European Journal of Operational Research 254 (2016) 957–967

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Decision support

A two-state partially observable Markov decision process with three actions

Tal Ben-Zvi[a,*], Tatyana Chernonog[b], Tal Avinadav[b]

[a] School of Business, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030, USA
[b] Department of Management, Bar-Ilan University, Ramat-Gan 5290002, Israel

ARTICLE INFO

Article history:
Received 31 July 2015
Accepted 28 April 2016
Available online 6 May 2016

Keywords:
Decision processes
Markov chains
POMDP
Control limits

ABSTRACT

A process can be in either a stable or an unstable state interchangeably. The true state is unobservable and can only be inferred from observations. Three actions are available: continue with the process (CON), repair the process for a certain fee – bring the process to the stable state (REP), and obtain the state of the process for a cost (INS). The objective is to maximize the expected discounted value of the total future profits. We formulate the problem as a discrete-time Partially Observable Markov Decision Process (POMDP). We show that the expected profit function is convex and strictly increasing, and that the optimal policy has either one or two control limits. Also, we show that “dominance in expectation” (the expected revenue is larger in the stable state than in the unstable state) suffices for a control limit structure.

1. Introduction

Consider a process that can be in either a stable or an unstable state during a certain period (Avinadav & Perlman, 2013; Ben-Zvi & Grosfeld-Nir, 2012). Although the true state of the process may be unknown, it can be inferred from observing the process output. Many studies have addressed such scenarios (see, for example, Anily & Grosfeld-Nir, 2006; Ben-Zvi & Grosfeld-Nir, 2013; Ben-Zvi & Nickerson, 2012; Douer & Yechiali, 1994; Grosfeld-Nir, 2007), and assumed that the decision maker can take one of two actions (decisions): continue process (CON action) or adjust/repair the process and continue (REP action).

The model proposed herein extends this stream of literature by making the following innovative contributions:

(1) The model we propose allows the decision maker to take a third type of action, namely, a thorough inspection (INS action) of the process. The purpose of the inspection is to reveal the state of the process, and it is associated with a given cost. Most of the studies which seek to maximize the expected profit of the process do not consider an inspection action (see literature above). We note that such an action is considered in models that aim to maximize the expected time until the process comes to a halt. For example, a study by Thomas, Gaver, and Jacobs (1991) assumes that it is possible to carry out a fully-reliable inspection, which involves stopping the normal running of the process. We, on the other hand, assume that the inspection's duration is negligibly small, and investigate how different inspection costs impact the problem and the decisions made by the decision-maker. We also discuss sensitivity analysis and the Value of Inspection in Section 4. In addition to analyzing perfect inspection, and in contrast to models in the literature that consider solely this type of inspection, we also consider imperfect inspection: an inspection subject to two types of errors.

(2) The model we introduce in this paper forms a Markov chain with an interchangeable transition between states: if the process is in the stable state during one period, there is a constant probability that it will deteriorate to be in the unstable state during the next period. However, in contrast to the studies cited above, which assume a deteriorating process (i.e., once the process enters the unstable state it remains there until a REP action is taken), in this work, we assume that an unstable state may shift to be stable, without any interference.

The latter contribution enables us to consider many applications that the current literature cannot address. One interesting example is the case of diagnosis of medical conditions such as strep throat. Strep throat is an infection in the throat and tonsils caused by group A streptococcus bacteria. Common symptoms of strep throat include sore throat, a fever, red and swollen tonsils,

* Corresponding author. Tel.: +12012168545.
E-mail address: tbenzvi@stevens.edu (T. Ben-Zvi).

http://dx.doi.org/10.1016/j.ejor.2016.04.062
0377-2217/

and enlarged lymph nodes (Centers for Disease Control and Prevention, 2015). This disease accounts for 15–30% of cases of acute pharyngitis in children and 5–10% of cases in adults (Kaplan, Top, Dudding, & Wannamaker, 1971; Komaroff, Pass, & Aronson, 1986). It is not easy to diagnose strep throat, since the clinical manifestations of group A streptococcal pharyngitis and those of nonstreptococcal pharyngitis (which is typically viral) overlap quite broadly, and only laboratory tests can confirm the diagnosis (Bisno, Peter, & Kaplan, 2002).

This application example fits our model: considering a certain patient, he or she may either have group A streptococcal pharyngitis (unstable state) or nonstreptococcal pharyngitis while presenting symptoms resembling those of streptococcal pharyngitis. The true state is unknown, but can be inferred from daily observations (symptoms of the disease). We note that the states may shift, as streptococcal pharyngitis is considered benign and remits spontaneously, even without treatment. In each time period (each day), the decision maker (the patient or his/her physician) needs to decide which of the following actions to take: (1) CON (i.e., continue monitoring the symptoms); (2) REP (i.e., "repair" the disease by taking antibiotics); or (3) INS (perform a certain medical test to find out whether the patient has the disease or not). As noted above, the INS action can refer to a perfect inspection (for example, performing a throat culture) or an imperfect inspection (performing an antigen detection test, which is subject to errors).

Although prescribing/taking antibiotics seems like an easy choice as soon as the physician or patient identifies symptoms related to the disease, it might do more harm than good: If the patient's symptoms are the results of a viral sore throat rather than strep throat, antibiotics will not alleviate the condition, and might present a risk to the patient (Butler, Rollnick, Pill, Maggs-Rapport, & Stott, 1998). Yet a policy of always carrying out a full inspection (throat culture) to identify whether antibiotics are needed might also be suboptimal: Many healthcare professionals believe that such a policy can lead to overtreatment of low-risk patients and to increased healthcare costs (Ebell, Smith, Barry, Ives, & Carey, 2000). Herein, we put forward a model that has the potential to prevent the common practices of prescribing unnecessary antibiotics and overtreating patients (Gunnarsson, Sundvall, & Gunnarsson, 2012; Wessels, 2011). Specifically, we formulate the problem as a discrete-time partially observable Markov decision process (POMDP), which provides a powerful probabilistic tool for decision-making. We suggest that this approach can be advantageous and provide superior outcomes at lower cost.

Two-state scenarios are not common only in the medical field (where a patient may have a certain disease or not) but also in other domains. We now briefly present two additional applications in other fields that can be modeled using our POMDP approach:

(a) Media: a TV series may be in a stable, "good" state (a hit) or an unstable, "bad" state (a flop) interchangeably. Observations are the show's weekly ratings (for example, Nielsen ratings), which could serve as an indication for the real state. However, weekly ratings alone do not perfectly reveal the true state of a show, as there might be reasons for high ratings for a particular episode (for example, a guest appearance of a famous star on the show) or bad ratings (the show is broadcast at the same time as an important event, such as the Super Bowl or the televised wedding or funeral ceremony of an important figure). Since network directors aim to maximize their revenue from advertisement and royalties, they need to decide whether to keep on broadcasting the show (CON), or cancel it and replace it with a new series (REP). They may also conduct a thorough analysis of the show with a team of media and marketing experts to figure out whether the show would attract many viewers (INS). This application was studied by Givon and Grosfeld-Nir (2008), who modeled the ratings of TV shows as a stochastic process that depends on the shows' state. However, their model considered only "deteriorating ratings", i.e., once a TV show enters the "bad" state, it remains there until the REP action is taken. The model presented in this paper is more general and allows a "bad" show to improve and become a "good" show, and vice versa. The sitcom *Seinfeld* is an example of a show that was able to recover from a "bad" state and peak ("good" state) in its last season. The scenario of deteriorating ratings is a special case in our model, where the transition probability from the "bad" state to the "good" state is 0.

(b) Pedagogy: a university professor may be either a good or a bad instructor. Observations are students' responses to the teaching quality surveys administered in each semester. Actions are to do nothing (CON), have the instructor go through teaching workshops and seminars to improve his teaching capabilities (REP), or interview the instructor and the students on their teaching and learning experiences, respectively, and obtain a clear idea of what type of instructor the professor is (INS). This example fits our two-state POMDP model, as the states "good instructor" and "bad instructor" are interchangeable. In fact, many professors in academia experience the following pattern: initially, they are good instructors, who make an ongoing effort to improve their classes. Mid-career faculty are more likely to get stuck in a particular approach in the classroom, which leads to fatigue as they repeatedly teach the same material semester after semester. Yet these same instructors may switch states again towards the end of their careers, as they are considered a great source of wisdom (Hammond & Morgan, 1991). However, there may be good mid-career instructors and bad young instructors. A course survey may provide an indication of an instructor's quality, but it does not perfectly reveal the true state, as even good instructors may have a bad semester, e.g., due to personal problems at home, and a bad instructor may receive very high evaluations due to easy grading, for example.

(c) Credit scorecards and delinquency status. Financial institutions grant credit (e.g., mortgages with monthly payments) to two major types of clients: "good" and "bad" risk. Observations are payments or non-payments over time. In a given time period (e.g., a month), the grantor needs to make a decision regarding a client to whom it has provided a loan that is currently outstanding; this decision is based on an analysis of that client's payment (or non-payment) history (see, for example, Hand & Henley, 1997; Rosenberg & Gleit, 1994; Thomas, 2000). The decision may be one of the following: continue (CON) monitoring the payments (whether one was received or not), recall the loan and cut its losses (REP), or conduct a thorough and costly investigation (INS) of the client's state (e.g., analyze his or her bank statements, paychecks, etc.). We note that payments (or non-payments) do not perfectly distinguish a "good" risk from a "bad" risk: a "good" risk client may still miss a couple of payments, and a "bad" risk client may still make a few consecutive payments. Moreover, a "good" risk client may lose his or her job and become a "bad" risk client, and vice versa. The goal of the grantor is to maximize its stream of revenues or profits by correctly identifying the clients it deals with.

Overall, as implied by the above examples, application of the proposed POMDP model has the potential to produce outcomes that are superior to those obtained simply by relying on the whim of the decision maker, which may not be based on an objective