Reinforcement Learning's Role in Human Wearable Exoskeletons: Upper and Lower Limb Applications

Varshitha Janavi*, Pranay Palem† and Puneet Sai Naru‡
Emails: vpurra@asu.edu*, ppalem1@asu.edu†, pnaru@asu.edu‡

***Abstract*—This paper explores the application of Reinforcement Learning (RL) in wearable exoskeletons, focusing on both the upper and lower limb systems. Traditional control methods, which are based on predefined models, often struggle with adaptability and real-time adjustments for individual users. RL, especially Deep Reinforcement Learning (DRL), overcomes these challenges by continuously learning user feedback and optimizing control strategies in real-time. This enables personalized assistance, reduced energy consumption, and improved stability during movement. The paper also presents mathematical models, such as the Twin Delayed Deep Deterministic Policy Gradient (TD3), that allow exoskeletons to adapt dynamically to muscle activation or gait patterns, offering a more flexible and efficient solution compared to conventional control systems.**

***Index Terms*—Reinforcement Learning (RL), Wearable Exoskeletons, Adaptive Control, Deep Reinforcement Learning (DRL), Twin Delay Deep deterministic Policy Gradient (TD3), Gait Assistance, Personalized Assistance, Upper-Limb Exoskeleton, Lower-Limb Exoskeleton.**

I. Introduction

Wearable exoskeletons are robotic devices that augment human movement by enhancing strength, endurance, and motor control. They are increasingly used in fields such as rehabilitation, assistive technology for individuals with mobility impairments, and industrial applications to reduce worker fatigue and injury. These devices are either active, providing powered assistance through actuators, or passive, augmenting movements using mechanical supports. A critical challenge in their design is to develop control systems that allow the exoskeleton to adapt to the body movements of the wearer seamlessly.

Traditional control methods, such as model-based and rule-based approaches, rely on predefined control policies, which are often limited in their adaptability to real-time, user-specific requirements. Reinforcement Learning (RL), particularly Deep Reinforcement Learning (DRL), has emerged as a promising alternative due to its ability to learn and adapt policies through interaction with the environment, enabling continuous improvement based on feedback from the system's performance. In the context of wearable exoskeletons, RL systems optimize control policies by receiving sensor data (e.g., electromyography (EMG), inertial measurement units (IMUs), or joint angles) and adjusting actuator outputs to optimize outcomes such as torque assistance, stability, or user comfort (Oghogho et al. 2022).

RL systems in exoskeletons can be broadly categorized into two main applications:

- **Upper Limb Exoskeletons**: These devices assist users with tasks involving arm, shoulder, or wrist movement, often used in rehabilitation for patients recovering from stroke or spinal cord injuries. RL allows these systems to personalize torque assistance and adapt to changes in the user's muscle activation patterns, improving the effectiveness of rehabilitation and reducing muscle fatigue during use. Recent work in this domain includes the application of DRL for EMG-based control, where RL algorithms adjust the level of assistance based on muscle activation, as demonstrated by (Oghogho et al. 2022).
- **Lower Limb Exoskeletons**: These systems are used to assist walking or standing, often for users with mobility impairments due to neurological conditions or injuries. RL helps optimize parameters such as gait stability, stride length, and balance, providing personalized control over time. For instance, (Sharifi et al. 2022) employed Proximal Policy Optimization (PPO), a popular RL algorithm, to enhance dynamic stability and adapt to real-world environments by learning from user-specific gait data.

II. What Reinforcement Learning Does in Wearable Exoskeletons

Reinforcement Learning (RL) plays a crucial role in enhancing the functionality and adaptability of both upper-limb and lower-limb wearable exoskeletons. These robotic devices are designed to augment human movement, assisting in various applications ranging from rehabilitation to industrial use.

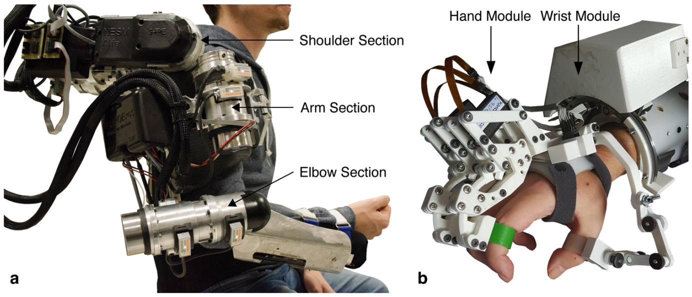

Fig. 1. An example of an upper-limb exoskeleton system

RL algorithms, such as Twin Delayed Deep Deterministic Policy Gradient (TD3), enable personalized torque adjustments based on real-time electromyography (EMG) signals for upper-limb exoskeletons. This allows the exoskeleton to dynamically adapt to the user's muscle activity, minimizing unnecessary effort and optimizing movement precision.

In lower-limb exoskeletons, RL is employed to optimize gait patterns and adapt to different walking conditions. The intelligent central pattern generator (iCPG) approach integrates RL with adaptable central pattern generators (ACPGs) to modify walking trajectories based on human-robot interaction (HRI) energy. This system can adapt to changes in user interaction, addressing the limitations of pre-defined walking trajectories commonly found in commercially available exoskeletons.

RL algorithms in both types of exoskeletons work towards:

- Personalizing assistance levels
- Optimizing energy consumption
- Improving stability and movement accuracy
- Adapting to real-time changes in user needs or environmental conditions

III. How Reinforcement Learning Works in Wearable Exoskeletons

Now we will explain how reinforcement learning (RL) works in the context of wearable upper exoskeleton control, based on the implementation described in the referenced paper (Oghogho et al. 2022). The authors employ the Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm to control an upper-limb exoskeleton using electromyographic (EMG) signals intelligently from the wearer's muscles.

A. Actor-Critic Architecture

The TD3 algorithm utilizes an actor-critic structure:

- The actor network maps states (joint angles and EMG signals) to actions (assistive gains for actuators).
- Two critic networks estimate the reward values for each state-action pair.

B. Q-Function Approximation

The critics learn to estimate the discounted sum of expected future rewards:

$$Q_\theta(s,a) = \mathbb{E}_{s_i \sim p_\pi, a_i \sim \pi}\left[\sum_{i=t}^{T} \gamma^{i-t}\left(r_i - \delta_i\right)\right] \tag{1}$$

This equation represents the Q-function, which estimates the expected cumulative discounted reward. Here, γ is the discount factor, r_i is the reward at time i, and δ_i is the temporal difference error.

C. Action Selection and Exploration

Actions are selected using the current policy with added exploration noise:

$$a \sim \pi(s) + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma) \tag{2}$$

This equation describes how actions are selected, combining the policy $\pi(s)$ with Gaussian noise ϵ for exploration.

D. Target Policy Smoothing

To prevent extreme action values, target policy smoothing is applied:

$$\tilde{a} \leftarrow \pi_{\phi'}(s) + \epsilon, \epsilon \sim \operatorname{clip}(\mathcal{N}(0, \tilde{\sigma}) - c, c) \tag{3}$$

This formula adds clipped noise to the target policy to prevent extreme action values.

E. Critic Optimization

The critics are optimized by minimizing the mean squared error between predicted and target Q-values:

$$\theta_i \leftarrow \min_{\theta_i} N^{-1} \sum \left(y - Q_{\theta_i}(s,a)\right)^2 \tag{4}$$

This equation shows how the critic networks are optimized by minimizing the mean squared error between predicted and target Q-values.

F. Actor Optimization

The actor policy is updated by maximizing the estimated Q-value:

$$\nabla_\phi J(\phi) = N^{-1} \sum \nabla_a Q_{\theta_1}(s,a)\Big|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s) \tag{5}$$

This formula represents how the actor policy is updated to maximize the estimated Q-value.

G. Target Network Updates

Target networks are updated using soft updates to improve stability:

$$\theta_i' \leftarrow \tau\theta_i + (1-\tau)\theta_i' \tag{6}$$

$$\phi' \leftarrow \tau\phi + (1-\tau)\phi' \tag{7}$$

These equations show how target networks are gradually updated to improve learning stability.

H. Reward Shaping for Assistive Control

A reward function is designed to identify optimal assistive gains for each joint:

$$\mathcal{R} = -\sum \left(\mu \left|e_i\right|^2 + \lambda EMG_{n_i}^4 + \mathcal{R}_p\right) \tag{8}$$

This function balances minimizing muscle effort (EMG activity) and trajectory accuracy (position overshoot).

I. Assistive Torque Generation

The learned gains generate assistive torques proportional to normalized EMG activities:

$$T_{a_i} = K_{a_i} EMG_{n_i} \tag{9}$$

This equation shows how assistive torque is generated, proportional to the normalized EMG activity and the learned gain K_{a_i}.

By implementing this TD3-based approach, the exoskeleton can learn personalized assistive gains for each user, optimizing the trade-off between muscle effort and motion accuracy in real-time.

IV. Why use Reinforcement Learning in Wearable Exoskeletons

Exoskeletons face multiple challenges, especially in ensuring safety, comfort, and efficiency. These challenges are heightened by the need for real-time adaptability. Let us discuss how reinforcement learning can help drastically overcome the challenges faced while using methods like ANN (Oghogho et al. 2022), DDPG (Mehr et al. 2023).

A. *Comparision to other algorithms*

In exoskeletons, it is crucial to ensure smooth and adaptive gait assistance is crucial. Traditional systems often rely on predefined gait patterns that do not account for individual user variations or changes in walking conditions. Although algorithms such as deep deterministic policy gradient (DDPG) have been used to control robots with 6 DOF, DDPG has been found to overestimate future reward and lead to non-optimal convergence (Mehr et al. 2023). Central pattern generators (CPG) have been preferred because of their ability to generate synchronized motions between different joints. But CPG has a fixed control strategy and defined trajectories which is undesirable when the exoskeleton is used on different people. Other studies have used adaptive CPG (ACPG) to address this issue. ACPG can be used to personalize the gait and control trajectories if the algorithm is initialized precisely and the user's behavior does not change over time. These requirements are considered as limitations in using ACPG's. Reinforcement learning, particularly the Twin delay deep deterministic policy gradient (TD3), has been applied along with CPG to solve this issue by learning optimal gait trajectories based on user-specific interactions (Sharifi et al. 2022).

B. *Results*

For upper-limb exoskeletons, precise control is required to assist with fine motor tasks. The exoskeleton is designed to assist individuals with weak muscles by providing additional support. Thus, it is essential for the skeleton to adapt to weak responses from the user. This can be quantised using the human robot interaction (HRI) energy. We have noticed how TD3 algorithm can be used to control the effective HRI energy $E_{\text{eff}}(t)$, to achieve the desired motion trajectory via iCPG. In Figure 2, we can notice that the RL agent amplified the interaction energy, $E(t)$ (brown dashed line) and suggested a higher effective energy value $E_{\text{eff}}(t)$ (solid blue line). This enables users with weak input responses to benefit optimally from using the exoskeleton(Mehr et al. 2023). Thus, RL minimizes user effort while maximizing the exoskeleton's support (Oghogho et al. 2022).

V. Future Work

While we explored multiple control algorithms which generate personalized motion trajectories in an exoskeleton, we need to further explore various hybrid RL methods and compare the learning time, adaptability to change in gait over time and complexity in initialization of parameters in each approach. We need to read literature to understand if there is any existing solution to mathematically defining complementing joint control to make the the RL approach easier. A further analysis needs to be conducted to take into account various real-time limitations with the physical components of the exoskeleton.

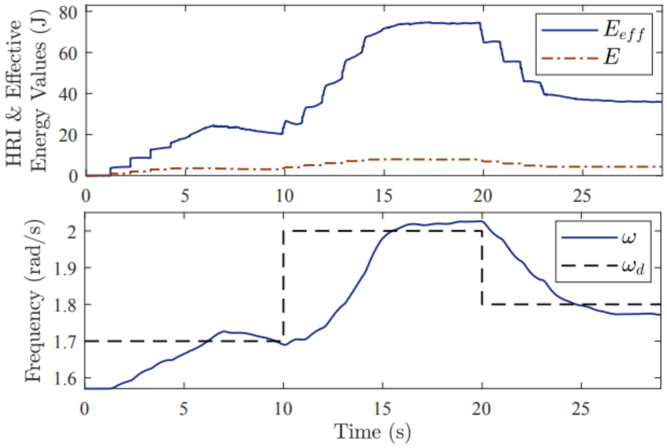

Fig. 2. Variation of HRI energy due to an RL agent selecting optimal energy for a user with weak muscles

VI. Conclusion

Reinforcement learning (RL) is revolutionizing the control systems of wearable exoskeletons by offering more adaptive, personalized, and efficient assistance. Algorithms such as the Twin Delayed Deep Deterministic Policy Gradient (TD3) and Proximal Policy Optimization (PPO) have shown significant improvements in both upper and lower limb exoskeleton performance. These advancements include enhanced stability, optimized HRI energy output, and reduced muscle effort, providing users with a more comfortable and effective experience. RL has proved to significantly reduced effort in initializing the exoskeleton as required by algorithms such as CPG. This enables faster deployment and adaptability when there is a significant change in gait/usage of the exoskeleton. Additional developments will aim to further integrate biofeedback mechanisms, such as electromyography (EMG) and motion tracking, into RL systems to improve real-time responsiveness and adaptability, creating a user-centered exoskeletons.

VII. Links

Here is a link to the presentation: Presentation Link

References

Reference	Description
Martin Oghogho et al. (2022). “Deep Reinforcement Learning for EMG-based Control of Assistance Level in Upper-limb Exoskeletons”. In: *2022 International Symposium on Medical Robotics (ISMR)*, pp. 1–7. DOI: 10.1109/ISMR48347.2022.9807562	This paper contributes to the discussion on TD3 for EMG-based upper-limb control, providing insights for the **Reinforcement Learning for Upper-Limb Exoskeletons** and **Experimental Results** sections.
Javad K. Mehr et al. (2023). “Deep Reinforcement Learning based Personalized Locomotion Planning for Lower-Limb Exoskeletons”. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5127–5133. DOI: 10.1109/ICRA48891.2023.10161559	Contributed to the mathematical formulations and DRL applications for lower-limb exoskeletons in the **Mathematical Models** and **Reinforcement Learning for Lower-Limb Exoskeletons** sections.
Mojtaba Sharifi et al. (2022). “Proximal Policy Optimization for Enhancing Dynamic Stability in Lower-limb Exoskeletons”. In: *IEEE Robotics and Automation Letters*	Sharifi et al. provide details on PPO for lower-limb exoskeletons, used in the **Reinforcement Learning for Lower-Limb Exoskeletons**, **Mathematical Models**, and **Experimental Results** sections.
Eddie Guo, Hyunseok Lee, and Mojtaba Akbari (2023). “Integrating Bioelectronics and Reinforcement Learning for Exoskeleton Control”. In: *2023 IEEE International Conference on Robotics and Automation (ICRA)*	This paper discusses the integration of bioelectronics with RL, which was referenced in the **Future Work** section of the report.