Boosting Robot Learning and Control with Domain Constraints

Hang Yin\textsuperscript{1,2}, Francisco S. Melo\textsuperscript{1}, Aude Billard\textsuperscript{2} and Ana Paiva\textsuperscript{1}

\textsuperscript{1}GAIPS, INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
\textsuperscript{2}Learning Algorithms and Systems Laboratory, École Polytechnique Fédérale de Lausanne

I. INTRODUCTION

In this short paper, we exploit robotics domain knowledge, be it acquired from humans or self-organization, to alleviate learning and control challenges from directly dealing with raw demonstrations or sparse reward signals. We take a unified latent variable perspective in incorporating domain constraints. The latent variables are regarded as task parameters or representations, which rationalize task observations with a generative model. The constraints can thus be specified with structured latent variables. Different from many related works, we explore latent structures that are computationally feasible and robotics-oriented to facilitate both task learning and control synthesis. The paper will briefly discuss adopted transfer learning framework also showcase successes by learning simulation-based tasks \cite{14,12} or a flexible policy initialization \cite{3}.

In this paper, we focus on imposing structures to an explicit representation of latent variables. This is realized by designing factored parameter distribution and general auxiliary objectives with a minor additional training cost. Moreover, unlike the transfer learning works, priors are usually not encoded as target policies but intermediate models such as cost functions, feature and modality transformations.

III. APPROACH AND CURRENT RESULTS

We assume there exists an abstract latent representation \( z \) which conditions the generation of \( x \) which denotes the observations of interests, e.g., motion or visual frames, formulating the likelihood \( p(x|z) \) and its variational lower bound as:

\[
    p(x) = \int p(x|z)p_0(z)dz
\]

\[
    \log p(x) \geq -E_q(z|x)(log p(x|z)) - KL(q(z|x)||p_0(z))
\]

where \( p_0(z) \) and \( q(z|x) \) denote the prior and approximate posterior distributions. The generative model \( p(x|z) \) is instanced as an energy-based model \( p(x|z) \propto e^{-\mathcal{J}(x,z)} \) \cite{17}.

Here \( z \) can be cast as the task parameter of \( \mathcal{J}(\cdot) \). In light of \( q(z|x) \), another way is to view \( z \) as a latent feature which maps raw \( x \) to a low-dimensional space. We discuss imposed constraints on \( z \) and obtained results in robot applications.

A. Parameter Dependency

We consider a quadratic \( \mathcal{J}(x) = (x - \mu)^T \Lambda (x - \mu) \) with \( z = \{ \mu, \Lambda \} \). The objective in effect encodes motion trajectories \( x \) with a reference \( \mu \) and tracking desirability \( \Lambda \). Meanwhile, \( \Lambda \) can be regarded as a force control proxy because of the popular heuristics correlating trajectory variance and compliance design \cite{2,8}. Here, we adopt a description of \( \Lambda \) which depends on \( \mu \), resulting a factored \( p_0 \):

\[
    p_0(z) = p_0(\Lambda|\mu)p_0(\mu)
\]

This can facilitate a sampling-based inverse optimal control to efficiently optimize the likelihood \cite{1}, avoiding dealing with ill-posed parameter constraints \cite{15}. Moreover, the dependency implies representing orthogonal control components in a local frame along the trajectory. This is known to be an important and intuitive task space decomposition in classic robotics, such as hybrid force control \cite{9}. In a robotic handwriting domain, a variable impedance controller is developed and the writing motion on a moving surface (Figure 1).
Fig. 1: Enforcing a dependency between task parameters to constrain the impedance representation: the arm-hand system performs compliant cursive handwriting, regulating the contact with a surface, whose orientation varies during the process.

B. Identical Modal Representation

We consider task examples \( x \) of different modalities, e.g., joint trajectory \( x_m \) and camera pixels \( x_v \) that form handwriting letters. The redundant description is similar to observing the pose of an object from different perspectives. This inspires an identical representation of \( z \) to abstract both \( x_m \) and \( x_v \). We enforce this domain constraint by adding a symmetrical KL-divergence to the variational objective in (1):

\[
q_m(z|x_m) = q_v(z|x_v) \Rightarrow \text{Diff}(q_m, q_v) = \text{KL}(q_m \| q_v) + \text{KL}(q_v \| q_m)
\]

This yields associative variational auto-encoders (VAE) when we parameterize posterior \( q \) and generative models \( p(x_m|z) \) and \( p(x_v|z) \) with neural models as the ones in [17]. The association can be exploited in control synthesis:

\[
x^*_m = \arg\min_{x_m} \| \text{Img}(x_m) - x_v \|^2 + \lambda \text{KL}(q(z|x_m)\|p_0)
\]

\[
x^{\text{Init}}_m \sim p(x_m|x_v) = \int p(x_m|z)q(z|x_v)dz
\]

where the shared prior \( p_0 \) and posterior \( q(z|x_m) \) provide informative initialization and bias in adapting motion \( x_m \) to match a novel letter image \( x_v \), with \( \text{Img}(\cdot) \) and \( \| \cdot \| \) denoting the obtained image and an Euclidean norm. The KL term encourages to search in alignment with the manifold and is weighted by \( \lambda \). As is shown in Figure 2 these constraints help to establish a rapid and stable convergence with less exploring trajectory samples.

C. Latent Dynamics

We consider task examples that can be temporally factored \( x = x_{0:T} \). This motivates to embed a latent dynamics constraint over \( z = z_{0:T} \) if \( z_t \) is taken as a belief representation of the observation history \( x_{0:t} \). Following the case of bridging modalities in (5), we capture the temporal relation by matching its estimation of the posterior and prior latent dynamics:

\[
q(z_t|z_{t-1}, x_t) = p_0(z_t|z_{t-1}) \Rightarrow \text{KL}(q(z_t)p_0)
\]

Fig. 3: Embedding a dynamical relation with \( z_t \) as an abstract representation of frames \( x_{0:t} \) (upper). The retained feature transformation and dynamics are exploited to learn a model-based control in the latent space (lower left): after accessing initial frames, a robot arm anticipates the rolling ball movement and score a goal without visual input (lower right).

IV. FUTURE PLAN

We envision future research steps of utilizing domain constraints in different aspects, including learning, control and data modality. To begin with, a linear latent dynamics may allow to analyze and introduce a stability regulation criterion, with the potential of characterizing more robust goal-directed behaviors. Secondly, we plan to explore auxiliary tasks, as well as associated learned representation and constraints like III-B and III-C to examine their impacts on the performance of addressing general motor control such as non-prehensile manipulation. Last but not the least, it would be interesting to investigate how rich data of other modalities, such as images or texts, can be leveraged by robots to reason about entity relations as well as the external world, so as to develop motor task skills from limited experience.
REFERENCES