Patient Journey Optimization Using A Multi-Agent Approach

Chung Ho CHOI

Abstract

With the increasing expectation from patients and the regulations enacted by the government, exploring ways to shorten patient journey has caught increasing attention. Patient journey optimization typically involves treatment operation scheduling at multiple medical units. The decentralized nature of the problem makes conventional centralized operation research methods inapplicable in general and motivates the use of the multi-agent approach. In this paper, we focus on cancer patient treatment. We model patients and medical units as autonomous agents which interact locally via a bidding process and a coordination process for patient journey optimization. With reference to a dataset containing more than five thousand different cancer patient journeys, the effectiveness of the proposed algorithm under different settings of implementation has been experimentally evaluated via experimental simulations.

1 Introduction

Shortening the length of patient journey is always an expectation of patients. Also, it is one of the vital key performance indicators for evaluating the effectiveness of healthcare service providers. In order to best utilize the existing resources to raise healthcare service level, careful application of scheduling algorithms is crucial [7]. While conventional operations research methods have been found effective for scheduling related problems found in different medical units [8, 9, 4], most of them take the centralized approach and are not designed to be applied to decentralized situations where multiple resources at multiple hospitals are to be coordinated [5, 7, 2].

The multi-agent approach is characterized by emphasizing on local interaction and self-organization of different entities being modeled. These properties make it especially suitable for tackling complex tasks with a lot of stakeholders [11, 1]. Multi-agent methods have found applications in a variety of problem domains, such as airport resource scheduling [3], load allocation in transportation logistics [6], supply chain management [10], etc. Recently, it has also been applied to patient scheduling in [5, 7] with some

initial success demonstrated. And yet, there are limitations. Paulussen *et al.*, in [5], assume that a quantified health state can be accurately derived as a utility measure for guiding the scheduling process. In [7], Vermeulen *et al.* did not consider the temporal constraints between the treatment operations during the scheduling process.

In this paper, we focus on the problem of improving the healthcare experience of cancer patients by optimizing their patient journeys. It is common that many developed cities have established a number of cancer centers to pool up resources for best utilization and yet to ensure that the centers are properly located to cater for the needs in different regions of the cities. However, the geographical distribution of the patient demand is dynamic in general and hard to be accurately predicted at different time points. The objective of this study is to explore the extent to which patient journey can be improved by better coordinating and mobilizing resources distributed at different cancer centers. In particular, we formulate the problem based on the healthcare system for cancer treatment in Hong Kong. We propose the use of a multi-agent approach where each patient and each unit of resources are modeled as autonomous agents which can interact with others to arrive an effective overall schedule with reduced amount of patient waiting times. To evaluate the effectiveness of the proposed approach, we made use of a patient identity anonymized data set collected by Hospital Authority in Hong Kong which contains 5819 cancer patients with an admission period spanning over 6 months, and have carried out simulations with the proposed approach given different settings of the environment.

The rest of the paper is organized as follows. The patient scheduling problem formulation is described in Section 2. Section 3 and Section 4 present the details of the proposed agent-based scheduling algorithm. Section 5 presents some preliminary experimental results and Section 6 concludes the paper.

2 PROBLEM FORMULATION

In this section, we first briefly describe the establishment of the cancer centers in Hong Kong. Then, we formulate the patient scheduling problem for cancer treatment as an optimization problem and explain how a particular multi-

agent approach can be adopted to address the distributed nature of the problem.

2.1 Cancer Patient Treatment - A Hong Kong Scenario

In Hong Kong, there are seven cancer centers. Figure 1 shows the geographical distribution of the seven cancer centers in Hong Kong. With the objective not to reveal the performance of individual centers, we denote the set of the seven centers as $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_7\}$. Currently, these cancer centers are essentially managed by themselves at the operational level. That is, on-demand information exchange among the centers for scheduling patients is not yet extensively used. That is why it is common for cancer patients to be scheduled for receiving a series of treatments all at only one cancer center, even though the same or some of them could be provided significantly earlier by other centers.
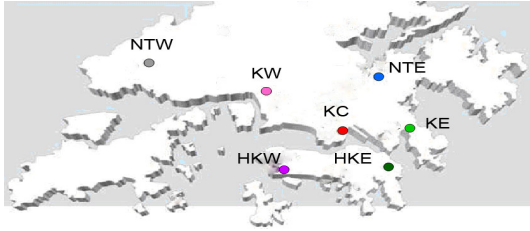
Figure 1. Seven geographically distributed cancer centers in Hong Kong.

Generally speaking, for any patient admitted to a hospital (cancer center), he or she first consults a medical doctor for diagnosis. Once the case is suspected to be cancer, the doctor will specify the patient a treatment plan which contains a sequence of treatment operations. For the application domain being chosen, we denote the set of treatment operations as $\Gamma = \{$radiotherapy planning, radiotherapy, surgery, chemotherapy$\}$.

To carry out the treatment operations, medical resources are needed. As resources are usually limited, related scheduling and allocation are by no means trivial. We denote the set of medical resources (or units) for our domain as $A = \{$radiotherapy planning unit, radiotherapy unit, operation unit, chemotherapy unit$\}$. We assume that one treatment operation can only be performed at one medical unit of the corresponding type. A patient journey is defined as the duration from the date of admission to the date of the last treatment operation completed. The goal of this work is to shorten the patient journey as far as possible.

2.2 Formulation

Let $K := A \times \mathcal{C}$ be the cartesian product of A and $\mathcal{C}$ giving the complete set of medical units, $M := K \to \Gamma$ be an one-to-one mapping between K and Γ specifying the treatment type of the medical units, and P be the set of cancer patients being scheduled.

Also, given a patient i, let N_Γ^i denote the number of treatment operations needed, D_0^i denote the admission date, D_j^i denote the date of the jth treatment operation where $1 \le j \le N_\Gamma^i$, $V_j^i \in K$ be the unit at which the jth treatment operation is performed, $Tr_j^i \in \Gamma$ be the type of treatment for the jth operation, C_k be the daily capacity (i.e., the maximum number of patients that could be treated) of medical unit $k \in K$, T_t be the duration (in days) of treatment type $t \in \Gamma$, and Z be the set of dates on which patient scheduling is being considered.

With the assumption that all the patients are being treated equally in terms of urgency, the scheduling problem can be formulated as:

$$\min_D \sum_{i=1}^{|P|} \sum_{j=1}^{N_\Gamma^i - 1} (|D_j^i - D_{j+1}^i|) \qquad (1)$$

with the following constraints to be satisfied:

$$D_{j+1}^i > D_j^i + T_{Tr_j^i} \qquad (2)$$

$$\forall d \in Z \quad \left| \{i : D_j^i = d \wedge V_j^i = k\} \right| \le C_k \qquad (3)$$

$$Tr_j^i = M(k) \qquad (4)$$

$$D_j^i > D_0^i > 0. \qquad (5)$$

The objective function in (1) is to minimize the time lags between treatment operations for cancer patients. Constraint (2) ensures the temporal constraints between treatment operations are not violated, constraint (3) is used to ensure all medical units are operating within their capacities. Constraint (4) ensures that only the right kind of treatments are assigned to each medical unit. Constraint (5) ensures that patients would only be scheduled to receive treatment operations after their admissions.

3 SCHEDULING FRAMEWORK

Theoretically, patient journeys "travelling" possibly different centers could be scheduled by directly optimizing the criterion function (1). However, it is impractical to do so as it is hard to assume that a cancer center is willing to share its real-time resource allocation related data (e.g., C_k) with other centers due to both technical and managerial reasons. So, on one hand, the information sharing infrastructure across different information systems of different centers' medical units has to be ready. Also, each medical unit

is used to be managed by its own staff using its own information system for resource allocation. Having the patient scheduling at different medical units to be effectively coordinated to optimize the patient journey experience poses challenges on how the resource should be managed accordingly even within a hospital, let alone crossing all the hospitals.

In this section, we argue that the use of the multi-agent approach which tries to model each stakeholder as an autonomous agent and allows only local interactions among the agents is particularly suitable for our application. It aims to minimize the information sharing requirement among the hospitals and yet to obtain a good enough suboptimal result for the (global) patient journey optimization. In our proposed framework, there are *patient agents* and *resource agents*. They interact via some designed protocol for achieving the aforementioned optimization.

3.1 Patient Agent

A patient agent is used to represent one cancer patient and is denoted as P_i with $i = 1, 2, ..., |P|$. It stores the patient's treatment plan. As it is common that some treatment operations have to be performed in prior to the others, the set of treatment operations to be received by a patient has to be ordered to satisfy certain temporal constraints. Hence, each patient agent P_i maintains an ordered set $Tr^i = \{Tr_1^i, Tr_2^i, ...Tr_{N_\Gamma^i}^i\}$ as its treatment plan.

3.2 Resource Agent

A resource agent is used to manage a specific medical unit. Here, we denote R_{ab} as a resource agent representing medical unit $a \in A$ at cancer center $b \in C$. Each resource agent has full access to the schedule of the medical unit it represents, but not those of the others.

3.3 Scheduling Algorithm

We adopt a two-phases scheduling algorithm similar to what being proposed in [5, 7]. For each newly admitted patient, a treatment plan is first designed and then the corresponding treatment operations are initially scheduled with the temporal constraints of the operations and the unit capacity taken into account (initial assignment phase). Then, a time-slot swapping process is enforced for further enhancement in patient journey quality (rescheduling phase). Here we assume that two patient agents with their treatment plans containing operations of the same type but scheduled at different time are willing to be matched by resource agents and to exchange the corresponding timeslots as far as none of their schedules is worsen (as suggested in [7]) and none of

the temporal constraints as specified in (2) are violated.[1] Algorithm 1 gives a high-level description of this two-phases scheduling algorithm.

Algorithm 1 Scheduling Algorithm

1: **for** every patient agent P_i **do**
2:     Initial assignment based on P_i's treatment plan
3:     **for** each P_i's treatment operation **do**
4:         Rescheduled to be performed earlier by exchanging timeslot with another patient agent with the help of the resource agent (rescheduling phase)
5:         **if** No involving parties are worsened in terms of their resulting overall schedules **then**
6:             The exchanging process is proceeded
7:         **end if**
8:     **end for**
9: **end for**

---

4 AGENT COORDINATION

In this section, more details about the scheduling algorithm are given, including (1) how the patient agents interact with the resource agents to establish candidate patient agent pairs for timeslot swapping consideration, and (2) how some "unnecessary" swappings can be rejected so as to further improve the scheduling optimality.

4.1 A bidding process for agent matchmaking

Figure 2 shows our proposed framework. As what have been introduced in Sections 3.1 and 3.2, there are two types of agents, namely patient agents and resource agents. In order to show clearly the coordination between agents, we further categorize patient agents during a matchmaking process into *initiating patient agents* and *target patient agents*. Initiating patient agents P_I are those patient agents who initiate a request for timeslot exchange. Target patient agents P_G are the others who are willing to participate in the exchanging process.

With the objective to shorten its patient journey, an initiating patient agent P_I first sends out a request of rescheduling one of the treatment operations to the corresponding resource agents R_{ab}. The request includes the earliest possible start date (*EPS*) and the latest possible start date (*LPS*) of its associated treatment operation. In order not to violate the temporal constraints between treatment operations, the *EPS* can be defined as:

$$EPS_j^I = D_{j-1}^I + T_{Tr_{j-1}^I} + \delta_1. \tag{6}$$

---
[1]This assumption may imply that some policy-wise incentive to be in place so that different medical units are willing to share their resources in this manner, which however is not the main focus of our study.
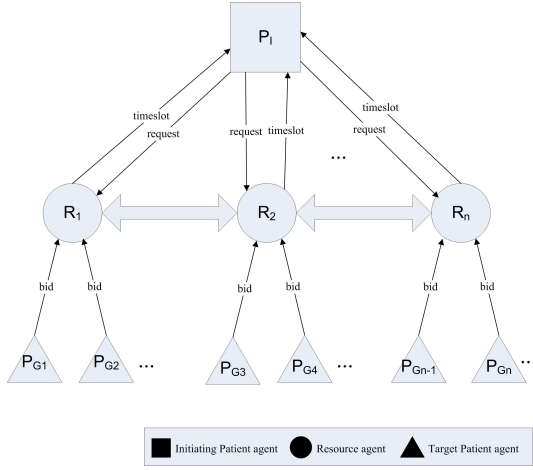
Figure 2. The proposed agent coordination framework for patient scheduling.

Note that δ_1 denotes how many days a patient should be admitted (if needed) before a treatment operation to be carried out. In our experiment, we set to be one. In practice, this value could be designated by healthcare providers in order to better suit their needs. With a similar argument, LPS is defined as:

$$LPS_j^I = D_j^I - 1. \quad (7)$$

Once a resource agent receives a request with EPS and LPS, it will first check whether there are available timeslots released by ceased patients which can fulfill the request. If yes, the released timeslot will be assigned to the initiating patient agent. If not, the resource agent will then pass the request to those patient agents (target patient agents, P_G) which reserved resources of the same type in the period from EPS to LPS. Those target patient agents who have received the request will submit a bid to the resource agent in response.

There are several factors needed to be considered in computing the bid value.

- First, the target patient agent should not have its last operation in its treatment plan to be exchanged, or its last operation has then to be performed later and thus it would end up with a lengthened journey.

- Second, as it is impractical to reschedule a patient's treatment operation without prior notification, we assume that the exchange of timeslots would not be considered if the initiating patient will have less than a week's time of notification.[2]

---

[2]In general, the time of notification can be adjusted according to the real situation.

- Third, the target patient agent also has to ensure that the temporal constraints between its treatment operations would not be violated after the exchanging process.

Taking into account the above considerations, the bid value submitted by a target patient agent P_G with its j_tth operation that matches with the j_ith operation of the initiating agent P_I as specified in the request (that is $Tr_{j_t}^G = Tr_{j_i}^I$) is formulated as:

$$Bid^G = (D_{j_t}^G - EPS_{j_i}^I) + Last + Noti + Temp, \quad (8)$$

where $Last$, $Noti$ and $Temp$ are three binary variables. $Last = 0$ if the j_tth operation is not the last one for P_G, or ∞ otherwise. $Noti = 0$ if a week's time of notification for the target patient agent to be notified, or ∞ otherwise. $Temp = 0$ if there are no temporal constraints violated, or ∞ otherwise.

Among all the target patient agents, the one with the lowest bid value will be accepted and the timeslot swapping between the initiating agent and target one will be confirmed. If two bids are found to be numerically identical, the resource agent will select one at random.

4.2 A coordination process for rejecting unnecessary swappings

A timeslot swapping confirmed as described in the previous section sometimes does not necessarily lead to ultimate improvement in patient journey. To illustrate that, suppose there is a patient agent with 3 treatment operations to be rescheduled. In case the last treatment operation could not be rescheduled to be performed earlier, any rescheduling of the first 2 treatment operations are essentially useless as the duration of the whole journey remains unchanged (see Figure 3(a)). As another example, even a shortened patient journey can be achieved, rescheduling of the first 2 treatment operations could also be useless if the rescheduling of the last treatment operation cannot be benefited from the rescheduling of the first two (see Figure 3(b)).

In order that these useless swappings can be rejected so as to be reserved for other potentially more useful swappings, the scheduling algorithm could be modified in such a way that a resource agent after identifying the most optimal bid among the target patient agents will not notify the initiating patient agent immediately. Instead, it will pass the bid to the resource agent which is responsible for the succeeding treatment operation of the initiating patient agent. Having received such a bid, the resource agent could derive a new EPS, denoted as $^{(new)}EPS_{j_i+1}^I$. Clearly, unnecessary swappings occur if that resource agent could not find a bid among those received from the target patient agent P_G such that $^{(new)}EPS_{j_i+1}^I \leq D_{j_t'}^G \leq EPS_{j_i+1}^I$, where
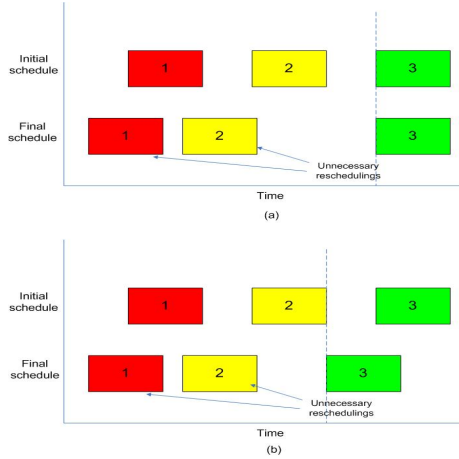
Figure 3. Unnecessary reschedulings.

$Tr^I_{j_i+1} = Tr^G_{j^t_t}$. In that case, the resource agent will notify its antecedent to discard the bid such that the corresponding timeslots would not be exchanged. In general, such a succeeding resource agent consultation process can be carried out in a recursive manner.

5 EXPERIMENTAL VALIDATION

To evaluate the effectiveness of the proposed multi-agent approach, we first obtained a dataset containing the scheduled treatment plans of 5819 cancer patients being treated at the seven cancer centers in Hong Kong within an admission period of 6 months (from 1/7/2007 to 31/12/2007). The average length of the patient journey among all cancer centers is 90.7 days. Based on the dataset, we have carried two groups of experimental simulations.

For the first group of simulations, we made use of the scheduled treatment plans in the dataset for the initial assignment and studied to what extent the multi-agent approach can improve the patient journey as a whole. As only swappings and taking timeslots of ceased patients were allowed in our simulations, the capacity requirement is essentially equivalent to that of the original data.

For the second group of simulations, we aim to make the simulation setting more flexible by making use of only the statistics of the scheduled treatment plans and the capacity of the medical units allocated as revealed in the dataset. The initial assignment strategy and the rescheduling strategy can both be specified by the user for evaluation. Also, we tried to increase the capacity value as revealed in the dataset by percentage so as to see to what extent the patient journey can be improved with more resources injected.

5.1 Simulations with initial assignment and unit capacity fixed (Exp. A)

Four different settings have been tested so as to demonstrate various aspects of the proposed algorithm. They include

Setting 1: Any patient agent is willing to explore the possibility to swap some or all of its treatment operation schedule with others'.

Setting 2: It is assumed that only 20% of the patients of each center are allowed to undergo timeslot swapping.

Setting 3: It is assumed that patients are reluctant to travel for a long distance even though some of their operations can be scheduled earlier, and thus only swappings between two nearby cancer centers are allowed. In particular, the neighborhood relationships are assumed to be

- $C_1 \rightarrow C_2$ or C_3
- $C_2 \rightarrow C_1$ or C_3
- $C_3 \rightarrow C_1$ or C_2 or C_4 or C_5 or C_6
- $C_4 \rightarrow C_3$ or C_6
- $C_5 \rightarrow C_3$ or C_7
- $C_6 \rightarrow C_3$ or C_4
- $C_7 \rightarrow C_5$

where $\alpha \rightarrow \beta$ implies that patients admitted in cancer center α would only be swapped to its neighboring cancer center β.

Setting 4: Timeslots released by ceased patients are allocated to the patient agents who have the longest patient journeys at a time point.

Figure 4a shows the average length of the patient journeys associated with the seven cancer centers in Hong Kong obtained given the four aforementioned settings.

The experimental results obtained shows that, on average, the average length of journey could be reduced by 9.8 days for those 5819 cancer patients if no restriction is imposed on the exchange of timeslots whenever there is a Pareto improvement (Setting 1). Given only 20% of patients per center are allowed for timeslot exchange (Setting 2), it is found that the average length of journey could still be reduced by an average of 6.1 days. This observation is interesting as even a small percentage of resources being released by each center could already achieve a significant improvement. With the geographical restriction on allowing only swappings between nearby centers as described in Setting 3, the average length of journey can also be reduced by 9.3 days.
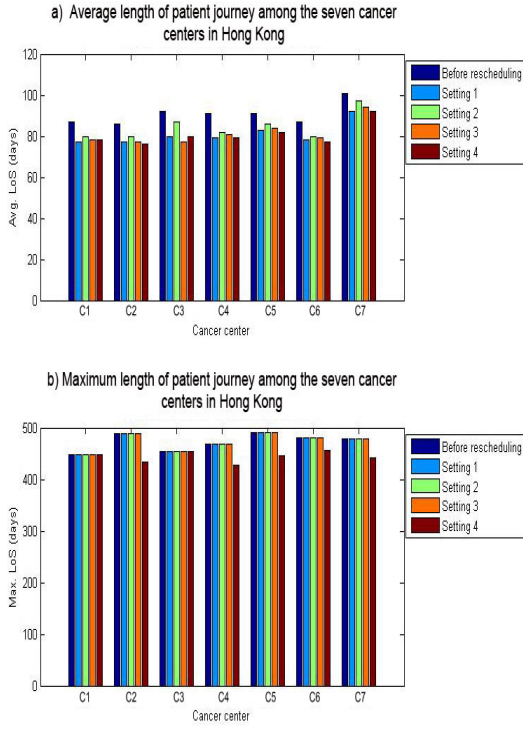
Figure 4. a) Average length of patient journey; b) Maximum length of patient journey among the seven cancer centers in Hong Kong under 4 different experiment settings in Exp A.

However, it should also be noted that according to Figure 4b, the maximum length of journey remains unchanged. The reason is obvious as no one is willing to swap with those with the longest length of journey and the last day of the treatment to be the end of the period being considered. Improvement on reduction on maximum length of journey can only be observed for Setting 4 where the released timeslots due to ceased patients are allocated to those with long length of journey.

5.2 Simulations revealing the effects of varying initial assignment and unit capacity (Exp. B)

Contrary to the group of simulations previously presented, we tried to simulate also the initial assignment process using again the treatment plan of the 5819 cancer patients. However, all the treatment operations were not scheduled according to the data but simulated based on the statistics of the inter-operation duration obtained from the dataset. In particular, during the initial assignment phase, patient agents would be assigned an initial schedule one by one based on their admission orders. For each treatment operation, each patient agent would be assigned with the next earliest available timeslot. However, it is obvious that there should be a minimum time lag between two subsequent treatment operations due to different medical reasons. We first used the average inter-operation duration computed based on the dataset to set the related parameters as shown in Table 1 and then performed the simulation. The average length of journey computed right after the initial assignment is found to be 83.3 days. Figure 5 shows that the improvements obtained due to the different settings are not very much different from those obtained in the previous section.

Table 1. Average time lags between treatment operations (RTP - Radiotherapy planning; RT - Radiotherapy; OT - Operation; CHEM - Chemotherapy).

Antecedent treatment operation (A)	Subsequent treatment operation(S)	Average time lag (days) between the start date of (A) and (S)
RTP	RT	21
RTP	OT	44
RTP	CHEM	23
RT	OT	94
RT	CHEM	39
RT	RTP	27
OT	RTP	55
OT	RT	67
OT	CHEM	54
CHEM	RTP	104
CHEM	RT	56
CHEM	OT	97

According to the dataset, the minimum days between any two treatment operations was found to be one only. This implies that treatment operation sometimes could be started one day after another if the resource is available. We have also tried to set the minimum days to be one in our initial assignment phase and compared the results obtained based on those presented in Table 1. By Figure 6, while we observed some improvements in performance, the enhancement however is not very significant. Hence, setting some reasonable time lags between treatment operations does not have too big an impact on lengthening too much the length of journey. Of course, there should always be medical reasons for setting a bigger minimum threshold.

For all the results presented so far, it is assumed that the capacity of each medical unit is fixed. To study the cost-effectiveness of increasing the units' capacities for patient journey optimization, we increased the capacity by the same percentage for all the medical units. According to Figure 7,
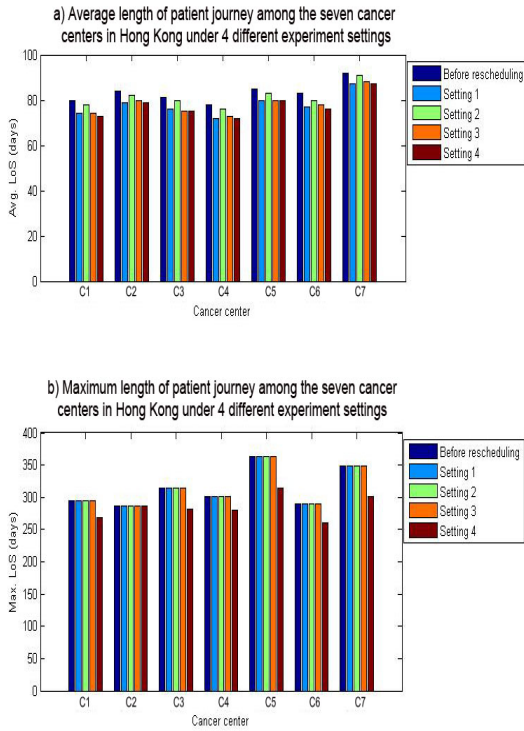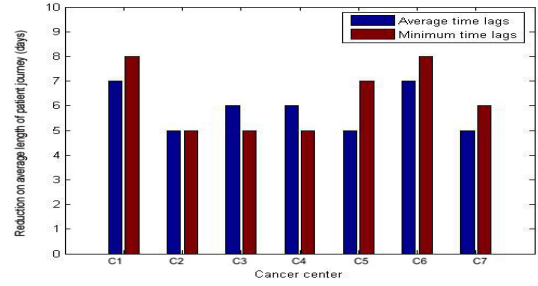
Figure 5. a) Average length of patient journey and; b) Maximum length of patient journey among the seven cancer centers in Hong Kong under 4 different experiment settings in Exp B.

Figure 6. Reduction on average length of patient journey (Setting 4) by varying the time lags between treatment operations.

Figure 7. Reduction on average length of patient journey (Setting 4) by varying capacities.

it is found that when all the resource capacities are increased incrementally (10%, 20%, 30%), the reduction on average length of patient journey will then drop accordingly. In particular, it is found that a higher increase in capacities will cause a greater drop as a result. In fact, such drop could be attributed to the fact that when the resource capacities are increased, patients would then be scheduled with less idle times between treatment operations; and hence with less chance to exchange timeslots with others.
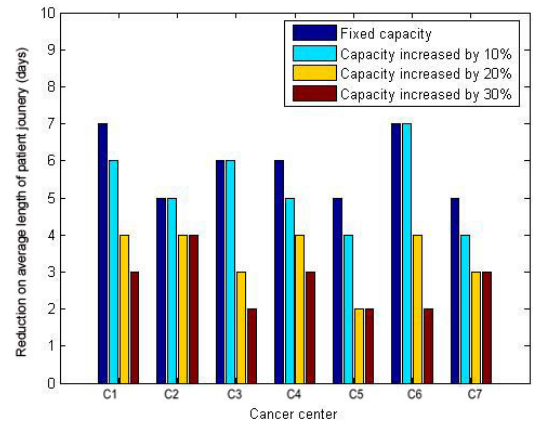
6   CONCLUSIONS

In this paper, a multi-agent approach was proposed for patient journey optimization. Particularly, by applying the approach, the shortening of the overall patient journey will not result in any single patient with its length of journey lengthened as a result. Also, all the temporal constraints among the treatment operations for each patient would not be violated during the scheduling process.

The effectiveness of the proposed approach has been demonstrated by applying it to a dataset containing 5819 scheduled treatment plans of cancer patients admitted to hospitals in Hong Kong. The effects of varying the initial assignment and the unit capacity on the overall reduction in length of patient journey are also studied.

Currently, since we are using a Pareto improvement approach, it is assumed that no single patient (agent) would get a lengthened schedule after swapping timeslots with another. In the future, we are going to see whether there would be a greater improvement in achieving a reduced overall length of patient journey when the above assumption is relaxed. In particular, we are going to see whether the bid as formulated in Section 4.1 could be defined in a more sophisticated way such that a greater reduction on length of patient journey could be achieved by the matchmaking process.

ACKNOWLEDGEMENT

References

[1] H. Czap and M. Becker. Multi-agent systems and microeconomic theory: A negotiation approach to solve scheduling problems in high dynamic environments. In *HICSS '03: Proceedings of the 36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 3*, page 83.2, Washington, DC, USA, 2003. IEEE Computer Society.

[2] K. Decker, J. Li, and Y. Demazeau. Coordinating mutually exclusive resources using gpgp. *Autonomous Agents and Multi-Agent Systems*, 3:200–0, 2000.

[3] X. Mao, A. Mors, N. Roos, and C. Witteveen. Coordinating competitive agents in dynamic airport resource scheduling. In *MATES '07: Proceedings of the 5th German conference on Multiagent System Technologies*, pages 133–144, Berlin, Heidelberg, 2007. Springer-Verlag.

[4] J. Patrick and M. Puterman. Reducing wait times through operations research: optimizing the use of surge capacity. *Healthc Q*, 11(3):77–83, 2008.

[5] T. O. Paulussen, I. S. Dept, K. S. Decker, A. Heinzl, and N. R. Jennings. Distributed patient scheduling in hospitals. In *Coordination and Agent Technology in Value Networks. GITO*, pages 1224–1232. Morgan Kaufmann, 2003.

[6] V. Robu, H. Noot, H. La Poutré, and W.-J. van Schijndel. An interactive platform for auction-based allocation of loads in transportation logistics. In *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 3–10, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[7] I. Vermeulen, S. Bohte, K. Somefun, and H. La Poutre. Improving patient activity schedules by multi-agent pareto appointment exchanging. In *CEC-EEE '06: Proceedings of the The 8th IEEE International Conference on E-Commerce Technology and The 3rd IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services*, page 9, Washington, DC, USA, 2006. IEEE Computer Society.

[8] J. Vissers and R. Beech. *Health operations management : patient flow logistics in health care*. Routledge, 2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN, 2005.

[9] J. Vissers, J. Bekkers, and I. Adan. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*, 16, 2005.

[10] M. Wang, J. Liu, H. Wang, W. K. Cheung, and X. Xie. On-demand e-supply chain integration: A multi-agent constraint-based approach. *Expert Systems with Applications*, 34(4):2683 – 2692, 2008.

[11] G. Weiss. *Multiagent systems : a modern approach to distributed artificial intelligence*. Cambridge, Mass. : MIT Press, 1999.