Scholar Publishing Group
Kinetic Mechanical Engineering
https://doi.org/10.38007/KME.2025.060102
ISSN 2790-2676 Vol. 6, Issue 1: 11-25

Research on Intelligent Classification and Typing Compression Algorithm for Machining State Data Based on Autocorrelation Coefficient and Stability Discrimination

Jiayi Sun[1, a*]

[1]*Smart Grid Department,Zhi Fang Design Co., Ltd.,Hongshan District 430070, Wuhan , Hubei , China*

Keywords: Autocorrelation coefficient analysis, Mechanical processing state monitoring, intelligent data classification, data compression by type

Abstract: Against the backdrop of the rapid development of intelligent manufacturing and the wide application of edge computing, the multi-source high-dimensional data generated during the operation of processing equipment has put forward higher requirements for the storage, communication and energy consumption control capabilities of the system. Especially when achieving preventive maintenance of key components, how to efficiently handle vibration signals with high collection frequency and large information volume has become the research focus. To solve the problem of low efficiency of traditional compression methods when dealing with complex signal structures, this paper proposes an automatic classification and hierarchical compression algorithm for edge-end operating state data based on sequence autocorrelation feature analysis and signal stability determination. This algorithm can perform dynamic identification and classification according to signal characteristics, thereby matching the optimal compression strategy. Specifically, the system first identifies the periodically prominent high-frequency vibration signals through the correlation analysis of the sampling sequence, and then further distinguishes the steady-state and slow-varying signals in the non-vibration data through stability analysis to achieve precise division of the data stream. In terms of the selection of compression methods, this paper designs a lossless compression mechanism combining difference transform and entropy coding for data sequences with relatively stable structures. For the slowly changing trend signal, an adaptive compression scheme integrating forward prediction and arithmetic coding was constructed; For vibration data with intense fluctuations, a joint modeling and quantization coding strategy in the time-frequency domain is introduced, achieving efficient lossy compression while maintaining feature retention. The experimental results show that the classification compression framework proposed in this paper effectively reduces resource overhead while improving the compression ratio, providing practical and feasible technical support for edge intelligent processing in the mechanical manufacturing process.

1. Introduction

As the global manufacturing industry continues to move towards high-end and intelligent development, the state's investment in the development of advanced manufacturing technologies is increasing day by day. Data has become the core resource driving the digital transformation of the manufacturing industry. The information infrastructure of various enterprises in the R&D, production, operation and service links is becoming increasingly complete, which puts forward new requirements for the efficient management and orderly flow of massive industrial data. During the digital transformation of the manufacturing industry, relevant departments are confronted with the major challenge of breaking down data silos and achieving full life cycle management of industrial data. Building a complete data chain covering data collection, transmission, storage and analysis has become the key to promoting a qualitative leap in intelligent manufacturing.

At present, the traditional compression methods for mechanical processing state information generally adopt a unified algorithm to process various signals, making it difficult to effectively optimize for different signal characteristics, resulting in large data redundancy and unsatisfactory compression efficiency. However, vibration signals, as typical high-frequency and variable signals, although frequency-domain compression methods based on discrete cosine transform, wavelet decomposition and Fourier transform have been widely applied, However, due to the high real-time requirements and limited computing resources in industrial sites, these methods still need to be further improved to enhance the compression effect and computing efficiency.

Edge computing technology, by deploying computing resources near the source of data generation, effectively alleviates the pressure on cloud processing, reduces network latency and energy consumption, and has become an important support platform for real-time data processing and intelligent analysis in the industrial Internet of Things. Utilizing edge computing to achieve intelligent classification and type compression of mechanical processing status data can not only improve data processing efficiency, It can also significantly reduce the network load and storage pressure of data transmission, thereby promoting the efficient utilization of industrial big data.

In response to the above problems, this paper designs an intelligent classification algorithm for machining state data that combines the autocorrelation characteristics of signals and the determination of steadiness. This algorithm can accurately distinguish vibration signals, steady-state signals and slow-changing signals, achieve differentiated compression processing, deploy the corresponding compression model at the edge end, and adopt lossless compression technology for slow-changing and steady-changing signals to maintain data integrity. For the vibration signal, a lossy compression scheme based on two-dimensional discrete cosine transform, polynomial fitting and floating-point binary residual coding was designed. It takes into account both the compression ratio and the retention of key signal features, significantly improving the transmission efficiency and storage performance of mechanical processing state data.

2. Relevant research

2.1. Mechanical Processing condition monitoring technology

In modern mechanical processing, achieving efficient monitoring of the processing status has become a key link to ensure product quality and stable operation of equipment, among which the identification and diagnosis of the processing status of lathes are particularly important. By deeply comparing the monitoring performance of vibration signals and acoustic signals under different cutting depths and spindle speeds [1], researchers found that both signal forms could effectively reflect the continuous downward trend of the characteristic frequencies of the motor and the spindle as the cutting load gradually increased, thereby verifying the reliability of frequency variation as a key characteristic parameter of the processing state. Although vibration signals show high accuracy and stability in state recognition due to their high sensitivity and strong anti-interference ability, their deployment usually relies on equipment contact sensors and multi-point deployment, which has certain limitations in complex processing environments. Relatively speaking, although acoustic signals are more susceptible to interference from dynamic factors such as environmental noise, structural resonance and cutting force, resulting in nonlinear fluctuations at the amplitude level, they still have important application value in complex industrial scenarios due to their advantages of enabling long-distance monitoring and flexible layout. Therefore, extracting robust features based on multimodal signal fusion has become an important direction for improving the performance of monitoring systems.

JJP Abadia designed a meta-learning framework combining ensemble learning and deep neural networks[2]. This method significantly improves the recognition accuracy of tool wear states in actual industrial environments by integrating multi-model learning strategies. While enhancing the generalization ability of the model, it reduces monitoring errors, thereby strengthening the reliability and intelligence level of the processing process. To address the migration difficulties caused by inconsistent data distribution in actual industrial scenarios, Z Huang[3] proposed a cross-domain tool wear monitoring method based on Deep Adversarial Domain Confusion Network (DADCN). In this framework, by constructing a structurally consistent feature extraction network for the source domain and the target domain and combining an independent adversarial learning mechanism, The extraction of dome-invariant features has been effectively achieved. Combined with the maximum mean difference (MMD) criterion, the adaptability of the model to different processing environments has been enhanced. The accuracy and robustness of this method in cross-platform tasks have been confirmed through experiments. Facing the modeling challenge caused by insufficient sample size, NS Ross[4] introduced transfer learning technology and systematically evaluated the recognition effects of various classical pre-trained models under different cooling processes (such as dry cutting, wet cutting, micro-lubrication and low-temperature cooling). Finally, the Epinction-V3 network with excellent performance was selected as the basic structure. By adjusting the hyperparameters to optimize the training process, the effective classification of tool states under the condition of small samples was achieved, showing a good application prospect.

Therefore, the current development of mechanical processing state monitoring technology shows a trend of integrating multimodal perception, adapting to cross-domain scenarios and enhancing the ability of small sample learning. Future research is expected to further combine reinforcement learning, self-supervised feature extraction and edge computing architecture to build a more accurate, stable and real-time responsive intelligent monitoring system. Thereby promoting the in-depth application of intelligent manufacturing technology in industrial practice

2.2. Data Compression Technology

With the continuous increase in the demand for high-speed data transmission and massive data storage in the industrial field, data compression technology has become the core means to ensure system performance and resource utilization efficiency. KC Muntha[5] designed and implemented a fibre channel communication architecture based on FPGA. This architecture combines the efficient temporal differential Neighborhood Index sequence (TSD^2 NIS) compression algorithm with the advanced Gigabit Transceiver high-speed transceiver module (GTH). The tests were completed on the Virtex-7 platform. The results show that the system not only achieves a data compression rate of 40%, but also realizes an extremely low transmission delay of 0.33 nanoseconds and an operating frequency of 321.425MHz, demonstrating the outstanding performance of this scheme in high-speed and low-delay data transmission and system scalability. It fully meets the strict requirements of modern industrial communication.

Aiming at the problems of high computational complexity and data requirements faced by monocular depth estimation models in autonomous driving, W Jun[6] proposed and applied a variety of synthetic data enhancement methods including Mask and CutFlip. Combined with the RMS optimization algorithm considering model delay, efficiency and resolution, the accuracy and real-time processing ability of the model were significantly improved. The measured results of this technology on the NVIDIA Jetson AGX Orin embedded platform show that through quantitative pruning, the model size has been reduced by 83.4% while still maintaining a relatively low relative error and latency, which greatly promotes the deployment of deep learning models in resource-constrained environments.

In response to the increasing demand for image data processing in power systems and industrial monitoring, M Liu[7] based on the traditional SPIHT image compression algorithm, by introducing block-level parallel wavelet transform and the parallel coding strategy based on the coding tree structure, achieved data-level and task-level parallel processing in the image compression process. This method significantly improves the efficiency of image compression and real-time processing capabilities, effectively solving the management problems of large-scale monitoring image data. Aiming at the huge data storage pressure caused by continuous monitoring in the Φ-OTDR optical fiber sensing system, FH Yu[8] designed a lossy data compression method based on quantization technology. This method can effectively reconstruct the signal without changing the hardware structure. Experiments show that This scheme is capable of compressing 128MB of data to 31.25MB within approximately 293.2 milliseconds, achieving a compression ratio of 16 times, significantly improving the efficiency of data storage, and ensuring the stable operation of long-term continuous monitoring.

Current data compression technologies, by integrating advanced algorithms and hardware implementation, have successfully addressed the challenges of massive data in industrial communication, visual perception, and large-scale sensor monitoring. In the future, with the continuous development of intelligent algorithms and computing architectures, these technologies will further enhance the real-time performance, reliability, and resource utilization of industrial systems.

3. Design and Experimental Verification of Multi-strategy Industrial Data Compression Method Based on Signal Characteristics Adaptation

3.1. Performance Analysis of the Cooperative Compression Strategy of Neural Network Prediction and Run-length Differential Coding

During the mechanical processing, the system first meticulously classifies the collected signal data into four major categories through classification algorithms: stable change signals, slow change signals, oscillation signals, and other types of signals. According to the characteristics of different signals, the algorithm selects the corresponding data compression strategy to significantly improve the data processing efficiency while ensuring the compression quality. For data such as stable change signals, since the numerical changes between adjacent elements are mostly consistent, there is a high degree of redundancy in the data. Therefore, the system adopts a run-length encoding method suitable for continuous repetitive data encoding for compression. The algorithm first performs the first-order difference operation on this type of stable signal, generates the difference sequence and saves the initial values, which provides the necessary information for restoring the original data during subsequent decompression. The system uses run-length encoding to perform length encoding on consecutive identical numerical segments in the difference sequence, thereby effectively reducing the data scale. The algorithm further compresses the result of run-length encoding through arithmetic encoding to achieve higher storage efficiency. Through this multi-stage compression process, the system not only guarantees the effect of data compression, but also significantly shortens the time of compression and decompression. Verified by experiments, this method shows superior performance when dealing with stably changing signals.

The slow-varying data compression method based on BP neural network fitting introduces the overall process of "prediction - quantization - entropy coding", effectively improving the efficiency and flexibility of time series data compression while ensuring the controllability of compression accuracy. This method draws on the successful experience of lossy compression in the fields of speech, image and video, and combines the significant advantages of neural networks in nonlinear modeling and time series prediction. Deep learning technology is embedded into the compression process, thereby constructing a prediction module with BP neural network as the core, and generating high-precision current predicted values by learning historical reconstruction data. To meet the requirements of causality, the system only takes historical known data as input during the training stage and feeds back the model output results for subsequent reconstruction, so that the entire encoding and decoding process maintains consistency and reversibility.

In specific operations, the BP neural network utilizes the dependency relationships among multiple variables in the input time series, integrates the historical data of the same variable and other variables at the previous moment, generates the current predicted value, and continuously optimizes the model parameters through adaptive training until the prediction error is lower than the preset threshold. When the predicted value is determined, the system further converts the prediction error into symbolic data through the uniform quantization process, thereby controlling the maximum absolute error introduced during the compression process. This quantification process adopts a fixed step size and limits the range of errors, thereby ensuring that the error between the reconstructed data and the original data is always within an acceptable range. For the error values that exceed the quantification range caused by extreme cases, the system has designed a retention mechanism to directly save the original values in floating-point format to avoid information loss.

In the final stage of the compression process, the system efficiently encodes the quantized error information through the arithmetic coding algorithm. The algorithm constructs a probability model, maps each symbol to sub-intervals with corresponding proportions within the interval [0,1), and compresses the entire data sequence into a real number by gradually subdividing the intervals, thereby greatly reducing the number of bits required for coding. During the recovery process, the decoder identifies the corresponding symbol sequence one by one based on the position of the real number within the interval, and at each step, uses the same prediction function and error value as the encoder to complete the data reconstruction, ensuring the accurate restoration of the entire compression and decompression process.

This compression method not only effectively improves the prediction accuracy of slow-changing data through the BP neural network, but also achieves a good balance between the compression ratio and data quality through the quantization strategy of precisely controlling the error range and the efficient entropy coding mechanism. It provides an intelligent, efficient and practically valuable solution for the compression of multivariable time series data.

3.2. Experimental Verification

The researchers built a complete machining process measurement system. This system integrates a three-axis vibration sensing module centered on the main shaft, a real-time motor current monitoring device, and displacement and power data channels exported by the CNC control system. With the help of a unified data acquisition strategy, During the continuous processing of the tool from the initial state to the failure process, dynamic signals and running trajectories are obtained respectively by two different sampling mechanisms of high frequency and medium frequency. In terms of specific collection forms, at fixed intervals, the system will extract a period of duration samples during the processing and store them respectively as CSV files with consecutive numbers to express the operating status at different stages of the tool life cycle.

In contrast, the data acquisition adopted by sample set B is more inclined towards a systematic comparative study under experimental control conditions. It uses a unified standard to observe the wear behavior of the cutting tool under various processing strategies, and vibration sensing units, acoustic emission sensors and current monitoring modules are deployed in multiple test cycles. The synchronous acquisition of key variables on the spindle end and the working platform has been achieved. Meanwhile, information such as processing parameters (such as feed rate, cutting depth, and material category) and state variables (such as cutting duration, tool number, and wear amount) has been supplemented, providing a rich and complementary data basis for modeling.

To evaluate the efficiency performance of the compression strategy in this paper when processing the above-mentioned signals, the researchers constructed a complete performance evaluation system from multiple dimensions such as storage space, compression ratio and data reconstruction fidelity. Among them, the compression factor Cf is defined as the ratio between the bit width of the original data and the bit width after compression, as shown in formula (1).

$$C_f = \frac{N_{unc}}{N_c} = \frac{S_{unc}}{S_c} \tag{1}$$

Among them, $\mathrm{N_{unc}}$ and $\mathrm{N_c}$ respectively represent the total number of bits before and after compression, and$\mathrm{S_{unc}}$ and $\mathrm{S_c}$ are the file sizes of the corresponding data. The compression ratio CR, as the reciprocal of this indicator, measures the number of bits required to compress data per unit and is defined as formula (2).

$$CR = N_{unc}N_c = S_{unc}S_c \tag{2}$$

Meanwhile, to visually reflect the proportion of space saved after compression, the study also introduced the space saving percentage index (%SS), which represents the degree of reduction in storage occupation of the compressed data compared to the original data. The higher the value, the more advantageous the compression strategy is in terms of space utilization. In this paper, these two typical industrial signal datasets are utilized to conduct comparative experiments under multiple algorithm Settings and parameter tuning schemes, in order to ensure that the proposed method still has good generalization ability and compression performance in multi-source and variable signal scenarios.

In the data experiment, the research team conducted classification and compression performance tests on the sensor signals obtained in multiple numerical control processing scenarios. During the initial identification process, the input signals were classified and processed based on the sampling stability determination mechanism and the autocorrelation feature analysis algorithm. The results showed that the data sequence of the first experimental source did not contain a stable trend, and therefore was not included in the verification process of the compression method. In contrast, the second experimental data source is of greater research value. This dataset was collected from 16 independent CNC metal cutting tasks, with a total data volume of 148MB, and was divided into four equal parts for use as compressed samples.

In this dataset, there are a total of 13 types of sensor channel signals reflecting the processing dynamics. After stability detection, 7 types of information were identified as having obvious long-term stable change characteristics, including task number, operation count, cutting depth, feed rate, material type, etc. The total data volume is approximately 59MB. For this type of low-frequency oscillation signal, an algorithm structure based on the combination of first-order derivative coding and repetitive segment compression mechanism was studied and constructed, and its systematic performance was compared with arithmetic encoders, ZIP archiving algorithms and the WinRAR method based on stroke compression.

The results show that the designed compression strategy performs close to mainstream compressors in terms of compression ratio, but has achieved obvious advantages in terms of space resource conservation. Due to the adoption of a linear difference structure, this method demonstrates extremely high efficiency during compression execution. The time required for compression is much lower than that of traditional archivers, and in some experimental samples, the processing speed is approximately twice as fast.

For another type of signal with a slower change rate, smaller amplitude and more predictable trend (such as current channels), the research team further introduced a backpropagation neural network compressor for data modeling. This neural network is capable of accurately approximating the sequence trend by adjusting the learning rate and the error tolerance threshold. In practical applications, by setting the upper limit of the maximum error at 0.001 and selecting five groups of typical numerical control data files (with a data volume of about 100MB per file) as input, the BP neural network was used for compression modeling, and the compression effect was evaluated laterally with two representative compressors, SZ and CA. The SZ algorithm is a representative among the current compression algorithms with controllable distortion tolerance, while CA is widely used in practical engineering environments due to its lower computational complexity.

Figure 1. Compression ratio performance of BP compressor under different error tolerances (Unit: %)

Figure 2. Compression ratio performance of CA compressor under different error tolerances (Unit: %)

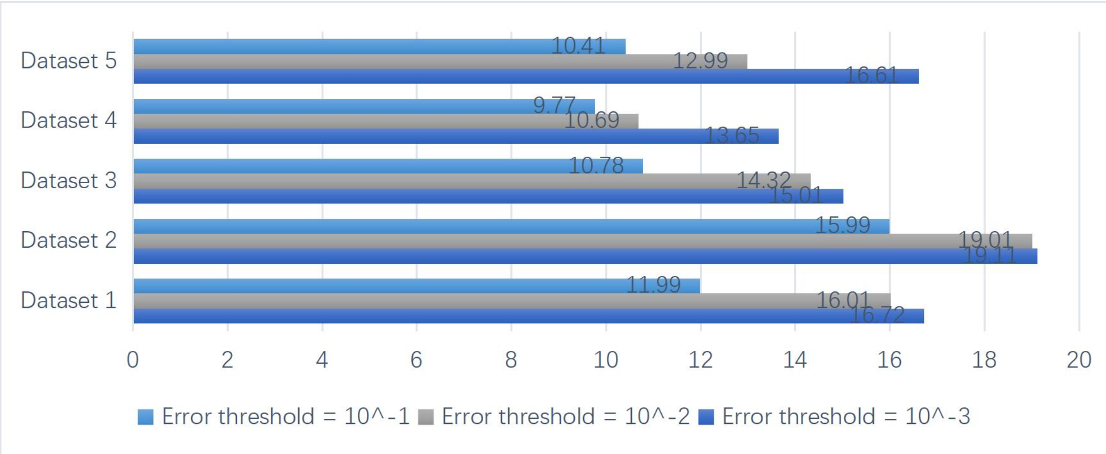

Figure 3. Compression ratio performance of the SZ compressor under different error margins (Unit: %)

It can be observed from the data in Figures 1, 2 and 3 that when the error tolerance threshold is relatively small (10^-3 or 10^-2), the overall performance of the BP compressor is the best. However, in the scenario where the error margin is expanded to 10^-1, the SZ compressor has more compression advantages for some data. This can be explained as follows: When SZ processes data with a relatively simple structure, the number of parameters that its encoder needs to store is lower than that of the neural network model, thereby bringing a higher compression ratio.

In terms of processing time efficiency, the researchers conducted time consumption tests on the three compression methods respectively under the same software and hardware environment. The results show that SZ has a compression time significantly lower than the other two models because the algorithm complexity is O(N) and the main execution logic is bit-level operation, with an average compression time of 4 to 8 seconds. The compression time of the BP neural network compressor fluctuates between 5 and 11 seconds. In some experiments, its running speed even exceeds that of SZ. Although the CA method has a small amount of calculation, it is not superior in terms of compression effect and speed.

Combining the comprehensive indicators of error control, compression effect and execution efficiency, the BP compression method shows a good multi-dimensional performance balance when facing slow-changing data. The RLED algorithm has dual advantages of compression efficiency and system resource conservation in steady-state sequence processing. The collaborative use of the two methods provides technical support for constructing an efficient and scalable industrial sensor data compression system.

4. Design and Optimization of Adaptive Lossy Compression Method for Mechanical Vibration Signals

4.1. Construction of a compression coding framework based on polynomial fitting and two-dimensional DCT

For the high-precision vibration signal data generated in the mechanical processing process, this paper proposes a lossy compression scheme combining polynomial fitting and quantization strategies. This scheme aims to significantly reduce the volume of the signal data by limiting the maximum error range. Since vibration signals are mostly represented in high-precision floating-point format, traditional lossless compression methods are difficult to achieve an ideal compression ratio, and industrial applications usually allow for a certain degree of accuracy loss. Therefore, the research team designed a compression process that allows users to customize the error threshold according to actual needs, in order to improve the overall data processing efficiency while ensuring the validity of the data.

In this paper, the collected vibration signals are first segmented for one-dimensional time series data according to the preset fixed length, and the original signals are converted into the form of two-dimensional matrices for subsequent frequency-domain analysis and processing. Then, the system applies two-dimensional discrete cosine transform to perform frequency-domain transformation on the signal matrix, thereby extracting the main spectral components of the signal and effectively achieving the concentration of signal energy and the reduction of redundant data.

After completing the frequency-domain transformation, the algorithm uses a prediction model based on polynomial fitting to perform piecewise quantization processing on the coefficients obtained from the transformation, ensuring that the quantization error is strictly limited within the preset allowable error range, thereby controlling the degree of distortion introduced during the compression process. Subsequently, the compressed quantization coefficients are further encoded through lossless encoders, such as Huffman encoders, thereby effectively reducing the volume of the final stored data.

In the decompression stage, the system first decodes the compressed encoded data to recover the quantized frequency-domain coefficients. Then, it converts the frequency-domain signal back to the time-domain signal matrix through the inverse two-dimensional discrete cosine transform. Subsequently, it performs point-by-point difference operations between the reconstructed signal and the original input signal to obtain the residual information. Considering that residual signals often have a large number of redundant characteristics, the algorithm utilizes binary compression optimization technology and implements secondary compression of residual information through bit-level binary reduction methods to minimize the data storage space to the greatest extent.

This multi-stage compression architecture not only effectively reduces the storage requirements of vibration signal data, but also ensures that the signal distortion is maintained within an acceptable range through an accurate error control mechanism, meeting the dual requirements of industrial sites for the quality and storage efficiency of vibration monitoring data. Through experimental verification, this method can flexibly adjust the compression strength according to different error tolerances, achieve a reasonable balance between data accuracy and compression performance, and demonstrate good application potential and promotion value.

To meet the adaptability requirements for diverse signal characteristics in mechanical vibration signal compression, researchers have designed a dedicated compression scheme that combines the frequency domain distribution characteristics of the signal itself. This scheme utilizes the discrete cosine transform (DCT) to achieve the mapping of the signal from the time domain to the frequency domain, thereby extracting the coefficients containing most of the energy for efficient compression. Specifically, mechanical vibration signals usually show rapid fluctuations and contain periodic components and noise in the time series. By dividing the original data into a two-dimensional matrix form with a size of L rows and M columns, the system applies two-dimensional DCT to transform this matrix to obtain the frequency domain coefficient matrix, thereby achieving the aggregation of frequency domain energy. This process effectively improves the correlation performance of the data. It is helpful for subsequent compression processing. This two-dimensional DCT transformation performs weighted summation calculations on the time-domain samples based on the double cosine function, ensuring that the low-frequency part after the transformation gathers the vast majority of the signal energy, thereby reducing information redundancy and improving the compression effect.

In the quantization stage, considering the insufficient performance of traditional uniform or non-uniform quantization methods on complex data, this paper introduces a prediction quantization strategy based on polynomial fitting. By flattening the two-dimensional DCT coefficients into one-dimensional sequences, the natural storage order of the data is utilized to retain local correlations and reduce additional conversion costs. The system uses polynomial functions to fit the local data. By minimizing the fitting error and dynamically adjusting the degree of the polynomial to meet the preset accuracy requirements, this method can not only capture the inherent trend of the data, but also significantly reduce the quantization error and coding complexity. When the fitting error exceeds the threshold, the system automatically increases the order of the fitting polynomial until it reaches the allowable range or the preset upper limit. The coefficients of the successfully fitted polynomial are saved as quantized output, while the unpredicted data points are processed separately through anomaly marking and arithmetic coding, thereby ensuring the accuracy and flexibility of compression. The parameters of polynomial fitting are obtained through the analytical solution of the least squares problem, ensuring the stability and unique solution characteristics of the fitting process.

For residual data that cannot be effectively predicted, the system further adopts the binary bit-level reduction technology based on the IEEE754 floating-point standard. By analyzing the sign, exponent and tail part of the residual, combined with the error tolerance and data dynamic range specified by the user, the minimum necessary significant bits in the tail are determined and irrelevant low bits are discarded, thereby effectively reducing the encoding bit width. In the specific implementation, the system first performs inverse transformation recovery on the quantized data, calculates the binary representation of the error, then calculates the lengths of the required retained sign bits and exponential bits based on the error boundary, and finally compresses the reduced bit stream through arithmetic coding. This method not only improves the compactness of the code stream, but also ensures that the error of the reconstructed data is within a controllable range, taking into account both the compression ratio and the signal quality.

This method divides the mechanical vibration signal into blocks and uses two-dimensional DCT to achieve a high concentration of energy in the frequency domain. Combined with polynomial fitting quantization technology, it conducts efficient prediction and compression of the data. Finally, it is supplemented by binary bit optimization based on floating-point format. Overall, it improves the compression effect and computational efficiency, while ensuring the accuracy and stability of data reconstruction. It is suitable for lossy compression requirements of real-time large-scale mechanical signals.

4.2. Performance Experimental Analysis

This paper focuses on the designed compression method. Firstly, a comprehensive analysis is conducted using the compression ratio (CR) as the evaluation index. Since this algorithm can effectively adjust the allocation of storage resources, an in-depth discussion is carried out around the effect of space-saving. For lossless compression algorithms, CR and compression factor (Cf) are sufficient to describe their performance. However, since this paper focuses on lossy compression techniques, there are certain differences between the decompressed data and the original input. Therefore, the error of a single data point is defined as e_i, and the average point-by-point relative error E is used to measure the distortion caused by compression. Its calculation method is expressed as follows formula (3).

$$E = \frac{1}{N}\sum_{i=1}^{N} \left| \frac{x_i^{'} - x_i}{x_i} \right| \tag{3}$$

Among them, x_i represents the i element of the original data, $x_i^{'}$ represents the corresponding value after decompression, and N is the total number of data samples. This indicator effectively reflects the information loss during the compression process and is an important basis for evaluating the quality of lossy compression. During the testing process, different error thresholds Δ were set respectively to compress the data, and the corresponding compression performances were recorded. The results show that with the increase of the allowable error Δ, the compression ratio shows an upward trend, indicating that the error tolerance is closely related to the storage compression efficiency. To further verify the superiority of the proposed method, this paper compares its effect with that of several popular compression programs. All experiments were completed under the same conditions. Figure 4 summarizes the average compression ratio performance of different methods for the tool vibration signal dataset when the error threshold is 0.0005.

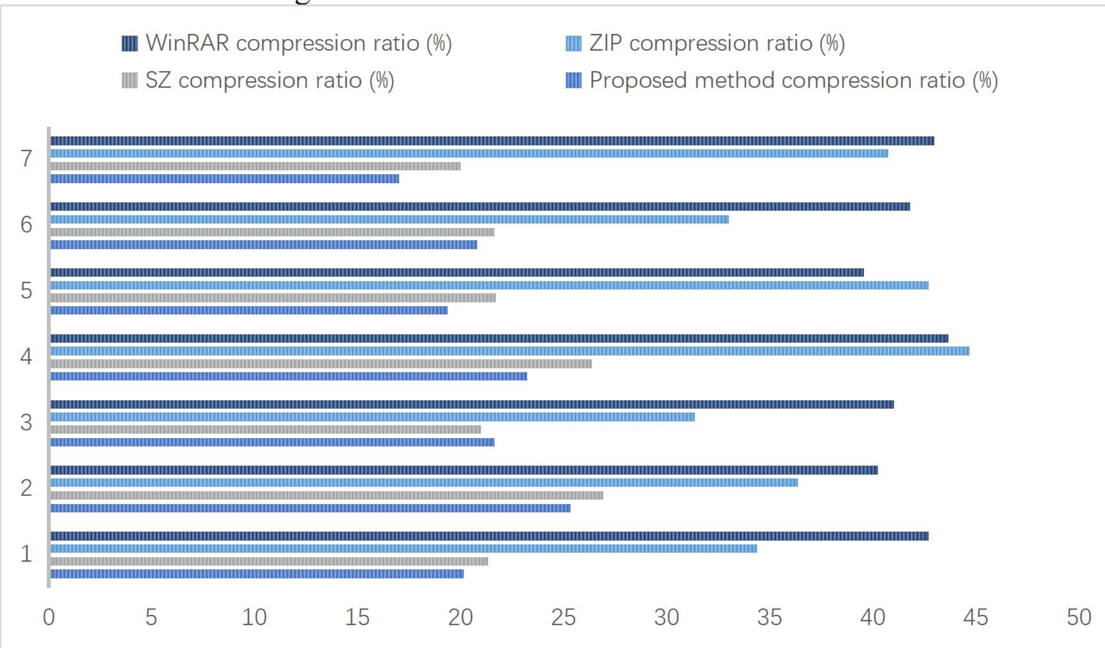

Figure 4. Average compression ratio data

The comparison results show that the compression scheme proposed in this paper has a better compression efficiency than general software such as ZIP and WinRAR within the range of controlling compression distortion, and its effect is similar to that of the dedicated and efficient compression tool SZ. Although the SZ algorithm has a relatively high compression ratio, its error fluctuates greatly and exceeds the acceptable range in some cases. Although the commercial compression software WinRAR is applicable to a variety of application scenarios, its performance on the data of this experiment is mediocre. It can be seen from this that the professional compression method designed for the vibration data of mechanical processing can significantly improve the data storage efficiency while meeting the accuracy requirements, and has strong engineering application potential.

5. Conclusions and Prospects

This study focuses on the state data in the mechanical processing process and proposes an intelligent classification method based on data autocorrelation and robustness discrimination. Moreover, differentiated compression strategies are designed for different types of data to address the challenges of real-time data processing and transmission in the manufacturing industry. In this paper, by analyzing the variation characteristics of various mechanical processing signals such as current signals, vibration signals and cutting parameters, they are classified into multiple categories, thereby achieving a targeted compression scheme. Among them, for the signals with small fluctuations and stability, the method combining differential coding and run-length coding is adopted in this paper to improve the compression efficiency. While for the data with slow changes and fluctuations, In this paper, the neural network model is used for fitting and prediction, taking into account both the compression ratio and the restoration accuracy. Meanwhile, for the vibration signals with prominent high-frequency characteristics in the frequency domain, a quantization method based on polynomial fitting is designed. Combined with the specialized binary data reduction technology, the best compression effect is achieved within the controllable error range. The experimental results show that the method proposed in this paper is significantly superior to the traditional general compression tools in the compression of mechanical processing data, especially achieving a higher compression ratio while ensuring a lower distortion. Nevertheless, the current data classification relies on limited statistical features, and more complex indicators need to be introduced in the future to improve the classification accuracy. Meanwhile, the neural network structure and training scheme also need to be optimized to reduce the prediction error. In addition, the compression technology of vibration signals should be improved in combination with more advanced signal processing methods to further enhance the performance. Overall, the intelligent classification and categorization compression framework based on signal characteristics constructed in this paper provides an effective solution for the efficient storage and transmission of mechanical processing state data, which is of great significance for improving the data management and intelligence level of the manufacturing process. Subsequent studies will continue to improve the classification algorithm and compression model. Promote the continuous improvement of data processing capabilities in the field of intelligent manufacturing.

References

[1] Xiu, Le. "Integration Innovation of Big Data Analysis and Artificial Intelligence in the Field of Education." International Journal of Social Sciences and Public Administration (2024): 5(2),175-182

[2] Xiu, Le. "The Challenges and Approaches of Generative Artificial Intelligence (GenAI) to promote Educational Innovation." International Journal of Computer Science and Information Technology (2024): 4(3),400-407.

[3] Pan Y. Research on the Design of Corporate Office Data Security Protection Systems[J]. Journal of Computer, Signal, and System Research, 2025, 2(3): 31-37.

[4] Yan J. Research on the Application of Spark Technology in Natural Resource Data Management[J]. Journal of Computer, Signal, and System Research, 2025, 2(3): 45-51.

[5] Chen, Anyi. "Identification of User Preferences and Personalized Recommendation Strategy in E-commerce Platforms." Journal of Computing and Electronic Information Management (2024): 15(13),100-103.

[6] Li B. Promoting the Effectiveness of Climate Policy through Data Analysis[J]. Journal of Education, Humanities, and Social Research, 2025, 2(2): 118-124.

[7] Ma Z. Practical Experience in Data-Driven Product Transformation and Revenue Growth[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 44-50.

[8] Xu, Yue. "Research on Software Development Amd Design of Computer Network Automatic Detection and Control System." Scientific Journal of Technology (2025): 7(3),77-83

[9] Gu Y. Research on Design Principles and Maintainability of High-Performance Web Applications[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 57-62.

[10] Ding M. Quantitative Analysis of the Quantitative Impact of Optimizing User Engagement through Content Design[J]. Journal of Media, Journalism & Communication Studies, 2025, 1(1): 36-41.
[11] Yuan S. Optimization of Vulnerability Detection and Repair Strategies Based on Static Application Security Testing[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 51-56.
[12] Ding J. Exploration of Process Improvement in Automotive Manufacturing Based on Intelligent Production[J]. International Journal of Engineering Advances, 2025, 2(2): 17-23.
[13] Zou Y. Design and Implementation of Computer Network Security Monitoring System[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 63-68.
[14] Xu H. Analysis of the Impact of System Deployment on the Digital Transformation of Supply Chain[J]. Economics and Management Innovation, 2025, 2(3): 34-40.
[15] Li J. Distributed Data Processing and Real-Time Query Optimization in Microservice Architecture[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 38-43.
[16] Fu Y. The Application of Big Data Analytics to Stock and Derivatives Trading Strategies[J]. Economics and Management Innovation, 2025, 2(3): 21-27.
[17] Wei X. Value of Machine Learning and Predictive Modeling in Business Decision-Making[J]. European Journal of Business, Economics & Management, 2025, 1(1): 78-85.
[18] An C. Application and Optimization of Data Analysis in Investment Risk Assessment[J]. Economics and Management Innovation, 2025, 2(3): 28-33.
[19] Huang J. Application of Sustainable Development Concept in the Brand Strategy of High-End Pet Products[J]. European Journal of Business, Economics & Management, 2025, 1(1): 92-98.
[20] Wu H. From Large Language Models to Innovative Applications of Blockchain AI in Web3[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 11-17.
[21] Jing X. Real-Time Risk Assessment and Market Response Mechanism Driven by Financial Technology[J]. Economics and Management Innovation, 2025, 2(3): 14-20.
[22] Zhang M. Optimization of Medical Device Software Lifecycle Management Based on DevOps[J]. Journal of Medicine and Life Sciences, 2025, 1(3): 8-13.
[23] Liu Y. The Latest Application and Security Analysis of Cryptography in Cloud Storage Data Audit[J]. Procedia Computer Science, 2025, 259: 984-990.
[24] Zhang Y. Efficient Storage and Access Management System for Network Databases Driven by Cloud Computing Technology[J]. Academic Journal of Engineering and Technology Science, 2024, 7(6): 123-127.
[25] Tang Z. Research on Resource Planning and Transmission Line Optimization System of US Power Grid Based on GIS Technology[J]. safety, 2024, 8(5): 53-59.
[26] Yuan S. Research on Anomaly Detection and Privacy Protection of Network Security Data Based on Machine Learning[J]. Procedia Computer Science, 2025, 261: 227-236.
[27] Yang, D., & Liu, X. (2025). Research on Large-Scale Data Processing and Dynamic Content Optimization Algorithm Based On Reinforcement Learning. Procedia Computer Science, 261, 458-466.
[28] Bukun Ren, Calculation Model and System Implementation of Similarity Measurement of Multiple Electronic Medical Records Based on Deep Learning and Multi-Modal Abstract Extraction, International Journal of World Medicine, 2025, 6(1), 9-18
[29] Zhao, Fengyi "Computer System Security and Power Data Network Integrated Security Strategy Analysis and Optimization." Journal of Network Computing and Applications (2025), 10(1): 14-19

[30] Chen, H., Zuo, J., Zhu, Y., Kabir, M. R., & Han, A. (2024). Polar-Space Frequency-Domain Filtering for Improved Pulse-echo Speed of Sound Imaging with Convex Probes. In 2024 IEEE Ultrasonics, Ferroelectrics, and Frequency Control Joint Symposium (UFFC-JS) (pp. 1-4). IEEE.

[31] Liu B. Data Analysis and Model Construction for Crew Fatigue Monitoring Based on Machine Learning Algorithms[J]. optimization, 2024, 8(5): 48-52.

[32] Zhu P. Construction and Experimental Verification of Automatic Classification Process Based on K-Mer Frequency Statistics[C]//The International Conference on Cyber Security Intelligence and Analytics. Cham: Springer Nature Switzerland, 2024: 391-400.

[33] Fan, Yijiao "Optimization Strategy and Implementation Effect Analysis of Unstructured Data Audit for Internal Audit of Commercial Banks in the Big Data Environment."Accounting, Auditing and Finance (2024), 5(3): 7-12

[34] Chen J. Design and Implementation of a Personalized Recommendation System Based on Deep Learning Distributed Collaborative Filtering Algorithm on Social Media Platforms[C]//2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS). IEEE, 2025: 1-5.

[35] Chen, H., Yang, Y., & Shao, C. (2021). Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding. Journal of Manufacturing Systems, 58, 306-315.

[36] Cao, Y., Cao, P., Chen, H., Kochendorfer, K. M., Trotter, A. B., Galanter, W. L., ... & Iyer, R. K. (2022). Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model. In Multimodal AI in healthcare: A paradigm shift in health intelligence (pp. 245-256). Cham: Springer International Publishing.

[37] Gu Y. Javascript Code Simplification And Optimization Based On Hybrid Static and Dynamic Analysis Techniques[C]//2025 IEEE 14th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, 2025: 826-833.

[38] Zhang J. Research on User Behavior Interest Feature Extraction and Accurate Advertising Recommendation Algorithm based on Deep Learning and Improved LDA Model[C]//2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS). IEEE, 2025: 1-5.

[39] Sheng, Cheng. "The Impact and Path of Financial Technology on Accounting and Financial Informatization under the Background of Technological Revolution." (2025). International Journal of Business Management and Economics and Trade, 6(1),52-61

[40] Sheng, Cheng. "Strategies and Practices of Enterprise Financial Accounting and Financial Management in the New Financial Environment." (2025). International Journal of Social Sciences and Economic Management, 6(1), 124-133

[41] Ma, K. (2024). Employee Satisfaction and Firm Performance: Evidence from a Company Review Website. International Journal of Global Economics and Management, 4(2), 407-416.

[42] Ma, K., & Shen, J. (2024). Interpretable Machine Learning Enhances Disease Prognosis: Applications on COVID-19 and Onward. arXiv preprint arXiv:2405.11672.

[43] Ma, K. (2024). Relationship Between Return to Experience and Initial Wage Level in United States. Frontiers in Business, Economics and Management, 16(2), 282-286.

[44] Varatharajah, Y., Chen, H., Trotter, A., & Iyer, R. K. (2020). A Dynamic Human-in-the-loop Recommender System for Evidence-based Clinical Staging of COVID-19. In HealthRecSys@ RecSys (pp. 21-22).

[45] Xu Q. Enhancing User Experience in Enterprise-Level Web Applications through Technological Innovation[J]. Journal of Computer, Signal, and System Research, 2025, 2(4): 90-96.

[46] Wu, Hongjun. "Optimization of Malicious Code Detection Model for Web3. 0 Binary Files Based on Convolutional Neural Network and Graph Convolutional Neural Network." (2025). International Journal of Multimedia Computing, 6(1), 29-43

[47] Fan Y. Financial Volatility Prediction Model Based On Denoising Autoencoder and Unstable Attention Mechanism[J]. Procedia Computer Science, 2025, 261: 45-52.

[48] Hao L. Research on Application of Deep Learning in Optimizing the Performance of Autonomous Vehicles[J]. European Journal of Engineering and Technologies, 2025, 1(1): 25-31.

[49] Li B. Research on the Application and Sustainable Development of Digital Twin Technology Based on Big Data Technology and Deep Learning in the Green Transformation of Manufacturing Industry[J]. Procedia Computer Science, 2025, 261: 191-198.

[50] Chen, H., Varatharajah, Y., de Ramirez, S. S., Arnold, P., Frankenberger, C., Hota, B., & Iyer, R. (2020). A retrospective longitudinal study of COVID-19 as seen by a large urban hospital in Chicago. medRxiv, 2020-11.

[51] Hao L. Test Scenario Design and Optimization of Automated Driving Lane Keeping System Based On PCA and Intelligent Algorithm[J]. Procedia Computer Science, 2025, 261: 237-246.

[52] Li, J. (2025). Research On Optimization Model of High Availability and Flexibility of Blockchain System Based on Microservice Architecture. Procedia Computer Science, 261, 207-216.

[53] Wang C. Application of Data Analysis in Bank Mortgage Loan Risk Assessment[J]. Financial Economics Insights, 2025, 2(1): 23-28.

[54] Huang J. Digital Technologies Enabling Rural Revitalization: The Practice of AI and BIM in the Adaptive Reuse of Historic Buildings[J]. International Journal of Architectural Engineering and Design, 2025, 2(1): 1-8.

[55] Xu H. Optimization and Allocation Model of Idle Resources in Shared Supply Chain Based on Differential Game Theory and Optimal Control Theory[J]. Procedia Computer Science, 2025, 261: 167-175.

[56] Tu, X. (2025). Research on the Mechanism of Data Intelligence-Driven Collaborative Management Across the Entire Value Chain in Manufacturing and a Hybrid Data Model. Procedia Computer Science, 261, 183-190.

[57] Zhou Y. Optimization of Multi dimensional Time Series Data Anomaly Detection Model Based on Graph Deviation Network and Convolutional Neural Network[J]. Procedia Computer Science, 2025, 261: 199-206.