Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011)
P. Qian, Y. Zhou, C. Rudin, eds.

Multi-Agent Patient Scheduling Through Auctioned Decentralized MDPs

Hadi Hosseini, Jesse Hoey, Robin Cohen
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, ON
{h5hosseini,jhoey,rcohen}@uwaterloo.ca

Abstract

In this paper, we consider a novel approach to a multiagent resource allocation problem that arises in medical settings in which a set of consumers (patients) require a finite set of resources and the overall state of all consumers in the system needs to be optimized. With many consumers and resources, it is intractable to compute an optimal solution. Yet, there is a measure of utility or consumer success (e.g. the health state of the patients) that can be leveraged towards determining preferable allocations. We propose an approximate solution in which each consumer is represented by a Markov Decision Process (MDP) based agent that models the expected progression of its utility given the resources, and has an approximate model of its expected access to resources given the global state of all competing agents. A global auction-based mechanism is then proposed that requires each agent to bid for resources based on its expected gains (e.g. its long-term health) from obtaining each resource. The allocation of resources proceeds by maximization over bids in an iterative multi-round auction. We are able to show that the global utility value across all consumers is closer to optimal using our approach compared to competing approaches. In addition, our approach is sufficiently flexible to support the integration of temporal constraints between resources.

1 Introduction

A hospital can be seen as a number of different departments with semi-autonomous functionalities [7]. This separation enables more efficient ways of organizing resources and gives more flexibility; however, it produces a vital need for coordination between different departments. To optimize these inter-related processes, not only should all the departments optimize their internal throughput but it is also necessary to coordinate all the resources and tasks to maximize the utility of the whole system.

In this paper, we approach the dynamic stochastic resource scheduling problem for domains in which the success of each task or agent is dependent stochastically on its ability to obtain a sequence of resources over time. In the medical domain, agents represent patients who require a set of treatments or diagnostic procedures, each of which can progress its health state (and conversely, lack of which can lead to decrease in health). In this medical setting, patients have different conditions (e.g. diseases), resources are shared between patients, resources have uncertain effects on health, but each patient cannot associate any monetary value with health.

We formulate this problem as a decentralized Markov decision process (dec-MDP), in which each consumer (patient) is represented by a single MDP that has a utility function over its success (health), and models the uncertain effects of resources on success. Agent MDPs also model the ability of the agent to obtain resources by exposing their expected regret related to that resource to a centralised decision maker using an iterative auction procedure. The actions of the consumer agents are then to bid for resources in this sealed-bid first-price iterative auction, or to wait until resources become free. We demonstrate empirically how this decentralised mechanism can come close to the optimal solution in the medical domain, considering situations with up to 20 agents and 10 resources.

Individuals have some limitations and preferences that should be taken into account without necessarily sharing private information with each other (due to privacy and confidentiality issues). Although one can think of having a "centralized coordinator agent", being able to consider all these issues and making the best plan for each agent, the naturally dynamic and distributed nature of agents makes this approach computationally infeasible, because this problem is shown to be NP hard. In such settings, multi-agent systems can provide a fair solution with low communication between agents, while maintaining the natural separation of authority. For example, resource allocation requires search within all feasible (and even infeasible) possible allocations, and is often too complex and time consuming to perform in a centralized manner when the environmental characteristics are both distributed and dynamic [8]. Therefore, we adopt the multi-agent resource allocation technique to address such scheduling and allocation problems.

We first overview related work, followed by a general exposition of the MDP and dec-MDP setting as applied to resource allocation. We then detail our multi-agent system, and show how it can be applied specifically to the medical domain. Our results follow on the medical domain.

2 Related Work

Vermeulen et al. in [11] proposed an adaptive approach to automatic optimization of appointment calendars that categorizes patients into different groups based on urgency and makes flexible reservations. However, it is still the human schedulers who are responsible for coordination between different groups and medical resources. In another work, to utilize the distributed and dynamic nature of multi-agent systems, Vermeulen et al. [10] introduced an algorithm based on negotiation and appointment exchange between patients. Considering Pareto-improver agents, meaning that no agent will agree to an exchange that will worsen its schedule, a multi-agent pareto-improvement exchange algorithm was proposed to effectively schedule patients' appointments on expensive resources. This approach does not consider emergency departments and highly critical patients as important role players in scheduling calendars.

Becker and Czap [1] considered the problem of scheduling centralized operating theaters in large hospitals. They represented every individual with a software agent being able to negotiate or make decision on their behalf. Traditional approaches did not take departmental interdependencies into account and fell short in considering personal or departmental dependencies (such as which teams work well together, departmental limitations, personal preferences). Becker et al. [1] introduced a two-stage scheduling system that sets a preliminary plan by just considering medical demands, and then uses a more advanced algorithm to enhance this plan. Their advanced algorithm is built upon a Nash-bargaining solution based on the concept of fairness where agents have to negotiate and come up with a Pareto-efficient solution. They added individual preferences by performing a conjoint analysis at the end of each round of negotiation where agents should answer a quick question about their preferences over different alternatives.

The decisions of patient admissions to different specialized hospital units or operating rooms are made autonomously by each department. To tackle this problem and also to provide a solution that outperforms previous attempts, Hutzschenreuter et al. [6] used a probabilistic graph to model the pathway of each patient. Unlike other approaches such as [9],[11],[3], this work considers the stochastic nature of treatment processes in the medical pathways and takes unplanned arrival of patients (such as emergency department) into account. It considers the patient's characteristics, uncertainty related to the duration of stay, medical rules, and preferences of involved units as important stochastic factors, and groups patients on the basis of their required treatment steps based on diagnosis related groups or expertise of medical specialists. However, this approach works based on the emergency level of each patient. It assumes that patients are admitted to a hospital unit only if resources are available. Therefore, if there are many requests for admissions it automatically rejects some less critical patients and admits those with higher levels of criticality.

In an attempt to minimize patient stay time in hospitals, Paulussen et al., in [9], proposed a distributed patient scheduling model. Their model is a patient-centered multi-agent based coordination system that uses market mechanism to properly allocate resources such as medical devices and physicians to patients. This market-based mechanism is suitable for environments where the situation is changing constantly and there is a need for an efficient solution with low communication needs. However, the required medical tasks are assumed to be interchangeable, meaning patients can use any order of their medical treatments to complete their medical pathway.

Markov decision processes have been used extensively in patient scheduling problems, but are usually restricted to only a single type of resources (e.g. CT scanners) and to a small number of resources and agents. For example Gocgun *et al.* model a dynamic patient arrival situation with only one or two resources using a single MDP [5]. Enabling such a method to scale to larger numbers of resources would require a dec-MDP approach as we propose here.

3 Multi-Agent Resource Scheduling Using Decentralized MDPs

We are motivated by the challenge of effectively allocating resources to patients in hospital settings. In this domain, there is usually uncertainty about diagnosis and treatment steps. Patients may need to prolong their stay in one unit, or need to take a different medical task while their treatment procedures are being carried on. Furthermore, patients should go through a series of tests and treatments in order to complete their medical pathways. Thus, a better scheduling system can decrease waiting time for patients and increase overall health.

The problems were are interested in involve a set of N *consumers*, each of whom requires some subset of a total of R *resources*. The consumers each have a measure of *quality* or *success* that the are trying to optimise, and this quality is influenced stochastically by the resources they require and by time. That is, consumers try to reach a higher *success* by obtaining a set of resources. However, resources are limited and the multi-agent system's goal is to achieve the highest global level of success across all agents by allocating resources to consumers in a the optimal temporal order. Further, each consumer has a resource *pathway* that represents the partial ordering in which they need the resources.

We consider the multi-agent system to be made up of two types of agents: a consumer agent acts on behalf of each consumer, and is responsible for calculating the utility of getting a resource at each time slot. Resource agents act on behalf of each resource, and are responsible for allocating their available time slots to the consumers by holding auctions. We model the uncertainties in the system using Markov decision processes (MDPs). One can model the whole resource allocation process using a single MDP consisting of a permutation of all the possible allocations to the patients. However, increasing the number of patients or resources may cause a dramatic increase in the size of state space, making it infeasible to solve. In the context of healthcare, finding a suitable allocation quickly is essential for improving the performance. Furthermore, patient scheduling needs to be done real-time due to the unknown patient arrival rates in different departments, especially emergency departments. Therefore, we model the patient scheduling problem using decentralized MDPs. Agents dynamically collaborate in order to find an optimal allocation while trying to maximize their local utility function based on the preferences and limitations.

3.1 MDP formulation of Consumers

Our model is a factored MDP represented as a set of variables and functions $\langle \mathbf{R}, H, P_T, \Phi A\rangle$ where $\mathbf{R}$ is a finite set of resource variables, each of which is $R \in \{n, v, d\}$ representing that the agent n =needs, v =has, or has d =had the resource in question. H is a variable measuring agent success (e.g. patient health in the medical domain). We use $S = \{\mathbf{R}, H\}$ to denote the complete set of state variables. $\Phi(H)$ is a reward function based on the success, P_T is a transition model that gives the probability of reaching state S' after having health state H and resources $\mathbf{R}$, and A is a set of actions. There is one action for each resource, plus a null action. The resource actions represent bids for the corresponding resource, as detailed in Section 3.2.

In the medical domain, we will consider undiscounted problems $\gamma = 1$, as the health state of patients does not warrant discounting. For example, if a discount factor $\gamma < 1$ is used, then it may be advantageous to postpone treatment of a patient only to benefit from a discounted cost in the future of a failure to treat, clearly an undesirable behaviour for the scheduling system. However, our formulation also admits such discounted problems.

A policy for the MDP is a function $\pi(\mathbf{R}, H) \mapsto A$ that gives an action for an agent to take in each state $S = \{\mathbf{R}, H\}$. The policy can be obtained in a number of ways, including by computing a value function $V^*(s)$ for each state $s \in S$, that is maximal for each state (i.e. that satisfies the Bellman equation [2]):

$$V^*(s) = \Phi(s) + \max_a \gamma \sum_{s' \in S} P(s'|s,a)V^*(s') \tag{1}$$

The policy is then given by the actions at each state that are the arguments of the maximization in Equation 1. Agents can also compute their expected *regret* for not obtaining a given resource as follows. First, the agent computes $V^*(s) = V^*(\mathbf{r}, h)$ (Equation 1), which in fact is two functions for a given resource, r_i: $V_i \equiv V^*(r_i, \mathbf{r}_\mathbf{i}^-)$ and $\bar{V}_i \equiv V^*(\bar{r}_i, \mathbf{r}_\mathbf{i}^-)$, where the first is the expected value if the agent has the resource r_i, and the second is the expected value if not (denoted $\bar{r}_i$), and $\mathbf{r}_\mathbf{i}^-$ is the set of all resources except r_i. Thus, the expected value for being in a health state h at time t, bidding for (denoted a_i) and receiving (or not receiving) resource r_i at time $t+1$ is $Q_i \equiv \sum_{h'} P(h'|h, \mathbf{r}, a_i)V_i$ ($\bar{Q}_i \equiv \sum_{h'} P(h'|h, \mathbf{r}, a_i)\bar{V}_i$). Thus, the expected regret for not receiving resource r_i when in health state h is $R_i(h) = Q_i - \bar{Q}_i$.

3.2 A Model for Resource-Based Agent Success

Figure 1 shows our model as a two time slice influence diagram. We make three primary assumptions. First, the agent success (H) is conditionally independent of the agent action (e.g. bid) given the current resources and the previous success. Second, the agent actions only influence the resource allocation, since the agent can only influence success indirectly by bidding for resources. Third, the reward is only dependent on the agent success, H. Thus, the transition function $P(\mathbf{r}', h'|\mathbf{r}, h, a)$ factors as $P(\mathbf{r}', h'|\mathbf{r}, h, a) = P(\mathbf{r}'|\mathbf{r}, h, a)P(h'|\mathbf{r}, h)$ where $P(\mathbf{r}'|\mathbf{r}, h, a)$ denotes the probability of getting the next set of resources given the current success, resources, and action. And $P(h'|\mathbf{r}, h)$ represents a dynamics model for the agent's success rate. For example, in the medical domain, this will depending on the condition (e.g. disease) of the person and on the criticality of the health state. While some conditions may cause a patient to deteriorate rapidly unless resources are obtained, others may have less of an effect.

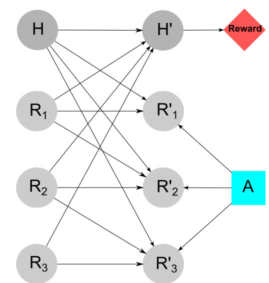


Figure 1: A consumer's MDP shown as a two time slice influence diagram

The success factor, $P(h'|\mathbf{r}, h)$, is a property of an agent's condition and can be esimated from global statistics about the nature of the conditions (e.g. diseases), independently of othe agents. On the other hand, the resource allocation factor, $P(\mathbf{r}'|\mathbf{r}, h, a)$, will be dependent on the current state of the multi-agent system. For example, the probability of an agent obtaining a resource will depend on the number of other agents currently bidding for that resource, and their success conditions. In general, we can make no assumptions about further conditional independencies in the resource allocation factor. That is, the probability of obtaining a resource R' at time t may depend stochastically on the set of resources at time $t-1$. However, in many domains, there may be further independencies that can be encoded in the model. For example, in Figure 1, resource R'_i is conditionally independent of all resources R_j where $j \notin \{i, i-1\}$ (for $i > 1$) and for $j \notin \{i\}$ for $i = 1$), so the resources are *ordered* according to the (linear) resource pathway of this particular consumer.

3.3 Coordination Mechanism

The coordination mechanism is to have each of N agents (e.g. N patients) able to estimate its own expected regret for not obtaining a given resource, given its best estimate of the future probabilities of getting resources (as given by the resource allocation factor $P(\mathbf{r}'|\mathbf{r}, h, a)$). The regret values of different agents can then be compared globally, and an allocation can be sought that minimizes the global regret. In the proposed auction mechanism, the consumer agents compete with each other over the time slots to achieve the objectives of their corresponding consumers [4].

To apply such coordination mechanism to our dec-MDP, we propose a simple auction-based system where agents can send their current estimates of regret to a central auctioneer, who allocates resources in an iterative auction. Our coordination mechanism is a multi-round auction mechanims based on first-price sealed-bid auction where each round is an auction of this type to identify the winner for each timestep of every resource. In each round, agents submit their bids, and auctioneer determines the winner and moves to the next step.

The auction proceeds in an iterative fashion, and each consumer agent bids on the resource with highest regret first. If an agent does not win the auction for its highest regret resource, it waits until the auction for its second highest regret resource becomes available. It can also decide to give up, and is then resigned to not having any resources in the next time step (which may be a better option than taking a resource ahead of time that will cause damage). While this iterative auction mechanism is not optimal, it works well in practice as shown in Section 5.

4 Medical Domain Example

In this model, the order of medical tasks is a key factor in allocating resources to patients. Patients have different medical needs based on their sickness and health conditions, and recovering from these health conditions requires series of medical tasks. However, these medical tasks are dependent to the results of the previous tasks (or tests). Therefore, the tasks should be carried on through series of tests and treatment steps in sequence.

We make three main assumptions in the following. First, we assume that task durations are identical (e.g. it always takes one unit of time to consume each resource). In order to model resource consumption times, we could extend our approach by modeling a longer resource as a linear chair of shorter ones. The second assumption is that each agent is only able to bid on a single resource at a time. Along with our iterative approach, this avoids conflicts in which an agent is indifferent between two or more resources. The third assumption is that all patients arrive at the same time ($t = 0$). This assumption could easily be relaxed, but would mean additional confusing factors in our simulations.

While in e-commerce scenarios, the human principals reveal their preferences through their willingness to pay, patients should not reveal their preferences through their willingness to pay for a specific time slot of a resource. The preferences rather have to be based upon severity of a patient's health state and the patient's stay time in waiting queues. The health state of patients decreases every timestep, if patients cannot take over their required resource.

Modeling the pricing mechanism in the absence of monetary units is done using a health-related utility function. As stated earlier, by finishing a medical pathway a patient (p) gains a reward equal to a utility function: $Reward^p(t) = H^p_{max} - H^p(t)$, where $Reward^p(t)$ is the reward of completing the medical pathway at the time t, H^p_{max} is the maximum achievable health state, $H^p_{max} = \max_t H^p(t)$, and $H^p(t)$ is the function representing the health state of patient p at time t. In this paper, we assume that maximum achievable health state after completing the medical pathway for all the patients is equal to 1, i.e., by finishing the medical pathway patients get 100% recovery.

Patients calculate their willingness to pay by solving their Markov representation of the needed tasks. Then, the regret $V(R', H') - V(\bar{R}', H')$ is used as the patients bidding price, giving the difference in value between getting (R') and not getting ($\bar{R}'$) the resource given current state R and the patient's current health state H'.

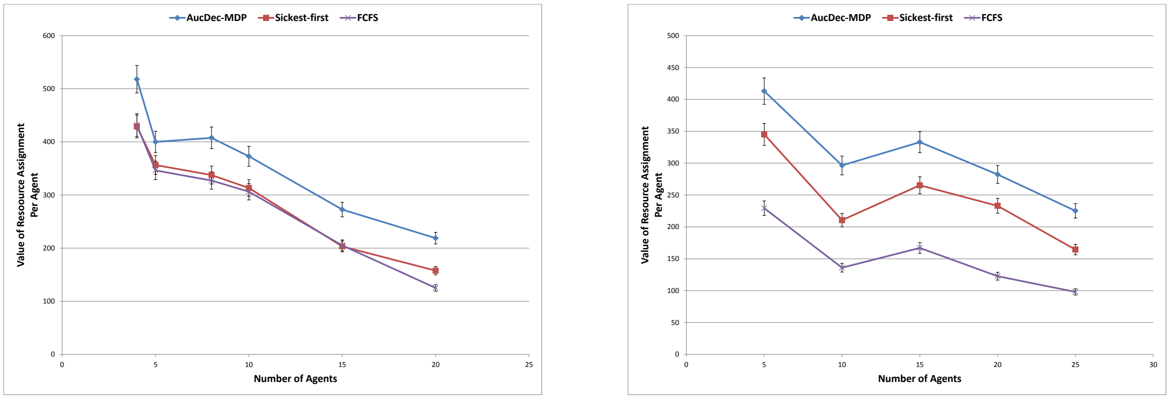
5 Experiments and Results

For simulation and testing purposes, we consider 3 to 4 different type of conditions (e.g. diseases). Each type of condition has a set of resources that are needed to be taken by the patient to complete the medical pathway. In the beginning of every simulation, we randomly set up the condition profiles for each type, and then, a random condition profile is assigned to each patient. A condition profile consists of an initial health state and a set of resources in a specific order. The experiment starts by initializing MDPs using a set of Dirichlet/Beta priors for each randomly created condition profile. Then, these MDPs are solved and assigned to the patient agents. We tested the bidding process using the optimal values and also the regret-based approach. These appear to return similar results due to the fact that our reward function is based on the initial health state of patients. The plots shown in this paper are using the optimal values of states.

For comparison purposes, we have simulated three different scenarios: our auction-based dec-MDP approach, the first-come-first-serve (FCFS) approach, and sickest-first approach where agents obtain required resources based on their initial health state. Patients all start at a fixed initial health condition and depending on the resources they acquire, their health state improves to healthy or degrades to the sick condition. Three trials are done for each randomly drawn set of disease profiles (and related MDPs), and this is repeated 10 times, and we present means and standard deviations over these 10 simulations. We present results with 4 (Figure 2a) and with 10 (Figure 2b) total resources available, in each case each agent requiring 4 randomly selected resources. We see that our approach consistently outperforms the FCFS and sickest-first approaches We find these initial results encouraging, and seek to explain in more detail some of the observed behaviours.

6 Conclusions

This paper presented a multi-agent decentralized mechanism for resource allocation in situations where resource consumers can measure their utility over the long term, and resources have stochastic effects on this utility. We model each consumer as a Markov decision process, and propose an iterative auction mechanism in which consumers bid for resources using their expected long-term regret. We apply this model to a general medical domain in which consumers are patients who require hospital resources (e.g. treatments, diagnostic tests), and each patient requires a partially ordered set of resources that have uncertain effects on health state. We show in simulation that the model can scale to a reasonable number of patients and resources. Our future work is to continue experimenting with the scalability of this method, and to remove various assumptions (e.g. task duration, agent arrival times). Further, we wish to address the fact that medical pathways of patients are often not stationary (i.e. they change over time).



(a) 4 available resources

(b) 10 available resources

Figure 2: Simulations results with three different allocation methods

References

[1] M. Becker and H. Czap. Artificial Software Agents as Representatives of Their Human Principals in Operating-Room-Team-Forming. *Multiagent Engineering*, pages 221–237, 2006.

[2] R.E. Bellman. *Dynamic programming*. Courier Dover Publications, 2003.

[3] K. Decker and J. Li. Coordinated hospital patient scheduling. In *Multi Agent Systems, 1998. Proceedings. International Conference on*, pages 104–111. IEEE, 2002.

[4] E. Durfee. Distributed problem solving and planning. *Multi-agent systems and applications*, pages 118–149, 2006.

[5] Yasin Gocgun, Brian W. Bresnahan, Archis Ghate, and Martin L. Gunn. A markov decision process approach to multi-category patient scheduling in a diagnostic facility. *Artificial Intelligence in Medicine*, 53:73–81, 2011.

[6] A.K. Hutzschenreuter, P.A.N. Bosman, I. Blonk-Altena, J. van Aarle, and H. La Poutré. Agent-based patient admission scheduling in hospitals. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track*, pages 45–52. International Foundation for Autonomous Agents and Multiagent Systems, 2008.

[7] S. Kirn, C. Anhalt, H. Krcmar, and A. Schweiger. Agent. HospitalHealth Care Applications of Intelligent Agents. *Multiagent Engineering*, pages 199–220, 2006.

[8] R. Mailler and V. Lesser. A Cooperative mediation-based protocol for dynamic distributed resource allocation. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 36(1):80–91, 2006.

[9] T.O. Paulussen, N.R. Jennings, K.S. Decker, and A. Heinzl. Distributed patient scheduling in hospitals. In *International Joint Conference on Artificial Intelligence*, volume 18, pages 1224–1232. Citeseer, 2003.

[10] I. Vermeulen, S. Bohte, K. Somefun, and H.L. Poutré. Multi-agent Pareto appointment exchanging in hospital patient scheduling. *Service Oriented Computing and Applications*, 1(3):185–196, 2007.

[11] I.B. Vermeulen, S.M. Bohte, S.G. Elkhuizen, H. Lameris, P.J.M. Bakker, and H.L. Poutré. Adaptive resource allocation for efficient patient scheduling. *Artificial intelligence in medicine*, 46(1):67–80, 2009.