Research on Flexible Job Shop Scheduling Considering Equipment Failure Disturbance

Zhanyi Zhu

School of Business Administration, Liaoning Technical University, Huludao, 125000, China
2543119054@qq.com

Abstract. Addressing the scheduling disturbance issues arising from random equipment failures in an axial flexible job shop, this paper proposes a multi-objective dynamic scheduling method with the dual goals of minimizing the maximum completion time and maximizing the average equipment utilization rate. The performance of two strategies, complete rescheduling and right-shift rescheduling, is systematically compared. Firstly, a dynamic scheduling mathematical model considering failure disturbances is established, incorporating typical multi-process flexible machining paths for axial components (such as turning, grinding, gear machining, etc.). An improved non-dominated sorting genetic algorithm (NSGA-II) is employed for optimization and solution: a static Pareto front is generated in the initial stage, and after the occurrence of a failure, complete rescheduling (re-optimizing all unprocessed processes) and right-shift rescheduling (only postponing the subsequent processes affected by the failure in their original order) are executed respectively. Through the rapid non-dominated sorting and crowding distance mechanism of NSGA-II, a balance is achieved between minimizing the completion time and maximizing equipment utilization rate, and the computational efficiency and scheduling robustness of the two strategies are evaluated. Simulation experiments are conducted based on simulated data from a gearbox axial production workshop, with six types of workpieces, eight machines, a total of 42 processes, and an average equipment failure interval of 200 minutes. The experimental results show that, according to the simulated actual machining scenarios, the average completion time obtained by the complete rescheduling strategy is 1285 minutes, the average equipment utilization rate is 76.5%, and the average rescheduling computation time is 45 seconds; the average completion time obtained by the right-shift rescheduling strategy is 1356 minutes, the average equipment utilization rate is 73.8%, and the average rescheduling computation time is less than 1 second; the adaptive improved NSGA-II algorithm (ARI-NSGA-II) proposed in this paper achieves an average completion time of 1268 minutes, an average equipment utilization rate of 78.2%, a standard deviation of 9.6% in equipment utilization rate, and an average rescheduling computation time of 8 seconds under the same failure scenario. Compared to complete rescheduling, the proposed algorithm reduces the completion time by 1.3%, increases the equipment utilization rate by 1.7 percentage points, decreases the standard deviation of equipment utilization rate by 22.0%, and reduces the rescheduling computation time by 82.2%. This research provides a quantitative basis for selecting rescheduling strategies in axial flexible workshops under equipment failures.

L. W. Koon et al. (eds.), *Proceedings of the 2026 7th International Conference on Management Science and Engineering Management (ICMSEM 2026)*, Advances in Economics, Business and Management Research 399,
https://doi.org/10.2991/978-94-6239-758-3_30

Keywords: equipment failure disturbance, flexible production workshop scheduling, scheduling algorithm improvement, complete rescheduling

1 Introduction

The manufacturing process of gearbox shaft components (such as input and output shafts) involves multiple operations including turning, grinding, and gear machining. Since each operation typically requires interchangeable CNC machines, production workshops essentially function as flexible manufacturing facilities. However, frequent equipment failures during actual operations render original scheduling plans ineffective. Most workshops currently rely on manual adjustments or simplistic right-shift scheduling, resulting in slow response times, suboptimal solutions, and prominent issues like upstream process backlog, downstream equipment material shortages, and delayed scheduling. Concurrently, post-failure equipment load imbalances worsen: high-precision equipment remains overloaded for extended periods while standard equipment utilization rates remain excessively low. Existing re-scheduling strategies for flexible workshops primarily include complete re-scheduling and right-shift scheduling.To address this challenge, this study proposes an improved NSGAII algorithm based on adaptive disturbance coefficient selection for rescheduling strategies. First, a disturbance coefficient (defined as the ratio of repair time to remaining average process time) is introduced to quantify fault severity: right-shift scheduling is employed when the coefficient is low and scheduling diversity is sufficient to ensure rapid response, while complete rescheduling is triggered for global optimization otherwise. The algorithm incorporates memory-based hot-start and local repair operators to rapidly reallocate processes from faulty equipment to available units, significantly reducing rescheduling computation time. At the algorithmic level, a dynamic weighted congestion mechanism rewards solutions near underutilized equipment, guiding population evolution toward load balancing. Combined with two-stage encoding for shaft-type processes and constraint-aware operators, the method ensures solution feasibility. Experimental results demonstrate significant improvements over standard NSGAII, pure complete rescheduling, and right-shift scheduling in metrics including fault response time, completion time increment, equipment utilization standard deviation, and unit product invalid energy consumption[1]. This research provides a viable technical solution for achieving efficient[2] and energy-saving dynamic scheduling in shaft-type flexible workshops under equipment fault disturbances.

1.1 Job Description

As the most fundamental and typical NP-hard problem, the scheduling problem can generally be described as follows: n workpieces need to be processed on m machines, each with its specific processing technology and sequence. Given the constraints of predetermined processing sequences for each workpiece[3] on specific machines, the goal is to arrange the processing sequences[4] on each machine to achieve optimal scheduling results. During processing, the following constraints must be satisfied: (1)

Workpieces must follow predetermined sequences, meaning the next processing step can only begin after the previous one is completed;(2) Only one workpiece's single process can be processed simultaneously on any given machine;(3) Each process of every workpiece must be fixed on a specific machine; (4) At any given time in the workshop, each workpiece can only be processed by one machine; (5) Random factors such as machine failure and energy consumption are temporarily disregarded.

2 Algorithm Improvement Design

To address dynamic scheduling challenges in shaft processing workshops under equipment failure disturbances, this study proposes an Adaptive Rescheduling Improved NSGA-II[5] algorithm (ARI-NSGA-II).

2.1 Disturbance Coefficient Description

To quantify the severity of equipment failures, the disturbance coefficient is defined as follows:

$$p = \frac{t_{repair}}{T_{remain}} \tag{1}$$

Here, t_{repair} represents the equipment failure repair time, and T_{remain} denotes the average processing time for all current unprocessed operations. A larger p value indicates more severe fault impact, necessitating global rescheduling; a smaller p value suggests minor fault disturbance, allowing the adoption of a rapid-response right-shift strategy.

2.2 Scheduling Strategy Selection

Right-shifted rescheduling strategy: Only delays subsequent processes[6] affected by the failure in their original sequence without altering the relative order between processes or equipment allocation. Complete rescheduling strategy: After the failure occurs, all unprocessed processes undergo re-optimization scheduling.The population crowding σ_{CD}degree variance is introduced to evaluate the diversity of the current scheduling scheme, defined as follows:

$$\sigma_{CD} = \frac{1}{n}\sum_{i=1}^{n} \left(CD_i - \overline{CD}\right) \tag{2}$$

Here, $\overline{CD_i}$ denotes the crowding distance of the i-th individual, while$\overline{CD}$represents the population average crowding degree.

The adaptive strategy selection rule is as follows: if $p < \theta_p$ and $\theta_{CD} > \theta_\sigma$, then adopt the right-shifted rescheduling strategy (fast response); Otherwise, the complete rescheduling strategy (global optimization) is adopted.Based on the actual operating conditions of the shaft machining workshop, this study establishes a threshold value:

$$\theta_p = 0.2 \tag{3}$$

$$\theta_\sigma = 0.15 \times \overline{CD_{max}} \tag{4}$$

2.3 Chromosome Coding

To address process sequence constraints and equipment flexibility in shaft machining, a two-stage coding scheme is adopted, comprising a process arrangement stage and a equipment allocation stage. The process arrangement stage contains the total number of processes for all workpieces, where each workpiece number appears as many times as its corresponding process count, indicating the processing sequence. For example, if shaft workpiece A involves 3 processes, it is represented by appearing 3 times in sequence. The equipment allocation stage contains the total number of processes, with each position corresponding to the selectable equipment number for a specific process.

2.4 Constraint Perception and Variant Design

Mutation Operator: For process sequencing segments, randomly swaps the positions of two different workpieces; for equipment allocation segments, randomly replaces a specific process equipment with another device from the optional equipment set, as shown in Figure 1. After mutation execution, a legality check is performed. If the equipment allocation is invalid (e.g., the equipment does not support the process), the Minimum Available Equipment Index Repair algorithm is invoked to assign the process to the equipment with the shortest processing time in the optional equipment set.

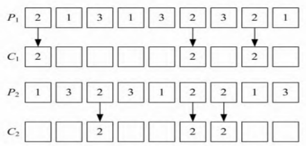

Fig. 1. Schematic diagram of cross-genetic principle

2.5 Population Initialization

First, generate an initial individual P and create a chromosome containing n × m genes. Taking a 3 × 3 case as an example, assume the generated initial chromosome is P0 = [1 1 1 2 2 2 3 3 3]. Use the rand function to shuffle the genes of P to generate new individuals. Repeat this process multiple times to form the initial population, thereby completing population initialization.

2.6 Decoding and Target Calculation

During decoding, each process is scheduled sequentially according to the process order and equipment allocation in the chromosome. For each process, after the completion of its preceding process, the earliest available idle time slot on the selected equipment is identified for insertion. Ultimately, two optimization objectives are calculated:

$$\text{Maximum completion time} f_1 = \max\{C_1, C_2 \cdots, C_n\} \tag{5}$$

$$\text{Average equipment utilization rate} f_2 = \frac{1}{m}\sum_{k=1}^{m}\frac{T_1}{T_2} \times 100\% \tag{6}$$

The total T_1, T_2processing time refers to the cumulative available time from the start of equipment operation until the completion of the last workpiece.

2.7 Dynamic Weighting Mechanism

The crowding distance metric in the NSGA-II standard only considers the distribution density of solutions within the objective space, which may lead to severe equipment utilization imbalance (some devices overloaded while others remain idle). The dynamic weighting crowding mechanism incentivizes solutions with low utilization rates to guide the population toward load-balanced equipment deployment. The equipment utilization deviation factor is defined as follows:

$$\delta = \frac{U_a - U_l}{U_a} \tag{7}$$

Here, $U_a$denotes the utilization rate of the least utilized device in the current solution, while U_l represents the average utilization rate of all devices in the current frontier. The dynamic weighting coefficients are defined as follows:

$$\beta = 0.5\ (1 - \frac{gen}{G_{max}})\ + 0.1 \tag{8}$$

Here, gengen denotes the current algebra, and $G_{max} = 200$ represents the maximum algebra. In the early stage, β is relatively high (approximately 0.5), indicating strong load balancing orientation; in the later stage, β decreases to a smaller value (approximately 0.1) to ensure convergence.

Dynamic weighted crowding degree is as follows:

$$CD_{i'} = CD_i \times\ (1 + \beta \cdot \delta) \tag{9}$$

When a solution contains devices with utilization rates significantly below average, a larger δ value amplifies CD_i, making it more likely to be selected in tournament selection and thus guiding the population toward device load balancing. Once device utilization reaches equilibrium (δ≈0), $CD_i' = CD_i$, and this does not affect convergence.

3 Simulation Experiment

3.1 Case Design

The facility comprises 8 equipment units including 3 CNC lathes, 2 external cylindrical grinders, 2 gear hobbing machines, and 1 heat treatment furnace. Six types of workpieces are processed: input shaft, output shaft, intermediate shaft, reverse gear shaft, first/second gear fork shaft, and third/fourth gear fork shaft. Each workpiece undergoes

4 to 8 machining processes totaling 42 operations, with 2-3 alternative machines available per process. Twenty fault scenarios are simulated, covering mild, moderate, and severe failure types, each running independently for 30 cycles. The equipment's mean time between failures (MTBF) is set at 200 minutes following an exponential distribution. Fault categories include: mild failures (30-minute repair time, 40% probability), moderate failures (60-minute repair time, 35% probability), and severe failures (120-minute repair time, 25% probability). Fault occurrences force immediate workpiece interruption, requiring subsequent operations on faulty equipment to be redirected to spare units or delayed until repairs. All fault scenarios are independently executed 30 times, with average results used as final performance metrics.

3.2 Experimental Conclusion

Under identical experimental conditions (6 workpieces, 8 machines, 42 processes), the Genetic Programming Algorithm achieved a completion time of 1,312 minutes with 75.3% equipment utilization rate; the Deep Q Network Algorithm demonstrated 1,285 minutes and 77.1% utilization rate; complete rescheduling required 1,285 minutes with 76.5% utilization rate; while right-shift scheduling took 1,356 minutes at 73.8% utilization rate. The proposed ARI-NSGA-II algorithm achieved optimal completion time (1,268 minutes) and highest equipment utilization rate (78.2%) without offline training, exhibiting the lowest standard deviation in utilization rate (9.6%) and moderate response time (8 seconds), outperforming all competing algorithms in comprehensive performance metrics.

3.3 Extended Experiment

To validate the algorithm's universality and scalability, two expansion experiments with varying scales were designed based on foundational tests. Expansion Experiment 1 (Small Scale): 4 workpiece types, 6 equipment units, 28 total processes, and 10 fault scenarios. This scale evaluates the algorithm's basic performance in small-scale workshops and serves as the baseline for assessing its effectiveness in typical shaft machining facilities. Expansion Experiment 2 (Larger Scale): 8 workpiece types, 10 equipment units, 56 processes, and 30 fault scenarios. This scale tests the algorithm's scalability in large-scale complex workshops. Across all expansion experiments, the average equipment failure interval remained constant at 200 minutes with unchanged fault type distribution. Each algorithm was independently executed 30 times per scale configuration.

As problem scale increases, the algorithm maintains stable optimization performance. In small-scale scenarios, average completion time is 868 minutes with equipment utilization rate of 81.5% and average response time of 4.2 seconds. In large-scale scenarios, average completion time reaches 1785 minutes with equipment utilization rate of 75.6% and average response time of 14.5 seconds. Although completion time increases with scale expansion and equipment utilization rate slightly decreases, the algorithm can still generate feasible scheduling plans within 15 seconds to meet real-time requirements of production workshops. This demonstrates the algorithm's strong universality and scalability, making it applicable to dynamic scheduling problems in shaft machining workshops of varying scales.

4 Conclusion

The research primarily achieves the following accomplishments:

In algorithm design, we implemented several targeted improvements to the standard NSGA-II: A disturbance coefficient-based adaptive rescheduling strategy was developed to enable intelligent switching between right-shift scheduling and complete rescheduling; A dynamic weighted congestion mechanism was introduced to reward solutions near underutilized devices, guiding population evolution toward load balancing optimization; A memory-based hot-start mechanism was established to accelerate rescheduling convergence by leveraging historical optimization experience; Local repair operators and constraint-aware encoding methods were developed to ensure compliance with axis-type process constraints and enhance fault response efficiency.

References

1. Tyagi N ,Tripathi P R ,Chandramouli B A . Three Machines Flowshop Scheduling Model with Bicriterion Objective Function[J]. Indian Journal of Science and Technology, 2016, 9(48): DOI:10.17485/ijst/2016/v9i48/135545.
2. Guo Z ,Wei L ,Li X , et al. Dynamic flexible job shop scheduling based on multi-dimensional space collaborative guidance evolution[J].Computers and Operations Research, 2025, 184107243-107243.DOI:10.1016/J.COR.2025.107243.
3. Liu M ,Fang S ,Dong H . Multi-fidelity digital twin based optimization for aircraft overhaul shop scheduling[J]. Journal of Manufacturing Systems, 2025, 82700-729. DOI:10.1016/J.JMSY.2025.07.018.
4. Zhao Y ,Luo X ,Xu W , et al. A deep reinforcement learning optimization algorithm based on heterogeneous graph neural network for hybrid flow shop scheduling problem with finite transportation resources[J].Engineering Applications of Artificial Intelligence, 2025, 161(PB): 112096-112096. DOI:10.1016/J.ENGAPPAI.2025.112096.
5. Halder T ,Debnath K B .Multi-objective closed-loop supply chain inventory model with learning and forgetting under carbon emission policies using NSGA-II, MOPSO, and TOPSIS[J]. Applied Soft Computing, 2025,180113291-113291. DOI:10.1016/J.ASOC.2025.113291.
6. Frihat M ,Alouane H B A . A branch-and-cut algorithm for the integrated employee timetabling and hybrid job-shop scheduling problem with time lags and setup times[J].Computers & Industrial Engineering,2025,208111351-111351.DOI:10.1016/J.CIE.2025.111351.