GENETIC SCHEDULING TO OPTIMIZE RESOURCE UTILIZATION FOR HOSPITALS

Kottalanka Srikanth¹, Dr. D. Arivazhagan²

¹(Research Scholar, AMET University, Chennai, India)  
²(Department: Information Technology, AMET University, Chennai, India)

ABSTRACT

In this Paper the main objective is how to reduce the waiting time. As it is commonly known that in the hospital minimum you need to wait for half an hour to one hour, which is a sheer waste and also the doctor is very busy who does not like to wait for his patients, since he will run out of his schedule if one of them fails to show up in time. The objective of this paper is to optimize balance between the patients’ waiting time, the doctor’s idle time and the overtime. The initial problem is to find an optimal appointment system for the hospital on per day basis. We suppose that there are eight types of patient. They all have a lognormal service time and their actual arrival time is drawn from a triangular distribution. A different type of patients is described. In the second column the average service time for every patient is shown. The next column shows the standard deviation of the service times. Column four and five give the parameters inserted in the lognormal distribution to achieve the values in the first two columns. Since the lognormal distribution has no upper bound it is truncated to a maximum service time of one hour.

Keywords: Forecasting, Health care management, Genetic Algorithm, Hospital, Resources, Scheduling, Doctor, Patient

1. INTRODUCTION

Real visiting time of patients is simulated through triangular distribution as shown in Table 1. Patients usually arrive early. Average coming time for normal patients is three minutes before their expected time. Lower limit is eight minutes before and the upper limit is two minutes after. Patients who are coming late, triangle are shifted a few minutes after. Average arrival time is to come exactly on time, lower limit is five minutes before and the upper limit is five minutes after.
The Doctor starts working exactly on time. He does not take a gap and works till last patient. Each day has eight hours. Four patients in each type are considered per day. Its average consulting times add up to 480 minutes which is equal to eight hours. Each day every patient comes up, and there are no walk-in patients.

To make our result reliable, a program simulates one thousand days per schedule. The outcome of each rule is calculated by adding up the average waiting time of patients, the average idle time of the doctor and the average overtime. It is fitness of the schedule. In this case the three variables are equally important, but it is possible to put a weight on each of the variables to indicate the importance.

2. LITERATURE SURVEY

The objective of outpatient scheduling is to find an appointment system for which a particular measure of performance is optimized in a clinical environment – it is an application of resource scheduling under uncertainty. The underlying problem applies to a wide variety of environments, such as general practice patient scheduling, scheduling patients for an MRI device, surgical scheduling.

The most primitive form of outpatient scheduling is single block scheduling. The single block rule assigns all patients to arrive at the same time. The patients are served on a first come first serve basis. Another, nowadays more common, form of appointment scheduling is the individual block rule. Patients are assigned unique appointment times that are spaced throughout the clinical session. Bailey (1952) was one of the first to analyze an individual block system. At that time in most clinics it was common practice to assign all patients to arrive at the same time.

Bailey combines single block and individual block scheduling. A number of patients is assigned the same arrival time at the beginning of the clinical session. The idea behind this is to keep an inventory of patients so that the doctor's risk of becoming idle is minimized if the first patient arrives late or fails to show up. All other patients are assigned unique appointment times spread throughout the clinical session.
Bailey used a Monte-Carlo simulation technique to find the number of patients to assign an appointment at the beginning of the session and the length of the intervals between the remaining appointment times.

From this he concluded that he should schedule two patients at the beginning of the session. The remaining patients are scheduled at intervals equal to the mean consultation time. This leads to a reasonable balance between the patient’s waiting time and the doctor’s idle time.

Bailey also found that shorter mean consultation times result in lower patient waiting times. Furthermore, he found that high variability of service times deteriorates both the patients’ waiting times and the doctor's idle time.

Assigning time blocks to surgeons on a first come first served basis to find a balance between the surgeons’s waiting cost, the idle cost of the facilities and operation room personnel is studied by Charnetski (1984) using simulation. The heuristic found distinguishes different types of procedures having different service time distributions and bases procedure times scheduled for a patient on a function of the mean and the standard deviation of the individual service times.

Ho and Lau (1992, 1999) and Ho, Lau and Li (1995) introduce a number of variable-interval rules and test their performance against traditional ones using simulation. Their best performing variable-interval rule increases appointment intervals toward the latter part of the session. They conclude that there is not one rule that performs well under all circumstances and provide a simple heuristic to assist in selecting an appointment rule for a clinic.

Their assessment of three environmental factors (no-show probability, variability of service times, and number of patients per clinical session) reveals that, among the three, the no-show probability is the major one that affects the performance and the choice of an appointment schedule.

Klassen and Rohleder (1996) classify patients based on their expected service time variability. They use simulation techniques to compare various ways of scheduling patients having a relatively high and relatively low service time variability, when appointment intervals are kept standing. They developed a rule that puts patients with lower service time variability before patients with higher service time variability, which performs better than Ho and Lau’s best performing rules. Later on, they consider the possibility that the scheduler can make errors and that not all patients accept every slot they are assigned to. However, they conclude that their rule mentioned above still performs well under these more realistic assumptions.

An appointment system for a multi-server queuing system, where doctors may arrive late, with constant intervals between two successive appointment times and multiple variable blocks is studied by Liu and Liu (1998). They try to minimize the total cost of the patient’s flow-time and the doctors’ idle time by developing a simulation search procedure to appoint the number of patients to assign to each block. They suggest a simple procedure to find an appointment rule for a given environment using the properties of the best rules, derived after simulating several environmental factors.

Swisher et al. (2001) developed a discrete-event simulation model to be applied for decision making in outpatient planning. By utilizing this model to a family practice clinic they observe that the results are quite sensitive to changes in the patient mix, patient scheduling, and staffing levels. The effect of patient scheduling is only studied in changing the instant of the appointment, rather than examining several appointment rules.

The majority of the studies mentioned above assume patients are homogeneous for scheduling purposes, and use independently and identically distributed service times for all patients. Furthermore, those studies do not take into account structural latecomers. Another disadvantage is that most studies focus on a particular problem so the solution is not suitable for any problem.
3. PROPOSED METHOD

First appointment scheduling rule that will be used is the individual block rule. Each patient is given a unique appointment time over the hospital session.

Each block
First appointment scheduling rule that will be used is the individual block rule. Each patient is given a unique appointment time over the hospital session.

Disjunctive graph distance
The rule two is studied is the Disjunctive graph distance rule. It is almost similar to the individual block rule except two patients are scheduled at the beginning of the hospital session.

The Disjunctive graph distance rule assumes patients is having independent identically distributed Gamma service times. It is considered that all patients come exactly on time.

Charnetski
The final rule to be calculated is Charnetski’s rule. It derived a heuristic that differentiate various types of procedures having different service time distributions and bases the amount of time scheduled for a patient on the mean and the standard deviation of the individual service times. Charnetski assumes that every patient come on time, and not take any overtime. The service times are normally distributed and truncated from below at 0. The main focus of the study is to establish a relationship between the scheduled time for a procedure and the average idle time of the doctor and the waiting time of the patients. A standardized prediction heuristic was used to schedule procedure times, given by

\[ d_i(h) = \mu_i + h\sigma_i \]

Where \( h \) is a scalar value held constant across procedures, \( \mu_i \) is the empirically determined mean for procedure i and \( \sigma_i \) is the procedure’s standard deviation. The value \( d_i(h) \) represents the amount of time scheduled for procedure i as a function of the factor h.

Critical Block and DG distance
In patient scheduling problem, we also could find the critical path. Critical path is defined as the longest paths taken from the first operation processed until the last operation leaves the workplace. All operations in this path are called critical operations. Plus, the critical operations on the same machine are called critical block.

We measure the distance between two jobs \( j1 \) and \( j2 \) by the number of differences in the processing orders of operation on each doctor. By doing so, it is just like summing the disjunctive arcs whose directions are different between jobs \( j1 \) and \( j2 \).

Genetic algorithm
Now we are using a genetic algorithm. The genetic algorithm finds its roots in artificial intelligence. This heuristic is used to generate useful solutions to optimization and search problems. Genetic algorithms belong to the larger class of evolutionary algorithms, which generate solutions to optimization problems using techniques inspired by natural evolution.

Initialization
From an initial population a large number of individual solutions are randomly generated. The solutions may be pointed to areas where optimal solutions to be found.
Selection

The outcome of every solution is calculated by a function called the fitness function. This fitness function points the performance of the specific solution compared to the other solutions. A lower score on the fitness function points a higher performance. Based on these fitness values the population is established, where solutions that are less fit are more likely to be selected for deletion.

Reproduction

Now to calculate a second generation population from the solutions of selected ones. It can be obtained through genetic operators like recombination and mutation. In recombination two existing solutions are combined to find two new solutions. Mutation nominally changes several randomly selected existing solutions. Occasionally, the solutions with the highest fitness are excluded from the evaluation process mentioned above to prevent the algorithm to get stuck at a local optimum. It is repeated until the population has obtained the same size to previous generation. These processes result in next generation population that is different from the initial generation. The average fitness has increased by this process for the population; the best solutions from the first generation are selected to find new solutions.

Termination

Selection as well reproduction steps are repeated till a termination condition reached.

Normal terminating conditions are:

- There might be a minimum criterion to the problem. Process can be terminated provided the minimum criterion is reached or a fixed number of generations is reached;
- The highest ranking solution's fitness is reaching or has reached a plateau such that successive iterations no longer produce better results.

Assumptions

The arrival process and the service time of a patient are modeled as a probability process. We have four types of treatments, with short and long service times, and large and small standard deviations. Few patients tend to arrive late. Patients are served according to the following rules.

- When there are no patients in the waiting room a newly arriving patient is served immediately;
- If there is one patient in the waiting room then the patient is served;
- If there is more than one patient in the waiting room, the patient with the earliest scheduled time is served;
- All patients are served, regardless of the scheduled finishing time of the doctor.

Parameters

Patients never arrive on time for appointment. Most patients arrive early; Many patients have a tendency to come late. Both early and late patients are modeled with a triangular arrival distribution, with a different mean. There are four types of treatments with either a relatively long or a short service time and a larger or a smaller standard deviation. Service times are drawn from a lognormal distribution.

Heuristic

The diagram below shows the steps of the simulation tool while simulating one day. This will be repeated for a large number of days. Preliminary, the following steps have to be taken.
1) The schedule to be measured has to be entered in the simulation tool;
2) For every patient in the schedule the actual arrival time has to be determined;
3) All patients must be sorted in order of their actual arrival time.

The doctor idle time is retained, as is the waiting time for every patient. At the end of the day, the tardiness is determined. The doctor idle time is the time in between two services. The waiting time of the patient is the difference between their appointment time and the time the patient is actually served. This excludes any waiting prior to appointment time, because additional waiting due to early arrival is voluntary and is not a consequence of the appointment schedule. If the patient is served before the appointment time the waiting time is zero. Late patients may consider some additional waiting as normal, being partly their own fault. Their waiting time is set to zero as well.

Output
The following output is generated from the simulation tool

- The mean and standard deviation of the waiting time of the patient;
- The mean and standard deviation of the idle time of the doctor;
- The mean and standard deviation of the tardiness.
4. EXPERIMENTAL STUDY AND COMPARISONS

The different algorithms we picked to compare to the genetic algorithm do not say that much about the order we have to schedule the patients in, by means of average service times, standard deviations, or arrival times. We have tried several scenarios that seem to be obvious using some knowledge from other literature. According to Bailey, shorter mean consultation times result in lower patient waiting times. Klassen and Rohleder concluded that patients should be scheduled in order of increasing standard deviation. It seems logical to schedule latecomers after punctual patients, because you do not want your system to get messed up already at the beginning of the day. We have to find the order of importance to rules mentioned above.

**Individual Block**

The time reserved for every patient is the average service time for that patient. For the individual block system we found the following results:

The patients are scheduled in order of:

1. Increasing service time;
2. Increasing standard deviation;
3. Increasing possibility of arriving late.

<table>
<thead>
<tr>
<th>Mean OT</th>
<th>6.6037</th>
<th>Variance OT</th>
<th>30.9676</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WT</td>
<td>3.1305</td>
<td>Variance WT</td>
<td>15.9294</td>
</tr>
<tr>
<td>Mean IT</td>
<td>6.4110</td>
<td>Variance IT</td>
<td>18.3622</td>
</tr>
<tr>
<td>Fitness</td>
<td>16.1476</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subsequently, patients are scheduled in order of:

1. Increasing standard deviation;
2. Increasing service time;
3. Increasing possibility of arriving late.

<table>
<thead>
<tr>
<th>Mean OT</th>
<th>6.9242</th>
<th>Variance OT</th>
<th>35.0269</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WT</td>
<td>2.9906</td>
<td>Variance WT</td>
<td>15.0648</td>
</tr>
<tr>
<td>Mean IT</td>
<td>6.4982</td>
<td>Variance IT</td>
<td>18.1916</td>
</tr>
<tr>
<td>Fitness</td>
<td>16.4131</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Finally, patients are scheduled in order of:

1. Increasing possibility of arriving late;
2. Increasing service time;
3. Increasing standard deviation.
Results for the individual block system

<table>
<thead>
<tr>
<th></th>
<th>Mean OT</th>
<th>Variance OT</th>
<th>Mean WT</th>
<th>Variance WT</th>
<th>Mean IT</th>
<th>Variance IT</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>OT</td>
<td>3.3455</td>
<td>24.7531</td>
<td>13.3588</td>
<td>106.8742</td>
<td>0.0537</td>
<td>0.1463</td>
<td>16.7580</td>
</tr>
</tbody>
</table>

Critical Block and DG distance

The Critical Block and DG distance is very similar to the individual block system. The only difference now is that we schedule the final patient at the same time as the first patient. From the above results we find that there is not that much difference in the performance of the different schedules, but the first scheme works slightly better than the following two, so we will use this scheme to apply disjunctive graph distance rule on.

First, we give the final patient from the schedule above an appointment at time zero.
Now we put the last patient with a short service time at time zero and shift all patients with a long service time ten minutes back.

<table>
<thead>
<tr>
<th></th>
<th>Mean OT</th>
<th>Variance OT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WT</td>
<td>6.9368</td>
<td>40.5728</td>
</tr>
<tr>
<td>Mean IT</td>
<td>0.8513</td>
<td>5.3766</td>
</tr>
<tr>
<td>Fitness</td>
<td>11.2109</td>
<td></td>
</tr>
</tbody>
</table>

Results for the disjunctive graph distance

Finally, we chop the day in half and pretend the morning is reserved for patients with a short service time, and patients with a long service time are served in the afternoon. Both in the morning and the afternoon, the final patient is put at time zero of the part of day.

<table>
<thead>
<tr>
<th></th>
<th>Mean OT</th>
<th>Variance OT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean WT</td>
<td>10.4235</td>
<td>92.7438</td>
</tr>
<tr>
<td>Mean IT</td>
<td>1.0056</td>
<td>4.8501</td>
</tr>
<tr>
<td>Fitness</td>
<td>15.3262</td>
<td></td>
</tr>
</tbody>
</table>

Results for the disjunctive graph distance, final patient with a short service time at time zero and final and first patient with a long service time at the same time.
Combination

Finally, the last patient with a short service time is put at time zero, and the patients with long service time are shifted back the amount of time that is scheduled for this patient.

The last patient with a short service time is put at time zero, and the patients with long service time are shifted back the amount of time that is scheduled for this patient.

The last patient with a short service time is put at time zero, and the patients with long service time are shifted back the amount of time that is scheduled for this patient.

Genetic algorithm

To find a better solution than the previous ones, we tried to optimize the fitness of the schedule using the genetic algorithm solver. The variables are the arrival times of the patients, which can vary between 0 and 480.

In scheduling, genetic algorithms represent schedules as individuals or patients. Each individual has its own fitness value which is measured by the objective function. The procedure works iteratively, and this iteration is a generation. Each generation consists of individuals who survive from the previous generations.

We started with a population of 20 initial solutions and the number of generations was equal to 100. The initial population contained 20 individuals with 32 random integers between 0 and 480. Unfortunately, this resulted in something worse than everything we found up until now. Subsequently, we started to increase the population size which improved our results, but we were still not able to find a fitness that exceeded the previous results.

Our next approach was to search in the neighborhood of the best results we found up until now. We created an initial population where every variable from every individual was in a range of 5 minutes of the optimal value up until now. The following results were found applying the genetic algorithm with a population size of 250 and the number of generations equal to 200.

Results for the Genetic Algorithm compared to Bailey & Welch.

In the graph above it can be seen that with the genetic algorithm we are able to improve the waiting time. Although the idle time has increased a little bit, we have managed to decrease the overall fitness almost two minutes by taking a lot of time of the average waiting time for the patients.
The image above shows the optimal solution found with the disjunctive graph distance approach, in blue, and the optimal solution found with the genetic algorithm in red. Every bar represents an appointment. The two higher bars at time zero represent two arrivals at the same time. It is remarkable that the genetic algorithm schedules patients that arrive in time very close to the critical block schedule, while patients that have a tendency to arrive late are scheduled several minutes later than critical block. We see that the third patient to arrive is shifted four minutes forward compared to the critical block schedule. This makes sense, since this patient has to wait for two patients to be served in front of him, instead of one.

5. CONCLUSION

With the results we can state that it is indeed possible to find a schedule with the genetic algorithm that outperforms the other algorithms mentioned. However it was not that easy as we thought beforehand. We got stuck in a local optimum several times, so we really had to push the algorithm in the right direction to get some satisfying results. We have seen that, although the disjunctive graph rule is very efficient, it is still a very good performing rule that could not be matched by Critical Block in this setting. Also we saw that it is wise to consider some scheduling rules instead of simply implementing the individual block system, since every rule we applied outperformed the individual block rule. Unfortunately, we noticed that the variations of the waiting time, idle time and overtime were pretty high, which means that the results we found may count for the long run, but we will probably not find the same results if we observe only a few days.

REFERENCES


