Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 8, No. 1, 2026

Flexible Workshop Scheduling Model and Analysis Based on the Improved NSGA-II Algorithm

Wenlu Zhang, Yun Xu, Hailin Liao, Tao Zhang

Sichuan University of Science and Engineering, Yibin, Sichuan, 644000, China

Abstract: To address the problems of low production efficiency and uneven equipment load in flexible manufacturing shop scheduling, this study constructs a multi-objective scheduling model. The model aims to minimize the maximum completion time (makespan) and maximize equipment load balancing. To efficiently solve this model, an improved NSGA-II algorithm is proposed. To address the tendency of the traditional NSGA-II to converge to local optima, the improved algorithm incorporates an adaptive local search strategy. This enhancement strengthens the algorithm's exploitation capability and improves the diversity of the solution set. To validate the effectiveness of the model and the algorithm, simulation experiments were conducted on the MATLAB platform using the Kacem international benchmark instances. The results obtained from the improved algorithm were then compared and analyzed against those from other algorithms. Experimental results show that the scheduling solutions obtained by the improved NSGA-II algorithm proposed in this study outperform others in terms of both maximum makespan and equipment load balancing. Compared with the traditional NSGA-II algorithm, the equipment load balancing index has improved by 20%, which fully demonstrates the effectiveness and superiority of the proposed scheduling optimization model and the improved algorithm.

Keywords: Flexible job shop scheduling; Improved NSGA-II algorithm; Load balancing

1. Introduction

Against the backdrop of the new global wave of technological revolution and industrial transformation, developing intelligent manufacturing has become a core strategy for enhancing national industrial competitiveness. As a critical component of intelligent manufacturing, the flexible job shop scheduling problem within manufacturing enterprises has consistently been a focal point of common concern in both academic and engineering communities. After decades of development, production scheduling theory has evolved into a relatively mature research framework. Among its various branches, Integrated Process Planning and Scheduling (IPPS), due to its effectiveness in addressing the high complexity and strong dynamics of modern manufacturing systems, has emerged as a significant research direction within this field. The research focus has gradually shifted from early-stage model construction to algorithm optimization and framework design for multi-constraint, multi-objective problems.

In terms of extending IPPS models, scholars continuously incorporate practical production constraints to enhance model applicability. Ba Li[1] addressed the insufficient consideration of lot sizing in the integrated process planning and scheduling problem by incorporating lot-splitting issues into the IPPS model and proposed a particle swarm optimization algorithm for its solution, thereby enhancing scheduling flexibility. Ba L[2] addressed the issue of significant process variations among workpieces in batch production by proposing an equal lot-splitting strategy. Furthermore, they incorporated transportation considerations into the Integrated Process Planning and Scheduling (IPPS) framework, establishing a mathematical model for the Batch Variation-IPPS (BV-IPPS) problem. This approach yielded an optimized solution better suited to practical production environments. Demir H I[3] extended the IPPS model by incorporating product delivery dates. They employed a hybrid approach combining stochastic search, genetic search, and other heuristic techniques, assigning differentiated penalty weights to each task within the objective function. Validation confirmed that their method can generate feasible schedules meeting predetermined deadlines while improving shop floor utilization. Regarding algorithmic advancements, meta-heuristic algorithms remain the predominant solution approach. Jin L[4] proposed an IPPS model that simultaneously considers the minimization of makespan and energy consumption. By employing a multi-objective memetic algorithm integrated with variable neighborhood search technology and the TOPSIS decision-making method, the approach determines the most satisfactory non-dominated solutions. Experiments demonstrate that this algorithm can effectively balance the makespan criterion and total power consumption criterion, thereby achieving energy-saving objectives. Qiao L[5] proposed a multi-level process network graph construction method based on product bill of materials and process characteristics. By improving the particle swarm optimization algorithm, invalid search paths in network subgraphs were avoided, enhancing algorithmic efficiency and addressing the scheduling problem in multi-shop production. Zhang L[6] building upon the AND/OR graph, proposed a general framework for implementing constructive meta-heuristics within a solution model. This framework encapsulates the computation of search boundaries and state transitions, providing two interfaces to accommodate different constructive search algorithms. Using the ant colony optimization algorithm as an example, they validated that this method demonstrates good effectiveness and high performance, achieving the minimization of the maximum completion time set.

In summary, the scheduling of multi-variety, multi-process, and technologically complex parts has become a prominent research focus. Increasingly, scholars are incorporating more practical production factors into the optimization objectives beyond minimizing makespan within the IPPS framework

based on FJSP. However, the issue of uneven equipment load in actual production remains insufficiently addressed. Therefore, this paper focuses on flexible manufacturing workshops, taking makespan and equipment load balancing as optimization objectives. It improves the NSGA-II algorithm by integrating a local search strategy to compensate for its insufficient local search capability, aiming to obtain scheduling solutions with superior performance.

2. Problem Description and Modeling

Traditional research on workshop scheduling has primarily focused on optimizing explicit metrics such as processing time, energy consumption, and total load, often overlooking the efficiency of dynamic matching between task demands and equipment resources in the temporal dimension. This oversight neglects the critical impact of equipment load balancing on the overall effectiveness of the production system. In practical flexible manufacturing workshops, it is common to process multiple types of workpieces concurrently, with significant differences in the type and capability requirements of processing equipment for different workpieces. If scheduling strategies fail to adequately consider this heterogeneity and achieve precise matching, it can easily lead to situations where the specific equipment necessary for processing certain workpieces is occupied by other tasks or not allocated in a timely manner. Consequently, this results in substantial load disparities among similar equipment. Such load imbalance not only directly triggers production bottlenecks, prolongs workpiece waiting times, and reduces overall equipment utilization but may also increase the risk of equipment failure due to localized overload, ultimately undermining the stability and efficiency of the production system.

Therefore, constructing an efficient intelligent scheduling solution must simultaneously address two core challenges. Firstly, to enhance production efficiency, it is essential to optimize process sequencing and resource allocation, while possessing the dynamic adjustment capability to respond to disturbances such as equipment failures and emergency order insertions, thereby minimizing the production cycle. Secondly, to improve equipment load balancing, it is necessary to fully consider the complex many-to-many matching relationships and dynamic constraints between equipment and workpieces, achieving intelligent and dynamic resource allocation. This approach aims to prevent equipment overload while reducing resource idle time.

Moreover, there often exists an inherent conflict or trade-off between the objectives of production efficiency and equipment load balancing. For instance, in pursuit of the shortest makespan, tasks may be preferentially allocated to a limited number of highly efficient machines, which can lead to severe load imbalance. Conversely, enforcing strictly equal load distribution might significantly prolong the total completion time. Therefore, optimizing individual performance metrics in isolation is unlikely to yield a globally optimal scheduling solution. It is imperative to employ a multi-objective collaborative optimization approach, which seeks an optimal balance among these competing objectives and provides decision-makers with a diverse set of high-quality scheduling alternatives.

Specifically, the Flexible Job-Shop Scheduling Problem (FJSP) addressed in this study can be formally described as follows: In a flexible manufacturing workshop comprising M processing machines, I different types of workpieces need to be processed. The manufacturing of each workpiece involves multiple operations, where each operation can be performed by any one machine from a given set of eligible machines. Both the sequence of operations for the same workpiece and the processing order of different workpieces on the same machine possess a degree of flexibility. The optimization objective of this problem is to assign a suitable processing machine for each operation of every workpiece and determine the processing sequence of all operations on each machine. This aims to collaboratively optimize the system's makespan and equipment load balancing index while satisfying all process constraints, thereby enhancing overall system performance. Given the process requirements of the workpieces, the following assumptions are made to reduce the complexity of the scheduling optimization model:

(1) Each machine can process only one operation at a time.
(2) Once started, an operation cannot be interrupted.
(3) Resources such as fixtures and cutting tools are sufficient, and the processing of different workpieces does not interfere with each other.
(4) The processing time for each operation on different machines is a fixed value, ignoring time fluctuations caused by factors such as tool wear.
(5) Machines of the same model have identical processing capabilities.
(6) Loading and unloading times of workpieces are ignored.

Based on the problem description and assumptions above, the parameters and objective functions are established as shown in Table 1:

Table 1 Parameter Definitions

Parameter	Description
M	Set of processing equipment indices
T	Set of processing equipment types
n_t	Number of equipment of type t
p_m	Load proportion of equipment m within its type t
M_m	Machine index
O	Set of operation indices
$O_{j,k}$	The k-th operation of workpiece j
$S_{j,k}^m$	Start time of the k-th operation of workpiece j on equipment m
$T_{j,k}^m$	Processing time of the k-th operation of workpiece j on equipment m
$C_{j,k}^m$	Completion time of the k-th operation of workpiece j on equipment m
C_{max}	Makespan (maximum completion time) of a batch of workpieces

(1) Minimize the makespan

$$X_{j,k}^m = \begin{cases} 1 & \text{The machining machine for operation } O_{j,k} \text{ is } m. \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$$C_{i,k}^{m} = S_{i,k}^{m} + \sum_{k=0}^{K} T_{i,k}^{m} \cdot X_{i,k}^{m} \quad (2)$$

$$C_{max} \geq \sum_{i=1}^{I} C_{i,k}^{m} \quad (3)$$

$$f_1 = min\ C_{max} \quad (4)$$

(2) Maximize the processing equipment load balancing degree

$$B_t = 1 - \sqrt{\frac{1}{n_t}\sum_{m \in M_t} \left(p_m - \frac{1}{n_t}\right)^2} \quad (5)$$

$$B_{machine} = \frac{1}{T}\sum_{t=1}^{T} B_t \cdot 100\% \quad (6)$$

$$f_2 = max\ B_{machine} \quad (7)$$

3. Improved NSGA-II Algorithm for Solving the FJSP

3.1. Chromosome Encoding and Decoding

To fully describe the complex decision variables within the scheduling model, a three-layer segmented chromosome encoding scheme has been designed. This encoding scheme can simultaneously and clearly define the three core decisions related to workpieces, operations, and processing equipment assignment, while ensuring consistency with process constraints. Specifically, the first layer is the workpiece scheduling layer, the second layer is the operation layer, and the third layer is the processing equipment assignment layer, which shares the same length as the preceding two layers.

Workpiece Scheduling Layer	1	2	3	1	1	3	2	3	2	2
Operation Scheduling Layer	101	201	301	102	103	302	202	303	203	204
Processing Equipment Assignment Layer	1	2	3	1	2	1	1	2	1	1

Figure 1. Encoding Scheme of the Improved NSGA-II Algorithm

The core idea of decoding is to transform the existing three-part encoding into an executable production line schedule. For the static scheduling problem, a forward scheduling decoding method is adopted, which is simple and efficient. Its fundamental principle is to first identify the workpiece to be processed based on the sequence defined in the workpiece scheduling layer. Then, the current operation to be scheduled is obtained from the second layer, and the processing machine for this operation is determined by the equipment assignment layer. The system searches for the earliest feasible time slot to start processing this operation, updates the machine's status accordingly, and repeats this process until the final operation is completed. The decoded schedule is typically visualized using a Gantt chart, which provides a clear and intuitive display of job sequences and time allocations, facilitating the verification of schedule feasibility.

3.2. Population Initialization

When solving scheduling problems with the NSGA-II algorithm, an initial population must be generated according to the encoding rules and constraints. The quality of the initial population is a key factor affecting the algorithm's convergence speed and the diversity of the solution set. Therefore, aiming at the optimization objective of processing equipment load balancing, generation rules are established based on group load tracking and the principle of assigning to the least-loaded machine. The specifics are as follows:

(1) Generating Workpiece Sequences by Random Permutation: Firstly, based on the number of operations for each workpiece, a sequence containing a corresponding number of workpiece identifiers is generated, which implicitly ensures the precedence constraints of operations. This sequence is then globally randomized. This approach maximizes the diversity of the initial population in the dimension of operation sequencing while satisfying hard constraints, providing a broad starting point for the evolutionary algorithm's search.

(2) Assigning Processing Equipment Using a Group and Processing-Time Based Load Balancing Strategy: Firstly, processing equipment is divided into several functional groups based on their processing capabilities, where equipment within the same group can process similar operations. Then, a group is randomly selected, and the available machine with the smallest accumulated processing time load within that group is assigned to the operation to be scheduled. After each assignment, the accumulated load of that machine is updated in real-time. This strategy achieves preliminary load balancing within groups and provides a good foundation for optimizing equipment load at a global level.

3.3. Chromosome Selection

In the NSGA-II algorithm, two selection operations are typically performed: parent selection and environmental selection. Parent selection involves choosing superior individuals from the population to serve as parents, which then undergo crossover and mutation to generate the offspring population, representing a pre-reproduction selection. Environmental selection, on the other hand, is a post-reproduction process aimed at controlling population size while preserving elite individuals to form the new generation. This paper employs the binary tournament selection method for both selection operations, with the following rules and steps:

(1) Two distinct individuals are randomly selected for comparison.

(2) The non-domination rank of the two individuals is compared. The individual with the lower (better) rank wins.

(3) If the non-domination ranks are equal, their crowding distance is compared. The individual with the larger crowding distance wins.

(4) If both the non-domination rank and crowding distance are equal, one individual is randomly chosen as the winner.

3.4. Crossover and Mutation of Chromosomes

The crossover operation is a key mechanism for maintaining population diversity and inheriting favorable genetic material. This paper designs a layered hybrid crossover strategy, employing distinct crossover methods for the operation sequence and the equipment assignment, respectively. For the Workpiece Sequence Encoding: To

strictly adhere to the inherent number and precedence constraints of operations for each workpiece, the Job-Order Crossover (JOX) method is employed. This method preserves the absolute positions of key workpieces from the parents, ensuring that offspring inherit favorable operation sequences while always satisfying process feasibility. For the Equipment Assignment Encoding: Since equipment load balancing is a critical factor affecting the production cycle and efficiency, it is beneficial for the crossover to inherently guide solutions toward balance. Therefore, a Load-Aware Uniform Crossover method is adopted. During crossover, this method considers the current processing load of each machine, preferentially selecting feasible machines with lower loads. This guides the evolution of offspring toward solutions with better load balance.

3.5. Local Search Optimization Strategy

The NSGA-II algorithm performs well in scheduling problems with two or fewer objectives, demonstrating fast convergence and evenly distributed solutions. However, in multi-objective problems involving three or more objectives, it exhibits shortcomings such as insufficient solution diversity in the middle and later stages of iteration and a tendency to converge to local optima. Therefore, an adaptive neighborhood local search strategy is introduced following the crossover and mutation operations. This strategy performs low-frequency, shallow, and constrained local search on elite solutions to achieve fine-tuned improvements, effectively enhancing the quality of the Pareto front while maintaining diversity and global search capability.

The adaptive nature of this strategy refers to the dynamic adjustment of the local search invocation frequency according to the evolutionary generation, balancing the algorithm's exploration and exploitation capabilities. Each time it is triggered, only a proportion of elite individuals are randomly selected from the first non-dominated front of the newly generated population for local search, thereby strictly controlling computational overhead. The triggering interval function for local search, L(g), is defined as follows:

$$L(g) = \begin{cases} I_{high}, & g \le \alpha G_{max} \\ I_{mid}, & \alpha G_{max} < g \le \beta G_{max} \\ I_{low}, & g > \beta G_{max} \end{cases} \quad (8)$$

3.6. Algorithm Termination Criterion

The algorithm terminates and exits the loop when either of the following conditions is met: (a) the iteration count reaches a predefined maximum, or (b) the improvement in the Pareto front is less than 1% for 20 consecutive generations. Upon termination, an approximate Pareto-optimal solution set is obtained.。

4. Simulation Experiments and Analysis

To validate the superiority of the improved NSGA-II algorithm in solving the static scheduling model, two international benchmark instances (8×8 and 10×10) proposed by Kacem et al. [35] were selected for verification. These benchmark instances are classic multi-objective FJSP problems, aligning with the DE scheduling problem studied in this paper. The 8×8 instance represents a set containing 8 workpieces to be processed and 8 machines, while the 10×10 instance contains 10 workpieces and 10 machines. Each instance was independently run 10 times, with the maximum completion time (makespan) and the equipment load balancing degree as the optimization objectives. The average values from these runs were used for comparison.

The improved NSGA-II algorithm was implemented in MATLAB. The simulation experiments were conducted using MATLAB R2022b. The algorithm parameters were set as follows: population size, maximum number of iterations, crossover probability, and mutation probability. As shown in Figures 2 and 3, the Gantt charts for the optimal solutions of the 8×8 and 10×10 instances, respectively, after running the improved NSGA-II algorithm are presented. The results obtained by the proposed improved NSGA-II algorithm were compared with those from: the traditional NSGA-II algorithm, the improved NSGA-II algorithm proposed by Cao Haitao[7], and the MOPSO algorithm proposed by Huang S[8]. The statistical results of this comparison are presented in Table 2.

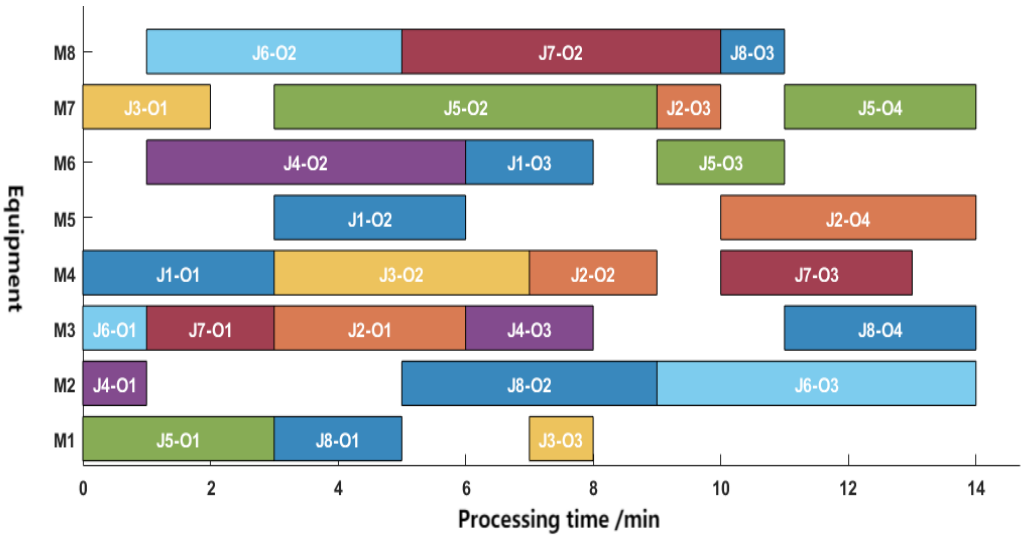


Figure 2. An Optimal Scheduling Solution for the 8×8 Instance

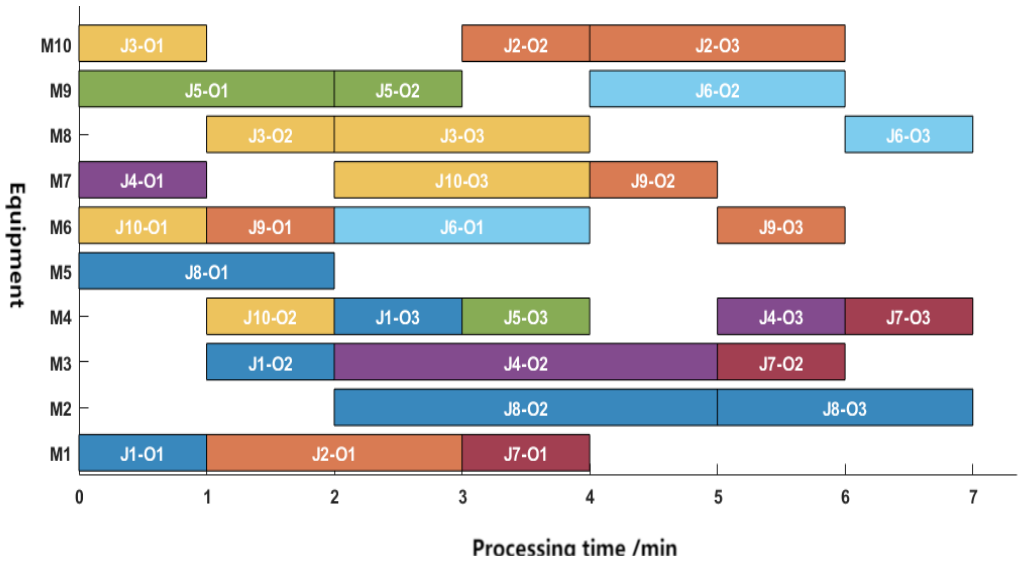


Figure 3. An Optimal Scheduling Solution for the 10×10 Instance

Table 2 Comparison of Results for the Kacem Instances

Instance	Objective	NSGA-II	MOPSO	Improved NSGA-II	This Paper's NSGA-II
Kacem	Makespan/min	17	16	16	14
8×8	load balancing degree	75%	77 %	77%	82%
Kacem	Makespan/min	12	8	7	7
10×10	load balancing degree	62%	63%	72%	82%

As shown in the comparative data in Table 2, for both Kacem benchmark instances, the improved NSGA-II algorithm proposed in this study obtains the optimal makespan. It also demonstrates an advantage in the equipment load balancing index, exhibiting a 20% improvement compared to the traditional NSGA-II algorithm.

5. Conclusion

This study constructed a static scheduling optimization model for flexible manufacturing workshops and proposed a corresponding improved NSGA-II solution algorithm. Firstly, a multi-objective optimization model was established with the core objectives of minimizing the maximum completion time (makespan) and minimizing equipment load standard deviation. This model integrates operation sequencing and machine assignment into a unified optimization framework. To address the characteristics of this model, a three-layer encoding and decoding mechanism was designed, comprising a workpiece scheduling layer, an operation scheduling layer, and a machine assignment layer. A population initialization method guided by load balancing was proposed to enhance the quality of initial solutions. Furthermore, an adaptive local search strategy was introduced to strengthen the algorithm's local exploitation capability. Simulation tests conducted on Kacem international standard benchmark instances verified the superior quality of solutions obtained by the improved algorithm. The results demonstrate that the proposed model and algorithm can effectively generate scheduling solutions characterized by tightly coupled operations and well-balanced resource loads.

References

[1] Ba Li, Li Yan, Yang Mingshun. Research on IPPS problem considering flexible lot splitting[J]. Mechanical Science and Technology for Aerospace Engineering,2017,36(03):426-435.DOI:10.13433/j.cnki.1003-8728.2017.0317.

[2] Ba L, Li Y, Yang M, et al. A mathematical model for multiworkshop IPPS problem in batch production[J]. Mathematical Problems in Engineering, 2018, 2018(1): 7948693.

[3] Demir H I, Kökçam A H, Erden C. Integrated Process Planning, Scheduling, and Due-Date Assignment[M]. CRC Press, 2023.

[4] Jin L, Zhang C, Fei X. Realizing energy savings in integrated process planning and scheduling[J]. Processes, 2019, 7(3): 120.

[5] Qiao L, Zhang Z, Huang Z. A scheduling algorithm for multi-workshop production based on BOM and process route[J]. Applied Sciences, 2021, 11(11): 5078.

[6] Zhang L, Wong T N. Solving integrated process planning and scheduling problem with constructive meta-heuristics[J]. Information Sciences, 2016, 340: 1-16.

[7] Cao Haitao, Research on multi-objective flexible job-shop scheduling problem based on improved NSGA-II [D]. Zhejiang University of Technology.,2019.

[8] Huang S, Tian N, Wang Y, et al. Multi-objective flexible job-shop scheduling problem using modified discrete particle swarm optimization[J]. SpringerPlus, 2016, 5(1): 1432.