Neurocomputing 131 (2014) 2–14

Contents lists available at ScienceDirect

Neurocomputing

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

Explorative learning of inverse models: A theoretical perspective

Matthias Rolf*, Jochen J. Steil

Research Institute for Cognition and Robotics (CoR-Lab), Bielefeld University, Germany

ARTICLE INFO

Article history:
Received 16 November 2012
Received in revised form
15 March 2013
Accepted 29 April 2013
Available online 12 November 2013

Keywords:
Inverse models
Exploration
Motor learning
Redundancy
Goal babbling

ABSTRACT

We investigate the role of redundancy for exploratory learning of inverse functions, where an agent learns to achieve goals by performing actions and observing outcomes. We present an analysis of linear redundancy and investigate goal-directed exploration approaches, which are empirically successful, but hardly theorized except negative results for special cases, and prove convergence to the optimal solution. We show that the learning curves of such processes are intrinsically low-dimensional and S-shaped, which explains previous empirical findings, and finally compare our results to non-linear domains.

1. Introduction

In many learning scenarios, agents perform actions in some action space, whereas outcomes are measured in a different observation space. We assume that these two spaces are connected by a forward function that turns actions into observations. This function can be executed by the agent, but is not explicitly known to the learner. This is the basic setup, for instance, for reinforcement learning, where the agent executes actions and can observe rewards. In order to achieve some desired behavior an inverse function is needed that returns an appropriate action. A standard example is motor learning, where the observation space consists of certain world states such as end-effector positions. Possible actions can range from joint angles or forces to entire sequences of movements. A forward function turns the actions into outcomes like effector positions. Learning a corresponding inverse without explicit knowledge of the forward function has to rely on exploration schemes that generate example data usable for supervised learning.

Explorative learning in such domains is an active field of research and far from trivial. One substantial challenge is to deal with the redundancy in such domains: often multiple actions are mapped on the same outcome, such as different joint angles of an arm resulting in the same hand position. If redundancy is present, learning cannot be phrased as standard regression problem anymore because multiple correct solutions exist for a learner. As a further escalation, non-linear problems with redundancy have non-convex solution sets which prohibit learning from different solutions [1]. Often, the action space is very high-dimensional, which makes exhaustive exploration unfeasible. Yet, a number of practically efficient exploration and learning schemes have been proposed. The key is to structure exploration in a goal-directed manner. This idea has been used for tuning of well initialized inverse functions in several models [2–4]. Goal-directed exploration is particularly beneficial for learning from scratch, because it is applicable in very high-dimensional spaces and allows us to avoid inconsistencies [5]. Unfortunately, only very few theoretical results are available why and when such schemes can be successful. To the opposite, Sanger [6] proved that certain explorative learning formulations can fail systematically even in simple domains.

This paper aims to deepen the theoretical understanding of such learning schemes in redundant domains. We first formalize the general problem and discuss its difficulties. Then, we provide a thorough analysis of the linear case with redundancy, which is applied to goal-directed exploration. To our knowledge, we thereby provide the first positive theoretical outcomes on such learning by proving convergence to an optimal solution if noise is added to the exploration process. We analyze temporal aspects of exploration and learning and show how learning follows temporally S-shaped learning curves along low-dimensional manifolds through the action space. We finally discuss our results in the light of phenomena and empirical results in non-linear domains.

2. Two spaces and their gradients

We consider an agent that can execute actions q in some action space $\mathbf{Q} \subseteq \mathbb{R}^m$. An action results in an outcome $x \in \mathbf{X} \subseteq \mathbb{R}^n$ in the

[*] Corresponding author.

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.04.050

observation space. The relation between both variables is defined by the forward function $f(q)=x$. For learning, we do not know the forward function, but can query the outcome by means of execution of an action by the agent. This lack of information is crucial to distinguish the exploratory approach from learning schemes that use knowledge of the forward function. The agent is asked to achieve some desired observation, or "goal" $x^* \in \mathbf{X}^* \subseteq \mathbf{X}$. The agent has to generate an action $\hat{q}$, such that the outcome $x=f(\hat{q})$ matches the goal x^*. The agent's selection of an action can be denoted by a function $g(x^*)=\hat{q}$. The learning task is to obtain a function g that can realize all goals:

$$f(g(x^*))=x^* \quad \forall \ x^* \in \mathbf{X}^* \tag{1}$$

Hence, g must be a *right-inverse function* of f on the set of goals $\mathbf{X}^*$. Inverse functions do not always exist, so we need to require that f is surjective with $n \le m$. For $n<m$ the problem is ill-posed since different actions must result in the same outcome, which is referred to as redundancy. An exemplary situation of action and observation space is shown in Fig. 1.

This formulation makes an instantaneous link between actions and observations, which can be done for many coordinative problems like inverse kinematics [5], parametrized throwing movements [7], facial expressions [8], or, hypothetically, playing golf [6]. The formulation neglects cases in which the outcome also depends on some internal state (for instance in inverse dynamics problems [9]), but results are still applicable to some extent if the problem can be localized (see Section 5.1).

2.1. *The learning task in the observation space*

In the observation space, obtaining a right-inverse function can be directly formulated as optimization problem. We parametrize the function g with adaptable parameters W. The *performance error* E^X naturally measures how much an inverse estimate g deviates from the solution in Eq. (1). For a finite set of goals $\mathbf{X}^* = \{x_0^*, \dots, x_{K-1}^*\}$ we can write

$$E^X(W,\mathbf{X}^*)=\frac{1}{2K}\sum_{k=0}^{K-1}\|f(g(x_k^*,W))-x_k^*\|^2 \tag{2}$$

Learning can, for instance, be formulated as gradient descent on this error functional [1,10]. The central difficulty is that computing the *performance gradient* $\partial E^X/\partial W$ requires analytic knowledge about the forward function: Since W appears inside $f(\cdot)$, differentiation of E^X with respect to W requires to know the derivative of f. In general, inverse problems do not provide a teacher that could indicate such optimal gradient directions.

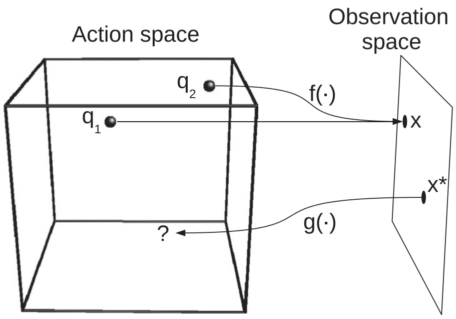

Fig. 1. Relation between action and observation space: a forward function f maps actions into a lower-dimensional outcome. An inverse g must suggest an action for a given target.

2.2. *Explorative learning in the action space*

If the performance gradient is not available, a feasible way to probe knowledge is to generate examples (x,q) by exploration [2–6,11,12]. The setup allows us to start in the action space by performing some q_l, and then observe the outcome $x_l=f(q_l)$. A data set $D=\{(x_l,q_l)\}_l$ can be used to learn in the action space. The *action error* on D measures how well the inverse estimate fits the data:

$$E^Q(W,D)=\frac{1}{2L}\sum_{l=0}^{L-1}\|g(x_l,W)-q_l\|^2. \tag{3}$$

Reducing this error can be approached by descending the *action gradient* $\partial E^Q/\partial W$. Importantly, this scheme is not a mere data-driven version of the optimization of E^X in observation space. In the action error we can replace $x_l=f(q_l)$ and see that the error evaluates on $g(f(q_l))-q_l$. Reducing this deviation corresponds to learning a *left-inverse function* g. This reflects the "reverse" sampling from the learner's output q_l to a correct input x_l. In contrast, standard supervised schemes, including explorative ones like active learning [13], require a mechanism to probe a correct output q^* for the learner, given an input x^*. Autonomous learning scenarios do not provide such an "oracle", which would reflect deep prior knowledge.

The learning task is to obtain a right-inverse function, which corresponds to minimizing E^X in observation space. Learning from exploratory data, however, minimizes E^Q in action space, which corresponds to learning a left-inverse function. Empirical results show that a right-inverse function can be learned by minimizing E^Q [5,2,11]. Why this is possible is not theoretically understood for the general case. In fact, this kind of learning largely depends on how the data set is chosen by exploration, whether f is linear or not, and whether the system contains redundancy: For the redundant case $n<m$, left-inverse functions do not exist on general data sets because different q_l can have the same outcome x_l. Trying to fit such inconsistent examples results in averaging. In non-linear domains such sets can be non-convex, so that averaging leads to invalid solutions [1]. Sanger [6] investigated the use of goal-directed exploration in the non-linear case without redundancy and showed that learning is not guaranteed to work.

This paper complements these previous, negative outcomes and investigates the role of redundancy. In order to disentangle the effects of redundancy, and effects of non-convex solution sets, we examine redundancy in linear domains. These domains have convex solution sets (linear subspaces) which allow us to study redundancy in isolation. Further, the linear case paves the way towards a localized understanding of non-linear problem domains. As a first positive result, we show that performance- and action-gradient have a non-negative relation in Section 2.3 and provide general fixpoint conditions. Using this framework, we re-investigate the setup of Sanger for the linear redundant case and show additional failure modes. An important outcome of this paper is that, if exploratory noise is added, learning will always converge to a valid solution, which is even optimal with respect to least-squares parameter values.

2.3. *Non-negative relation in linear domains*

In the linear domain, the relation between actions $q \in \mathbf{Q} \subseteq \mathbb{R}^m$ and outcomes $x \in \mathbf{X} = f(\mathbf{Q}) \subseteq \mathbb{R}^n$ is given by the linear forward function:

Definition 1 (*Linear forward function*). We define the forward function as $f(q)=M \cdot q$ where M is a real-valued matrix $M \in \mathbb{R}^{n \times m}$ with $n \le m$ and $rank(M)=n$.

Requiring M to have full rank implies surjectivity of f and thus solvability of the right-inverse problem. In correspondence to a