APPLYING SET PARTITIONING METHODS IN THE CONSTRUCTION OF OPERATING THEATRE SCHEDULES

Rowse, E., Lewis, R., Harper, P., and Thompson, J.
Cardiff University School of Mathematics, Cardiff CF10 4AG, WALES.
RowseEL\LewisR9\Harper\ThompsonJM1@cf.ac.uk

ABSTRACT

A major factor contributing to the high number of cancelled operations in hospitals is the unavailability of beds on hospital wards for post-operative recovery. In this research, we model the effects of the operating theatre timetable on the demand for beds and show that a timetable can be produced that leads to a reduced number of cancelled operations whilst also levelling the demand for beds throughout the week. For this work, a set partitioning formulation has been developed to assign surgical specialties to operating theatres, and an extension of the model has been used to incorporate constraints on the demand for post-operative beds. Results are presented to highlight the potential of this model, and the use of a robust optimization approach is also investigated in order to address the stochastic nature of the factors affecting operating theatre scheduling.

KEYWORDS

Scheduling, Timetabling, Set partitioning, Optimization, Health Informatics.

1. INTRODUCTION

The scheduling of hospital operating theatres (OTs) has proven to be a very challenging and active area of research over the past 20 years. The practical importance of the problem is due to the surgical suite being one of the major departments within a hospital. Surgical suites have high costs associated with their function and surgical patients provide a significant proportion of demand for other hospital departments, both before and after surgery. It is therefore desirable, from the hospital manager’s point of view that the OTs are utilised in the best way in order to make the most of these valuable resources.

OT scheduling is also an interesting and challenging research problem due to the large number of variables that can affect operations. According to Van Oostrum et al. (2008), the main uncertainties related to the scheduling of operations are the stochastic duration of surgical procedures, personnel availability, the no-show of patients and the occurrence of emergency patients. Santibanez et al. (2007) also discuss the potential benefits of a systematic approach to surgery scheduling, including the increased efficiency of the OTs, increased patient throughput, lower wait times for both patients and surgical staff, and increased transparency and fairness in the allocation of time to different surgical specialties.

There are several stages of OT scheduling that are used for different planning horizons. This paper concerns the scheduling of operations at the tactical level which is used for medium-term planning of elective patients. This involves the construction of a Master Surgery Schedule (MSS) which is a cyclic weekly timetable used by all theatre staff. The MSS specifies the surgical specialty that has priority in each OT during each session of the week. A novel extension to the construction of the MSS is considered in this paper which takes into account the effect the MSS has on the demand for beds on wards for post-operative recovery. By constructing the MSS in this way, demand for beds on wards can be levelled throughout the week resulting in fewer elective operations being cancelled.

A more in-depth review of previous approaches for construction of MSSs is given in Section 2. The new set partitioning model that incorporates bed constraints is then presented in Section 3, followed by a
discussion on the methods used to incorporate these constraints in Section 4. Results of this new model are discussed in Section 5, with conclusions and a discussion on further work appearing in Section 6.

2. LITERATURE REVIEW

As mentioned previously, OT planning and scheduling continues to be a challenging and popular topic for research. The reader is directed to a particularly thorough literature review paper by Cardoen et al. (2010) for an overview of the different stages of OT scheduling. Their literature review found that the most common technique used for the construction of the MSS is mixed integer programming (MIP). Here we review a selection of papers as examples of previous approaches to constructing the MSS.

Blake and Carter (1997) commented on the scope of OT scheduling research and advised that techniques that integrate the OTs with other hospital departments are urgently required. The majority of papers still only consider an isolated OT; however some chose to integrate other resources such as wards and the intensive care unit. A reason why the scope of research projects may have been simplified to exclude other departments could be the resulting increased formulation and computational complexity of the scheduling models.

A small number of papers in the literature consider constructing the MSS for only one surgical team or only one OT. Vissers et al (2005) use MIP to construct an MSS for the cardiothoracic surgery department with a four week cycle time. A number of resources are considered in the model as constraints, specifically intensive care beds and nursing staff. Kuo et al. (2003) also use intensive care beds as constraints in their model, and use linear programming to allocate multiple specialties to multiple OTs in order to maximise surgeon revenue. Their results indicate a 15% increase in revenue.

Van Oostrum et al. (2008), meanwhile, use an MIP model to construct an MSS that uses a column generation technique to find a solution. The stochastic nature of the duration of surgical procedures is considered and planned slack is built into the timetable in order to account for this. Their MIP model aims to maximise the OT utilisation as well as levelling the subsequent hospital bed requirements.

Belien and Demeulemeester (2007) have also proposed a number of MIP and quadratic programming models for constructing the MSS. They evaluate these methods by considering the resulting bed occupancy after surgery, with the aim of levelling it as much as possible. They build a model that minimises the total expected bed shortage with constraints on the demand for OT blocks for each surgical group, and constraints on the capacity of the number of available OT blocks each day. Belien et al. (2009) subsequently discuss a decision support system for the implementation of these models in a large hospital. They find that the different models provide slightly different schedules, and conclude that it is the responsibility of the manager to then choose the “best” schedule among these.

As an alternative to MIP-based methods, Vanberkel et al. (2011) use a queuing theory approach to build the MSS in such a way that demand on downstream hospital departments is predicted and taken into account in the MSS.

3. SET PARTITIONING MODEL

The set partitioning problem (SPP) is a binary integer programming problem that involves partitioning a set of elements into a set of mutually exclusive subsets subject to the satisfaction of some constraints. This model has been successfully used for the modeling of scheduling and rostering problems (Ryan & Foster, 1981), and also vehicle routing problems (Balinski & Quant, 1964). In general, the SPP is NP-hard; however, in some cases exact approaches can be used to determine globally optimal solutions (Marsten & Shepardson, 1981), even for instances involving hundreds of millions of variables and hundreds of constraints (Ryan 1992). Structural properties of the constraint matrix can also be taken advantage of in order to solve larger instances of the problem in many cases (Balas & Padberg 1976).

The traditional formulation of the SPP model is as follows:

\[ \min \ z = c^T x \] \hspace{1cm} (1)
subject to

\[\begin{align*}
Ax &= e \\
x &\in [0,1]^n
\end{align*}\]  

where \(A\) is an \(m \times n\) binary matrix, \(c\) is an arbitrary vector of costs, and \(e = (1,1,1,\ldots,1)^T\) is an \(m\)-vector. The decision variables \(x_j, j = 1,\ldots, n\), can be thought of as the probability that the \(j\)th column is included in a solution.

Set partitioning models have previously been used in healthcare applications to assign elective surgical patients to OT slots taking into account constraints relating to OT and surgeon availability (Fei et al., 2009), and to schedule anaesthesiologists for surgery based on the matching of skills with specific tasks (Ernst et al., 1973). However, it is not apparent that set partitioning models have been used for a more tactical level of OT scheduling such as the construction of the MSS.

The proposed set partitioning model for the construction of the MSS is outlined herein. The aim of the model is to select a subset of possible “plans” for each surgical specialty in order to assign exactly one specialty to each OT session and to ensure that the number of predicted patients in post-op recovery does not exceed the number of beds available on wards. A plan for a specialty defines which OT the specialty has use of on which day of the week and during which session (morning/afternoon/whole day), and reflects the specialty's preferences of theatres, days and distribution of sessions. All possible plans are generated using an enumeration algorithm whilst taking into account these preferences. The solution of the model will provide one plan for each of the surgical specialties which, when put together, will form the MSS.

The basic formulation of the proposed set partitioning model for the construction of the MSS is defined by Equation (1) subject to

\[\begin{align*}
Ax &= e \\
B &\leq d \\
x &\in [0,1]^n
\end{align*}\]  

where \(x_j, j = 1,\ldots, n\), are decision variables that indicate whether or not a plan is selected in the final solution, and \(c\) is a vector of the cost of each plan.

The “cost” of each plan reflects the difference between the total number of beds available for post-operative care and the predicted bed occupancy. The objective is to then minimise this cost: i.e. it is desirable to minimise the number of empty beds on the wards.

Here, \(A\) is an \(m \times n\) binary matrix where columns represent possible plans for each surgical specialty. The first \(s\) rows of \(A\) represent generalised upper bound (GUB) constraints that relate each plan to a specific specialty; the remaining rows represent constraints for each OT session. These rows of the \(A\) matrix have elements \(a_{ij} = 1\) if operating room session \(i\) is used in plan \(j\), and \(a_{ij} = 0\) otherwise.

The right-hand side values of the OT constraints are given in the vector \(e\) where \(e = (1,1,1,\ldots,1)^T\). This indicates that only one plan must be selected in the solution for each specialty, and that only one specialty can occupy an operating room session at any one time.

The matrix \(B\) has entries that are determined from the plans in the \(A\) matrix and represent the number of surgical patients who require beds for each plan on each ward on each day. The elements of the \(B\) matrix, \(b_{kj}\), represent the number of beds required on ward \(k\) on day \(l\) for plan \(j\). If there are \(p\) wards and the model considers bed usage on \(q\) days, then \(B\) is a \((p \times q) \times n\) matrix. Constraints are constructed so that the number of beds required is less than or equal to the number of beds available. The number of beds available on ward \(k\) on day \(l\), \(d_{kl}\), form the right-hand side values of the bed constraints.

In our model the bed constraints are treated as elastic in order to allow for the “sharing” of beds between specialties on different wards. Both slack and surplus variables are added to the constraints so that they can be under- or over-satisfied. In this particular application, slack variables represent empty beds, and surplus variables represent additionally required beds. In order to be able to specify how many beds are transferred between which particular wards, consider a \(p \times p\) matrix, \(Z_{(j)}\), whose elements \(z_{kj}\) are decision variables that specify how many beds are moved from ward \(k\) to ward \(v\) on day \(l\). By definition, it follows that the sum of the elements of row \(k\) of \(Z_{(j)}\) represents the number of empty beds on ward \(k\) on day \(l\). Similarly, following from the definition of the surplus variables for this application, the sum of the elements of column \(k\) of \(Z_{(j)}\) represents the number of extra beds for ward \(k\) on day \(l\).
In order to control which slack and/or surplus variables are used in the bed constraints, a matrix is used to define allowable movements of patients between wards. This transition matrix, \( W \), is a \( p \times p \) matrix that is not necessarily symmetric. \( W \) is constructed based on prior knowledge obtained from the hospital on which wards each specialty can use, and so \( W \) is assumed to be constant for each day \( l \). The elements of \( W \) are defined as \( w_{kv} = 1 \) if patients meant to be on ward \( k \) are permitted to use beds on ward \( v \), and \( w_{kv} = 0 \) otherwise.

Combining elements in \( W \) and \( Z^{(l)} \) gives the total slack and surplus for each ward, and results in Constraint (6) being re-written as follows:

\[
\sum_{j=1}^{n} b_{kj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} + \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = d_k^{(l)} \quad \forall k = 1, \ldots, p, \ l = 1, \ldots, q
\]  

An additional constraint is that the sum of the surplus variables across all wards does not exceed the sum of the slack variables across all wards on each day. This is needed in the model to prevent the total number of beds in the hospital from being exceeded.

The complete formulation of the set partitioning model for the construction of the MSS is therefore:

\[
\min \sum_{j=1}^{n} c_j x_j
\]

subject to

\[
\sum_{j=1}^{n} a_{ij} x_j = 1 \quad \forall i = 1, \ldots, s
\]  

\[
\sum_{j=1}^{n} a_{ij} x_j \leq 1 \quad \forall i = s + 1, \ldots, m
\]  

\[
\sum_{j=1}^{n} b_{kj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} + \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = d_k^{(l)} \quad \forall k = 1, \ldots, p, \ l = 1, \ldots, q
\]  

\[
\sum_{k=1}^{p} \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} \leq \sum_{k=1}^{p} \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} \quad \forall l = 1, \ldots, q
\]  

\[
x_j \in \{0, 1\} \quad \forall j = 1, \ldots, n
\]

\[
z_{kv}^{(l)} \geq 0 \text{ and integer} \quad \forall k = 1, \ldots, p, \ v = 1, \ldots, p, \ l = 1, \ldots, q
\]

4. GENERATING THE B MATRIX

As mentioned, the A matrix defines in which theatre and at what time each surgical specialty will operate. Using the A matrix, the B matrix is then generated by filling in the required number of beds on each ward on each day for each plan. In the model, patients are sent from theatre to either a ward or the critical care unit for post-operative recovery. Patients who are sent to critical care are subsequently moved to a ward. Data concerning patients’ post-operative length of stay (LoS) in each ward and critical care for all specialties can be used to determine how long patients will remain in beds in the model.

In our case, two different methods based on the use of the LoS data are used to populate the B matrix. These methods are:

- Single scenario of the sampled LoS
- Multiple scenarios of the sampled LoS
Obviously, the latter is a generalisation of the former.

Let the post-op length of stay for a patient be denoted by the random variable $T$. More specifically, $T$ denotes the duration of time after surgery until the patient either leaves hospital (end of the spell in hospital) or moves to the care of a different specialty (end of episode). In either case, $T$ is effectively the time taken for the patient to “recover” from surgery.

**Generate B Matrix for a Single Scenario**

```plaintext
For each plan (column) in A do
  Look-up which specialty the plan refers to
  For each session (row) of A do
    If the specialty is scheduled in the OR session then
      Enter number of new arrivals in the row of the B corresponding to the weekday that the OR session is on
    For each day in the length of stay distribution for this specialty do
      For each remaining arrival do
        Generate a random number $r$ between 0 and 1
        If $r \leq P(\text{leaving hospital on this day, given that the patient has been in hospital this long})$ then
          Decrease the number of remaining arrivals by 1
        End-if
      End-for
    Update B with the number of remaining arrivals on this day
  End-if
End-for
End-for
```

Figure 1. Algorithm for populating the B matrix for a single scenario

In the first method of filling in the B matrix, an “example” of the number of beds required is generated based on each plan. The conditional probability of failure is used from survival analysis to determine the probability that a patient leaves on day $d$ given that the post-op LoS has reached $d$ days. The Kaplan-Meier estimate of the survivor function is used because we have found that no parametric distribution can be suitably fitted to our LoS data. Using the random variable $T$ as the post-op LoS, the survivor function, denoted $S(t)$, is the probability that the random variable $T$ takes a value longer than a specified time $t$, i.e. $S(t) = P(T > t)$. An estimate of $S(t)$ can be calculated using the Kaplan-Meier method, sometimes known as the product-limit estimate. The Kaplan-Meier estimate of the survivor function, $\hat{S}(t)$, at time $t$ is given by:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{m_{(j)}}{n_{(j)}} \right)$$

where $t_j$ are the ordered survival times/length of stays for $j = 0, 1, 2,..., k$.

Using information from the LoS data such as the number of patients that have each distinct length of stay, $m_{(j)}$, and the number of patients that could have left at each length of stay time, $n_{(j)}$, the Kaplan-Meier estimate is calculated as:

$$\hat{S}(t_{(j)}) = \frac{n_{(j)} - m_{(j)}}{n_{(j)}}$$

The conditional probability of failure, $L(t_{(j)})$, is the probability that the event (patient leaves hospital) occurs in a small time interval $h$ after time $t$, and is defined as $L(t_{(j)}) = P(t < T < t + h \mid T > t)$. $L(t_{(j)})$ can be estimated when finding the Kaplan-Meier estimate of $S(t)$ as follows:

$$\hat{L}(t_{(j)}) = \frac{m_{(j)}}{n_{(j)}}$$

The B matrix is then populated using the conditional probability of failure estimate, $\hat{L}(t_{(j)})$, according to the steps shown in Figure 1.

Our second method of populating the B matrix is very similar to the first, however instead of using just one scenario, there are multiple, say $t > 1$, scenarios. This method essentially generates $t$ different B matrices using the same technique as discussed above, and appends them to create more bed constraints for the
optimisation model. The idea behind using multiple scenarios is that the more constraints that an optimal schedule can satisfy, the more likely the schedule will be able to cope with different realisations of uncertain bed demands when implemented in a hospital. The number of scenarios, $t$, is chosen by the user, however $t$ must be chosen with care as there is a trade-off between including more scenarios to give in a more “robust” schedule, and having too many constraints so that feasible solutions do not exist. If $t$ scenarios are generated, then there becomes $t \times p \times q$ bed constraints, where $p$ wards and $q$ days are being modelled. Constraints (11) and (12) are consequently re-formulated as follows to reflect the multiple scenarios in this method for all scenarios $g = 1,...,t$, wards $k = 1,..., p$ and days $l = 1,..., q$.

$$\sum_{j=1}^{n} b_{gkj}^{}X_j^{} - \sum_{v=1}^{p} w_{kv}^{}z_{gvk}^{(l)} + \sum_{v=1}^{p} w_{vk}^{}z_{gkv}^{(l)} = d_{gk}^{(l)}$$ (15)

$$\sum_{k=1}^{p} \sum_{v=1}^{p} w_{kv}^{}z_{gvk}^{(l)} \leq \sum_{k=1}^{p} \sum_{v=1}^{p} w_{vk}^{}z_{gkv}^{(l)}$$ (16)

5. RESULTS

In our research we used a case study arising in a large teaching hospital in Wales in which over 25,000 surgical operations are performed annually. Currently, around 18% of operations in this hospital are cancelled, with non-clinical reasons such as lack of available beds accounting for 54% of these. In addition, 31% of patients are also currently placed on a ward that does not necessarily have specialist nurses or equipment for the surgical specialty post-surgery due to shortfalls in bed capacity on the specialist ward.

The surgical suite at the case study hospital has 14 OTs used by 18 surgical specialties. Elective operations are scheduled during two operating sessions per day (morning/afternoon) over a five day week (Monday - Friday). There is one dedicated emergency theatre in which no elective operations are scheduled. Data on the preferences on theatres and sessions has been obtained from the hospital managers and has been used to generate the A matrix. The enumeration of all possible plans for this case study results in an $A$ matrix with almost 1500 plans/columns (and hence 1500 binary decision variables) and 158 OT constraints. Ten wards and a critical care unit are modelled for seven days a week, resulting in a $B$ matrix with 77 bed constraints for a single scenario (method 1). This optimisation problem was solved using the commercial software FICO Xpress using between 1 and 10 scenarios. This resulted in runtimes of approximately 30 seconds to 5 minutes respectively on a 3.0 GHz Windows 7 machine with 3.87 GB RAM. Simulations of the resulting MSS were then performed in order to obtain a measure of their robustness. A simulation of the optimal schedule is intended as a test to determine how the schedule will cope when different aspects of the uncertainty are realised i.e. how many patients will require a bed on each day in each ward compared to the number of beds available. If more patients require a bed than there are beds available it is equivalent to a violated bed constraint occurring in the optimisation model. It is of interest to examine the percentage of simulations of the optimal schedule for which at least one of the bed constraints would be violated.

Results from 1000 runs are presented in Figure 2. As can be seen, the average percentage of simulations in which the optimal schedule would have resulted in at least one of the bed constraints being violated decreases as the number of scenarios increases. This is as expected, due to more bed constraints being included in the optimisation acting as realisations of uncertainty; hence the resilience of the optimal schedule to uncertainty in the simulations will increase with the number of bed constraints included.

It can also be seen that as the number of scenarios increases, and hence the number of bed constraints in the model increases, the percentage of problem instances for which no feasible solution exists increases. This is also as expected due to the problems becoming more constrained as scenarios are added to the model.

6. CONCLUSIONS AND DISCUSSION

In this paper a set partitioning based optimisation model has been developed that incorporates constraints on the demand for beds. The model constructs an MSS by assigning specialties to OTs subject to post-operative
bed constraints. An increasing number of scenarios have been used to generate bed constraints for the optimisation model and simulation of the resulting optimal MSSs has then been performed in order to obtain a measure of the robustness of the MSS. The results presented here indicate that the more scenarios that are included in the bed constraints, the more robust the resulting MSS is with respect to the percentage of simulations that have violated bed constraints. In addition, the more scenarios that are included, the lower the percentage of simulations that have violated bed constraints. However, there is a trade-off between the number of scenarios to include and the likelihood of finding a feasible solution, since the number of problem instances for which no feasible solution exists increases as the number of scenarios increases.

![Graph showing the results of 1000 runs](image)

The method of populating the B matrix using multiple scenarios of sampled LoS can be used to achieve a more robust scheduling approach. A robust schedule is desirable from the hospital’s point of view because it will give more confidence that the schedule can withstand the variations in demand for beds when different aspects of this uncertainty are realised. Some of our current work is focused on adapting the SPP model presented here into a more robust formulation in the sense of Bertsimas and Sim (2004).

As we have seen, the production and simulation of schedules in this way has the potential to make more efficient use of valuable hospital resources. Another advantage is that this automated approach allows hospital administrators to try out various what-if scenarios to help inform future strategic decisions. For example, we might investigate the implications of adding extra beds or operating rooms, or we may choose to examine the effects of increasing the number of sessions per day, extending the working week, or using a fortnightly schedule. We are currently working with hospital administrators in order to trial our MSS at the University Hospital Wales. This will allow us to gauge its effects on the number of cancellations, the lengths of waiting lists, and the number of patients that are not assigned to the correct specialist ward post-surgery.

A further avenue of future research concerns the algorithms used to produce the MSS. To date, exact methods have been used to construct optimal schedules under the given constraints for our case study hospital. However, it is unlikely that reasonable run times will always be experienced by these methods, particularly for larger problem instances, perhaps suggesting the need for heuristics and/or approximation algorithms in some cases. Interestingly, many parallels can be drawn between this problem (where we seek to assign specialties to theatres and sessions) and the well-studied NP-hard university timetabling problem (where teaching “events” need to be assigned to rooms and timeslots) (Lewis & Thompson, 2015). It is likely that many of the local search and metaheuristic-based techniques developed for university timetabling will be applicable to this current problem though, of course, suitable modifications will also need to be made,
particularly to the objective function of the current problem which would involve both the production of suitable B matrices and subsequent simulations of all candidate solutions encountered during a run.

ACKNOWLEDGEMENTS

The first author would like to thank Professor David Ryan and other members of the Department of Engineering Science, University of Auckland, namely Associate Professor Andrew Mason and Dr Michael O'Sullivan, for their support and advice on this work. The authors would also like to give thanks to Tony Turley and Andrew Nelson at the case study hospital for providing information and data related to this problem.

REFERENCES