Data-Driven Imitation Learning of Human Motion Style for Robotic Motion Control

Guoyao Zhang, Zhongcheng Lei, Wenshan Hu, Hong Zhou
Department of Artificial Intelligence and Automation, Wuhan University, Wuhan 430072, P. R. China
E-mail: zhongcheng.lei@whu.edu.cn

1 Introduction

Humans rely on imitation as a fundamental learning mechanism throughout their development [1]. With the advancement of robotics and artificial intelligence, enabling robots to acquire imitation learning capabilities similar to those of humans has become a major challenge in the field of intelligent robotics [2].

Traditional control approaches for robots, which are typically based on precise mathematical models and predetermined motion planning strategies, have historically been successful in achieving efficient locomotion for quadrupedal, bipedal, and humanoid robots [3]. However, these methods often depend on highly accurate environmental models, which poses considerable challenges in terms of robustness and generalization.

Imitation learning offers robots a more natural and flexible pathway for acquiring skills. By imitating human demonstrations, robots can learn complex motor skills [4]. However, most of the existing work is focused on the field of computer graphics[5], particularly in character animation, and often lacks consideration of real-world physical constraints. As a result, when these methods are transferred to real robotic platforms, there is a significant gap between simulation and reality.

To address the challenges outlined above, this paper proposes a skill reuse framework based on generative adversarial imitation learning (GAIL) to tackle the challenge of enabling robots. Through this approach, robots can extract reusable skills that comply with their own physical limitations from large-scale, unstructured motion data, without the need for complex manual annotations or editing.

2 Problem Formulation

In the skill reuse framework based on GAIL, we formulate the robot motion control problem as a Markov Decision Process (MDP). The robot interacts with the environment according to a control policy π to optimize a given objective function, thereby training an appropriate motion strategy.

At each time step t, the robot observes the system state s_t, then samples and selects an action a_t from the control policy π. In this context, the state s represents the observation space, which includes the position and velocity of each joint of the robot. The action a corresponds to the action space, which consists of the target positions for the proportional derivative (PD) controllers of each joint. The PD controllers compute the torque for each joint's motor based on the specified PD gains. After executing the action, the environment transitions the robot to the next state s_{t+1} based on the state transition function p, and a reward r_t is received. The robot's goal is to maximize the expected cumulative reward R, i.e., to maximize the objective function $J(\pi)$, as shown in (1).

$$J(\pi) = \mathbb{E}_{p(\tau|\pi)} \left[\sum_{t=0}^{T-1} \gamma^t r_t \right] \tag{1}$$

where $p(\tau|\pi) = p(s_0) \prod_{t=0}^{T-1} p(s_{t+1}|s_t, a_t)\pi(a_t|s_t)$ represents the likelihood of the trajectory distribution $\tau = \{s_0, a_0, r_0, s_1, \ldots, s_{T-1}, a_{T-1}, r_{T-1}, s_T\}$ under the control policy π. $p(s_0)$ denotes the initial state distribution. T is the maximum time step, indicating the time horizon of the trajectory, and $\gamma \in [0, 1]$ is the discount factor, used to balance the weighting between immediate and future rewards.

3 Control Method

In this section, we describe the two main components of the skill reuse framework based on GAIL: motion retargeting and control strategy.

3.1 Motion Retargeting

Motion retargeting refers to the process of mapping a human motion dataset onto another human model. The human motion data used in this paper comes from two main sources: one is motion data collected through motion capture suits, and the other is from open-source character animation datasets available on the internet. Due to the difference in the number of joints between human motion data and those of a robot, it is necessary to retarget the human motion data to the data of the robot's skeletal architecture. The fundamental principle of motion retargeting involves matching the skeletal structures between the source and target models. Subsequently, optimization techniques are employed to map the motion data from the source model onto the target model.

3.2 Control Strategy

The skill reuse framework based on GAIL consists of two parts: low-level skill training and high-level task training. The low-level skill strategy focuses on teaching the robot various atomic motion skills and movement styles through imitation learning. This strategy does not involve the implementation of specific task objectives. In contrast, the high-level task policy uses reinforcement learning to design the specific task as a reward function, invoking actions from the low-level skill policy based on the task goals, thereby efficiently completing the predetermined task.

By combining the goal of motion imitation and the unsupervised skill discovery objective, the low-level skill policy encourages the development of a unique set of skill fea-

tures that generate behaviors with a style similar to that of the given dataset. The reward function at each time step for the low-level skill policy is shown in (2). This reward function consists of two parts: the first part is based on the GAN objective, which primarily encourages the policy to generate motion styles similar to those in the dataset; the second part is the objective for skill discovery, which encourages the low-level skill policy to produce different behaviors for different latent variables z, ensuring that the skills are reusable.

$$r_t = -\log\left(1 - \mathcal{D}(s_t, s_{t+1})\right) + \beta \log q(z_t \mid s_t, s_{t+1}) \quad (2)$$

where $\mathcal{D}(s_t, s_{t+1})$ denotes the output of the discriminator network, indicating the robot's movement quality. and $q(z|s, s')$ is the encoder, β is the weight for the encoder term. The goal of skill discovery is to ensure that each latent variable z produces different actions, so that the encoder can recover the specific z responsible for generating a particular action.

The high-level task policy $w(z|s, g)$ is trained based on the low-level skill policy, conditioned on a specific task goal g. It outputs the latent variable z to leverage the low-level policy's actions to complete the task.

The reward function for the high-level task policy is formulated by combining the task reward function with the movement style reward function from the discriminator, as shown in (3). This combined reward structure effectively guides the high-level task policy to generate smooth and natural movements, thereby enhancing the overall performance and adaptability of the robotic system across various tasks.

$$r_t = w_G r^G(s_t, a_t, s_{t+1}, g) - w_S \log\left(1 - \mathcal{D}(s_t, s_{t+1})\right) \quad (3)$$

In this formulation(3), w_G and w_S are manually specified weight coefficients, $r^G(s, a, s', g)$ represents the task reward function, The discriminator $\mathcal{D}(s, s')$ from the low-level skill policy as an evaluator of the robot's movement quality, encouraging the high-level task policy to produce latent variable sequences Z that exhibit minimal variation between consecutive time steps. This approach ensures that the robot's actions remain continuous during task execution.

The control strategy is analogous to the human nervous system, where the high-level task policy functions like the brain, responsible for task planning and scheduling, while the low-level skill policy resembles the cerebellum, responsible for executing specific actions. Through this approach, the robot can more efficiently learn and imitate human motion skills and movement styles, enabling it to autonomously select appropriate actions and successfully complete tasks when facing complex situations.

4 Results

To evaluate the effectiveness of the GAIL framework in acquiring human motion skills from extensive unstructured human motion data and executing tasks with human-like motion styles, we conducted tests on robot fall recovery from arbitrary states. In the fall recovery test, the robot starts by free-falling from a height of 10 meters, and we observe whether it can stand up with human-like motion styles.

The experimental results demonstrate that the robot is able to leverage its acquired motion skills to adapt to environmental changes in real-time and perform complex goal-oriented movements. It is worth noting that the dataset used as a reference does not include instances of the robot falling from a high place and recovering to a standing posture. Instead, the robot autonomously learns human-like motion styles and realizes seamless transitions between different states. This indicates that the robot is not only capable of effectively imitating human motion skills but can also autonomously execute advanced tasks, demonstrating a high degree of adaptability and flexibility.

5 Conclusion

This paper constructs a skill reuse framework based on GAIL, providing an efficient and flexible solution for imitating human motion skills and styles. The framework is a highly scalable data derived method that enables the learning of reusable motion skills from large-scale unstructured human motion data. The framework does not require complex manual annotation or data clipping. However, the current methods still face challenges such as mode collapse and insufficient policy robustness, which limit their application in more diverse and dynamic real-world scenarios.

Future research will focus on the following directions: further optimizing the policy learning paradigm to alleviate mode collapse, improving the diversity and stability of learned policies; and integrating multimodal data fusion with advanced control theory to advance the framework's application in both theoretical research and engineering practice.

References

[1] Emma Flynn and Andrew Whiten. Dissecting childrens observational learning of complex actions through selective video displays. *Journal of Experimental Child Psychology*, 116(2):247–263, 2013.

[2] Kourosh Darvish, Luigi Penco, Joao Ramos, Rafael Cisneros, et al. Teleoperation of humanoid robots: A survey. *IEEE Transactions on Robotics*, 39(3):1706–1727, 2023.

[3] Nicola Scianca, Daniele De Simone, Leonardo Lanari, and Giuseppe Oriolo. Mpc for humanoid gait generation: Stability and feasibility. *IEEE Transactions on Robotics*, 36(4):1171–1188, 2020.

[4] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. Deepmimic: example-guided deep reinforcement learning of physics-based character skills. *ACM Trans. Graph.*, 37(4), July 2018.

[5] Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. Ase: Large-scale reusable adversarial skill embeddings for physically simulated characters. *ACM Transactions On Graphics (TOG)*, 41(4):1–17, 2022.