Journal of Mathematical Psychology 79 (2017) 29–43

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Model-based estimation of subjective values using choice tasks with probabilistic feedback

Kentaro Katahira [a,b,c,*], Shoko Yuki [d], Kazuo Okanoya [b,d]

[a] Center for Evolutionary Cognitive Sciences, The University of Tokyo, Meguro, Tokyo, 153-0041, Japan
[b] Behavior and Cognition Joint Research Laboratory RIKEN Brain Science Institute, Wako, 351-0198, Saitama, Japan
[c] Department of Psychology, Graduate School of Informatics, Nagoya University, Nagoya, Aichi, 464-8601, Japan
[d] Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, 153-004, Japan

HIGHLIGHTS

- A novel method for estimating subjective value from choice behavior is proposed.
- The proposed method employs a new choice task using probabilistic feedback.
- The proposed method performs the model-based estimation based on a reinforcement learning theory.
- The validity and limitations of the proposed method are investigated.
- The proposed method is demonstrated using actual choice data from rats.

ARTICLE INFO

Article history:
Received 10 July 2016
Received in revised form 5 May 2017

Keywords:
Subjective value
Model-based estimation
Reinforcement learning
Choice behavior
Random feedback

ABSTRACT

Evaluating the subjective value of events is a crucial task in the investigation of how the brain implements the value-based computations by which living systems make decisions. This task is often not straightforward, especially for animal subjects. In the present paper, we propose a novel model-based method for estimating subjective value from choice behavior. The proposed method is based on reinforcement learning (RL) theory. It draws upon the premise that a subject tends to choose the option that leads to an outcome with a high subjective value. The proposed method consists of two components: (1) a novel behavioral task in which the choice outcome is presented randomly within the same valence category and (2) the model parameter fit of RL models to the behavioral data. We investigated the validity and limitations of the proposed method by conducting several computer simulations. We also applied the proposed method to actual behavioral data from two rats that performed two tasks: one manipulating the reward amount and another manipulating the delay of reward signals. These results demonstrate that reasonable estimates can be obtained using the proposed method.

1. Introduction

Subjective values (or utilities) assigned to positive or negative events by living systems in general differ from their objective value (e.g., amount of money). Rewards with larger amounts and less delay are basically preferable, but the subjective values are not linearly related to objective, measurable values such as amount and delay (e.g., Kahneman & Tversky, 1979). Investigations into the valuation systems of living systems have gained significant attention in various fields such as psychology, neuroscience, and psychiatry (e.g., O'Doherty, 2014; Rangel, Camerer, & Montague, 2008). For example, some psychiatric disorders (e.g., depression) can be characterized by altered subjective values (for a review, see Chen, Takahashi, Nakagawa, Inoue, & Kusumi, 2015). Thus, the validity of animal models of a psychiatric disorder may be evaluated based on the subjective values of the subjects.

Traditional econometric methods of estimating subjective value cannot be applied to animals because they rely on verbal instruction (e.g., Kable & Glimcher, 2007; Kahneman & Tversky, 1979). Several methods have been used to estimate subjective values or preferences in animal studies. A typical procedure is to have the subjects learn the relationship between a specific response (e.g., pressing a lever or remaining in a specific location) and the resulting outcome, from which the subjective value is measured (e.g., Green & Estle, 2003). This approach requires sufficient training so that the animals learn the relationships among all of the

* Corresponding author at: Department of Psychology, Graduate School of Informatics, Nagoya University, Nagoya, Aichi, 464-8601, Japan.
E-mail address: katahira@lit.nagoya-u.ac.jp (K. Katahira).

http://dx.doi.org/10.1016/j.jmp.2017.05.005
0022-2496/

items and the choice behavior reaches the steady state. Another common method utilizes the law of how animals distribute their responses depending on the reinforcement, i.e., the matching law (Miller, 1976). Both approaches rely on the pairwise comparison of preferences for two items. Thus, to measure the subjective values of several items, the researcher must examine the preferences for multiple combinations of items. This method requires much time and sophisticated experimental considerations.

In the present study, we propose a novel method for estimating subjective values especially from animal behaviors using novel behavioral tasks and reinforcement learning (RL) model-based analysis. RL is usually formulated as an algorithm that attempts to maximize the total reward that a decision-maker can obtain. Recent studies, however, have begun to use the RL framework to model human behavior that does not necessarily lead to reward maximization (Neiman & Loewenstein, 2011; Shteingart & Loewenstein, 2014). For example, basketball players tend to choose to make a 3-point shot immediately after an experience of success; however, this dependence decreases the success rate. This choice behavior is modeled using an RL model (Neiman & Loewenstein, 2011, 2014). Additionally, RL models have been important data analysis tools for experiments involving value-based, decision-making tasks (Corrado & Doya, 2007; Daw, 2011; O'Doherty,Hampton & Kim, 2007).

Standard RL theory assumes that there is an increased probability of choosing an option that has been reinforced in the immediate past. The magnitude of dependence decays exponentially with the passage of time (trials) (Katahira, 2015). The main idea of the proposed method is to utilize this property. The RL theory also assumes that the larger the subjective value of an outcome, the more frequently the decision-maker repeats the same choice in the immediate future. Using the model parameter fit of RL models to trial-by-trial data, one can effectively estimate the subjective values of various decision-outcomes. The proposed method takes advantage of transient, trial-level dynamics of behavior, whereas other conventional methods examine only steady-state behavior. By using the transient effect of outcome on subsequent choices, it can estimate the value of multiple types of outcomes in a single experiment consisting of only two options.

The remainder of this paper is organized as follows. First, we describe the proposed method, which consists of the novel experimental design and RL model-based analysis. Next, we examine the validity and several properties of the proposed method based on synthetic data. We then apply the proposed method to actual behavioral data from rats. In the demonstration, we examined the rats' subjective values regarding amounts of rewards and delays of the reward (and no-reward) signal. Finally, we discuss the advantages and limitations of the proposed method.

2. Proposed method

The proposed method consists of novel experimental tasks and RL model-based trial-by-trial analysis of behavioral data. In the following, we describe the basic task structure, the RL models, and the statistical analysis procedure.

2.1. Basic task properties

The proposed choice task has the following structure. First, the outcome of choice (decision-outcome) should contain at least one appetitive outcome. This point is particularly crucial in animal studies to ensure that there is an incentive that will motivate animals to engage in the task. Second, the task must have contingency between the valence of outcome (appetitive, neutral, or aversive) and the animal's choice, as in conventional decision-making tasks. Contingency is required because it provides the animal with an incentive to learn the value of its actions. Within the outcome valence, however, the outcomes may be randomly chosen, irrespective of the animal's choice. Third, the contingency between choice and outcome valence must change during the task so that the animals' choice does not converge with the same option. Although a class of RL models, as employed in the present study, does not converge to deterministic choice behavior, actual animals' choice behavior often becomes deterministic if they are exposed to the constant contingency condition. This also occurs with other RL models, such as actor-critic learning (Sakai & Fukai, 2008). Using dynamic changing contingency prevents subjects from converging to a deterministic choice behavior, which is less useful for estimating subjective values. Specific task examples are presented in the following simulation and in experiments using rats.

2.2. Reinforcement learning model

In this section, we introduce RL models (Sutton & Barto, 1998). Specifically, we consider several variants of Q-learning with a single state (Watkins & Dayan, 1992), which is the model most commonly used in the model-based analysis of choice behavior. The model assigns each action an action value denoted as $Q_i(t)$, where i is the index of the action and t is the index of the trial. In the common setting, the initial action values are set to zero, i.e., $Q_i(1) = 0$ for all i. Based on the outcome, the action values for the action i are updated as follows:

$$Q_i(t+1) = Q_i(t) + \alpha_L\left(r(t) - Q_i(t)\right) \tag{1}$$

where α_L is the learning rate that determines the degree to which the model updates the action value depending on the reward prediction error, $r(t) - Q_i(t)$. The range of the learning rate is restricted between 0 and 1. For the unchosen action option $j\,(i \neq j)$, the action value is updated as follows:

$$Q_j(t+1) = (1-\alpha_F)\,Q_j(t)\,, \tag{2}$$

where α_F is the forgetting rate (Erev & Roth, 1998; Ito & Doya, 2009). In a typical RL model, the action value of the unchosen option is not updated. This convention can be represented by setting $\alpha_F = 0$. We call this the standard Q-learning model. The model with an identical learning rate and forgetting rate ($\alpha_L = \alpha_F$) is called Q-learning with forgetting (F-Q-learning). We also consider the model in which the learning rate and forgetting rate are allowed to differ ($\alpha_L \neq \alpha_F$) and the learning rate can take a non-zero value. This is called Q-learning with differential forgetting (DF-Q-learning).

We suppose that there are at least two different types of decision-outcomes. To estimate the subjective value of the outcomes, we propose two methods to represent the subjective value. One method is non-parametric and assigns a single parameter for each outcome type. The other method uses a parametric function, which represents subjective values as a function of the objective quantity of outcomes (e.g., amount, or delay).

In the non-parametric method, we set the value of outcome m (m is the index of decision-outcome) as κ_m, and $r(t) = \kappa_m$ if outcome m appears at trial t (Katahira, Fujimura, Matsuda, Okanoya, & Okada, 2014; Katahira, Fujimura, Okanoya, & Okada, 2011). For example, the index of outcome indicates the reward amount (in the following simulations and Task 1 of the rat experiments) and the delay in reward (no-reward) signals (Task 2 of the rat experiments). We denote the value of the reference outcome (such as absence of reward) as κ_0. This is often fixed at zero, but we also examine the case in which κ_0 is estimated as a free parameter. We call $r(t)$ and thus κ_ms the reward value. We assume that the reward value reflects the subjective value of the corresponding outcome.

When the outcome types are quantifiable, one can parameterize the value function. In the parametric method, for example, the