



Decision Support

Robust decomposable Markov decision processes motivated by allocating school budgets



Nedialko B. Dimitrov^a, Stanko Dimitrov^{b,*}, Stefanka Chukova^c

^a Graduate Program in Operations Research & Industrial Engineering, The University of Texas at Austin, USA

^b Management Sciences, University of Waterloo, Canada

^c School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, New Zealand

ARTICLE INFO

Article history:

Received 17 January 2014

Accepted 2 May 2014

Available online 17 May 2014

Keywords:

Markov processes

Dynamic programming-optimal control

School funding

ABSTRACT

Motivated by an application to school funding, we introduce the notion of a robust decomposable Markov decision process (MDP). A robust decomposable MDP model applies to situations where several MDPs, with the transition probabilities in each only known through an uncertainty set, are coupled together by joint resource constraints. Robust decomposable MDPs are different than both decomposable MDPs, and robust MDPs and cannot be solved by a direct application of the solution methods from either of those areas. In fact, to the best of our knowledge, there is no known method to tractably compute optimal policies in robust, decomposable MDPs. We show how to tractably compute good policies for this model, and apply the derived method to a stylized school funding example.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Allocation of school funding is critical to improving school performance. Unfortunately, there is no consensus on how this limited resource should be allocated. As such, the allocation of school funding is a recurring topic of political discussion (Shutt, 1979; Fensterwald, 2013; Blume, 2013; Garber, 1997). For example, the state of California passed an initiative in 2012 to raise taxes specifically to fund schools. In 2013, a major debate was how to allocate this schools funding: should it be allocated by population or should poorer schools receive more funding?

Motivated by allocating funding to school, in this paper we propose a new method of allocating finite resources, based on control theory. Our method extends previous work on using Markov decision processes (MDPs) for budget allocation. The main contributions of this paper are (1) to introduce the concept of robust, decomposable MDPs, (2) to propose a computationally tractable method of computing good policies in such MDPs and (3) to illustrate, through a stylized example, that funding allocation based on these MDPs outperforms strategies inspired by a current real-world school funding policy, No Child Left Behind.

For a school district composed of multiple schools, we would like to find a funding allocation policy that maximizes the total

discounted expected reward over a finite planning horizon. Each school can be in one of a finite number of performance states—bad, good, excellent, *etc.*—and transitions between those states based on funding allocated by the school district. Given an allocation of funding, each school transitions independently between its performance states. The school district, does not know the exact transition probabilities for each school, but instead knows an uncertainty set within which the transition probabilities reside. Further, the school district has a finite budget to distribute between the schools. The funding allocation problem of the school district can be modeled as a large MDP, where a state is a vector of performance states for the individual schools; and the action set corresponds to a funding allocation. The school district's funding allocation choice couples the transitions of the individual schools. We model and compute good funding allocation policies by decomposing the large school district MDP into smaller MDPs through a novel Lagrangian relaxation. Because both the transitions are only known through uncertainty sets in both the large and small MDPs, we call the resulting model a robust, decomposable MDP.

Because our model is motivated by school funding, for completeness, in Appendix E we discuss the existing literature on school funding and improving school performance. In Section 2 we motivate the robust, decomposable MDP model in words. Related control theory work is discussed in Section 3. Sections 4 and 5 mathematically present the model and show how it can tractably produce a funding policy. In Section 6 we compare the

* Corresponding author. Tel.: +1 5198884567x38966.

E-mail addresses: ned@alumni.cs.utexas.edu (N.B. Dimitrov), sdimitro@uwaterloo.ca (S. Dimitrov), schukova@msor.vuw.ac.nz (S. Chukova).

funding policy generated from our model to a stylized version of an existing policy, namely No Child Left Behind. We conclude the paper with some discussion in Section 7.

2. The model

In this paper, we use a robust, decomposable, finite horizon MDP to compute funding distribution policies for a school district. While the tools derived are generally applicable to robust, decomposable, finite horizon MDPs, they are motivated by an application to school district funding. In this section, we begin with an English language description of the basic problem and motivation for the model we develop.

Consider a *school district* composed of a set of *schools*. While we use this phrasing, school district could be interpreted in a state-level sense, as a group of smaller districts. In either case, we are concerned with a large institution—which we call the school district—composed of smaller institutions—which we call the schools.

Each school has a *proficiency level*, which, in practice, is determined annually by the school's students' performance on a standardized exam. The school district wants to maximize the proficiency of its schools. The main tool the school district has to improve the proficiency of its schools is distributing a limited amount of funding amongst the schools. In this model we assume all schools have a base operating cost that they receive regardless of their performance, and thus we normalize that cost to zero. The funding allocated by our model is any additional funding a school receives to help improve its performance. Of course, not all schools are the same. Schools have varying numbers of students, and varying capabilities to effectively use allocated funding. The school district makes an annual decision on how to allocate its limited funding amongst the individual schools. Realistically, the school district can only make a funding allocation plan for a decade or so in the future.

Each school transitions into a state of higher or lower proficiency randomly, but the transition probabilities are based on the funding the school receives from the district. Unfortunately, it is impossible for the school district to know these transition probabilities—which depend on the school in question, its current proficiency state, and its level of funding. Even historical data on the school's transitions would not help identify specific transition probabilities, since the situation at an individual school changes quickly, as compared to the annual funding decision cycle. Instead, the school district only has an uncertainty set—a range of possible transition probabilities—for the school.

We formalize the school district's funding allocation problem into a mathematical model as follows. Each individual school can be modeled as an MDP, where the states describe the school's proficiency, the actions describe the school's funding level, and the rewards describe the benefit the school district receives from the school's proficiency level. The rewards for each school capture the school's importance to the district, which could, for example, be proportional to the number of students in the school. The exact transition probabilities for each school's MDP are not known, but instead for each state action pair, we know an uncertainty set describing a set of possible transition probabilities. We term an individual school's MDP as a *little MDP*.

The school district's funding allocation problem is also an MDP. The school district's MDP has states that are the cross product of the states of the individual schools. The individual schools MDPs are coupled by the joint funding allocation decision. The only actions available to the school district are those that satisfy the district's common budget. But, given a distribution of the common funding, the transitions at each of the schools are independent. The school district has a finite horizon, about a decade, for which

to plan funding allocation decisions. We term the school district's MDP as the *big MDP*.

The proposed method answers the basic question: *How should the school district distribute its limited funding to maximize the discounted total proficiency of its schools over its planning horizon?* We model this decision using a robust, decomposable, finite horizon MDP. The term robust comes from the fact that the transition probabilities for each school are not known, but are only known through uncertainty sets. The term decomposable comes from the fact that the big MDP is made of a cross product of little MDPs, coupled only by the limited funding to take actions across them. The term finite horizon comes from the approximately decade long planning period for the district. We formalize the robust, decomposable, finite horizon MDP with mathematical notation in Section 4. In the next section, we discuss some related control theory work.

3. Related work

Markov decision processes have provided a powerful modeling tool in solving problems related to planning and decision-making under uncertainty, see (Puterman, 2005). For example recent applications of MDPs include: credit card profitability (So & Thomas, 2011), transshipment between warehouses (Seidscher & Minner, 2013), infectious disease spread (Yaesoubi & Cohen, 2011), air-cargo revenue management (Han, Tang, & Huang, 2010), etc. However, often in practice an MDP model ends up with very large state and action spaces. The computational difficulties in analyzing MDPs with high dimensionality stimulated research on developing techniques to deal with this problem. Typical approaches to address high dimensionality in MDPs are: function approximation, reachability considerations and aggregation techniques (Meuleau et al., 1998; Powell, 2007).

Nilim and Ghaoui (2005) study the sensitivity of the optimal solutions to Markov decision problems with respect to the state transition probabilities. Typically, in practice, the MDP transition probabilities are unknown and the errors in their estimations often impose limitations in using MDPs as a modeling tool. The authors consider a finite-state, finite-action MDP, and model the uncertainty in the transition matrices by using uncertainty sets. They successfully use a modification of the classical dynamic programming algorithm to solve this robust MDP problem.

Adelman and Mersereau (2008) focus on stochastic dynamic programming problems that are suitable for relaxation via decomposition. The problems they consider consist of a number of sub-problems that are independent of each other except for a set of coupling constraints on the action space. They use Lagrangian relaxation and linear programming to approximate the dynamic programming formulation of the problem and provide comparisons between the two relaxations and the optimal solution. Also, they conclude with a useful discussion on how to select the appropriate relaxation.

The model we develop, while it has pieces from both Adelman and Mersereau and Nilim and El Ghaoui, is novel. Specifically, solution methods for our model do not follow directly from either paper. This is demonstrated by the fact that the big MDP is not solvable by standard MDP methods, even using an exponentially sized state space, as the corresponding big MDP of Adelman and Mersereau. Moreover, it cannot be solved by the methods of Nilim and El Ghaoui, because the uncertainty sets we address have fundamentally different structure than their robust MDPs. We provide additional discussion on this in Appendix B.

Sisikoglu, Epelman, and Smith (2011) develop a learning algorithm for solving a discounted homogeneous MDP with unknown transition probabilities. These transition probabilities are learned either through simulation or direct observation of the

Download English Version:

<https://daneshyari.com/en/article/479800>

Download Persian Version:

<https://daneshyari.com/article/479800>

[Daneshyari.com](https://daneshyari.com)