, 92%, and 92%, respectively. The product qualification rates of the CG are 82%, 75%, 88%, 83%, and 87%, respectively. The EG completes the same orders in a shorter time, which has a higher product qualification rate. It is better than the CG. The proposed enterprise intelligent manufacturing data analysis technology has a better application effect. The management personnel are invited from the experimental

enterprise to rate the data analysis results of the experimental and control groups. The total score is ten. Table 2 displays the test results.

According to Table 2, the average scores of enterprise management personnel on the clarity, guidance, rationality, and predictability of the EG data analysis results are 9.3 points, 9.3 points, 9.4 points, and 9.2 points, respectively. The average scores of enterprise management personnel on the clarity, guidance, rationality, and predictability of the CG data analysis results are 8.0 points, 8.2 points, 8.2 points, and 8.0 points, respectively. The scores of the EG exceed the CG. The proposed enterprise intelligent manufacturing data analysis technology has better performance than traditional data analysis technology. Based on all the above experimental results, the proposed STKmeans algorithm based on BD analysis has superior performance. The data analysis technology for enterprise intelligent manufacturing is feasible.

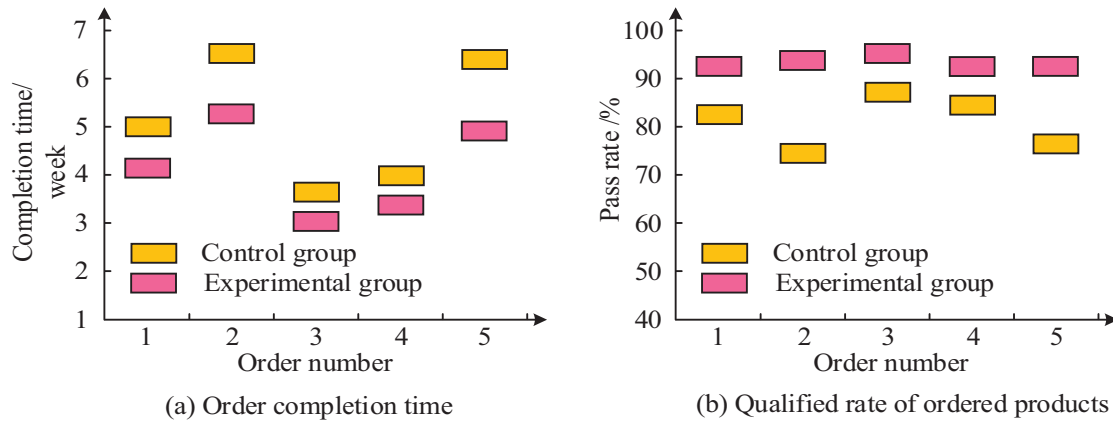


Fig. 10. Order completion time and pass rate.

Table 2. Scoring of enterprise managers.

Serial number	Index type	Executive 1		Executive 2		Executive 3	
		Experimental group	Control group	Experimental group	Control group	Experimental group	Control group
1	Articulation	9.3	8.1	9.0	8.0	9.6	7.8
2	Instructive	9.5	8.4	9.5	8.4	9.0	7.9
3	Rationality	9.1	8.0	9.6	8.3	9.4	8.4
4	Predictability	9.0	7.7	9.1	8.1	9.5	8.0

5 Conclusion

In the information age, BD has a profound impact on various industries, especially in the enterprise intelligent manufacturing. Traditional data analysis is difficult to store and analyze a large amount of data in the intelligent manufacturing process of enterprises. To address this issue, based on BD analysis and combined with cluster analysis, enterprise intelligent manufacturing data analysis technology is designed. To verify the performance, comparative experiments are conducted. The accuracy is 97%, and the F1 value is 70%, both of which exceed the comparison algorithm. The error and running time are 6% and 5s respectively, which are below the comparison algorithm. At the same time, the improved algorithm has the highest accuracy on these three datasets. Subsequently, the effectiveness of the proposed enterprise intelligent manufacturing data analysis technology is tested. The results show that the DB value of the EG is lower than that of the CG. The completion time of orders in the EG is shorter than the CG. The product qualification rate in the EG is higher than that in the CG. The average scores of enterprise management personnel on the clarity, guidance, rationality, and predictability of data analysis results in the EG are 9.3 points, 9.3 points, 9.4 points, and 9.2 points, respectively, which exceed those in the CG. In summary, the proposed data analysis technology for enterprise

intelligent manufacturing based on big data analysis and cluster analysis can analyze and mine data in enterprise intelligent manufacturing, providing scientific data support and decision-making reference for enterprises. However, during the experimental process, the study only analyzes the production process of the enterprise and does not fully understand the data of all parties in the enterprise. In future research, comprehensive real-time data research will be conducted on intelligent manufacturing enterprises.

Conflict of Interest

The authors have no relevant financial or non-financial interests to disclose.

Data availability statement

All data generated or analyzed during this study are included in this article.

Author contribution statement

Wenle Wang analyzed the data and Qilong Li helped with the constructive discussion. Wenle Wang, Qilong Li, and Fuwen Zhu made great contributions to manuscript preparation. All authors read and approved the final manuscript.

References

1. B. Wang, F. Tao, X. Fang, C. Liu, Y. Liu, T. Freiheit, Smart manufacturing and intelligent manufacturing: a comparative review, *Engineering* **7**, 738–757 (2021)
2. Y. Fu, Y. Hou, Z. Wang, X. Wu, K. Gao, L. Wang, Distributed scheduling problems in intelligent manufacturing systems, *Tsinghua Sci. Technol.* **26**, 625–645 (2021)
3. H. Dai, H. Wang, G. Xu, J. Wan, M. Lmran, Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies, *Enterprise Inform. Syst.* **14**, 1279–1303 (2020)
4. C. Zhang, G. Zhou, H. Li, C. Yan, Manufacturing blockchain of things for the configuration of a data-and knowledge-driven digital twin manufacturing cell, *IEEE Internet Things J.* **7**, 11884–11894 (2020)
5. M. Barma, U. Modibbo, Multiobjective mathematical optimization model for municipal solid waste management with economic analysis of reuse/recycling recovered waste materials, *J. Comput. Cogn. Eng.* **1**, 122–137 (2022)
6. L. Li, J. Zhang, Research and analysis of an enterprise E-commerce marketing system under the big data environment, *J. Organizational End User Comput.* **33**, 1–19 (2021)
7. R. Li, J. Rao, L. Wan, The digital economy, enterprise digital transformation, and enterprise innovation, *Manag. Decis. Econ.* **43**, 2875–2886 (2022)
8. G. Du, Y. Lin, Brand connection and entry in the shopping mall ecological chain: evidence from consumer behavior big data analysis based on two-sided markets, *J. Cleaner Product.* **364**, 1–12 (2022)
9. O. Kulkarni, S. Jena, V. Sankar, MapReduce framework based big data clustering using fractional integrated sparse fuzzy C means algorithm, *IET Image Process.* **14**, 2719–2727 (2020)
10. Z. Chen, Y. Meng, R. Wang, T. Chen, Water quality big data analysis of the river basin with artificial intelligence ADV monitoring, *Membrane Water Treat.* **13**, 219–225 (2022)
11. H. Thai, J. Huh, Optimizing patient transportation by applying cloud computing and big data analysis, *J. Supercomput.* **78**, 18061–18090 (2022)
12. S. Sasikala, S. Gomathi, V. Geetha, L. Murali, A proposed framework for cloud-aware multimodal multimedia big data analysis toward optimal resource allocation, *Comput. J.* **64**, 880–894 (2021)
13. M. Agersted, K. Babak, Y. Liu, W. Melle, T. Klevjer, Application of an unsupervised clustering algorithm on in situ broadband acoustic data to identify different mesopelagic target types, *ICES J. Mar. Sci.* **78**, 2907–2921 (2021)
14. E. Zhang, H. Li, Y. Huang, S. Hong, L. Zhao, C. Ji, Practical multi-party private collaborative k-means clustering, *Neurocomputing* **467**, 256–265 (2022)
15. D. Luo, H. Liu, E. Qi, Recognition and labeling of faults in wind turbines with a density-based clustering algorithm, *Data Technol. Appl.* **55**, 841–868 (2021)
16. A. Cupak, G. Kaczor, Regionalization of low flow for chosen catchments of the upper Vistula river basin using non-hierarchical cluster analysis, *Idojaras* **126**, 27–45 (2022)
17. X. Tang, X. Ji, J. Liu, Predicting aircraft taxiing estimated time of arrival by cluster analysis, *IET Intell. Trans. Syst.* **16**, 252–262 (2022)
18. A. Brintrup, J. Pak, D. Ratiney, T. Pearce, P. Wichmann, P. Woodall, Supply chain data analytics for predicting supplier disruptions: a case study in complex asset manufacturing, *Int. J. Product. Res.* **58**, 3330–3341 (2020)
19. H. Henderi, T. Wahyuningsih, E. Rahwanto, Comparison of Min-Max normalization and Z-score normalization in the K-nearest neighbor (kNN) algorithm to test the accuracy of types of breast cancer, *Int. J. Inform. Inform. Syst.* **4**, 13–20 (2021)
20. P. Anitha, M. Patil, RFM model for customer purchase behavior using K-means algorithm, *J. King Saud Univ. Comput. Inf. Sci.* **34**, 1785–1792 (2022)

Cite this article as: Wenle Wang, Qilong Li, Fuwen Zhu, Enterprise intelligent manufacturing data analysis technology based on big data analysis, *Int. J. Simul. Multidisci. Des. Optim.* **15**, 5 (2024)