Computers & Operations Research 67 (2016) 90–101

Contents lists available at ScienceDirect

Computers & Operations Research

journal homepage: www.elsevier.com/locate/caor

Multi-class, multi-resource advance scheduling with no-shows, cancellations and overbooking

Mahshid Salemi Parizi, Archis Ghate*

Department of Industrial and Systems Engineering, University of Washington, Seattle, WA, USA

ARTICLE INFO

Available online 8 October 2015

Keywords:
Approximate dynamic programming
Markov decision processes

ABSTRACT

We investigate a class of scheduling problems where dynamically and stochastically arriving appointment requests are either rejected or booked for future slots. A customer may cancel an appointment. A customer who does not cancel may fail to show up. The planner may overbook appointments to mitigate the detrimental effects of cancellations and no-shows. A customer needs multiple renewable resources. The system receives a reward for providing service; and incurs costs for rejecting requests, appointment delays, and overtime. Customers are heterogeneous in all problem parameters. We provide a Markov decision process (MDP) formulation of these problems. Exact solution of this MDP is intractable. We show that this MDP has a weakly coupled structure that enables us to apply an approximate dynamic programming method rooted in Lagrangian relaxation, affine value function approximation, and constraint generation. We compare this method with a myopic scheduling heuristic on eighteen hundred problem instances. Our experiments show that there is a statistically significant difference in the performance of the two methods in 77% of these instances. Of these statistically significant instances, the Lagrangian method outperforms the myopic method in 97% of the instances.

1. Introduction

We present a formal abstraction, a Markov decision process (MDP) model, and an approximate dynamic programming (ADP) solution algorithm for a class of problems motivated by recent literature reviews about patient scheduling [6,8,17,23]. To understand this class of problems concretely, imagine a scheduler at a hospital, who dynamically and stochastically receives appointment requests for different types of elective surgeries. Distinct types of surgeries require different kinds and amounts of resources. For example, one type of surgery may require a two-hour block of time in the operating room, one surgeon, two surgeon's assistants, a nurse and a certain medical equipment. Another type of surgery may require a two-hour block of time, two surgeons, no assistants, and one nurse. Indeed, Gupta and Denton [23] stated that surgeries require "clinical assistants, nurses, anesthesiologists, surgeons, operating rooms, diagnostic devices, surgical tools and other equipment". Since these surgeries are elective, the requests may be rejected, or scheduled to a future day in a booking horizon. The scheduler may also receive cancellations for surgeries that were booked in advance. Similarly, some scheduled surgeries may turn out to be no-shows "at the last minute." To mitigate the adverse effects of such cancellations and no-shows on system-performance, the scheduler may overbook surgeries with the hope that additional resources may be summoned if and when deemed necessary. The hospital may receive different rewards for performing different types of surgeries and incur costs/penalties for rejecting arriving request. It may also incur an indirect waiting cost depending on the delay between when the request arrives and the future day for which it is scheduled. Finally, if additional resources are requested to tackle overbookings, the hospital may incur a cost of overtime. The scheduler is interested in booking these dynamic, stochastic surgery requests so as to optimize an appropriate metric of performance over an infinite horizon because there is no known time of extinction for this system. We call this type of problems multi-class, multi-resource, advance scheduling with cancellations, no-shows and overbooking.

These advance scheduling problems also arise in healthcare operations other than elective surgery scheduling. For example, a clinic that employs one or more physicians must also book stochastically and dynamically arriving patient requests and plan for no-shows and cancellations (estimates of no-show rates vary widely in the literature depending on several factors: for instance, 42% [37]; 10% [7,12,61]; 15%–30% [14]; 20% [30]; 3%–80% [54]). Similarly, a radiological facility that houses different kinds of imaging machines also needs to book requests for distinct classes of diagnostic exams [47,58].

* Corresponding author at: Department of Industrial and Systems Engineering, University of Washington, Box 352650, Seattle, WA 98195, USA.
E-mail address: archis@uw.edu (A. Ghate).

http://dx.doi.org/10.1016/j.cor.2015.09.004
0305-0548/

Beyond healthcare, such advance scheduling problems arise in revenue management [16]; at maintenance and repair facilities [48]; in manufacturing [39,57]; in military reconnaissance and combat planning [1]; in communication networks [3,52,60]; and in high-performance computing [59].

Despite the prevalence of these problems in a broad range of applications, research on their mathematical models and dynamic stochastic optimization is sparse. In the healthcare context, Cayirli and Veral [8] noted that the use of patient classification in scheduling problems which consider dynamically and stochastically arriving requests was very limited in the literature. They discussed potential advantages of classifying patients by relaxing the usual assumption of homogeneous patients. They also stated that no-shows is a major exogenous factor that affects the performance of any appointment scheduling system. They commented "no rigorous research exists which investigates possible approaches to adjusting the appointment systems in order to minimize the disruptive effects of no-shows, walk-ins, and/or emergencies." Daggy et al. [13] discussed the importance of considering no-shows and overbooking in patient scheduling models. Gupta and Denton [23] stated that cancellations and no-shows are open challenges in appointment scheduling and that overbooking could be used to tackle this difficulty.

The main reason why most researchers seem to have balked from building mathematical models and developing optimization algorithms for these problems is the well-founded belief that any such model would be inevitably large-scale and hence their exact solution computationally difficult. This concern is appropriately rooted in the three well-known curses of dimensionality in dynamic programming (DP) – the curse of state-space, the curse of action-space, and the curse of expectation [49]. Indeed, Truong [58] stated that "advance scheduling problems have generally been deemed intractable".

In this paper, we attempt to fill this gap in the literature by making the following contributions:

1. we formally introduce a group of multi-class, multi-resource, advance scheduling problems with no-shows, cancellations, and overbooking; to the best of our knowledge, this class of problems has not yet been studied in the existing literature;
2. we provide an MDP model for this class of problems;
3. we show that this MDP is weakly coupled (see Adelman and Mersereau [2]) – that is, the state- and action-spaces are Cartesian products of finite sets over different job classes, the immediate expected rewards are additively separable over job classes, transition probabilities are multiplicatively separable over job classes, and the only feature that links different job classes is the resource constraints;
4. we argue that this weakly coupled MDP is large-scale; that is, a naive implementation of the standard Lagrangian relaxation method of Adelman and Mersereau [2] would be computationally challenging;
5. we then show how Lagrangian relaxation with constraint generation and affine value function approximation, which was originally developed by Gocgun and Ghate [20] for solving general large-scale weakly coupled MDPs, can be applied to this class of problems;
6. we compare the resulting Lagrangian policy with a myopic heuristic via extensive numerical experiments that include eighteen hundred problem instances;
7. our numerical experiments suggest that the Lagrangian solution outperforms the myopic solution with a statistically significant margin on a large majority of the problem instances.

This paper is organized as follows. We review related literature in the next section. Our class of advance scheduling problems is formally introduced in Section 3. A weakly coupled MDP formulation for this class of problems is described in Section 4. An ADP algorithm based on Lagrangian relaxation of this MDP is presented in Section 5. Computational experiments are included in Section 6. We close with some concluding remarks in Section 7.

2. Literature review

The problems considered in this study are categorized as advance scheduling problems because stochastically and dynamically arriving jobs are scheduled to future slots within a booking horizon. The first part of our literature survey (Section 2.1) therefore focuses on papers that model advance scheduling in this manner; the second part (Section 2.2) reviews papers that do not model advance scheduling the way we envision here, but do include no-shows and/or overbooking. These two parts focus on papers that consider healthcare as their main application domain. The advance scheduling problems we study do share some similarities, albeit tenuous, with other problems studied in application areas such as revenue management, machine scheduling, admission control in queues, and inventory allocation. We do not review those papers here but refer the reader to various surveys about these application areas: Talluri and van Ryzin [56] for revenue management; Leung [38] for machine scheduling; Green and Savin [22] for admission control in queues; and Ha [24,25] for inventory allocation. Our advance scheduling framework is in contrast with allocation scheduling, where an arriving job is either rejected or put in a queue and jobs in the queue are dynamically selected for service later [43]. We do not review the vast literature on allocation scheduling but instead refer the reader to Truong [58] for an up-to-date survey.

2.1. Literature on advance scheduling

Patrick et al. [47] presented an MDP model and a linear programming based ADP method to book stochastically and dynamically arriving patient requests to future slots at a diagnostic resource in a public healthcare setting. That paper considered a single resource. Gocgun and Ghate [20], in a paper that was based on Gocgun's doctoral dissertation [19], extended the model in Patrick et al. [47] to include multiple resources. Gocgun and Ghate [20] showed that their problem can be formulated as a weakly coupled MDP and proposed an extension of the Lagrangian relaxation method in Adelman and Mersereau [2] and in Hawkins [28] for approximately solving this MDP. Neither the model in Patrick et al. [47] nor that in Gocgun and Ghate [20] included cancellations, no-shows, or overbooking. Saure et al. [55] applied the method in Patrick et al. to scheduling problems in radiation therapy. Gocgun and Puterman [21] applied an extension of Saure et al. [55] to chemotherapy appointment scheduling. This extension considered target dates for appointments and a tolerance around that target in which the scheduler has the flexibility to book an appointment. Again, these papers did not incorporate no-shows, cancellations, overbooking or multiple resources.

An application of advance scheduling to multi-class, multi-resource surgeries is developed in the Master's thesis of Astaraky [4]. That model aims to minimize a combination of waiting time till the surgery date, overtime in the operation room, and congestion in hospital wards. Astaraky's model does not consider no-shows and uses patient-discharge from the booking schedule that can be viewed as a type of cancellation. The state in his model includes the number of patients booked in the booking horizon, the number of patients in post-operative recovery, and the demand waiting to be scheduled. This state is different from what we need in this paper. For instance, Astaraky does not need to