Menemui Matematik (Discovering Mathematics) 46(2) (2024) 81-89

Menemui Matematik
(Discovering Mathematics)

journal homepage: https://myjms.mohe.gov.my/index.php/dismath/

Simulated Annealing Approach for an Overbooking Appointment Scheduling Problem

Nur Irdina Husna Khairuddin[1], Nor Aliza Ab Rahmin[2*] and Athirah Nawawi[3]
[1,2,3]*Department of Mathematics and Statistics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor*

[1]197101@student.upm.edu.my, [2]aliza@upm.edu.my, [3]athirah@upm.edu.my
*Corresponding author

Received: 1 June 2024
Accepted: 18 October 2024

ABSTRACT

No-shows are patients that do not show up for scheduled appointments or cancel at the last minute, preventing the health centre from filling the slot. Due to missed appointments, the health centre may lose time and money, and patient care may be compromised. In this study, we concentrate on patient no-show behaviour to reduce resource idle time, resource overtime, and patient waiting time. We aim to improve the overbooking appointment scheduling problem by applying the simulated annealing method after implementing the heuristics procedure. We also discover the effects of multiple patient assignments in the same slot, where we tend to find the greatest number of patients per slot that may be allocated to reduce costs. Our findings indicate that when using a large dataset of patients, simulated annealing performs slightly better than heuristics methods, and as a result, the maximum number that may be assigned to the same time slot is four.

Keywords: Overbooking, Multiple assignment, Heuristics procedure, Simulated annealing

INTRODUCTION

The goal of a medical appointment scheduling is to efficiently allocate resources, such as doctors, operators, and equipment, to patients in order to meet success criteria (Alizadeh et al., 2020). The problem of missed appointments, or when patients who have appointments do not show up, has long been a source of worry for medical professionals. Due to the lengthy waiting list, unattended time slots not only result in inefficient resource utilisation and performance but also prevent patients from receiving timely medical exams and treatments.

In this research, we focus on the patient no-shows that is one of the most critical barriers to effective appointment scheduling in hospitals. Overbooking schemes, in which the number of booked patients exceeds the number of available time slots, are a common strategy for mitigating the negative impact of patient no-shows. Overbooking is also a common strategy used by the airline industry to boost revenue (Chen et al., 2018). However, in healthcare, it is different in the fact that overshowing patients must be addressed by service providers, resulting in more effort and a poor patient experience (Lawley and Muthuraman, 2008).

The purpose of this research is to reduce the cost of the outpatient department by minimizing a weighted combination of three performance measures: resource overtime, resource idle time, and patient waiting time. Heuristic procedure and simulated annealing method are applied in this study. To discover which strategy delivers the best solution, the performance of simulated annealing method will be compared with the result of heuristic procedure in constructing near-optimal overbooking appointment schedule under patient no-show conditions.

In a recent study published in Chen et al. (2018), they looked at the connection between the time slot structure and the best overbooking solution. When compared to pre-defined time intervals, they discovered that flexibility in appointment start times might deliver a better patient experience while keeping the same service provider efficiency. In this research, the overbooking model is a single-server model, which implies that only one resource is available throughout the day and only one patient may be seen at any given time. We focus on finding the near-optimal overbooking appointment schedule with pre-defined time intervals which is fixed-length slot structure under patient no-show behaviour.

NOTATION

Table 1: Sets

N	Set of patients 1; 2; : : : ; n to be scheduled for the session
i	Index of patient
J	Set of time slots in the session
j	Index of time slot
S	Set of scenarios
s	Index of scenario
N_s^1	Set of patients who show up at the health care unit under scenario s
N_s^2	Set of patients who do not show up at the health care unit under scenario s

The total number of patients booked for a single session is represented by n. Since a reasonable overbooking level of 2 patients is used as the baseline, the value of n is set to n = 14 in the model. The model considers a finite number of scenarios, indicated by S, to approximate those stochastic components. According to the no-show rate for each scenario, each patient either shows up or does not.

Table 2: Parameters

b_j	Beginning time of time slot j
d_{is}	Service duration for the i^{th} patient under scenario s
w^{ot}	Penalty for each unit of resource overtime
w^{wait}	Penalty for each unit of patient waiting time
w^{idle}	Penalty for each unit of resource idle time
E	Close time for the healthcare facility

The starting time b_j of each slot j is fixed since the overbooking model we studied uses pre-defined time slots. Each time slot has a fixed time interval.

$$x_j^i = \begin{cases} 1, & \text{if the } i^{th} \text{ patient in the schedule is assigned to the } j^{th} \text{ time slot} \\ 0, & \text{otherwise} \end{cases}$$

Table 3: Decision variables

a_i	Appointment time of the i^{th} patient in the schedule
$wait_{is}$	Waiting time of the i^{th} patient for receiving the medical service under scenario s
$idle_{is}$	Idle time of the resource between the $(i)^{th}$ and the $(i + 1)^{th}$ services under scenario s
ot_{is}	Overtime of the resource under scenario s
x_j	Number of patients assigned to the j^{th} time slot
m	Maximum number of patients which can be assigned in one time slot
z_{is}^{start}	Start time of the medical service provided to the i^{th} patient under scenario s
z_{is}^{end}	End time of the medical service provided to the i^{th} patient under scenario s

OVERBOOKING MODEL

Objective function: To minimize the penalty cost of resource overtime, resource idle time and patient waiting time.

$$\min w^{ot} \sum_{s \in S} ot_s + w^{idle} \sum_{i \in N, s \in S} idle_{is} + w^{wait} \sum_{i \in N_s^1, s \in S} \frac{wait_{is}}{N_s^1}$$

Subject to:

$$a_i = \sum_{j \in J} b_j x_{ij}, \forall i \in N \quad (1)$$

$$\sum_{j \in J} x_{ij} = 1, \forall i \in N \quad (2)$$

$$a_{i+1} \geq a_i, \forall i \in N \backslash \{n\} \quad (3)$$

$$z_{is}^{start} = 0, \forall s \in S \quad (4)$$

$$z_{is}^{start} = a_i + wait_{is}, \forall i \in N \backslash \{1\}, s \in S \quad (5)$$

$$z_{is}^{start} = z_{(i-1)s}^{end} + idle_{(i-1)s}, \forall i \in N\{1\}, s \in S \quad (6)$$

$$z_{is}^{start} + d_{is} = z_{is}^{end}, \forall i \in N, s \in S \quad (7)$$

$$z_{ns}^{start} + idle_{ns} - ot_s = E, \forall s \in S \quad (8)$$

$$z_{is}^{start}, z_{is}^{end}, wait_{is}, idle_{is}, ot_s \geq 0, \forall i \in N, s \in S \quad (9)$$

$$x_{ij} \in \{0,1\}, \forall i \in N, j \in J \quad (10)$$

$$x_j < m, \forall j \in J \quad (11)$$

$$2 \leq m \leq 4, m \ integer \quad (12)$$

$$\sum_{j \in J} x_j = n \quad (13)$$

Constraint (1) is used to allocate the appointment time for the i^{th} patient in the sched- ule's j^{th} time slot. Because it is assumed that patients are punctual, each patient's arrival time is equal to their appointment time. Constraint (2) shows that each patient can only be allocated to one time slot. Constraint (3) guarantees that the booked appointments are in the right order. Constraint (4) sets the starting start time of the medical treatment supplied to the first patient to zero. The patient waiting time, service start time, and service completion time for each patient are calculated using constraints (5) and (7). The resource idle time between the $(i-1)^{th}$ and i^{th} services are calculated using constraint (6). The resource overtime is calculated using constraint (8). The non-negativity requirement of the variables is fulfilled by constraint (9). Constraint (10) demonstrates that patient assignment is binary.

Constraints (11) and (12) are intended to investigate the impact of numerous assignments in the overbooking model. Constraint (11) guarantees that the number of patients given to the j^{th} time slot does not exceed the maximum number that can be assigned to a single time slot. Before booking the appointment, constraint (12) specifies that the maximum number of patients that may be allocated in one time slot is set to m = 2, 3, 4. Because 2 is the least number of patients that may be allocated to one slot in an overbooking model, and it is not acceptable to assign more than 4 patients to one slot, the value of m is set in the interval [2,4]. Constraint (13) ensures that the total number of patients assigned to all time slots equals the total number of patients allocated for a single session, consequently $x_1 + x_2 + \ldots + x_{12} = n$.

PATIENT NO-SHOW MODEL

Notation:
α = no-show probability of patient.
Procedure:

1. $\forall i \in N$ and $s \in S$, generate a random number Uis which is uniformly distributed in the interval (0,1).
2. $N_{s1} = \{i \in N : U_{is} > \alpha\}$ and $N_{s2} = \{i \in N : U_{is} \leqslant \alpha\}$. i.e., (N_{s1}, N_{s2}) is a partition of N, where Ns1 and Ns2 respectively denote under scenario s the sets of patients who show up for the appointment and do not.
3. For $s \in S$ and $i \in N_{s1}$, $\{dis : i \in N, s \in S\}$ are generated according to their empirical probability distributions. For $s \in S$ and $i \in N_{s2}$, dis = 0.

Taken the mean value of no-show probability from the empirical data from the paper Chen et al. (2018), they obtain the value $\alpha = 0.176$. As a result, we may deduce that if $U_{is} > 0.176$, the patient shows up for his or her appointment, whereas if $U_{is} \leqslant 0.176$, the patient does not.

DATA GENERATION

Table 4: Summary statistics from empirical data

Random variable	Mean	Standard deviation	Distribution
Examination time per test (min)	12.70	8.09	0.5+87*BETA(2.3,12.7)
Setup time for each test (min)	6.40	5.17	-0.5+LOGN(7.01,6.43)

Referring to Chen et al. (2015) paper, the probability distributions as shown in Table 4 were estimated by them through the collection of data related to the appointments in a medical imaging center from January 2015 to March 2015.

We first set a total number of data that we want to generate. Then, a set of data is generated which probability that larger than 0.176 indicates the patients show up. A total number of patients that show up, n is then calculated. A set of data of setup time and examination time is generated according to the number of patients that show up. Setup time and examination time are set to zero for patient who do not show up.

We generate three different sets of data which include D = 1, D = 20 and D = 100 using uniform distribution. These sets of data will be use in solving the overbooking appointment problem. Each set contains data of 14 patients. For example, dataset D = 1 indicates one set of data which contains 14 patients' data. Whereas dataset D = 20 indicates 20 sets of data which contains 20 times 14 patients' data that is 280 patients' data and D = 100 contains data of 1400 patients.

HEURISTIC PROCEDURE

For each value of m, the heuristic approach is performed to discover one optimal solution and determine the maximum number of patients who may be allocated to the same time slot. This is used to determine which sets of solutions will be utilized as the starting solutions in the simulated annealing method.

We begin by considering the first option of solutions. We expect that patients who show up for appointments will be on time, therefore we set arrival time equal to appointment time. We determine the duration of treatment for each patient in the corresponding solution by adding set up time and examination time. We expect that those who show up for their appointments will be on time, therefore we set the arrival time to equal the appointment time. After that, we compute end time of patient treatment.

Afterwards, if the patient does not show up for an appointment, we set waiting time to zero, and if patient does show up for an appointment, we compute the waiting time. If the end time of current patient treatment is earlier than or equal to the arrival time of the following patient, we set the waiting time of the next patient to zero and compute idle time. In contrast, if the end time of the current patient is larger than arrival time of the following patient, we set idle time to zero.

SIMULATED ANNEALING (SA) METHOD

For each dataset D, SA method is applied to produce a near-optimal overbooking solution. The solution that has the lowest objective value is taken from the heuristic procedure and it will be set as the initial temperature to initiate the SA process. We use Microsoft Excel to randomize the number of patients assigned to each slot as the solutions for SA method. Integer representation is used to encode the solutions. It shows the slot number assigned to each patient, instead of showing the total number of patients assigned to each slot as in heuristic procedure. This is done to guarantee that the total number of patients assigned to each session stays at 14 throughout the SA process.

As we obtained the objective values for each solution, we applied SA method to find the optimal solution. We begin by setting the list of solutions as and then assigning the initial temperature equal to the best objective value from heuristic procedure and the current state equal to the first solution from the neighbourhood. Next, we set the current temperature equal to the initial temperature and choose the second solution from the neighbourhood to compute the cost difference between the current state and the neighbourhood. The process will stop when the current temperature is greater than the final temperature. If the difference is greater than 0, we conclude that the solution is equal to the current solution. If the difference is less than zero, we generate a random number between the interval [0,1] and compute the probability of e−Δ/current temp. If the probability is greater than the random number, we set the solution equal to current solution. Otherwise, we gradually decrease the current temperature and choose the next solution from the neighbourhood. The process stops when the final temperature is greater than the current temperature and when the neighbourhood is fully explored.

RESULT AND DISCUSSION

We compare the result of heuristic procedure with SA method to see which approach gives a better performance. The minimum value which corresponds to the best performance of each measure among heuristic procedure and SA method is highlighted.

Dataset D=1:

Table 5: Comparison between heuristic procedure and SA method.

	Heuristic	Simulated Annealing
Best solution (k)	3	6
Total objective value	21.9793	21.7460
Total waiting time	352.920	312.920
Total idle time	0.000	0.000
Total overtime	31.62	31.62
CPU time	0.473 s	0.678 s

SA method gives a lower total waiting time compared to heuristic procedure hence yields to lower total objective value. This indicates SA method has a better performance than heuristic procedure. The best solutions of heuristic procedure and SA method are shown as in Table 6 and Table 7.

Table 6: Best solution of heuristic procedure (k=3)

Slot	1	2	3	4	5	6	7	8	9	10	11	12
No. of Patient	3	1	1	1	1	1	2	1	1	2	0	0

Table 7: Best solution of SA method (k=6)

Slot	1	2	3	4	5	6	7	8	9	10	11	12
No. of Patient	3	1	1	1	1	0	2	1	2	2	0	0

Dataset D=20:

Table 8: Comparison between heuristic procedure and SA method.

	Heuristic	Simulated Annealing
Best solution (k)	6	5
Total objective value	27.5275	27.7007
Total waiting time	43.3087	45.8788
Total idle time	0.92380	0.92380
Total overtime	33.1085	33.1085
CPU time	0.233 s	0.338 s

Both methods give the same average of idle time and overtime however heuristic procedure has a lower mean of waiting time. Both give almost the same objective values. However, heuristic procedure shows a slightly better performance. The best solutions of heuristic procedure and SA method are shown as in Table 9 and Table 10.

Table 9: Best solution of heuristic procedure (k = 6)

Slot	1	2	3	4	5	6	7	8	9	10	11	12
No. of Patient	3	1	1	2	1	1	1	1	1	2	0	0

Table 10: Best solution of SA method (k = 5)

Slot	1	2	3	4	5	6	7	8	9	10	11	12
No. of Patient	3	1	2	1	1	1	1	1	2	1	0	0

Dataset D=100:

Table 11: Comparison between heuristic procedure and SA method.

	Heuristic	Simulated Annealing
Best solution (k)	2	8
Total objective value	39.2779	39.2495
Total waiting time	22.7502	22.2132
Total idle time	6.8009	6.8017
Total overtime	14.6507	14.6611
CPU time	0.224 s	0.660 s

Both methods have a slight difference in their mean of idle time and overtime. SA method has a lower mean of waiting time compared to heuristic procedure. Since SA method has a slightly lower objective value, we conclude that SA method delivers a better performance compared to heuristic procedure. The best solutions of heuristic procedure and SA method are shown as in Table 12 and Table 13.

Table 12: Best solution of heuristic procedure (k = 2)

Slot	1	2	3	4	5	6	7	8	9	10	11	12
No. of Patient	4	1	1	1	1	1	2	1	1	1	0	0

Table 13: Best solution of SA method (k = 8)

Slot	1	2	3	4	5	6	7	8	9	10	11	12
No. of Patient	4	1	1	1	1	1	1	2	1	1	0	0

CONCLUSION

The goal of this research is to find a solution to the problem of overbooking appointment scheduling by implementing simulated annealing approach in order to reduce resource over- time, resource idle time, and patient waiting time. In this research, we focus on single server overbooking model with fixed length time slot and the only uncertainty we considered is the no-show condition of the patients. Hence, we assumed that patients who show up will be punctual. We also generate variety of patients data to test the algorithm. Referring to the distributions provided from previous research, we generate the no-show probability, setup time, and examination time for three different datasets. Besides, we also study the effect of multiple assignments to determine the maximum number of patients that may be allocated to the same time slot in order to limit the potential for conflict when patients arrive at the same time.

To achieve the objectives of our research, we solved the overbooking appointment scheduling problem by using heuristic procedure and simulated annealing method. Both methods are coded using C programming. Based on the outputs, we found that the best condition for the maximum number of patients that can be assigned to the same time slot is 3 for a smaller data involved. Meanwhile, for a larger data, we found that it is best to assign the maximum of 4 patients in the same time slot to achieve the best performance in minimizing the cost of the medical center. We discovered a trend of three or four patients allocated to the first slot, with the final two slots remaining empty and unoccupied. For other slots, only one or sometimes two patients are assigned. This overbooking scheduling pattern can minimize the cost of resource overtime, resource idle time and patient waiting time.

In term of CPU time, we observed that heuristic procedure gives a shorter time taken to compute compared to simulated annealing method. This is due to the accuracy of checking step by step throughout the neighbourhood according to the simulated annealing algorithm. Hence, it took a longer time to compute compared to heuristic procedure. However, the range of time taken is still acceptable as the difference between the two method is not too much. To sum up, we can conclude that simulated annealing approach gives a better performance in solving the overbooking problem that involving a large set of data compared to heuristic procedure. The optimal solution acquired from the heuristic approach is improved or at the very least maintained by the simulated annealing method.

REFERENCES

Alizadeh Foroutan, R., Zeidi, J., Abedi, M., & Chiong, R. (2020). A modified genetic algorithm for non-emergency outpatient appointment scheduling with highly demanded medical services considering patient priorities. *Computers and Industrial Engineering*, 139, 106106.

Chen, Y., Kuo, Y.-H., Fan, P., & Balasubramanian, H. (2018). Appointment overbooking with different time slot structures. *Computers and Industrial Engineering*, 124.

Lawley, M., & Muthuraman, K. (2008). Stochastic overbooking model for outpatient clinical scheduling with no-shows. *RCHE Publications*, 40.

Chen, Y., Kuo, Y.-H., Balasubramanian, H., & Wen, C. (2015). Using simulation to examine appointment overbooking schemes for a medical imaging center.