Ecological Modelling 351 (2017) 77–86

Contents lists available at ScienceDirect

Ecological Modelling

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

Time to fly: A comparison of marginal value theorem approximations in an agent-based model of foraging waterfowl

Matt L. Miller[a,*], Kevin M. Ringelman[b], John M. Eadie[c], Jeffrey C. Schank[a]

[a] University of California, Psychology Department, One Shields Ave., Davis, CA 95616, United States
[b] Louisiana State University, School of Renewable Natural Resources, LSU Agricultural Center, Baton Rouge, LA 70803, United States
[c] University of California, Wildlife, Fish, and Conservation Biology, One Shields Ave., Davis, CA 95616, United States

ARTICLE INFO

Article history:
Received 17 August 2016
Received in revised form 13 February 2017
Accepted 14 February 2017
Available online 9 March 2017

Keywords:
Optimal foraging theory
Decision-making
Algorithm
Simulation

ABSTRACT

One of the fundamental decisions foragers face is how long an individual should remain in a given foraging location. Typical approaches to modeling this decision are based on the marginal value theorem. However, direct application of this theory would require omniscience regarding food availability. Even with complete knowledge of the environment, foraging with intraspecific competition requires resolution of simultaneous circular dependencies. In response to these issues in application, a number of approximating algorithms have been proposed, but it remains to be seen whether these algorithms are effective given a large number of foragers with realistic characteristics. We implemented several algorithms approximating marginal value foraging in a large-scale avian foraging model and compared the results. We found that a novel reinforcement-learning algorithm that includes cost of travel is the most effective algorithm that most closely approximates marginal value foraging theory and recreates depletion patterns observed in empirical studies.

1. Introduction

Optimal patch selection describes how foragers should be distributed across heterogeneous landscapes with respect to food items and to each other. Under simplifying assumptions, optimality theory predicts that foragers consuming continually-replenishing resources might select patches according to the ideal free distribution (Fretwell and Lucas, 1969), in which animals distribute themselves in the patches proportionate to the gain rate of resources. However, for many animals, patch selection is dynamic: when resources are depleted in one area, animals must find new patches containing those resources. In 1976, Charnov proposed marginal value theorem (MVT), an analytical solution for determining when to leave a foraging patch that predicts that a forager should depart when the intake rate for that patch falls below the long-term average intake rate across all available patches. Charnov's theoretical result sparked a flurry of empirical and modeling studies that suggested other patch-departure rules including fixed-time, fixed-intake, minimizing, maximizing, and inter-reward intervals (e.g., Cowie, 1977; Hodges, 1981; Iwasa et al., 1981; McNair, 1982; Green, 1984; McNamara and Houston, 1985).

There remains some uncertainty as to how well any animal fits the predictions of MVT and whether any animal could actually gather the information required to implement MVT (Stephens and Krebs, 1987; Stephens et al., 2007). Nonetheless, animals likely use some foraging strategy to be able to thrive in their environment, whether through evolution, learning, development, or some combination of these processes. We may expect such strategies to approximate MVT.

Since MVT has broad application, including agent-based models (ABMs) that explore how individual behaviors scale up to create patterns at larger scales (Grimm et al., 2005; DeAngelis and Mooij, 2005; McLane et al., 2011) and non-biological systems such as robot foraging (Ulam and Balch, 2004), finding the best algorithmic approximation for MVT is of critical importance for applied foraging contexts. In this paper, we explore several avenues by which behavioral and biological researchers can optimally program foraging agents that approximate MVT, using overwintering waterfowl as a test species. In particular, we are interested in which algorithms provide the longest survival times in an environment with depleting resources. We examine two approaches suggested in the robotics literature, subsequently modified to account for

* Corresponding author.
E-mail addresses: hzmiller@ucdavis.edu (M.L. Miller), kringelman@agcenter.lsu.edu (K.M. Ringelman), jmeadie@ucdavis.edu (J.M. Eadie), jcschank@ucdavis.edu (J.C. Schank).

http://dx.doi.org/10.1016/j.ecolmodel.2017.02.013
0304-3800/

Table 1
Abbreviations used in this paper.

Abbreviation	Meaning
ABM	Agent-based model
DT50M	Days to 50% mortality
DTD	Days to deficit
GUD	Giving-up density
IFD	Ideal free distribution
LTFA	Long-term-forward average
MVT	Marginal value theorem
OMVT	Online MVT algorithm
RLMVT	Reinforcement-learning MVT algorithm
XOMVT	Extended OMVT algorithm
XRLMVT	Extended RLMVT algorithm

travel time, and examine how these different approximations to MVT affect energy intake and survival in our modeled system.

1.1. Marginal value theorem

We consider three different implementations of marginal value theorem in this paper: Charnov's (1976) original MVT, or *classical MVT*; an algorithm approximating MVT based on reinforcement learning (Wawerla and Vaughan, 2009), or *reinforcement-learning MVT*; and an algorithm approximating MVT using continuously-updated estimates (Wawerla and Vaughan, 2010), or *online MVT* (abbreviations are summarized in Table 1). In the latter two cases, we consider both the original techniques suggested by Wawerla and Vaughan as well as modified versions of these methods that take into account differential costs of travel between foraging areas, or *extended reinforcement-learning MVT* and *extended online MVT*.

1.2. Classical MVT

In 1976, Charnov proposed a mathematical model for the amount of time a forager should remain in a patch of a given quality. If we consider that patches can be assigned a *patch type* based on patch quality, identified by p, and a forager has a net gain in a patch of type p of $g_p(T_p)$ if it spends T_p amount of time in that patch type, then the marginal rate of intake as length of time in the patch type increases is

$$\frac{\partial g_p\left(T_p\right)}{\partial T_p}. \quad (1)$$

(See Table 5 for notation). Similarly, if we are given the travel time between patches, t', and the cost of travel E_{travel}, we can calculate the net intake rate for each patch type,

$$\frac{g_p\left(T_p\right) - t'E_{travel}}{t' + T_p}. \quad (2)$$

If we know the proportion of each patch type in the environment, π_p, we can then calculate the average net intake rate for the whole environment,

$$\frac{\sum_p \pi_p g_p\left(T_p\right) - t'E_{travel}}{t' + \sum_p \pi_p T_p} \quad (3)$$

Charnov showed that T_p is optimized when the marginal rate was equal to the average net intake rate for the environment. That is, more time spent in the patch would be wasted effort and less time would fail to exploit useful resources.

Calculating this rate requires omniscient knowledge of patch quality across the environment. The gain function g must be sufficiently characterized to calculate its rate of change as time-in-patch increases, and for heterogeneous patches, g must be characterized for all patches in the environment. Even in the case of a single forager it is uncertain that the gain function can be adequately characterized (see, for example, Stephens and Krebs, 1987) because average intake rate depends on the time in patches, and time in patches depends on average intake rate. With multiple foragers and exploitation competition, both marginal and net intake rates change as a function of the number of foragers in the patch; an ideal forager would not only resolve its own circular dependency in MVT, but would also have to solve it for every other forager. Ignoring the fact that animals cannot gather this information in the real world (Bartoń and Hovestadt, 2013), the problem of circularity in applying MVT has suggested that computation even with perfect information is intractable (Wawerla and Vaughan, 2010). These difficulties in applying MVT are not surprising since MVT was derived as a theoretical optimum that behavior might approach in the limit, not as a strategy for decision-making and not under real-world conditions such as resource competition and depletion (Stephens and Krebs, 1987; Wajnberg et al., 2006).

1.3. Reinforcement-learning MVT

Because of these difficulties in developing optimal foraging algorithms that satisfy MVT, algorithms that approximate MVT have been proposed to determine time in patches, often relying on very simple rules (Gibb, 1958; Krebs, 1973; Krebs et al., 1974; McNamara, 1982). One of the more promising approaches was proposed by Wawerla and Vaughan (2009) who estimated the optimal patch departure time by simulating reinforcement learning, which we refer to as *reinforcement learning MVT*, or RLMVT.

RLMVT is based on an approach to simulating reinforcement learning by implementing the *n-armed bandit* using *softmax* to overcome local optima (Sutton and Barto, 1998; see Supplemental material C.1). This algorithm can be used to optimize foraging behavior. Consider that the essential problem in implementing MVT is determining the energy gain rate at which the agent should switch patches (the *switching threshold*). An agent can choose to switch at too high a rate, which will result in spending too little time in any given patch; it can choose to switch at too low a rate, which will result in spending too much time in any patch; or it can choose the optimal rate, which should converge with the predictions of MVT. If we consider each of the switching thresholds as one of the slot machine's n arms and we equate the net gain across all patches using that switching threshold as its reward, we can use the n-armed bandit algorithm to find the optimal rate at which to switch patches under MVT. That is, the expected reward for a given switching threshold (corresponding to an arm) is the average net gain rate experienced in patches for which that threshold was used. The softmax algorithm helps greedy optimization systems like the n-armed bandit from settling into local basins of attraction by making choices that currently appear less than optimal more likely to be selected; it also prevents the expected reward values for less-likely thresholds from becoming outdated by allowing them to be selected occasionally (and thus updated) throughout the simulation. Softmax uses a parameter called *temperature* (τ), in which higher temperatures make less-optimal-appearing choices increasingly likely, going to equal probabilities for all choices regardless of expected reward at infinitely-high temperatures (see Supplemental material C.1 for softmax details).

There are two complications: (1) switching threshold is a continuous variable, while the arms on the slot machine represent discrete values; and (2) the optimal switching threshold may change as food resources are depleted. Wawerla and Vaughan (2009) conceptualized this mapping between MVT and the n-armed bandit and addressed the first complication by generalizing Sutton and Barto's algorithm for a continuous action space (detailed in Supplemental material C.1). The second complication arises as a result of agents foraging within a finite environment: classical MVT