

Research on Parallel Distributed Clustering Algorithm Applied to Milling Parameter Optimization

Xudong Wei

Harbin University of Science and Technology

Qingzhen Sun (✉ weixudong686@163.com)

Harbin University of Science and Technology

Xianli Liu

Harbin University of Science and Technology

Caixu Yue

Harbin University of Science and Technology

Steven Y. Liang

Georgia Institute of Technology

Lihui Wang

KTH Royal Institute of Technology

Research Article

Keywords: Big data, Data mining, Distributed clustering, T.K-means algorithm, MapReduce framework, Milling parameters Optimization

Posted Date: August 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-808329/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Research on parallel distributed clustering algorithm applied to milling parameter optimization

Xudong Wei^a Qingzhen Sun^{a,*} Xianli Liu^a Caixu Yue^a Steven Y. Liang^b Lihui Wang^c

^a Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China;

^b George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, 30332, USA;

^c KTH Royal Institute of Technology, Stockholm, 25175, Sweden

Abstract

In the big data era, traditional data mining technology cannot meet the requirements of massive data processing with the background of intelligent manufacturing. Aiming at insufficient computing power and low efficiency in mining process, this paper proposes a improved K-means clustering algorithm based on the concept of distributed clustering in cloud computing environment. The improved algorithm (T.K-means) is combined with MapReduce computing framework of Hadoop platform to realize parallel computing, so as to perform processing tasks of massive data. In order to verify the practical performance of T.K-means algorithm, taking machining data of milling Ti-6Al-4V alloy as the mining object. The mapping relationship among milling parameters, surface roughness and material removal rate is mined, and the optimized value for milling parameters are obtained. The results show that T.K-means algorithm can be used to mine the optimal milling parameters, so that the best surface roughness can be obtained in milling Ti-6Al-4V titanium alloy.

Keywords Big data; Data mining; Distributed clustering; T.K-means algorithm; MapReduce framework; Milling parameters Optimization

Declarations

Funding

This research is supported by:

(1) International (regional) cooperation and exchange program of National Natural Science Foundation of China under Grant No. 51720105009.

(2) National key R&D plan. Network collaborative manufacturing and smart factory special project: “Complex Tool Monitoring and Full Life Cycle Intelligent Management and Control Technology” under Grant No. 2019YFB1704800.

Conflicts of interest/Competing interests

The authors declare that there is no conflict of interests regarding the publication of this article.

Availability of data and material

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability (Not applicable)

Authors' contributions

Xudong Wei contributed to the conception of the study; Xianli Liu contributed significantly to analysis and manuscript preparation; Xudong Wei, Caixu Yue performed the experiment and data analyses; Steven Y. Liang, Lihui Wang helped perform the analysis with constructive discussions.

Ethics approval

The content studied in this article belongs to the field of metal processing, does not involve humans and animals. This article strictly follows the accepted principles of ethical and professional conduct.

Consent to participate

My co-authors and I would like to opt in to In Review.

Consent for publication

I agree with the Copyright Transfer Statement.

1 Introduction

In recent years, it has become a hot issue to optimize milling parameters considering multi-index. Surface roughness and material removal rate of milling titanium alloy are the key concerns. In the past decades, many researchers have studied the multi-objective optimization in cutting process, and the research methods can be divided into theoretical-experimental and data analysis method.

Optimizing cutting parameters of machine tools by theory and experiment method has been gradually applied. Chen et al. [1] proposed an energy consumption model of milling operation considering both the electrical energy consumption and embodied energy consumption of cutting tools and cutting fluids is. And a multi-objective optimization model is established to achieve maximizing energy efficiency and machining efficiency. Warsi et al. [2] combined individual responses to formulate a multi-objective function using gray relational analysis augmented with analytic hierarchy process. Multi-objective function was optimized using regression analysis and response surface optimization to determine optimum machining parameters during high-speed turning of Al 6061 T6 alloy. Rao et al. [3] used teaching learning based optimization (TLBO) technique coupled with 3D finite element method (FEM) simulation to optimize the cutting parameters such as cutting speed, feed and depth of cut. Sharma et al. [4] used response surface methodology to optimize the spindle speed, cutting depth and feed speed with material removal rate

and end-friction activity immobility as performance indexes.

Wang et al. [5] established a process dimension tree and a complete dimension model, which represented the process dimension tree expresses the machining sequences of each surface and the process datum of each process dimension. It simplifies the change or optimisation of process route and indicates the influence of the dimensioning datum of a blank on rough machining allowances and dimensional accuracy. Ma et al. [6] proposed an effective approach for parameters optimization of complex end milling process based on virtual machining for actual tool path generation and feedrate scheduling based on material removal rate. A machining simulator based on octree and instantaneous force model is developed to evaluate feasibility of the given numerical control (NC) programs. The optimization process is controlled by the efficient global optimization method to find global optimal solution with fewer simulations and less computation time. The milling parameters optimization based on data analysis method has been widely used. Diyaley et al. [7] studied six metaheuristic algorithms, in the form of artificial bee colony optimization, ant colony optimization, particle swarm optimization, differential evolution, firefly algorithm and teaching-learning-based optimization techniques applied for parametric optimization of a multi-pass face milling process. Nguyen et al. [8] used a type of neural network entitled the radius basic function (RBF) to render the relationships between milling inputs and performances measured. The adaptive simulated annealing (ASA) algorithm was applied to obtain the optimal values. Zhang et al. [9] presented a new analytical energy consumption model and the related optimization of cutting parameters to reduce the energy consumption of micro milling process. With the developed prediction model of energy consumption, a hybrid cuckoo search and grey wolf algorithm is used to determine the optimum cutting parameters for minimizing the total energy consumption. Xiao et al. [10] established a multi-objective optimization model with milling speed, feed per tooth and spindle speed as optimization variables, and a weight coefficient is introduced to facilitate the solution to convert it into a single objective optimization model. The combinatorial optimization algorithm is compared with the NSGA-II, the results show that the combinatorial optimization algorithm has better performance in solving speed and optimization accuracy. Deng et al. [11] proposed a data-driven method for machining parameter planning by learning the parameter planning knowledge from the high-quality historical processing files. Graph neural networks (GNN) are established to learn the planning models for machining parameters. The proposed method provides an end-to-end strategy for constructing machining parameter planning models thus human interactions can be greatly reduced and the performance of the models are able to be improved as the increase in historical processing files.

With the exponential growth of production data stored in factories, the performance of traditional algorithms is increasingly unable to meet the demand. Therefore, it is urgent to introduce new methods to deal with large-scale data

calculation. Distributed framework is an effective way to solve the problem of massive data storage and computing, which makes big data processing possible [12, 13]. Hadoop integrates Hdfs distributed storage [14] and MapReduce distributed computing framework [15], which originated from Apache Nutch project, started in 2002, and is one of the subprojects of Apache Lucene [16]. At present, the popular big data analysis frameworks are based on Hadoop, and different functions are added to realize function expansion. The representative ones are FusionInsight Hadoop, Hortonworks HDP [17], Cloudera CDH [18] and Transwarp TDP [19], etc. In this paper, the Cloudera CDH platform is built for data analysis and mining, which can make various mining algorithms, management and control tools, data query tools developed in the form of plug-ins, and facilitate the integration of functional frameworks.

In view of the large amount of running data in mass and single production for titanium alloy, this paper proposes a distributed clustering method based on cloud computing platform Hadoop [20], which explores the relationship between workpiece surface roughness, material removal rate and cutting parameters to optimize cutting parameters. The limited computing resources are integrated in Hadoop data analysis platform, and the massive data mining ability is improved. K-means clustering algorithm is improved by introducing the concept of distributed clustering. A new clustering algorithm T.K-means is formed, which improves clustering efficiency and readability of mining results. The algorithm can be used to mine the optimal combination of cutting parameters in historical processing data. Therefore, it keep a compromise between workpiece surface roughness and cutting efficiency, which has guiding significance for actual production and processing.

2 A new algorithm based on distributed clustering and parallel computing

2.1 Partition-based Clustering method and K-means algorithm

Partition-based clustering is to divide an individual into multiple regions, and each region represents a group that meets specific conditions or characteristics [21]. K-means algorithm is a classical clustering algorithm, and its data mining process is mainly divided into the following four steps.

(1) K data points are randomly selected as the initial clustering center.

(2) All the sample data are assigned to the nearest class to the K center points, and the new clustering center (average value of sample points within the class) is recalculated.

(3) The distances from the data point to the new clustering center are calculated. If the total square sum of distance decreases, it means that the clustering center has changed, and the data points need to be reassigned to the new cluster.

The least square error D is defined as:

$$D = \sum_{i=1}^K \sum_{x \in C_i} |x - x_i|^2 \quad (1)$$

Where K is the total number of classes, x_i is the average value of class C_i .

(4) This process will continue until iteration termination or:

$$|D_{i+1} - D_i| < \varepsilon \quad (2)$$

Where ε is minimum, D_{i+1} and D_i are standard measure function values of two adjacent iterations.

Partition-based clustering aggregates similar datasets, which including performance index A_i and cutting parameter B_i . If a certain amount of B_i corresponds to A_i , it indicates that A_i will occur when B_i occurs. Therefore, it is an important method to find the cutting parameters corresponding to the optimal performance index in this paper.

2.2 K-means algorithm and improved method

The traditional K-means algorithm has two shortcomings when dealing with massive data. On the one hand, the number of packets cannot be known in advance by K-means algorithm, which leads to inaccurate grouping [22]. When the algorithm runs, it will produce a large number of suboptimum performance index, which are meaningless to the mining target. On the other hand, it is designed to run on a single node with too little memory and insufficient computing power. Its limited data processing ability makes it difficult to complete the task of mining massive data.

Aiming at the above problem, B-K-means algorithm is obtained by splitting and merging operations. It has iterative self-organizing ability and realizes automatic identification of classification number. In addition, distributed clustering method is used to reduce irrelevant dataset by getting the mapping relationship between optimized cutting parameters and optimal performance index. The above algorithm is combined with MapRduce framework in Hadoop platform to realize parallel computing. Finally, a distributed clustering algorithm T-K-means is formed, which will be applied to cutting parameters optimization for milling titanium alloy.

B-K-means algorithm is improved on the basis of K-means algorithm, which can realize dynamic clustering or iterative self-organizing data analysis. By splitting and merging operations, it can identify the essential attributes of data in the clustering process. The following parameters of B-K-means algorithm need to be determined as: K —preset clustering number, θ_s —the maximum sample standard deviation for each class, θ_c —the minimum distance among clustering center, θ_N —the minimum sample number for all classes, L —the maximum logarithm of classes that can be merged in one iteration, I —the maximum iteration times.

B-K-means algorithm process is mainly divided into the following five steps.

(1) k arbitrary samples are randomly selected from all data as initial clustering centers ($z_1, z_2, z_3, \dots, z_k$). Because the cluster number changes dynamically, there are differences between k and K . Above six indexes are set, and all data samples are distributed to the classes closest to k center points.

(2) The cluster center is corrected by calculating sample mean. The average distance H_j and the total average distance \bar{H} among all samples and corresponding cluster center of each class are calculated.

(3) Splitting or merging the last clustering results. Splitting can solve the problem that classification number is too small or sample number in a class is too large, so as to find more cluster centers. If the data samples are too close and has a

strong correlation with each other, there can be obvious differences among classes by merging.

Splitting operation:

The standard deviation vector σ_j of sample distance in each cluster and its maximum component σ_{\max} are calculated. If $s_{\max} > \theta_s$, $H_j > \bar{H}$ and $H_j > 2(q_N + 1)$, that is, the sample number in C_j exceeds the specified value by more than one time or $k \leq K/2$. Then z_j is split into two new cluster center z_j^+ and z_j^- , as shown in formula (3).

$$\begin{cases} z_j^+ = z_j + \rho\sigma_{\max} \\ z_j^- = z_j - \rho\sigma_{\max} \end{cases}, 0 < \rho < 1 \quad (3)$$

Merging operation:

The distance of all cluster centers are calculated by formula (4).

$$H_{ij} = \|z_i - z_j\|, \begin{cases} i = 1, 2, K, k-1, \\ j = i+1, K, k \end{cases} \quad (4)$$

If $H_j < \theta_c$, the values of H_{ij} are arranged in ascending order according to the minimum distance, and the two cluster center with minimum distance are merged to obtain a new center as shown in formula (5):

$$z_i^* = \frac{N_i z_i + N_j z_j}{N_i + N_j} \quad (5)$$

Where two cluster center are weighted by sample number of each class, and k is reduced by 1 for each merging.

(4) Repeat iterative operation and continue the above process until the clustering result converges.

Through splitting and merging operation of B-K-means algorithm, the automatic clustering is realized, which provides condition for the subsequent accurate search of optimized datasets in each class.

2.3 Hadoop platform and MapReduce parallel computing framework

Parallel computing is the main way to solve the problem of big data analysis and processing. As shown in Fig.1, Hadoop system mainly includes two parts: HDFS distributed storage and MapReduce parallel computing [23]. Hadoop platform has high reliability, efficiency, fault tolerance and scalability.

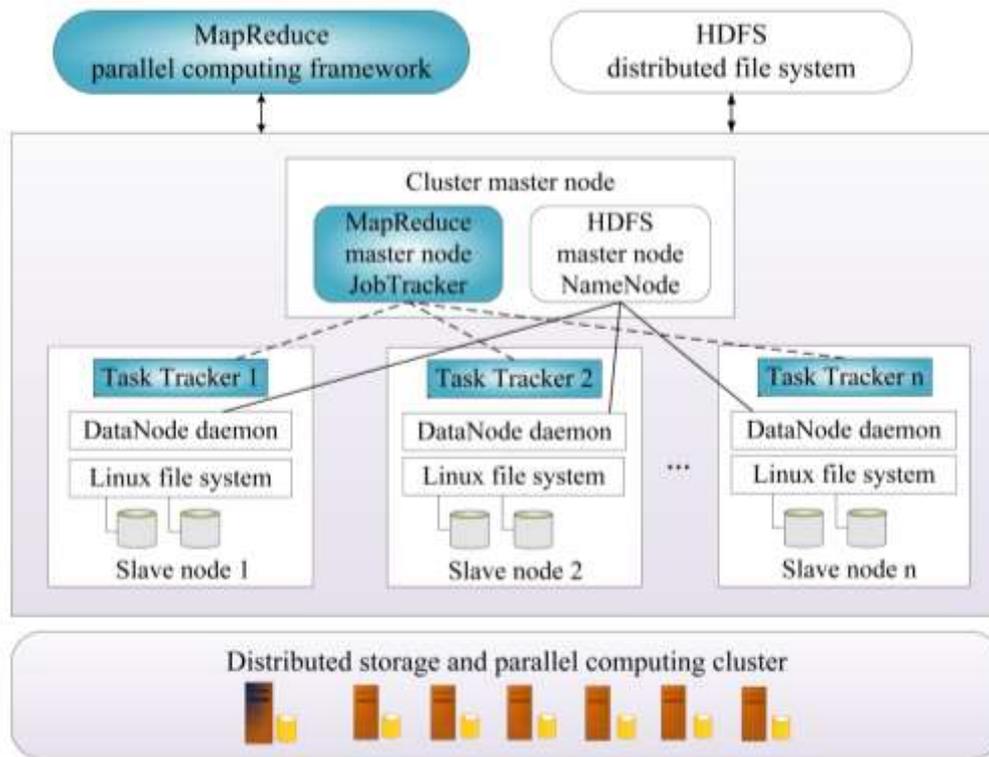


Fig.1 Framework diagram of Hadoop

Distributed file storage system HDFS is a highly fault-tolerant system, which can detect and deal with hardware failure and run on low-cost universal hardware. HDFS builds the local files on each node into a logical whole file system, which improves data storage capability, thus providing an effective solution for the storage of massive data.

MapReduce is a distributed computing model, which can realize parallel computing, and it is the main way of big data processing at present. Some classical data mining algorithms have been combined with MapReduce, and become parallel algorithms [24-26]. MapReduce shields the details of distributed computing framework, and abstracts computing to 'Map' and 'Reduce'. The intermediate result in the form of 'key-value' pairs is generated by 'Map' processing. While all the 'values' of the same 'key' in the intermediate result are regulated by 'Reduce' to get the final result. The flow of MapReduce is shown in Fig.2.

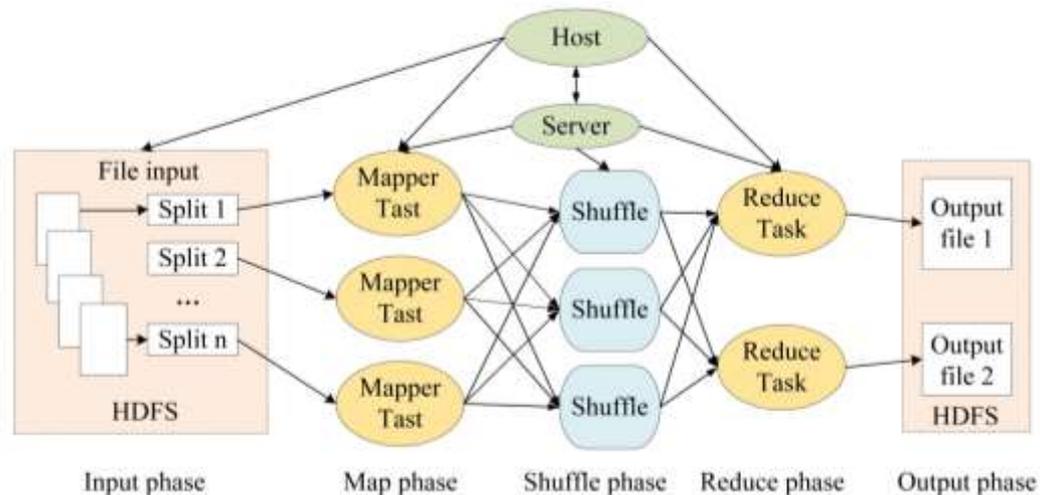


Fig.2 Flowchart of MapReduce

2.4 Workflow of T.K-means algorithm

Combining B-K-means algorithm with MapReduce parallel computing framework, the algorithm is parallelized. T.K-means algorithm reduces the scale of irrelevant data and improves the readability of clustering results by progressive clustering for performance index and cutting parameters. Repeated retrieval of database is avoided, and mining efficiency is greatly improved.

The concrete steps of T.K-means algorithm are as follows:

(1) Extract the performance index from database and cluster them by B-K-means algorithm. If the suboptimum performance index and corresponding cutting parameter set are eliminated, the data scale in original database is reduced and a new dataset will be formed.

(2) In HDFS, the new dataset is divided into several subsets with equal size, which will be randomly assigned to random working node.

(3) In Map phase, the subset is scanned at corresponding nodes. The algorithm is used to generate partial clustering center and samples, thus generating key-value pairs $\langle key, value \rangle$. The *key* represents clustering center and the *value* represents the sample of clustering center.

(4) In Reduce phase, a new clustering center is generated by calculating mean value of data samples corresponding to the same *key* in $\langle key, value \rangle$, and its data samples are output.

(5) Judging whether the clustering center is no longer changed, if the convergence condition is reached, jumping to step 6, otherwise jumping to step 3.

(6) According to the clustering result generated last time, the original data is classified to get the final clustering center and corresponding data sample.

The flow chart of T.K-means algorithm is shown in figure 3.

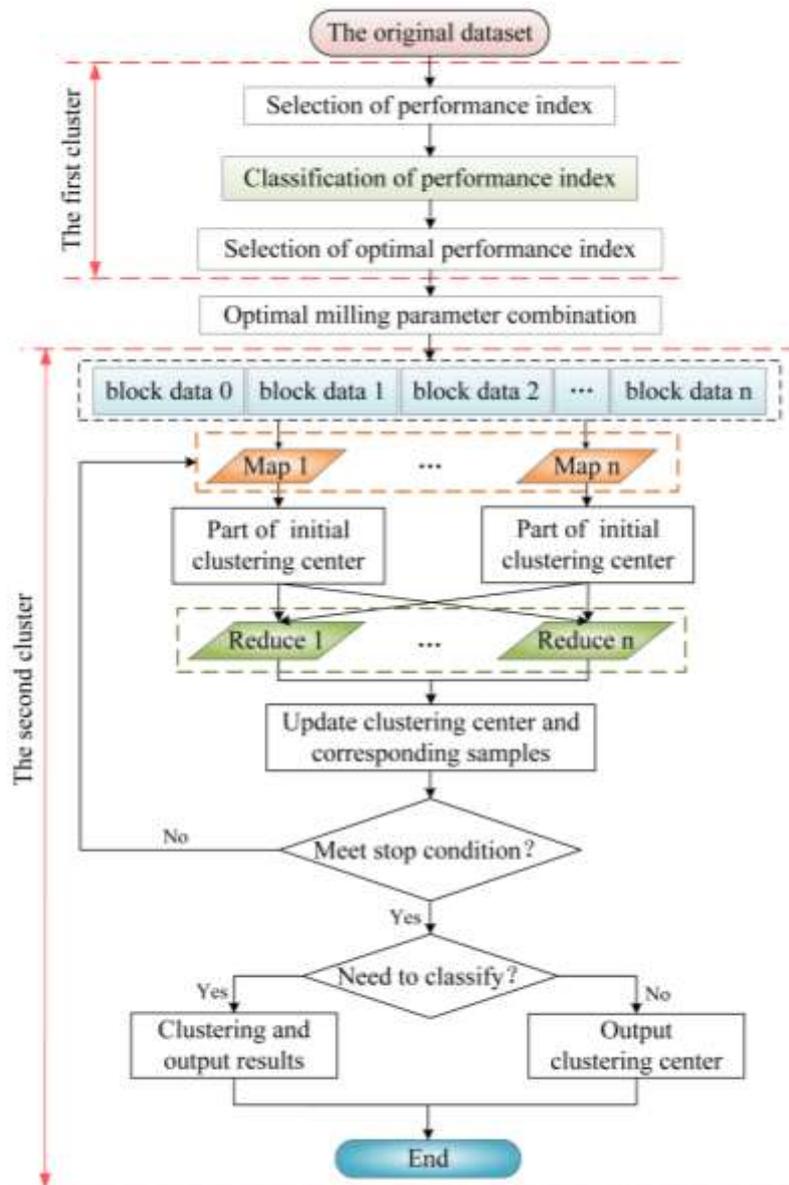


Fig.3 Flow chart of T.K-means algorithm

3 Application of T.K-means algorithm for cutting parameter optimization

Based on data mining technology, a large number of data accumulated in actual production are mined, and real cutting parameters that affect the performance index are obtained. The mining results reflect the cutting parameters corresponding to accessible value of historical optimal performance index. In this paper, the optimal cutting parameter is found in massive running data combined with T.K-means distributed clustering algorithm based on Hadoop data analysis platform. The mining results show that this method is highly representative and reliable in practical application.

3.1 Determination of mining objectives and parameter optimization framework

For milling titanium alloy, surface roughness directly affects the mechanical properties, reliability, working performance and service life of the workpiece. Material removal rate also has an important impact on enterprise benefits, so the

machined surface roughness and material removal rate are both selected as performance index.

T.K-means algorithm is applied to cutting parameter optimization for milling titanium alloy, and the corresponding clustering rule is obtained. According to this rule, the optimal performance index can be obtained by using cutting parameter, which can be used as the optimization target value to guide actual machining.

The selection of cutting parameters should meet the following two conditions: (1) It is closely related to workpiece surface roughness. (2) It can be adjusted and controlled in actual running, which can guide machining. Therefore, the following cutting parameters can be selected: cutting speed v_c , feed f_z , radial cutting depth a_e and axial cutting depth a_p . Fig.4 shows cutting parameters optimization framework for milling Ti-6Al-4V alloy based on data mining technology.

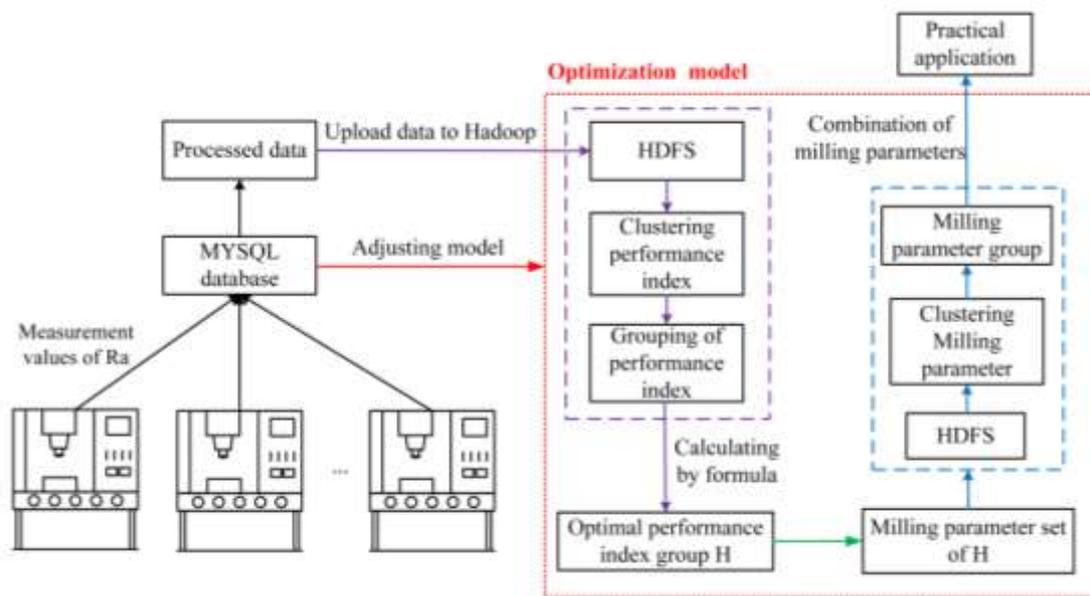


Fig.4 Architecture diagram of titanium alloy milling cutting parameters optimization based on data mining technology

3.2 Source of data

At present, there is a lack of a large number of actual machining data for analysis. In this paper, the mathematical model of cutting parameters and performance index is established by response surface methodology, and a large number of virtual data generated by this model are used to replace real data for analysis [27, 28]. The 3-Axis vertical CNC machining center is VDL—1000E. A ST210-R4-20030 carbide plain cutter is used for flank milling. Ra is measured by TR-200 portable instrument. The workpiece material is Ti6Al4V commonly used in aerospace industry. The milling method is climb milling, and the cooling method is dry cutting. The experimental site and measurement instrument are shown in Fig.5:



Fig.5 The experimental site and measurement instrument

Mathematical model of cutting parameters and surface roughness was established by response surface methodology:

$$\begin{aligned}
 y = & 2.17 - 9.45X_1 - 1.7 \times 10^{-3} X_2 - 0.04X_3 - 2.65X_4 - \\
 & 0.15X_1X_2 - 0.025X_1X_3 + 8.15X_1X_4 - 8.19X_2X_3 + \\
 & 9.18X_2X_4 + 0.048X_3X_4 + 87.02X_1^2 + 7.95 \times 10^{-5} X_2^2 + \\
 & 4.17 \times 10^{-3} X_3^2 + 0.41X_4^2
 \end{aligned} \quad (6)$$

Where y is predicted value of surface roughness, X_1 , X_2 , X_3 and X_4 are feed f_z , cutting speed v_c , radial cutting depth a_e and axial cutting depth a_p , respectively.

Results indicated that the regression model has a high degree of fit ($R^2 = 0.8832$), the lack of fit is not significant ($P_0 = 0.0556 > 0.05$), P -value is less than 0.0001, which indicates that the fitting level is excellent. The model can better reflect the relationship between workpiece surface roughness and four cutting parameters.

The formula for material removal rate is:

$$Q = \frac{a_p a_e v_f}{1000} \quad (7)$$

In machining process, there are some problems such as machine tool vibration, various tool quality, etc. In this paper, the random number m is set to solve the influence of above problems on workpiece surface roughness, and ensure Ra value is within a reasonable range.

$$y_1 = y + m\%y \quad (8)$$

The range of four milling parameters is $v_c \in [80, 120]$, $f_z \in [0.08, 0.14]$, $a_e \in [0.5, 1.5]$ and $a_p \in [8, 12]$. The random number $m \in [-10, 10]$. The above five parameters are set within assigning range, v_c , f_z , a_e , a_p are taken at an interval of 40/19, 0.01, 0.5, 4/19, respectively, and m is taken at random. They are substituted in formula (5) and (6), a total of 80,000 pieces of data are generated, which are used to provide a large amount of data for mining analysis.

3.3 Weight coefficients distribution and evaluation criteria

It is necessary to comprehensively consider the relative importance of two performance index and the influence of range on milling parameters when distributing weight coefficients. They are expressed by relative importance weight coefficient (γ) and range weight coefficient (η) respectively. The γ of surface roughness and material removal rate is set by analytic hierarchy process (AHP), and the ratio is 7:3. Different performance index have different range. The larger the range is, the faster the change speed is, and the greater the influence on parameter optimization. In order to reduce the influence of range on the optimization results, the smaller the range of the performance index is, the larger corresponding weight is. In this paper, the variation range of surface roughness is $0.2\sim 1.3\mu\text{m}$, and the variation range of material removal rate is $1.168\sim 16.488\text{cm}^3/\text{min}$. The η of workpiece surface roughness and material removal rate are 0.683 and 0.317.

After determining γ and η , the final weight coefficients λ_i of the two performance index are calculated by formula (9):

$$\begin{cases} \lambda_i = \alpha\gamma + \beta\eta \\ \alpha + \beta = 1 \end{cases}, \alpha \in (0,1), \beta \in (0,1) \quad (9)$$

Where α is the percentage of γ in λ_i , and β is the percentage of η in λ_i . In contrast, γ is more important than η , so α is set to 0.7 and β is set to 0.3.

After calculating, the final weight coefficients of workpiece surface roughness and material removal rate are $\lambda_1=0.7649$ and $\lambda_2=0.2351$.

In order to find the optimal clustering group, it is necessary to evaluate clustering results, as shown in formula (10):

$$w = \frac{\lambda_1 R_a}{\lambda_2 Q} \quad (10)$$

The smaller surface roughness is, the larger material removal rate is, and the smaller the calculated value w is.

3.4 Mining results and analysis

The calculated values of λ_1 , λ_2 , R_a and Q are calculated to obtain the optimal clustering group ($\langle 0.2172, 0.2436 \rangle$, $\langle 7.0431, 8.5347 \rangle$). The remainder are deleted, and the simplified optimal performance index and corresponding milling parameter sets are obtained. The milling parameters are reduced from 80,000 to 2,327, and the data scale is greatly reduced. On Hadoop platform, data group number is set to 10, the maximum iteration number is 100 times, data mining is completed for the simplified parameter sets. Finally, seven milling parameter data group corresponding to the optimal performance index grouping is obtained, as shown in Table 1.

Table 1 Optimization results of cutting parameters

组	$v_c(\text{m}/\text{min})$	$f_z(\text{mm}/\text{z})$	$a_e(\text{mm})$	$a_p(\text{mm})$
1	$\langle 117.8947, 120 \rangle$	$\langle 0.11, 0.12 \rangle$	0.50	$\langle 8.0, 8.4211 \rangle$
2	$\langle 115.7895, 117.8947 \rangle$	$\langle 0.11, 0.12 \rangle$	0.50	$\langle 8.0, 8.4211 \rangle$
3	$\langle 113.6842, 115.7895 \rangle$	$\langle 0.11, 0.12 \rangle$	0.50	$\langle 8.0, 8.2105 \rangle$
4	$\langle 111.5789, 113.6842 \rangle$	$\langle 0.11, 0.12 \rangle$	0.50	$\langle 8.0, 8.2105 \rangle$

5	<117.8947,120>	<0.11,0.12>	0.50	<8.0, 8.4211>
6	<115.7895,117.8947>	<0.11,0.12>	0.50	<8.0, 8.2105>
7	<111.5789,113.6842>	<0.11,0.12>	0.50	<8.0, 8.2105>

According to data mining results of milling Ti6Al4V alloy, the significance of above clustering is that when cutting parameters are within the grouping range in Table 1, there is a greater probability that surface roughness and material removal rate of the workpiece will be within the optimal range. Each group result is the target value of cutting parameters, and the center value is selected as the optimization value to visually display the results. The center value of two optimal performance index is (0.2351, 7.5343). The results are shown in Table 2.

Table 2 Optimization results of cutting parameters

组	v_c (m/min)	f_z (mm/z)	a_e (mm)	a_p (mm)
1	119.32	0.117	0.500	8.31
2	116.64	0.120	0.500	8.21
3	114.51	0.114	0.500	8.13
4	111.57	0.110	0.500	8.15
5	119.64	0.115	0.500	8.24
6	116.84	0.114	0.500	8.11
7	114.91	0.111	0.500	8.17

The parameters are adjusted according to optimization value in Table 2. Machined surface roughness is in the range of 0.2172~0.2436 μm , and material removal rate is in the range of 7.0431~8.5347 cm^3/min . The optimization value of two optimal performance index are 0.2351 μm and 7.5343 cm^3/min . The average value of 80,000 pieces of data is 0.5374 μm and 5.6170 cm^3/min . By comparison, workpiece surface roughness is reduced by 0.3023 μm and material removal rate is increased by 1.9173 cm^3/min . The above optimization results of cutting parameter are substitute into formula (6). In order to ensure the rationality of calculation, the average value m is 0. The surface roughness values are in the range of 0.2172~0.2436 μm .

4 Data-driven cutting parameters adjustment method

The above results show that there are many adjustable parameter combinations for obtaining satisfactory surface roughness and material removal rate. The optimization process of cutting parameters is completed. However, there are unstable factors in the actual machining, which may result in poor surface roughness. Therefore, it is necessary to monitor machining status in real time and adjust cutting parameters.

Changing cutting parameters is a complex dynamic problem, which involves the interaction and fusion of various machining information. Therefore, an effective method is needed to solve this problem. Data-driven method is a simulation process that reflects the whole life cycle of equipment by using physical model, real-time transmission information of sensors and historical operating data [29]. It can realize surface roughness prediction online by inputting data in real time, and correct cutting

parameters by using prediction model and parameters adjusting strategy. Panda et al. [30] established a surface roughness prediction model by monitoring vibration signal data and using multiple quadratic regression (MQR) models, which can accurately predict surface roughness in real time and monitor its stability trend. If surface roughness is more than preset threshold, the machining parameters need to be adjusted. Model and data are the core of data-driven method. Static historical data is used to build prediction model, and dynamic machining parameter is used to predict surface roughness in digital space. In the process of monitoring processing signals in real time, if the prediction result is not ideal, the milling parameters are adjusted and fed back to the machine tool in physical space. In this way, the surface roughness can be adjusted to ensure the machining stability.

This paper proposes a cutting parameter adjustment method for surface roughness based on data-driven method as shown in Fig.6.

The digital driving method concludes the following steps:

(1) Historical machining data is used to establish surface roughness prediction model, which is verified through experiments. The trained model will be embedded into data-driven system.

(2) The real-time signal data is monitored in physical space, which is transmitted to prediction model to realize display status.

(3) A virtual model is built to show the real-time machining state in the form of simulation.

(4) When the predicted surface roughness cannot meet actual machining requirements, the signal of changing cutting parameters will be sent and explicit instruction is given in the simulation.

(5) Cutting parameter combination is changed automatically to ensure that the surface roughness is within the required range.

(6) The process is executed circularly until machining is finished.

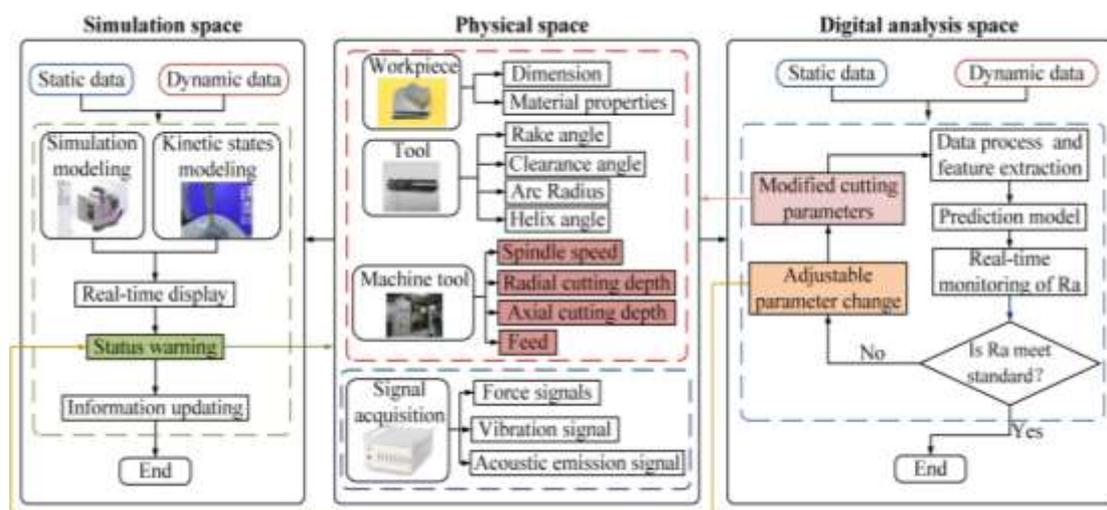


Fig.6 Flow chart of cutting parameter adjustment method for surface roughness based on data-driven method

5 Conclusion

Based on the cloud computing environment, this paper introduces the concept of distributed clustering to improve K-means algorithm. The proposed algorithm is combined with MapReduce framework to realize parallel computing, and finally forms a new efficient clustering algorithm. T.K-means algorithm is applied to cutting parameters optimization aiming at reducing workpiece surface roughness and improving material removal rate.

The following two conclusions are obtained:

(1) T.K-means algorithm is used to cluster performance index and cutting parameter respectively, which reduces retrieval times in database and avoids useless clustering group. The clustering efficiency and results readability are improved, and the algorithm is combined with MapReduce parallel computing framework. It integrates limited computing resources and solves the problem that the traditional algorithm has insufficient ability for data mining.

(2) T.K-means algorithm is used to calculate virtual titanium alloy milling data, and the clustering rule of 'optimized cutting parameters to optimal surface roughness' is stipulated. The range of adjustable cutting parameters corresponding to optimal surface roughness is obtained. The mined cutting parameters reflect the maximum accessible value of workpiece surface roughness in all data. It can be demonstrated that the clustering results can be used to guide cutting parameters optimization for milling titanium alloy when there are a lot of real machining data.

Finally, a data-driven cutting parameter adjustment method is proposed to monitor machining state and ensure machining stability in real time.

References

- [1] Chen X, Li C, Jin Y, et al. Optimization of cutting parameters with a sustainable consideration of electrical energy and embodied energy of materials[J]. *The International Journal of Advanced Manufacturing Technology*, 2018.
- [2] Warsi S S, Agha M H, Ahmad R, et al. Sustainable turning using multi-objective optimization: a study of Al 6061 T6 at high cutting speeds[J]. *International Journal of Advanced Manufacturing Technology*, 2019, 100.
- [3] Rao K V. Power consumption optimization strategy in micro ball-end milling of D2 steel via TLBO coupled with 3D FEM simulation - ScienceDirect[J]. *Measurement*, 2019, 132: 68-78.
- [4] Sharma A, Islam A, Sharma K, et al. Optimization techniques to optimize the milling operation with different parameters for composite of AA 3105[J]. *Materials Today: Proceedings*, 2021.
- [5] Wang, YL, XH, et al. Establishment and application of a process dimension tree[J]. *INT J PROD RES*, 2016.
- [6] Diyaley S, Chakraborty S. OPTIMIZATION OF MULTI-PASS FACE MILLING PARAMETERS USING METAHEURISTIC ALGORITHMS[J]. *Facta Universitatis Series Mechanical Engineering*, 2019, 17(3): 365.
- [7] Ma H, Liu W, Zhou X, et al. An effective and automatic approach for parameters optimization of complex end milling process based on virtual machining[J]. *Journal of Intelligent Manufacturing*, 2020, 31.

- [8] Nguyen T T, Nguyen T A, Trinh Q H. Optimization of Milling Parameters for Energy Savings and Surface Quality[J]. *Arabian Journal for Science and Engineering*, 2020: 1-15.
- [9] Zhang X, Yu T, Dai Y, et al. Energy consumption considering tool wear and optimization of cutting parameters in micro milling process - ScienceDirect[J]. *International Journal of Mechanical Sciences*, 178.
- [10] Xiao Y, Jiang Z, Gu Q, et al. A novel approach to CNC machining center processing parameters optimization considering energy-saving and low-cost[J]. *Journal of Manufacturing Systems*, 2021, 59(5-8): 535-548.
- [11] Deng T, Li Y, Liu X, et al. A Data-driven Parameter Planning Method for Structural Parts NC Machining[J]. *Robotics and Computer-Integrated Manufacturing*, 2021, 68: 102080.
- [12] Ramesh B. *Big Data Architecture*[M]. Springer India, 2015.
- [13] Zhong Y, Huang X D, Liu D, et al. NoSQL storage scheme for large-scale equipment monitoring data[J]. *Computer Integrated Manufacturing System*, 2013(12): 3008-3016.
- [14] Kaseb M R, Khafagy M H, Ali I A , et al. An improved technique for increasing availability in Big Data replication[J]. *Future generation computer systems*, 2019, 91(FEB.): 493-505.
- [15] Li P, Guo S, Yu S, et al. Cross-cloud MapReduce for Big Data[J]. *IEEE Transactions on Cloud Computing*, 2015: 1-1.
- [16] Zhai ZW. *Hadoop core technology: Core Hadoop*[M]. Mechanical Industry Press, 2015.
- [17] Singh S, Jagdev G. Execution of Big Data Analytics in Automotive Industry using Hortonworks Sandbox[C] // 2020 Indo-Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN). 2020.
- [18] Bhathal G S, Dhiman A S. Big Data Solution: Improved Distributions Framework of Hadoop[C] // 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS). 2018.
- [19] Cao X. Design and implementation of Sophon platform sharing system of Star Ring Technology [D]. Nanjing University, 2019.
- [20] Jun, Lee, Kim. Cloud-based big data analytics platform using algorithm templates for the manufacturing industry[J]. *International Journal of Computer Integrated Manufacturing*, 2019.
- [21] Kanungo T, Mount D M, Netanyahu N S, et al. An efficient k-means clustering algorithm: analysis and implementation[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, 24(7): 881-892.
- [22] Zhou A W, Ya-Fei Y U. The Research about Clustering Algorithm of K-Means[J]. *Computer Technology and Development*, 2011.
- [23] Huang Yihua. Depth understanding of big data: Big Data processing and programming practices [M]. Beijing: Machinery Industry Press, 2014.
- [24] Lv Z, Hu Y, Zhong H, et al. Parallel K-Means Clustering of Remote Sensing Images Based on MapReduce[J]. 2010.
- [25] He Q, Zhuang F, Li J, et al. Parallel Implementation of Classification Algorithms

Based on MapReduce[C]// Rough Set & Knowledge Technology-international Conference. DBLP, 2010.

- [26] Failure H, Failure H, Access S D , et al. The Hadoop Distributed File System: Architecture and Design[J]. Hadoop Project Website, 2007, 11(11): 1 - 10.
- [27] Ji W, Yin S, Wang L. A big data analytics based machining optimisation approach[J]. Journal of Intelligent Manufacturing, 2019: 1-13.
- [28] Ren L, Zhang L, Zhang Y, et al. Resource virtualization in cloud manufacturing[J]. Computer Integrated Manufacturing Systems, 2011, 17(03): 511-518.
- [29] Zhao Z, Wang S, Wang Z, et al. Surface roughness stabilization method based on digital twin-driven machining parameters self-adaption adjustment: a case study in five-axis machining[J]. Journal of Intelligent Manufacturing, 2020:1-10.
- [30] Panda A, Sahoo A K, Panigrahi I, et al. Prediction models for on-line cutting tool and machined surface condition monitoring during hard turning considering vibration signal[J]. Mechanics and Industry, 2020, 21(5): 520.