Computers & Operations Research 77 (2017) 240–253

Contents lists available at ScienceDirect

Computers & Operations Research

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

Least squares approximate policy iteration for learning bid prices in choice-based revenue management

Sebastian Koch

Chair of Analytics & Optimization, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany

ARTICLE INFO

Article history:
Received 1 October 2015
Received in revised form
3 May 2016
Accepted 20 July 2016
Available online 21 July 2016

Keywords:
Revenue management
Capacity control
Approximate dynamic programming
Approximate policy iteration

ABSTRACT

We consider the revenue management problem of capacity control under customer choice behavior. An exact solution of the underlying stochastic dynamic program is difficult because of the multi-dimensional state space and, thus, approximate dynamic programming (ADP) techniques are widely used. The key idea of ADP is to encode the multi-dimensional state space by a small number of basis functions, often leading to a parametric approximation of the dynamic program's value function. In general, two classes of ADP techniques for learning value function approximations exist: mathematical programming and simulation. So far, the literature on capacity control largely focuses on the first class.

In this paper, we develop a least squares approximate policy iteration (API) approach which belongs to the second class. Thereby, we suggest value function approximations that are linear in the parameters, and we estimate the parameters via linear least squares regression. Exploiting both exact and heuristic knowledge from the value function, we enforce structural constraints on the parameters to facilitate learning a good policy. We perform an extensive simulation study to investigate the performance of our approach. The results show that it is able to obtain competitive revenues compared to and often outperforms state-of-the-art capacity control methods in reasonable computational time. Depending on the scarcity of capacity and the point in time, revenue improvements of around 1% or more can be observed. Furthermore, the proposed approach contributes to simulation-based ADP, bringing forth research on numerically estimating piecewise linear value function approximations and their application in revenue management environments.

1. Introduction

During the last decades, revenue management and particularly capacity control, which is at the heart of modern revenue management systems, have evolved into one of the most successful application areas of operations research. For a long time, academic research on revenue management assumed that demand is strictly associated with a product and thus not dependent on the market conditions and other products offered. However, the integration of customer choice behavior is today considered the most important trend in recent years.

We therefore reconsider the well-known revenue management problem of capacity control under customer choice behavior. The problem can be briefly stated as follows: a firm offers differently priced products that are provided using a number of shared resources with a fixed capacity. The products are possibly linked to sales restrictions and other conditions to segment the market. Customers arrive successively and stochastically over a given booking horizon with each customer purchasing at most one unit of a product. The chosen product may depend on the available products. Service provision occurs at the end of the booking horizon. Any capacity remaining at the end is worthless and overbooking of the given resources' capacity is not allowed. Capacity control now continuously addresses the following decision problem throughout the booking horizon: Which products should the firm offer for sale such that the overall expected revenue is maximized?

The described problem can be solved by dynamic programming, using a recursively formulated value function (Bellman equation). However, an exact solution is difficult due to the multi-dimensional state space which comprises the resources' remaining capacity. Therefore, approximate dynamic programming (ADP) techniques are often used. In doing so, the multi-dimensional state space is usually encoded by a small number of basis functions that sufficiently describe the current booking status, leading to a (parametric) approximation of the value function. Virtually all approximations make use of additive bid prices that represent the marginal value of capacity such that the state space decomposes by resources.

E-mail address: sebastian.koch@wiwi.uni-augsburg.de

http://dx.doi.org/10.1016/j.cor.2016.07.015
0305-0548/

Broadly speaking, two classes for learning value function approximations exist: mathematical programming and simulation. So far, the literature on capacity control largely focuses on mathematical programming. In contrast, simulation-based ADP—which we consider in this paper—is quite underrepresented. In the language of capacity control, the central idea of simulation-based ADP is to simulate thousands of booking horizons by Monte Carlo simulation and thereby iteratively learn the value function approximation. In what follows, a sample path refers to the stochastic exogenous information of a simulated booking horizon. While stepping forward in time following a specific sample path, decisions are made given the current approximation, and the information gathered is used to update the approximation.

In this context, our main contribution is to develop a least squares approximate policy iteration (API) approach for learning bid prices. Thereby, we investigate value function approximations that are linear as well as piecewise linear (concave) in the resources' remaining capacity. Because the approximations are also linear in their parameters, we can estimate them via linear least squares regression. Thereby, we mimic either known or heuristic properties of the optimal expected revenue and of bid prices by enforcing structural constraints on the parameters. This advancement of common simulation-based ADP approaches turns out to have a critical impact on performance.

In a simulation study, we first show that the approach leads to good value function approximations in examples with only one resource. Then, we show that the approach is competitive with the dynamic programming decomposition (DPD) approach in network settings, which is state-of-the-art to approximately solving dynamic programs in a capacity control environment, and even outperforms it in many cases. Revenue improvements of around 1% or more are observable particularly when capacity is scarce or when booking horizons reflecting a later point in time are considered. Revenue improvements in this range are widely acknowledged to have a big impact on the overall profit.

From a practical point of view, the proposed approach is particularly interesting because it is rather independent of the presumed customer choice behavior. Furthermore, the paper contributes to the research on simulation-based ADP giving a rough guideline how to design corresponding approaches in the field of revenue management. Moreover, we provide a simple and intuitive approach for numerical estimation of piecewise linear value function approximations that can be easily adapted to other applications.

The remainder of the paper is structured as follows: In Section 2, we formalize the problem statement, review the relevant scientific literature and position our work. Based upon this, we develop our approach in Section 3. In Section 4, we present the results obtained in our simulation study. After a discussion of managerial and theoretical insights in Section 5, we conclude the paper and give an outlook on future research in Section 6.

2. Background and previous research

We first formalize the problem statement from Section 1, describing the setting and notation (Section 2.1). We then discuss current solution approaches (Section 2.2) and review the most relevant scientific literature on simulation-based ADP (Section 2.3).

2.1. Problem formulation

There is an extensive literature on revenue management models that allow for the automation of capacity control. Overviews are found in the textbooks by Talluri and van Ryzin [38] and Phillips [30].

Gallego et al. [9], Talluri and van Ryzin [37], and Liu and van Ryzin [20] established capacity control under a general discrete choice model of demand, allowing customers to substitute among several more or less suitable products. We follow their setting and consider a firm that disposes of resources $h = 1, \ldots, m$ which are jointly used by products $j = 1, \ldots, n$. The products are associated with revenues $\boldsymbol{r} = (r_1,\ldots,r_n)^T$. The capacity consumption of one unit of a product j is given by $\boldsymbol{a}_j = \left(a_{1j},\ldots,a_{mj}\right)^T$ with $a_{hj} = 1$ if product j uses resource h or $a_{hj} = 0$ else. The resources' initial capacity is described by $\boldsymbol{c}^0 = \left(c_1^0,\ldots,c_m^0\right)^T$. The vector $\boldsymbol{c} = (c_1,\ldots,c_m)^T$ denotes resources' remaining capacity, and selling a product j reduces capacity to $\boldsymbol{c} - \boldsymbol{a}_j$.

The booking horizon is discretized into sufficiently small time periods $t = 1, \ldots, T$, such that in each period t at most one customer arrives who purchases at most one unit of one product. The periods are numbered forward in time. In each period t, the firm selects the subset of products available for sale, called the offer set. A possible offer set is captured by the vector $\boldsymbol{x}_t = (x_{t1},\ldots,x_{tn})^T$ of binary decision variables, with $x_{tj} = 1$ if product j is offered in period t and $x_{tj} = 0$ else. Depending on $\boldsymbol{x}_t$, product j is sold with probability $p_j(\boldsymbol{x}_t) \geq 0$, and no purchase is made with probability $p_0(\boldsymbol{x}_t) \geq 0$. The probabilities satisfy $p_j(\boldsymbol{x}_t) = 0$ if $x_{tj} = 0$ as well as $\sum_{j=1}^n p_j(\boldsymbol{x}_t) + p_0(\boldsymbol{x}_t) = 1$. Although one can easily include time-varying purchase probabilities into the model presented, we follow the literature and assume time-homogenous purchase probabilities to ease the notation.

Let $V_t(\boldsymbol{c})$ denote the optimal expected revenue-to-go in period t with capacity $\boldsymbol{c}$ and let $\Delta_j V_{t+1}(\boldsymbol{c}) = V_{t+1}(\boldsymbol{c}) - V_{t+1}\left(\boldsymbol{c} - \boldsymbol{a}_j\right)$ denote the opportunity cost of selling one unit of product j in period t. Furthermore, let $\boldsymbol{\psi} = (\psi_1,\ldots,\psi_T)^T$ represent a policy of the firm. The policy is a function of the state $\boldsymbol{c}$ that gives the decision (i.e., the offer set) to make in that state: $\boldsymbol{x}_t = \psi_t(\boldsymbol{c})$. The firm's decision problem now is to determine an optimal policy $\boldsymbol{\psi}^*$ or, equivalently, to choose an optimal offer set $\boldsymbol{x}_t^*$ in each period t, such that the overall expected revenue $V_1\left(\boldsymbol{c}^0\right)$ is maximized. The value function $V_t(\boldsymbol{c})$ is computed recursively by (DP)

$$V_t(\boldsymbol{c}) = \max_{\boldsymbol{x}_t \in \{0,1\}^n} \left\{ \sum_{j=1}^n p_j(\boldsymbol{x}_t) \cdot \left(r_j - \Delta_j V_{t+1}(\boldsymbol{c})\right) \right\} + V_{t+1}(\boldsymbol{c}) \;\forall\, t, \boldsymbol{c} \geq \boldsymbol{0} \tag{1}$$

with the boundary conditions $V_t(\boldsymbol{c}) = -\infty$ if $\boldsymbol{c} \ngeq \boldsymbol{0}$ and $V_{T+1}(\boldsymbol{c}) = 0$ if $\boldsymbol{c} \geq \boldsymbol{0}$. Please note that the optimal policy and, thus, the revenue-maximizing offer sets, are given by the solution of the decision problem in each period t (i.e., the maximization in (1)):

$$\boldsymbol{x}_t^* = \psi_t^*(\boldsymbol{c}) \in \arg\max_{\boldsymbol{x}_t \in \{0,1\}^n} \left\{ \sum_{j=1}^n p_j(\boldsymbol{x}_t) \cdot \left(r_j - \Delta_j V_{t+1}(\boldsymbol{c})\right) \right\} \tag{2}$$

2.2. Current solution approaches

Roughly speaking, two issues make the exact solution of DP (1) difficult: First, recursively computing the value function $V_t(\boldsymbol{c})$ for all states $\boldsymbol{c}$ is—apart from some simple cases—not possible due to the multi-dimensional state space. Second, the decision problem (2) inherent in each period t is an assortment optimization problem over 2^n possible offer sets.

To address the first issue, virtually all ADP approaches approximate opportunity cost additively by means of bid prices $\pi_{th}(c_h)$ that represent the value of one unit of capacity of resource h at time t, given that the remaining capacity is c_h: