

# Nurse Scheduling Algorithms based on Different Scenarios

Zhichao Zhang

Southwest Minzu University, Chengdu 610000, China

**Abstract:** This algorithm is based on a combination of integer programming, genetic algorithm, and decision tree to study nurse scheduling management problems in different scenarios. Firstly, the usage scenarios are classified based on the number of department members and energy levels. Then, the initial solution is obtained through integer programming based on information such as the actual energy level of nurses in the hospital. Finally, genetic algorithm is used to optimize the initial solution to obtain a final solution that is more in line with practical requirements. This algorithm can automatically classify or manually set use scenarios, and automatically obtain shift scheduling information by entering personalized parameters, which can meet the personal shift scheduling needs of nurses to the greatest extent, and can greatly reduce the high-frequency shift scheduling workload of Matron.

**Keywords:** Integer Programming; Evolutionary Algorithms; Decision Tree; Nurse Scheduling.

## 1. Introduction

With the continuous deepening of China's medical and health system reform, as well as the improvement of the medical and health system and the various needs of patients, hospitals need to meet the needs of various patients as much as possible with limited medical resources. The workload of medical workers is becoming increasingly huge, and many hospitals use manual methods for nurse scheduling. Relying solely on the past experience of managers not only has low efficiency, but also cannot achieve fairness and justice, It can no longer meet the growing demand for information management in hospitals. Therefore, how to make reasonable use of nurse resources has become the focus of hospital management. Nurse scheduling is an important part of nursing management, which is the foundation for completing nursing work and meeting the multi-dimensional needs of patients [1]. It directly determines the efficiency and quality of nursing work [2].

Early research on nurse scheduling can be traced back to the 1960s. But in China, the research on nurse scheduling started relatively late. Shen Yindong et al. solved the scheduling problem of 28 nurses within a week by designing a series of dynamic transformation rules and solving a constrained nurse scheduling model[3]. WARNER [4,5] developed a mathematical planning model for the scheduling of recurrent nurses based on the original nurse scheduling model. WARNER's model aims to reduce nurses' complaints about scheduling. During the implementation process, it was considered how to minimize the cost caused by the shortage of nurses due to the minimum number of registered nurses, professional nurses, and all personnel working on shifts of 3-4 days per week. The shortened scheduling time makes it difficult for the algorithm to consider the special requirements of nurse scheduling. Zeng Fenghua used integer programming and Evolutionary algorithm to realize automatic typesetting, but the conditions considered were relatively few, far from the reality. On the basis of Zeng Fenghua's algorithm, this paper sorted out the constraint conditions that are more suitable for the actual conditions of Chinese hospitals through field research, and modified the

more suitable and customizable objective function to solve the initial solution, which can not only be more practical, but also reduce the irrationality of Evolutionary algorithm that needs to be corrected, and improve the Algorithmic efficiency to a certain extent.

This paper plans to design an automatic nurse scheduling algorithm based on integer programming, Evolutionary algorithm and decision tree. A reasonable and fair scheduling scheme is more conducive to nurses to deal with patient needs, reduce the workload of nurses to work more attentively, and can also greatly reduce the Matron's scheduling management time, and pay more attention to nurses and patients; By using decision trees to classify personnel and automatically recommend the most suitable constraint selection, the difficulty of usage is greatly reduced, achieving the goal of standardized management and improving work efficiency.

## 2. Methodology

### 2.1. Integer Programming Model Parameters

Table 1. Parameter Definition Table

Parameter	meaning
M, m	M represents a collection of m nurses, $m=1,2,3,\dots$
J, j	J represents the set of dates within a week, $j=1,2,3,4,5,6,7,\dots$
K, k	K represents the set of nursing staff (nurses) shift types, $k=1, 2, 3$
G, g	G represents the set of competency levels for nursing staff (nurses), $g=1,2,3,4,\dots$
NightMax	The maximum number of night shifts per nurse per week
DayMax	The maximum number of night shifts per nurse per week
DayMin	Minimum number of employees per day
WeekMax	Maximum number of working hours per nurse per week
WeekMin	Minimum value per nurse per week
Y1	The number of A shifts per day, with day shifts replacing those below A shifts
Y2	Number of daily P shifts, below P shifts, use before midnight shift instead
Y3	Number of N shifts per day, below N shifts, use the small hours instead
A	A represents a set of leave requests and directed scheduling
B	B represents the set of constraint selection, $B=1,2,3,\dots$

Integer programming is usually used to solve optimization

problems with integer variables. By establishing objective functions, such as minimizing the number of personnel or working time, and setting different constraints, including reasonable resource consumption, effective working time, etc., the optimal solution that meets work needs is obtained [6].

The Table 1 are the parameters used in this model and their meanings.

## 2.2. Variables and Their Definitions

$$X_{mjkg} = \begin{cases} 1, & \text{If nurse } m \text{ has a shift type of } k \text{ and a level of } g \text{ on day } j \\ 0, & \text{Other case} \end{cases}$$

## 2.3. Hard Constraints

Hard constraints are constraints that must be followed.

1. The maximum number of night shift days per person per week is 1 day

$$\sum_{m=1}^M \sum_{j=1}^J \sum_{g=1}^G X_{mj3g} \leq \text{NightMax}$$

2. There must be a high-energy nurse on duty every day on the day shift

$$\sum_{m=1}^M \sum_{k=1}^k \sum_{g=1}^G X_{mjkg} \geq 1, \forall j \in J$$

3. Customization of the corresponding relationship between nurse level and shift level

$$\sum_{g=g_1}^G \sum_{j=1}^J \sum_{k=1}^K X_{mjkg} = 0, \forall m \in M$$

4. Setting of Desired Vacation and Desired Shift

The desired shift is that nurses must be on a certain type and level of shift on a certain day and the Desired Vacation is that the nurses must take a day off.

$$X_{mjkg} = 1, X_{mjkg} = 0, \forall j \in J, m \in M, k \in K, g \in G$$

5. Cannot take on day shift after the small hours

$$\sum_{g=1}^G X_{mj3g} + \sum_{g=1}^G X_{m(j+1)1g} \leq 1, \forall j \in J, m \in M$$

6. Each person can only work one shift per day

$$\sum_{k=1}^K \sum_{g=1}^G X_{mjkg} \leq 1, \forall j \in J, m \in M$$

7. Cannot connect to the day shift on weekends in the next week or next night on Mondays

$$\sum P + \sum_{g=1}^G X_{m11g} \leq 1, \forall m \in M$$

Among them, the scheduling cycle for a single week starts on Monday and ends on Sunday, where P is the historical data for last week.

8. The number of people on the day shift is Y1, the number of people on the night shift is Y2, and the number of people on the night shift is Y3

$$\begin{aligned} \sum_{j=1}^J \sum_{g=1}^G X_{mj3g} &= Y3, \sum_{j=1}^J \sum_{g=1}^G X_{mj2g} \\ &= Y2, \sum_{j=1}^J \sum_{g=1}^G X_{mj1g} = Y1, \forall m \\ &\in M \end{aligned}$$

9. Every time period has people working

$$\sum_{m=1}^M \sum_{k=1}^K \sum_{g=1}^G X_{mjkg} \geq 1, \forall j \in J$$

10. Scope of work per person per week

$$\begin{aligned} \sum_{k=1}^K \sum_{g=1}^G \sum_{j=1}^J X_{mjkg} &\leq \text{DayMax}, \sum_{k=1}^K \sum_{g=1}^G \sum_{j=1}^J X_{mjkg} \\ &\geq \text{DayMin}, \forall m \in M \end{aligned}$$

11. Daily Workforce Range

$$\begin{aligned} \sum_m^M \sum_k^K \sum_g^G X_{mjkg} &\leq \text{WeekMax}, \sum_m^M \sum_k^K \sum_g^G X_{mjkg} \\ &\geq \text{WeekMin}, \forall m \in M \end{aligned}$$

## 2.4. Soft Constraints

Soft constraints are the conditions that need to be satisfied as much as possible. If the optimal solution cannot be obtained after 1000 optimizations, the current solution is considered the optimal solution.

1. Each person's vacation days should be as continuous as possible.

2. Each person's night shift should have three consecutive night shifts followed by two rest periods.

3. Meet the personal work tendency of nurses

## 2.5. Partial Objective Functions for Integer Programming

$$\min Z = \sum_{m=1}^M N_m \sum_{m=1}^M \sum_{j=1}^J \sum_{k=1}^K \sum_{g=1}^G X_{mjkg}$$

$N_m$  is the tendency of each nurse to work

## 2.6. Genetic Algorithm

Step 1. Encode the scheduling obtained from integer programming to make it a binary string.

Step 2. Slice the encoded results, separate binary strings belonging to different nurses, divide a large chromosome into multiple small chromosomes, calculate fitness for these small chromosomes, use roulette wheel algorithm for selection, and then use these small chromosomes for crossover and mutation operations. This slicing operation can greatly avoid violating certain hard constraints, thereby accelerating iteration efficiency, Automatically removing most of the offspring that will greatly reduce fitness.

Step 3. Merge the slices into a whole and calculate the total fitness.

Step 4. Store the above results in the population to expand the population.

Step 5. Check if the iteration conditions are met. If yes, select the output with the highest fitness. Otherwise, use the roulette wheel algorithm to decide whether to choose from the

population with high or low fitness, and then proceed to step 2 for cyclic iteration.

## 2.7. Classification Decision Tree

### 2.7.1. The Concept of Decision Trees:

Decision Tree is a common algorithm in machine learning, with a very simple idea, just like the process of making decisions through selection. Decision tree is a basic classification and regression method, called a classification tree when used for classification and a regression tree when used for regression. In general, a decision tree contains a root node, several internal nodes, and several leaf nodes. The basic principle of its classification is to cluster up layer by layer based on the similarity of each category, with each clustering point forming a leaf node. Under the condition of predicting the sample information of each land feature category, the node can also be further subdivided downwards, dividing the sample data into several subsets; By analogy, clustering layer by layer up until the root node is reached, completing the division and extraction of different categories.

### 2.7.2. Establishment of Classification Rules:

Based on the above constraints, a total of three recommendation modes are established: 1. Human resource abundant mode: daily workload is constant, and the total workload of each nurse per week is within the range; 2. Insufficient human resources mode: The total workload of each nurse per week is a fixed value, and the daily workload is a range; 3. The human resource shortage mode that can meet the level arrangement: the total workload of each nurse per week is a fixed value, the daily workload is a range, and it can ensure the presence of high-level nurses in each shift type. Based on multiple scheduling results generated multiple times, using randomly generated parameters such as the energy level relationship between the number of people and the corresponding number, scheduling is carried out through a program based on the constraints of three different modes. A scheduling table is generated based on the constraints that can be met to determine the classification results of this group of parameters. Analyze the parameters and quantify them into three features: the first feature is whether the number of nurses is less than 10, with values of yes and no; the second feature is whether the number of nurses is odd or even, with values of odd and even; the third attribute is the energy level distribution, with values of the proportion of the number of nurses at each energy level.

### 2.7.3. Decision Tree Construction Method:

The ID3 (Iterative Dichotomiser3) algorithm was proposed by J.R. Quinlan in the 1980s. Its core idea is to use information gain as the basis for selecting splitting attributes, and select the attribute with the highest information gain as the splitting attribute of node N. ID3 algorithm has a fast classification speed and testing speed, and this feature is divided into discrete eigenvalues, which effectively does not need to consider the problem of attribute continuity, so ID3 algorithm is used here.

Information entropy is the most commonly used indicator to measure the purity of a sample set. Assuming that the proportion of the k-th class of samples in the current sample set D is, then the information entropy of D is

$$Ent(D) = - \sum_{k=1}^{|y|} P_k \log_2 P_k$$

The smaller the value of Ent(D), the higher the purity of D.

Information gain:  $Gain(D, a) = Ent(D) - \sum_{v=1}^V \frac{|D^v|}{|D|} Ent(D^v)$

Generally speaking, the greater the information gain, the greater the "purity improvement" obtained by using attribute a for partitioning.

## 3. Results and Discussion

### 3.1. Basic Data

This algorithm is based on the real scene of a department in a hospital to build a nurse scheduling algorithm based on integer programming and Evolutionary algorithm. Through communication with the hospital, we obtained information about the actual scheduling situation and the nurses in a certain department over a period of time, as shown in Table 2.

Table 2. Nurse Information Table

Level	Number of people
N0	6
N1	7
N2	5
N3	8

Explanation: The levels of N0-N3 are gradually increasing, and their abilities are becoming stronger. Only nurses with N1 and above can work night shifts alone.

At present, the most commonly used scheduling modes in China are traditional weekly scheduling mode, APN continuous scheduling mode, SHE scheduling mode, and flexible scheduling mode [8]. This algorithm adopts the APN scheduling mode. The APN scheduling time is divided into two types: (8-4, 4-12, 12-8) and (8-3, 3-10, 10-8).

### 3.2. Empirical Integer Programming

Using the data in this article, a total of 2184 variables and 2673 constraints need to be solved. The algorithm is programmed using Python, and the output results are mapped to a table for display. The results are shown in Table 3.

From the analysis in the table, it can be seen that the output results fully comply with the hard constraint requirements, and it is also evident that the number of N shifts is lower than other shift types, and the gap in the number of AP shifts is controlled within a certain range, which is a usable layout data.

### 3.3. Demonstration of Evolutionary Algorithm

Evolutionary algorithm is used to optimize the initial solution of integer programming, and the results are shown in Table 4.

After calculation, the penalty value of this optimization result is the lowest after 10000 iterations. Compared with the results in 3.2, the number of consecutive breaks has been improved to a certain extent.

### 3.4. Evaluation of Decision Tree Classification Effect

For the training and validation of decision trees, there are multiple ways to partition the dataset.

1. Common method: Divide the dataset into two parts: the training set and the validation set.

2. Reservation method: Divide the dataset into two mutually exclusive sets, one of which serves as training set S and the other as testing set T. After training the model on S, evaluate the testing error on T.

**Table 3.** The initial results after integer programming

Number	Name	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	Nurse A	A4.0	N2.0	/	A1.0	A2.0	/	N3.0
1	Nurse B	A2.0	A3.0	/	A2.0	A1.0	A2.0	/
2	Nurse C	A2.0	A2.0	A2.0	N1.0	/	/	A2.0
3	Nurse D	A1.0	A1.0	A1.0	/	A1.0	/	P1.0
4	Nurse E	N3.0	/	A2.0	A2.0	A1.0	A3.0	/
5	Nurse F	A2.0	/	P2.0	P2.0	A1.0	/	A1.0
6	Nurse G	/	A1.0	A2.0	A1.0	/	A1.0	A1.0
7	Nurse H	N2.0	/	A2.0	/	A1.0	A2.0	A1.0
8	Nurse I	A2.0	A1.0	/	A1.0	/	A3.0	A1.0
9	Nurse j	P2.0	N1.0	/	N2.0	/	A2.0	P2.0
10	Nurse K	/	/	A1.0	A2.0	A2.0	A1.0	A1.0
11	Nurse L	/	P2.0	A2.0	A2.0	A1.0	N2.0	/
12	Nurse M	A2.0	A2.0	/	A2.0	A2.0	A1.0	/
13	Nurse N	/	A2.0	A3.0	A3.0	A2.0	A2.0	/
14	Nurse O	/	A2.0	/	A1.0	A3.0	A3.0	N3.0
15	Nurse P	A1.0	P1.0	A1.0	P1.0	/	/	A1.0
16	Nurse Q	A3.0	A2.0	A3.0	/	P2.0	N1.0	/
17	Nurse R	A3.0	A1.0	A2.0	/	A4.0	A2.0	/
18	Nurse S	A3.0	A4.0	A4.0	A4.0	N2.0	/	/
19	Nurse T	/	A2.0	A1.0	A1.0	/	A4.0	A4.0
20	Nurse U	P1.0	A1.0	A1.0	/	N1.0	/	A1.0
21	Nurse V	A1.0	A1.0	P1.0	/	/	A1.0	A1.0
22	Nurse W	A1.0	A1.0	A1.0	A1.0	/	/	A1.0
23	Nurse X	/	/	A1.0	A2.0	A2.0	A2.0	A1.0
24	Nurse Y	A1.0	/	N1.0	/	P1.0	P1.0	A1.0
25	Nurse Z	/	A4.0	N3.0	/	A2.0	P2.0	A3.0

**Table 4.** The results after Evolutionary algorithm optimization

Number	Name	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	Nurse A	A4.0	N2.0	/	A1.0	A2.0	/	N3.0
1	Nurse B	A2.0	A3.0	A2.0	/	A1.0	/	A2.0
2	Nurse C	A2.0	A2.0	/	A2.0	A1.0	A2.0	/
3	Nurse D	A1.0	A1.0	A1.0	N1.0	/	/	P1.0
4	Nurse E	N3.0	/	A2.0	A2.0	A1.0	A3.0	/
5	Nurse F	A2.0	/	P2.0	P2.0	A1.0	/	A1.0
6	Nurse G	/	A1.0	A2.0	A1.0	/	A1.0	A1.0
7	Nurse H	N2.0	/	A2.0	A1.0	/	A2.0	A1.0
8	Nurse I	A2.0	A1.0	/	/	A1.0	A3.0	A1.0
9	Nurse j	P2.0	N1.0	/	N2.0	/	A2.0	P2.0
10	Nurse K	/	/	A1.0	A2.0	A2.0	A1.0	A1.0
11	Nurse L	/	P2.0	A2.0	A2.0	A1.0	N2.0	/
12	Nurse M	A2.0	A2.0	/	A2.0	A2.0	A1.0	/
13	Nurse N	/	A2.0	A3.0	A3.0	A2.0	A2.0	/
14	Nurse O	/	A2.0	/	A1.0	A3.0	A3.0	N3.0
15	Nurse P	A1.0	P1.0	A1.0	P1.0	/	/	A1.0
16	Nurse Q	A3.0	A2.0	A3.0	/	P2.0	N1.0	/
17	Nurse R	A3.0	A1.0	A2.0	/	A4.0	A2.0	/
18	Nurse S	A3.0	A4.0	A4.0	A4.0	N2.0	/	/
19	Nurse T	/	A2.0	A1.0	/	N1.0	A4.0	A4.0
20	Nurse U	P1.0	A1.0	A1.0	A1.0	/	/	A1.0
21	Nurse V	A1.0	A1.0	A1.0	/	/	A1.0	A1.0
22	Nurse W	A1.0	A1.0	P1.0	A1.0	/	/	A1.0
23	Nurse X	/	/	A1.0	A2.0	A2.0	A2.0	A1.0
24	Nurse A	A1.0	/	N1.0	/	P1.0	P1.0	A1.0
25	Nurse B	/	A4.0	N3.0	/	A2.0	P2.0	A3.0

**Table 5.** The Decision Trees under Different Dataset Classification Method's performance measure

Method	$P_1$	$R_1$	$F1_1$	$P_2$	$R_2$	$F1_2$	$P_3$	$R_3$	$F1_3$	$P$
Normal method	0.91	0.9	0.91	0.77	0.5	0.6	0.76	0.99	0.86	0.83
hold-out	0.96	0.93	0.95	0.75	0.49	0.6	0.74	0.99	0.85	0.86
cross alidation	0.92	0.92	0.93	0.8	0.47	0.52	0.75	0.98	0.84	0.84

3. Cross validation method: First, divide the dataset  $D$  into  $k$  mutually exclusive subsets of similar size. Then, each time,

use the union of  $k-1$  subsets as the training set, and the remaining subset as the test set; In this way,  $k$  sets of training/testing sets can be obtained, allowing for  $k$  training and testing sessions, ultimately returning the mean of these  $k$  test results.

#### 4. Conclusion

This study has completed the implementation of the nurse scheduling algorithm, which can generate a weekly nurse scheduling table based on the preset number of nurses, nurse abilities, scheduling coverage requirements, and historical data. At the same time, it can also meet some special needs, such as the number of night shifts, willingness to work, and willingness to take consecutive breaks, to a certain extent, making scheduling more humane. From the final results, it can be seen that the algorithm implemented by Python produces expected results, with all hard constraints met and soft constraints satisfied as much as possible and the lowest penalty value obtained until the final optimal solution is obtained. By evaluating the constructed decision tree, it can be seen that it has excellent classification ability and can classify the required constraint selection based on parameters, reducing the difficulty of algorithm use.

#### Acknowledgments

This work was financially supported by Southwest Minzu University Graduate Innovative Research Project (Project No. ZY2022733) fund.

#### References

- [1] Wang Huili, Liu Meijuan. Comparison of nursing quality and satisfaction of nurses and patients under different scheduling models [J]. *Journal of Nursing Administration*, 2015, 15(3): 222-223.
- [2] Li Yufeng, Li Xiaolan. Application of humanized management in nurse scheduling [J]. *China Medicine and Pharmacy*, 2019, 9(11): 161-163.
- [3] Mei Xun, Ye Chunming. Nurse scheduling studies that consider fairness and nurse preference [J]. *Computer Engineering and Applications*. 2019, 55(04): 263-270.
- [4] Howell J P. Cyclical scheduling of nursing personnel [J]. *Hospitals*, 1966, 40(2): 77.
- [5] Warner D M, Prawda J. A Mathematical Programming Model for Scheduling Nursing Personnel in a Hospital [J]. *Management Science*, 1972, 19(4-Part-1): 411-422.
- [6] Yang Kai, Cao Shuqing. Study on Nurse Shift Scheduling Model of Operation Room Based on Optimization Algorithm of Integer Programming and Transformation Rule [J]. *Hospital Management Forum*, 2020, 37(07): 65-68.
- [7] Zeng Fenghua. Study on the weekly rostering algorithm of nurses and its realization [D]. *South China University of Technology*, 2018.
- [8] Huang Yufen, Xiang Yujun, Li Jiahuan, Chen Yangyang, Zhao Weiyi. Application of Qualification Positioning and Grouping Management Model in Nurse Rostering in Emergency Department of Children's Hospital [J]. *Nursing and Rehabilitation Journal*, 2021, 20(09): 94-97.