Neural Networks 22 (2009) 348–361

Contents lists available at ScienceDirect

Neural Networks

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

Learning and generation of goal-directed arm reaching from scratch

Hiroyuki Kambara [a,b,*], Kyoungsik Kim [b,c], Duk Shin [d], Makoto Sato [a], Yasuharu Koike [a,b]

[a] Tokyo Institute of Technology, Precision and Intelligence Laboratory, Yokohama, 226-8503, Japan
[b] Japan Science and Technology Agency CREST, Saitama, 332-0012, Japan
[c] Tokyo Institute of Technology, Department of Computational Intelligence and Systems Science, Yokohama, 226-8502, Japan
[d] Toyota Central R&D Labs., Inc., Aichi, 480-1192, Japan

ARTICLE INFO

Article history:
Received 20 September 2007
Received in revised form 14 July 2008
Accepted 18 November 2008

Keywords:
Reaching
Reinforcement learning
Feedback-error-learning
Internal model
Trajectory planning

ABSTRACT

In this paper, we propose a computational model for arm reaching control and learning. Our model describes not only the mechanism of motor control but also that of learning. Although several motor control models have been proposed to explain the control mechanism underlying well-trained arm reaching movements, it has not been fully considered how the central nervous system (CNS) learns to control our body. One of the great abilities of the CNS is that it can learn by itself how to control our body to execute required tasks. Our model is designed to improve the performance of control in a trial-and-error manner which is commonly seen in human's motor skill learning. In this paper, we focus on a reaching task in the sagittal plane and show that our model can learn and generate accurate reaching toward various target points without prior knowledge of arm dynamics. Furthermore, by comparing the movement trajectories with those made by human subjects, we show that our model can reproduce human-like reaching motions without specifying desired trajectories.

1. Introduction

When we move our hand from one point to another, the hand paths tend to gently curve and the hand speed profiles are bell-shaped (Abend, Bizzi, & Morasso, 1982; Atkeson & Hollerback, 1985; Uno, Kawato, & Suzuki, 1989). Since humans show these highly stereotyped trajectories among an infinite number of possible ones, it has been suggested that the central nervous system (CNS) is optimizing arm movements so as to minimize some kind of cost function (Flash & Hogan, 1985; Harris & Wolpert, 1998; Uno et al., 1989). Cost functions specify movement-related variables that should be minimized during or after the movement. Meanwhile, several computational control models have been proposed to explain the way the CNS generates a set of motor commands that could minimize cost functions (Flash, 1987; Gribble, Ostry, Sanguineti, & Laboissiere, 1998; Hogan, 1984; Miyamoto, Nakano, Wolpert, & Kawato, 2004; Todorov & Jordan, 2002; Wada & Kawato, 1993). The hand trajectories predicted by these models are in strong agreement with experimental data. The purpose of these models, however, is to predict well-learned reaching movements themselves and not to describe the process of learning. In order to reproduce the movements, the control models were designed using detailed knowledge about the dynamics of musculoskeletal systems.

The purpose of this paper is to propose a motor control model that can learn the control law for reaching movements while actually controlling the arm. Let us call this type of model a "motor control-learning model". From observing infants' inaccurate and jerky motions (Konczak & Dichgans, 1997; Zaal, Daigle, Gottlieb, & Thelen, 1999), the motor skill to generate accurate and smooth adult-like movements seem to be acquired through motor learning performed in our daily life. However, this kind of learning is not as simple as general supervised learning problems. Since there is no explicit "teacher" that can provide the CNS with correct motor commands, the CNS has to learn how to control the body in a trial-and-error manner, through interaction with the environment.

Reinforcement learning has attracted much attention as a self-learning paradigm for acquiring optimal control strategy through trial-and-error (Sutton & Barto, 1998). In particular, the actor-critic method, one of the major frameworks for the temporal difference learning, has been proposed as a model of learning in the basal ganglia (Barto, 1995; Doya, 1999). We adopt the actor-critic method (Doya, 2000) in order to acquire a feedback controller for multi-joint reaching movements. Although we are not the first to apply the actor-critic method to a reaching task, the previous model only explained a reaching movement toward one particular target (Izawa, Kondo, & Ito, 2004). In our daily life, we are not

* Corresponding author at: Tokyo Institute of Technology, Precision and Intelligence Laboratory, Yokohama, 226-8503, Japan. Tel.: +81 45 924 5054; fax: +81 45 924 5016.
E-mail address: hkambara@hi.pi.titech.ac.jp (H. Kambara).

0893-6080/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2008.11.004

always reaching to the same target. The CNS should be learning how to generate reaching movements toward various targets in the workspace. However, it is difficult to realize various movements with high accuracy using a single feedback controller. Since the gravitational force acting on the arm depends on the posture of the arm, the force required to hold the hand at the target varies with the target position. Furthermore, the magnitude of muscle tension varies with the posture of the arm even if a level command signal is sent to the muscle. For these reasons, there is no guarantee that a single feedback controller trained for a particular target would generate accurate reaching movements to other targets.

Here we introduce an additional controller called an inverse statics model, which supports the feedback controller in generating reaching movements toward various targets. It handles the static component of the inverse dynamics of the arm. That is, it transforms a desired position (or posture) into a set of motor commands that leads the hand to the desired position and holds it there. Note that the arm converges to a certain equilibrium posture when a constant set of motor commands is sent to the muscles because of the spring-like properties of the musculoskeletal system (Feldman, 1966). If the inverse statics model is trained properly, it can compensate for the static forces (e.g. gravity) at the target point. Therefore, accurate reaching movements toward various target points are realized by combining the inverse statics model and the feedback controller that works moderately well within the workspace. To acquire an accurate inverse statics model in a trial-and-error manner, we adopt the feedback-error-learning scheme (Kawato, Furukawa, & Suzuki, 1987). In this scheme, inverse dynamics (or statics) models of controlled objects are trained by using command outputs of the feedback controller as error signals. This learning scheme was originally proposed as a computational coherent model of cerebellar motor learning (Kawato et al., 1987). The original model, however, did not explain how to acquire the feedback controller for arm movements. In our model, the actor-critic method is introduced to train the feedback controller. Therefore, our model gives a possible solution to the problem of feedback controller design in the feedback-error-learning scheme.

In addition to the feedback controller and the inverse statics model, we introduced a forward dynamics model of the arm into our motor control-learning model. The forward dynamics model is an internal model that predicts a future state of the arm given outgoing motor commands. It has been proposed that the CNS is utilizing the forward dynamics model to internally predict the state of the arm during the control process (Miall & Wolpert, 1996; Wolpert, Miall, & Kawato, 1998). The existence of the forward dynamics model in the CNS is also supported by psychophysical experiments (Bard, Turrell, Fleury, Teasdale, Lamarre, & Martin, 1999; Wolpert, Ghahramani, & Jordan, 1995). The forward dynamics model can be trained in a supervised learning manner since the teaching signal can be obtained from somatosensory feedback. In the literature of automatic control, the strategy to combine system identification with reinforcement learning succeeded in autonomously controlling machines with complex dynamics such as helicopters (Abbeel, Coates, Quigley, & Ng, 2007). In our model, the forward dynamics model is designed to predict the state of the arm at a future time instant so as to compensate for delay of motor output caused by graded development of the muscle force. The predicted future states are then utilized to determine command outputs of the feedback controller.

In the present study, we apply our motor control-learning model to a point-to-point reaching task in the sagittal plane. By simulating the learning process of the reaching task, we show that our model can accurately control the arm to reach toward various target points without assuming prior knowledge of the arm dynamics. In addition, we compare reaching movements simulated by our model with those of human subjects, and show that our model can reproduce features of both hand path and speed profile in human reaching movements without planning desired trajectories.

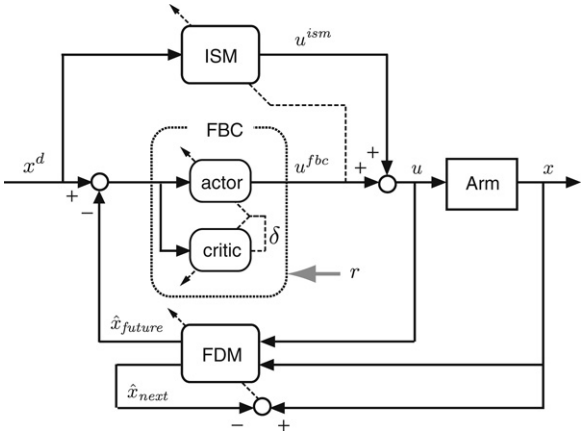

Fig. 1. The architecture of motor control-learning model: the model has three main modules, Inverse Statics Model (ISM), Feedback Controller (FBC), and Forward Dynamics Model (FDM). The FBC is composed of actor and critic units, which correspond to a controller and value function estimator respectively in the actor-critic method. The ISM generates a feed-forward motor command $\boldsymbol{u}^{ism}$ that shifts the equilibrium state of the arm to the desired state $\boldsymbol{x}^d$. On the other hand, the FBC generates a feedback motor command $\boldsymbol{u}^{fbc}$ that reduces the error between the desired state $\boldsymbol{x}^d$ and the future state $\hat{\boldsymbol{x}}_{future}$ predicted by the FDM. The error signal for the ISM is the feedback motor command $\boldsymbol{u}^{fbc}$. Meanwhile, the teaching signal for the FDM is the state of the arm $\boldsymbol{x}$ observed at next time instant. The FBC is trained by the actor-critic method so as to maximize the cumulative reward r. The temporal difference error δ related to the reward r is used as the reinforcer and error signal for the actor and critic units, respectively.

2. Motor control-learning model

Fig. 1 illustrates architecture of the motor control-learning model for a reaching task. The model consists of three main modules, inverse statics model (ISM), feedback controller (FBC), and forward dynamics model (FDM). The FBC is composed of actor and critic units, which correspond to a controller and value function estimator respectively in the actor-critic method.

At the beginning of each trial, a target point of reaching is given as a desired state $\boldsymbol{x}^d$ to the model. This $\boldsymbol{x}^d$ is kept constant at the target point throughout the trial. The ISM receives $\boldsymbol{x}^d$ as an input and generates a time-invariant motor command $\boldsymbol{u}^{ism}$. If the ISM were trained correctly, $\boldsymbol{u}^{ism}$ shifts the equilibrium of the arm to the target point. On the other hand, at time t, the FBC receives a state error between desired state $\boldsymbol{x}^d$ and future state $\hat{\boldsymbol{x}}_{future}(t-\Delta t)$ predicted by the FDM Δt second before time t. The FBC, then, transforms the state error into a feedback motor command $\boldsymbol{u}^{fbc}(t)$. The sum of $\boldsymbol{u}^{ism}$ and $\boldsymbol{u}^{fbc}(t)$ is sent to the arm as a total motor command $\boldsymbol{u}(t)$. Based on the total motor command $\boldsymbol{u}(t)$ and the state $\boldsymbol{x}(t)$, the FDM predicts next state $\hat{\boldsymbol{x}}_{next}(t)$ and also future state $\hat{\boldsymbol{x}}_{future}(t)$.

The three modules improve their performance in the following way. A teaching signal for FDM's prediction $\hat{\boldsymbol{x}}_{next}(t)$ is given by observing the actual state at time $t+\Delta t$. Therefore, the FDM can be trained in normal supervised learning manner, in which the error signal is determined as

$$E_{fdm}(t) = \boldsymbol{x}(t+\Delta t) - \hat{\boldsymbol{x}}_{next}(t). \tag{1}$$

On the other hand, the ISM is trained with the feedback-error-learning scheme in which the error signal for ISM's output $\boldsymbol{u}^{ism}$ is FDM's output, that is,

$$E_{ism}(t) = \boldsymbol{u}^{fbc}(t). \tag{2}$$