Neural Networks 53 (2014) 52–60

Contents lists available at ScienceDirect

Neural Networks

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

# Assist-as-needed robotic trainer based on reinforcement learning and its application to dart-throwing

Chihiro Obayashi *, Tomoya Tamei, Tomohiro Shibata

*Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara, Japan*

ARTICLE INFO

*Article history:*
Received 14 December 2012
Received in revised form 16 January 2014
Accepted 24 January 2014

*Keywords:*
Assistive robotics
Assist-as-needed
Motor skill learning
Reinforcement learning

ABSTRACT

This paper proposes a novel robotic trainer for motor skill learning. It is user-adaptive inspired by the assist-as-needed principle well known in the field of physical therapy. Most previous studies in the field of the robotic assistance of motor skill learning have used predetermined desired trajectories, and it has not been examined intensively whether these trajectories were optimal for each user. Furthermore, the guidance hypothesis states that humans tend to rely too much on external assistive feedback, resulting in interference with the internal feedback necessary for motor skill learning. A few studies have proposed a system that adjusts its assistive strength according to the user's performance in order to prevent the user from relying too much on the robotic assistance. There are, however, problems in these studies, in that a physical model of the user's motor system is required, which is inherently difficult to construct. In this paper, we propose a framework for a robotic trainer that is user-adaptive and that neither requires a specific desired trajectory nor a physical model of the user's motor system, and we achieve this using model-free reinforcement learning. We chose dart-throwing as an example motor-learning task as it is one of the simplest throwing tasks, and its performance can easily be and quantitatively measured. Training experiments with novices, aiming at maximizing the score with the darts and minimizing the physical robotic assistance, demonstrate the feasibility and plausibility of the proposed framework.


## 1. Introduction

Acquiring expertly skilled movements is generally a difficult task. Moreover, instructing novices in acquiring expertly skillful movements is also inherently difficult, because such movements are generated by unseen muscle (d'Avella & Bizzi, 2005; d'Avella, Saltiel, & Bizzi, 2003; Murai, Yamane, & Nakamura, 2009) and neural activity (Hemami & Dariush, 2012; Ijspeert, 2008; Mylonas et al., 2012; Williamson, 1998).

This paper proposes a novel robotic trainer for motor skill learning. It is user-adaptive inspired by the assist-as-needed principle well known in the field of physical therapy (Cai et al., 2006; Emken & Reinkensmeyer, 2005; Jezernik, Schärer, Colombo, & Morari, 2003).

Most previous studies in the field of the robotic assistance of motor skill learning have used predetermined desired trajectories, and it has not been examined intensively whether these trajectories were optimal for each user (Crespo & Reinkensmeyer, 2008; Duschau-Wicke, Brunsch, Lunenburger, & Riener, 2008; Emken & Reinkensmeyer, 2005).

Furthermore, the guidance hypothesis states that humans tend to rely too much on external assistive feedback, resulting in interference with the internal feedback necessary for motor skill learning (Schmidt & Wrisberg, 2008).

MIT-Manus (Krebs et al., 2004) and MIME (Lum et al., 2006) pioneered impedance control to rehabilitation robotics, and MIT-Manus (Krebs et al., 2003) include impedance selection based on the user's performance. The impedance selection, however, has not been automatized, and instead physical therapists do the selection based on their knowledge and experience.

A few studies have proposed a system that adjusts its assistive strength according to the user's performance in order to prevent the user from relying too much on the robotic assistance (Crespo & Reinkensmeyer, 2008; Emken & Reinkensmeyer, 2005). There are, however, problems in these studies, in that a physical model of the user's motor system is required, which is inherently difficult to construct.

In this paper, we propose a framework for a robotic trainer that is user-adaptive and that neither requires a specific desired trajectory nor a physical model of the user's motor system, and we achieve this using model-free reinforcement learning.

We chose dart-throwing as an example motor-learning task as it is one of the simplest throwing tasks, and its performance can be easily and quantitatively measured. Training experiments with novices aimed at maximizing the score with the darts and

* Corresponding author. Tel.: +81 0743 72 5982; fax: +81 0743 72 5989.
*E-mail addresses:* chihiro-o@brain.kyutech.ac.jp, chihiro-o@is.naist.jp (C. Obayashi), tomo-tam@is.naist.jp (T. Tamei), tom@is.naist.jp (T. Shibata).

http://dx.doi.org/10.1016/j.neunet.2014.01.012
0893-6080/

minimizing the physical robotic assistance. We demonstrate the feasibility and the plausibility of the proposed robotic trainer through experiments by comparing the results of four conditions: (1) without robot; (2) with non-adaptive fixed stiffness robot; (3) with adaptive robot; and (4) with non-adaptive decreasing stiffness robot.

This paper is organized as follows. Section 2 outlines the framework for our assist-as-needed robotic trainer. Section 3 describes how we applied this framework to develop the training system for dart-throwing. Section 4 describes our training experiments to validate the plausibility and the feasibility of the proposed training method. Section 5 presents the experimental results. Section 6 discusses the results while concluding remarks are provided in Section 7.

## 2. Assist-as-needed robotic trainer

The key points of the framework are enumerated as follows.

*Task-goal oriented*. In general, it is not trivial at all to predetermine some desired trajectory for motor skill learning because each person has their own motor control system. Since one of the most important aims of motor skill learning is to accomplish a task, the aim of the robotic trainer should be task-goal oriented, which requires a means of measuring the user's achievement (performance) on the task.

*Assist-as-needed*. The guidance hypothesis tells us that humans tend to rely too much on external assistive feedback, resulting in interference with the internal feedback necessary for motor skill learning. Therefore the robotic trainer should adjust its assistive strength according to the user's performance on the task; i.e., it should decrease its assistive strength when the user's performance increases, and vice versa.

*Model-free*. It is nontrivial to define the optimal throwing trajectory for each user in advance, owing to individual differences in body dynamics and in the neural controller. Furthermore, the prior dynamics between assistive strength and the user's performance are also different in each person. Therefore the robotic trainer should employ a model-free assistance algorithm.

*Minimum constraint*. It is also nontrivial to determine the optimal assistance policy for each user in advance. The robotic trainer should attempt to minimize the constraints on the user's motion, which also increases the safety of the system. One way to do this is to reduce the number of contact points between the user and the robot, while another way is to make the constraints compliant, allowing the user to move the robot easily.

In this study, we propose to employ a policy-gradient type of reinforcement learning algorithm (Peters & Schaal, 2008) as the core of the assist-as-needed robotic training. The aim of the learning algorithm is to maximize the task achievement and simultaneously to minimize the assistive strength of the robot. An advantage of the policy-gradient method is that the state and policy representations can be chosen to be meaningful for the task and can incorporate domain knowledge, which often requires fewer parameters in the learning process than in value-function-based methods (Sutton & Barto, 1998). Comparing the motor behaviors of experts and novices is one promising way of finding the state and policy representations in a low dimensional space. Another advantage is that the policy-gradient method is a model-free approach. Because of these advantages, this technique has been applied to robot learning studies including human–robot interaction studies (Mitsunaga, Smith, Kanda, Ishiguro, & Hagita, 2008; Tamei & Shibata, 2009; Tapus, Ţăpuş, & Matarić, 2008).

## 3. Training system for dart-throwing

In this section, we describe an application of the proposed framework to learning dart-throwing. We chose dart-throwing as our motor learning task because it is one of the simplest throwing tasks. More detailed reasons are as follows. First, throwing darts is usually performed by fixing the body trunk, primarily driven by one of the upper limbs, whose motion is mostly constrained in the sagittal plane. Second, its performance can be easily and quantitatively measured by a numerical score. Third, a dart is light-weighted, so the acceleration required at the tip of the hand for throwing a dart is much smaller than, for example, in the case of the baseball. From this fact, we could expect that some kinematic assistance would be helpful (Obayashi, Tamei, Imai, & Shibata, 2009) and that the effect of muscle fatigue could be ignorable through experiments where participants are required to throw darts many times.

### 3.1. Comparison of experts and novices in darts throwing

Here we describe the motion comparison of experts and novices in dart-throwing to find quantitative indices to be used in the state and policy representations in the actual training phase that distinguish their motions.

Twenty-eight healthy adults (14 females, 14 males, mean age 25.1 ± 3.4 years, 2 males had experience in dart-throwing in at least the last 2 years, the others did not have any experience of dart-throwing) participated in this motion comparison task.

The soft-tip darts throwing setting was used throughout this study, following the official rules of the World Darts Federation. The distance between the center of the dartboard and the dart on the dartboard, $d$, was semi-automatically measured by the two Web cameras. In the rest of this paper, we use the normalized score to measure dart-throwing performance of subjects. The normalized score is defined as $(d_{\max} - d)/d_{\max}$ where $d_{\max}$ is the radius of the dartboard. Thus the normalized score is bounded by 0 and 1; 1 and 0 indicate that the dart hit the board center and the rim of the board, respectively.

We used the MAC3D System (Motion Analysis Corp. USA) to measure upper-limb and trunk motion. The measurement frequency was 200 Hz. The markers for optical motion measurement were attached to each subject's upper limb (shoulder, elbow, and hand) and trunk according to the Helen-Hayes marker set. The measured marker positions were low-pass filtered using a second-order Butterworth filter with a cutoff frequency of 10 Hz. We used the elbow displacements $d_e$ and shoulder displacement $d_s$ as an index,

$$d_s = \|\boldsymbol{P}_s(t_{\text{release}}) - \boldsymbol{P}_s(t_{\text{end}})\|, \quad (1)$$

$$d_e = \|\boldsymbol{P}_e(t_{\text{release}}) - \boldsymbol{P}_e(t_{\text{start}})\|, \quad (2)$$

where $t_{\text{start}}$, $t_{\text{end}}$ $t_{\text{release}}$, $\boldsymbol{P}_s$ and $\boldsymbol{P}_e$ are start time of take-back, end time of take-back, release time, shoulder marker position and elbow marker position for each subject (Fig. 1). The throwing period was defined as the period in which the hand marker's height was higher than the shoulder marker's height. The end time of take-back was defined as the time at which the sign of the $z$-axis velocity changed around ±50 ms of the time at which the throwing hand was closest to the throwing arm's shoulder. The start time of take-back was defined as the time at which the sign of the $z$-axis velocity changed during the period from the end time of take-back to before 500 ms of the end time of take-back. The release time was defined as the time at which the sign of the $z$-axis velocity changed during the period from the end time of take-back to after 250 ms of the end time of take-back.

The subjects were instructed to shoot for the bull's eye as much as possible with their preferred rhythm. The subjects were first asked to throw darts 30 times as a warm-up and were then motion-captured while throwing darts over 16 trials. In each trial, subjects initially held three darts with their right hand and threw them one by one. Then they walked to the dartboard, picked up the thrown