2020 IEEE International Conference on Robotics and Automation (ICRA)
31 May - 31 August, 2020. Paris, France

Data-Driven Reinforcement Learning for Walking Assistance Control of a Lower Limb Exoskeleton with Hemiplegic Patients

Zhinan Peng[1†], Rui Luo[1†], Rui Huang[1], Jiangping Hu[1], *Senior Member, IEEE*, Kecheng Shi[1], Hong Cheng[1], *Senior Member, IEEE*, Bijoy Kumar Ghosh[1,2], *Fellow, IEEE*

***Abstract*— Lower limb exoskeleton (LLE) has received considerable interests in strength augmentation, rehabilitation and walking assistance scenarios. For walking assistance, the LLE is expected to have the capability of controlling the affected leg to track the unaffected leg's motion naturally. An important issue in this scenario is that the exoskeleton system needs to deal with unpredictable disturbance from the patient, which requires the controller of exoskeleton system to have the ability to adapt to different wearers. This paper proposes a novel Data-Driven Reinforcement Learning (DDRL) control strategy to adapt different hemiplegic patients with unpredictable disturbances. In the proposed DDRL strategy, the interaction between two lower limbs of LLE and the legs of hemiplegic patient are modeled in the context of leader-follower framework. The walking assistance control problem is transformed into a optimal control problem. Then, a policy iteration (PI) algorithm is introduced to learn optimal controller. To achieve online adaptation control for different patients, based on PI algorithm, an Actor-Critic Neural Network (ACNN) technology of the reinforcement learning (RL) is employed in the proposed DDRL. We conduct experiments both on a simulation environment and a real LLE system. Experimental results demonstrate that the proposed control strategy has strong robustness against disturbances and adaptability to different pilots.**

***Index Terms* - Data-driven Control, Reinforcement Learning, Leader-Follower Multi-Agent System, Lower Limb Exoskeleton, Hemiplegic Patients, Actor-Critic Neural Network.**

I. INTRODUCTION

Lower limb exoskeleton (LLE) has become an interesting topic thanks its wide applications in both strength augmentation [1], [2], [3], [4], walking assistance [5], [6] and rehabilitation [7], [8]. In the fields of rehabilitation and walking assistance, most exoskeletons are developed for assisting paraplegic patients whose lower body are both disabled [9], [10]. It should be noted that stroke has gradually become a global health-care problem, some exoskeleton researchers have focused on walking assistance or rehabilitation case for hemiplegic individuals [11], [12], [13], [14].

In the early studies on rehabilitation and gait recovery of hemiplegia, researcher proposed Ankle-Foot Orthosis (AFO) to achieve good recovery performance [15], [16]. In order to provide active power assistance for hemiplegic patients, many powered orthosis with active motors have been developed, such as active AFO developed by Blaya [17] and Series Elastic Remote Knee Actuator (SERKA) developed by Sulzer [18]. However, these kinds of orthosis are designed for repairing local motion function of hemiplegic patients in particular scenarios, such as the SERKA is design for stroke patient with stiff-knee gait (SKG).

With the development of exoskeleton technology, exoskeleton has gained more interests in both rehabilitation and walking assistance for hemiplegic patients [19], [20]. Y. Sankai developed a single leg exoskeleton system for hemiplegic patients based on the Hybrid Assistive Limb (HAL) [21]. For the studies on the HAL system with single leg case, motion information of the unaffected side is generated to synchronize gait of the affected side [22]. Note that the single leg based HAL system should be redesigned as the wearer has different disabled side. In [23], a powered exoskeleton was used to improve patients with hemiparesis walking function via robot assisted gait training. In [24], the authors proposed a control approach of a LLE to provide walking assistance, without giving desired joint angle trajectory, for facilitating recovery. Recently, a learning-based control strategy was proposed for a LLE with hemiplegia [25]. In fact, disturbances caused by system or external environment will affect the control performance of system, which should be considered in controller designs. Moreover, most existing control methods rely on accurate system model (*model-based* methods), which is not true in practical systems. Therefore, the motivation of this paper aims to address these problems.

This paper presents a Data-Driven Reinforcement Learning (DDRL) Control algorithm for walking assistance of lower exoskeleton with hemiplegic patients. First, the interaction relations between the both two low limbs of LLE and hemiplegic patient are formulated as a *Leader-Follower Multi-Agent System* (LFMAS) framework. A Policy Iteration (PI) algorithm is utilized to obtain optimal controller. Further, in order to improve adaptive performance for walking assistance with different hemiplegic patients, a Reinforcement Learning (RL) method, namely Actor-Critic Neural Network (ACNN), is proposed to achieve better control performance, where the learning process only depends on measurement motion data from the LLE system. The main contributions of this paper can be summarized as follows:

1) A DDRL control strategy based on PI algorithm is designed to learn the optimal walking assistance controller. The proposed method is provided in a model-free manner without using accurate dynamics model of the exoskeleton system and system identification.

†Both authors contribute equally to this work.

[1]Z. Peng, R. Luo, R. Huang, J. Hu, H. Cheng, K. Shi and B. K. Ghosh are with the Center for Robotics, School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China.

[2]B. K. Ghosh is with the Department of Mathematics and Statistics, Texas Tech University, Lubbock, USA.

2) To improve the adaptability of exoskeleton's controller, an online-learning based ACNN structure is employed in the controller design, which is aim at performing adaptability performance for different patients and achieving good robust against disturbances.

Moreover, the proposed DDRL method is tested on a two degree-of-freedom (2-DOF) simulation environment, and then it is successfully applied on a real LLE system with healthy subjects who simulate paraplegia. Simulation case and experimental results illustrate that the proposed control strategy has a good robustness performance against disturbances and have adaptive ability for different wearers or even the same wearer with different gait patterns.

In the following, Section II first introduces the modeling process of LFMAS for exoskeleton system with hemiplegic patients, and the details of the dynamics of the exoskeleton is given. Then, Section III presents the PI based control strategy. Section IV proposes the data-driven adaptive control algorithm via reinforcement learning framework based on PI method. In Section V, we will illustrate our method in simulation situation and apply it to an actual exoskeleton system with healthy people who simulate hemiplegic patients. This paper ends with conclusion and future work in Section VI.

II. Modelling and Formulation for Walking Assistance Control Problem

In this section, the modelling process for the LLE with hemiplegic patients, namely LFMAS, is given to describe the information interaction relations among both lower limbs of LLE and patients' legs. Then, an information exchange scheme is introduced for the LFMAS.

A. Modeling Exoskeleton System as LFMAS

In this paper, we aim at designing adaptive controller of a LLE system with both lower extremities to asist hemiplegic individual walking. For hemiplegic patients, it should be noted that one of the two legs usually loses walking ability and the other one is normal. Before giving the controller designs, it is important that how to model the interaction relation among them appropriately such that the both low limbs and the LLE can achieve their mutual communication.

Inspired by the area of cooperative distributed control, *leader-follower* mechanism has been successfully applied on multi-agent systems [26], which allows information interaction among agents in a distributed way. In this paper, the unaffected leg of the hemiplegic patients and the both lower extremities of exoskeleton are modeled as a LFMAS. Fig. 1 (a) gives the structure of the LFMAS for exoskeleton system with hemiplegic individuals, where the exoskeleton with hemiplegia is divided into three components: one Leader and two Follower agents. That is, the unaffected leg of patient is treated as the Leader of LFMAS, along with an Inertial Measurement Unit (IMU) sensors which is used to measure its joints' states. Furthermore, both two lower extremities of the LLE system are defined as two Follower agents, which can be described as follows:

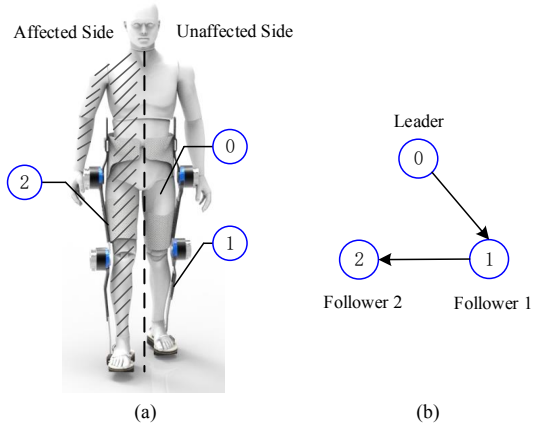


Fig. 1. (a) The framework diagram of the LFMAS modelling (The notations 0, 1, and 2 denote Leader, Follower 1, and Follower 2, respectively); (b) Communication topology network of the LFMAS structure.

1) *Follower 1* is the exoskeleton leg of unaffected side, which synchronizes the leader agent's (unaffected side of patient's leg) motion immediately.
2) *Follower 2* is the other side of exoskeleton's limb with the disabled leg of patient, the patient's disabled leg is tightly connected with the exoskeleton.

In the framework of LFMAS, considering that there is a phase difference between the motion of the affected side and the unaffected side, naturally, Therefore, Follower 2 is expected to track the Leader's trajectory motion after half gait cycle interval.

To achieve better assistance control performance, the information interaction scheme should be designed for both lower extremities and patient's legs, which allows them transmit their information (LLE's state and control signal) with their neighbors. Thus, we introduce an information exchange rule to describing the evolution of the agents' communication.

1) Information Evolution Rule: The information update for Follower agent i ($i = 1,2$) includes combining its own information with those received from its neighbors, and Leader can transmit its information to Follower. Assume that each agent has a weight vector $a_i = [a_{ij}]$, in which each element a_{ij} represents that agent i assigns to the information obtained from a neighboring agent j. Fig. 1 (b) denotes the communication topology network between agents where arrows indicate the direction of information flow.

2) Weight Rule: Let $\mathcal{N}(i)$ be the neighbors set of the i^{th} Follower agent. For arbitrary $i \in \{1,2\}$, if $j \in \mathcal{N}(i)$, $a_{ij} > 0$; if $j \notin \mathcal{N}(i)$, $a_{ij} = 0$. Let $\sum_{j\in\mathcal{N}(i)} a_{ij} = d_i$ be the sum of the neighbors' weights for agent i.

B. Dynamic Model and Control Objective

In this paper, the dynamics of the LLE system is described as a general nonlinear mechanical system (i.e., Euler-lagrange system). Therefore, the dynamics of the both lower extremities, i.e., Follower 1 ($i = 1$) and Follower 2 ($i = 2$) of the exoskeleton are described as follows:

$$M_i(\theta_i)\ddot{\theta}_i + C_i(\theta_i,\dot{\theta}_i)\dot{\theta}_i + G_i(\theta_i) = \tau_i, \quad i = 1,2 \qquad (1)$$

where $\theta_i = (\theta_{ih}, \theta_{ik})^\top \in R^2$ denotes the joints' angle of the LLE, θ_{ih} and θ_{ik} represent the hip joint and knee joint, respectively. $M_i(\theta_i)$ denotes inertia matrix, $C_i(\theta_i, \dot{\theta}_i)$ represents the centripetal and coriolis matrix. $G_i(\theta_i)$ denotes the the gravitation term, $\tau_{ie} = (\tau_{iu}, \tau_{id})^\top$ are the input torques generated by up and down motors for hip and knee joint. Further, we can rewrite Eq. (1) as a state-space form:

$$\begin{bmatrix} \dot{\theta}_i \\ \ddot{\theta}_i \end{bmatrix} = \begin{bmatrix} \dot{\theta}_i \\ -M_i{}^{-1}(C_i\dot{\theta}_i + G_i) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ M_i{}^{-1} \end{bmatrix} \tau_i$$

or equivalently,

$$\dot{q}_i(t) = f_i(q_i(t)) + g_i(q_i(t))u_i, \tag{2}$$

where $q_i(t) = [\theta_i^\top, \dot{\theta}_i^\top]^\top$, $g_i(q_i(t)) = \begin{bmatrix} \mathbf{0} \\ M_i{}^{-1} \end{bmatrix}$, $f_i(q_i(t)) = \begin{bmatrix} \mathbf{0} & I \\ \mathbf{0} & -M_i{}^{-1}C_i \end{bmatrix} q_i(t) + \begin{bmatrix} \mathbf{0} \\ -M_i{}^{-1}G_i \end{bmatrix}$, $\tau_i = u_i$.

Furthermore, the dynamic of the Leader (the motion trajectory of patient's unaffected leg) is expressed by:

$$\dot{q}_r(t) = f(q_r(t)), \tag{3}$$

where $q_r(t)$ indicates the joint angle collected from human via an IMU sensors matched on the pilot's leg.

Design Objective: The goal is to generate the controller strategy u_i to ensure the trajectory $q_i(t)$ generated by Eq. (2) can track the trajectory $q_r(t)$ in Eq. (3). That is, it is desired to make the following tracking error index go to zero:

$$\lim_{t\to\infty} \|q_i(t) - q_r(t)\| = 0. \tag{4}$$

In order to achieve control objective, the local neighbor tracking errors of dynamics (2) for follower i are defined as

$$\varepsilon_i(t) = \sum\nolimits_{j\in\mathcal{N}(i)} a_{ij}(q_i(t) - q_j(t)) + c_i(q_i(t) - q_r(t)), \tag{5}$$

where $\mathcal{N}(i)$ and a_{ij} have been defined in II-A. $c_i > 0$ denotes the pinning gain, which means agent i can obtain the Leader's information.

Taking the derivation of Eq. (5), combining Eq. (2) and Eq. (3), the dynamics of the tracking errors are written as

$$\dot{\varepsilon}_i(t) = f_{\varepsilon_i} + (d_i + c_i)g_i(t)u_i(t) - \sum\nolimits_{j\in\mathcal{N}(i)} a_{ij}g_j(t)u_j(t), \tag{6}$$

where $f_{\varepsilon_i}(t) = \sum_{j\in\mathcal{N}(i)} a_{ij}(f_i - f_j) + c_i(f_i - f)$, d_i indicates the sum of the weights of the i^{th} follower agent.

III. Policy Iteration based Walking Assistance Controller

Based on the system modelling and problem formulation, in this section, the state-of-the-art algorithm called policy iteration (PI) will be proposed to obtain the optimal controller $u_i^*(t)$ for walking assistance problem.

From the perspective of RL control [27], [28] and inspired by RL methods [30], [29], [31], we use a Q-function (cost function) to evaluate the long-term learning performance, which is defined as follows:

$$Q_i(\varepsilon_i(t)) = \int_t^\infty r_i\big(\varepsilon_i(s), u_i(s), u_{(j)}(s)\big)ds, \tag{7}$$

where $u_{(j)}(t)$ denotes the neighbors' control of Follower agent i, and $r_i\big(\varepsilon_i(t), u_i(t), u_{(j)}(t)\big) = \varepsilon_i^\top(t)Q_{ii}\varepsilon_i(t) + u_i^\top(t)R_{ii}u_i(t) + \sum_{j\in\mathcal{N}(i)} u_i^\top(t)S_{ij}u_{(j)}(t)$ is the reward function, where the $Q_{ii} > 0$, $R_{ii} > 0$ and $S_{ij} > 0$ are symmetric positive definite weighting matrices, respectively. For the notation simplification, we set $r_i(\varepsilon_i(t), u_i(t), u_{(j)}(t)) = r_i(\varepsilon_i(t), u_i(t))$.

In this paper, the walking assistance control problem is transformed into an optimal control problem, which aims at designing a distributed controller to guarantee the *Design Objective* as well as minimizing the local cost function Eq. (7).

Further, the Hamilton function is represented as

$$H_i\big(\varepsilon_i(t), u_i(t), Q_i(\varepsilon_i(t))\big) = r_i(\varepsilon_i(t), u_i(t)) + \nabla Q_\varepsilon^\top \dot{\varepsilon}_i(t), \tag{8}$$

where $Q_i(0) = 0$, $\nabla Q_\varepsilon = \partial Q_i(\varepsilon_i(t))/\partial\varepsilon_i(t)$ is a partial differential part.

Using the stationary condition for Eq. (8), i.e., let $\partial H_i/\partial u_i = 0$, the optimal controller $u_i^*(t)$ is obtained as

$$u_i^*(t) = -\frac{1}{2}(d_i + c_i)R_{ii}^{-1}g_i^\top(t)\nabla Q_\varepsilon. \tag{9}$$

The optimal cost function $Q_i^*(\varepsilon_i(t))$ satisfies the following coupled Hamilton-Jacobi-Bellman (HJB) equation:

$$H_i\big(\varepsilon_i(t), u_i^*(t), Q_i^*(\varepsilon(t))\big) = r_i(\varepsilon_i(t), u_i^*(t)) + \nabla Q_\varepsilon^{*\top}\dot{\varepsilon}_i(t) = 0. \tag{10}$$

Since the coupled HJB equation Eq. (10) exists the partial differential part, which makes it hard to be solved analytically. Therefore, the PI algorithm [32], [33], is introduced to deal with this issue in an iterative way.

Let $u_i^l(t)$ and $Q_i^l(\varepsilon_i(t))$ represent iterative control and iterative Q-function, respectively, with l is iteration index. The PI algorithm consists of two components, *policy evaluation* and *policy improvement*. The iteration process can be summarised as follows:

1. ***Policy Evaluation***: Update the iterative Q-function by

$$H_i\big(\varepsilon_i(t), u_i^l(t), Q_i^l(\varepsilon(t))\big) = r_i(\varepsilon_i(t), u_i^l(t)) + \nabla Q_\varepsilon^{l\top}\dot{\varepsilon}_i(t) = 0. \tag{11}$$

2. ***Policy Improvement***: Compute the control law u_i^l by

$$u_i^{l+1}(t) = -\frac{1}{2}(d_i + c_i)R_{ii}^{-1}g_i^\top(t)\nabla Q_\varepsilon^l. \tag{12}$$

3. If $|Q_i^l(\varepsilon_i) - Q_i^{l-1}(\varepsilon_i)| \le \delta$ (δ is a small positive constant), end. Else, let $l = l + 1$, go to step 1.

The PI algorithm is an effective method to solve the various optimal control problems. It has been proved that the iterative Q-function and the iterative control strategy in PI will converge to the optimal values $Q_i^\star(t)$ and $u_i^\star(t)$ through iterations [34], [35].

However, it is worth noting that the PI algorithm needs the knowledge of system models, i.e., $g_i(t)$ exists in the controller Eq. (12). Therefore, system identification is needed normally, but it is not suitable for the actual exoskeleton system with different hemiplegic patient. Because for different patients, the identification process needs to be reconstructed. To overcome this difficulty, the following section will present a data-driven adaptive control strategy in a online-learning

fashion, and it should be emphasized that this method without needing the knowledge of the accurate system dynamics, and no system identification is needed [37], [38].

IV. Data-driven Reinforcement Learning Controller with Actor-Critic Neural Networks

In this section, base on PI algorithm, we will present the DDRL algorithm to achieve online-learning control and better adaptive performance for different patients via the framework of RL called ACNN. In the ACNN, actor network is used to approximate controller and critic network is introduced to estimate Q-function online, respectively. The detailed descriptions are given as follows.

A. The Critic Networks Modular

First, the critic networks are adopted to approximate the Q-function $Q_i(t)$ as follows:

$$\hat{Q}_i(t) = \hat{W}_{ci}^{\top}(t)\psi_{ci}(z_{ci}(t)), \tag{13}$$

where z_{ci} is an input information of the critic modular and information from ε_i, u_i, and $u_{(j)}$, $\psi_{ci}(z_{ci})$ denotes the activation function, and $\hat{W}_{ci}$ is the weight vector of the critic network modular.

Then, at every time step, the Hamilton function (8) can be approximated as follows:

$$\mathfrak{e}_{ci}(t) = \int_t^{t+T} r_i(\varepsilon_i, u_i)ds + \hat{W}_{ci}^{\top}\big(\psi_{ci}(z_{ci}(t+T)) - \psi_{ci}(z_{ci}(t))\big), \tag{14}$$

where T denotes the sampling period. Then, the Eq. (14) is utilized to define the approximation error for the critic NNs. Thus, the residual error function to be minimized is defined as $E_{ci} = \frac{1}{2}\mathfrak{e}_{ci}^2(t)$. Then, according to gradient descent based weight update rule [39], the tuning weight can be adopted as follows

$$\begin{aligned}\dot{\hat{W}}_{ci}(t) &= -\kappa_{ci}\frac{\partial E_{ci}(t)}{\partial \mathfrak{e}_{ci}(t)}\frac{\partial \mathfrak{e}_{ci}(t)}{\partial \hat{Q}_i(t)}\frac{\partial \hat{Q}_i(t)}{\partial \hat{W}_{ci}(t)} \\ &= -\kappa_{ci}\psi_{ci}(z_{ci})\big(\hat{W}_{ci}^{\top}\Delta\psi_{ci}(z_{ci}) + \int_t^{t+T} r_i(\varepsilon_i, u_i)ds\big),\end{aligned} \tag{15}$$

where $\Delta\psi_{ci}(z_{ci}) = \psi_{ci}(z_{ci}(t+T)) - \psi_{ci}(z_{ci}(t))$, κ_{ci} is the learning rate of the critic network modular for agent i.

B. The Actor Networks Modular

Define the actor neural networks, which is employed to approximate the control strategy, as follows:

$$\hat{u}_i(t) = \hat{W}_{ai}^{\top}(t)\psi_{ai}(z_{ai}(t)), \tag{16}$$

where z_{ai} is an input vector of the actor network including ε_i of agent i, $\psi_{ai}(z_{ai})$ denotes the activation function, and $\hat{W}_{ai}$ is the weight matrix.

Then, in order to obtain the desired approximation optimal controller to minimize the Q- function $\hat{Q}_i$, the error function of the actor network is given by

$$\mathfrak{e}_{ai}(t) = \hat{Q}_i(\varepsilon_i(t)) - U_d, \tag{17}$$

where U_d is the ultimate utility function. From perspective of the RL, the value of the U_d according to different applications. The objective function to be minimized in the actor network is given by $E_{ai}(t) = \frac{1}{2}\mathfrak{e}_{ai}^2(t)$.

Algorithm 1 DDRL Walking Assistance Control Algorithm

1: **Initialization**
2: The initial values of critic weight $\hat{W}_{ci}(0)$ and actor weight $\hat{W}_{ai}(0)$ are all chosen as zero;
3: Select the weight learning rates of the critic network and actor network κ_{ai} and κ_{ci}.
4: Choose a small enough computation precision δ;
5: **repeat**
6: Calculate the actor network Eq. (16) to estimate the control strategy $\hat{u}_i$;
7: Calculate the critic network Eq. (13) to estimate the Q-function $\hat{Q}_i$;
8: According to the available system data q_i and q_r, compute the error ε_i by Eq. (5);
9: Calculate the objective function E_{ci};
10: Update the weights in the critic NNs using Eq. (15);
11: Calculate the objective function E_{ai};
12: Update the weights in the actor NNs using Eq. (18);
13: **until** $\|\hat{W}_{ci}' - \hat{W}_{ci}\| \le \delta$ ($\hat{W}_{ci}$ and $\hat{W}_{ci}'$ denote the weight of the current time and previous time);

Similarly, using the gradient descent rule, the following updating rule of the actor network is utilized:

$$\begin{aligned}\dot{\hat{W}}_{ai}(t) &= -\kappa_{ai}\frac{\partial E_{ai}(t)}{\partial \mathfrak{e}_{ai}(t)}\frac{\partial \mathfrak{e}_{ai}(t)}{\partial \hat{Q}_i(t)}\frac{\partial \hat{Q}_i(t)}{\partial z_{ci}(t)}\frac{\partial z_{ci}(t)}{\partial \hat{u}_i(t)}\frac{\partial \hat{u}_i(t)}{\partial \hat{W}_{ai}(t)} \\ &= -\kappa_{ai}\psi_{ai}(z_{ai})\hat{W}_{ci}^{\top}\nabla\psi_{ci}(z_{ci})\xi_i\psi_{ci}^{\top}(z_{ci})\hat{W}_{ci},\end{aligned} \tag{18}$$

where $\xi_i = \partial z_{ci}/\partial \hat{u}_i$, $\nabla\psi_{ci}(z_{ci}) = \partial\psi_{ci}(z_{ci})/\partial z_{ci}$and κ_{ai} is the learning rate of the actor NN for agent i.

The procedure of the data-driven adaptive control strategy is presented in Algorithm 1. It should be noted that only the measured system data, i.e., ε_i and u_i are required in the design of the DDRL algorithm. Thus, the proposed method is a data-driven/model-free approach [36], which improves the potential application of the proposed method.

V. Experiments and Discussions

The proposed data-driven control strategy is validated both on two scenarios: *2-DOF manipulator* in simulation platform and *walking assistance* experiments on an actual LLE system, respectively.

A. 2-DOF System Simulation

1) Dynamic Model of 2-DOF Manipulator System: For simulation, the simulation environment is set up in Simulink-Matlab. The dynamics of the 2-DOF systems are the same as Eq. (1), and the system matrices are given as follows [40]: $M_i = \begin{bmatrix} m_{i1}+m_{i2}+2m_{i3}\cos(\theta_{i2}) & m_{i2}+m_{i3}\cos(\theta_{i2}) \\ m_{i2}+m_{i3}\cos(\theta_{i2}) & m_{i2}\end{bmatrix}$, $C_i = \begin{bmatrix} -m_{i3}\dot{\theta}_{i2}\sin(\theta_{i2}) & -m_{i3}(\dot{\theta}_{i1})+\theta_{i2}\sin(\theta_{i2}) \\ -m_{i3}\dot{\theta}_{i1} & 0\end{bmatrix}$, and the $G_i = \begin{bmatrix} m_{i4}g\cos(\theta_{i1})+m_{i5}g\cos(\theta_{i1}+\theta_{i2}) \\ m_{i5}g\cos(\theta_{i1}+\theta_{i2})\end{bmatrix}$, $\tau_i = [\tau_{i1}, \tau_{i2}]^{\top}$, m_{ip} ($p = 1,2,3,4,5$) are the masses. Note that, in simulation case, the given dynamic system can be used to product system data needed in DDRL algorithm.

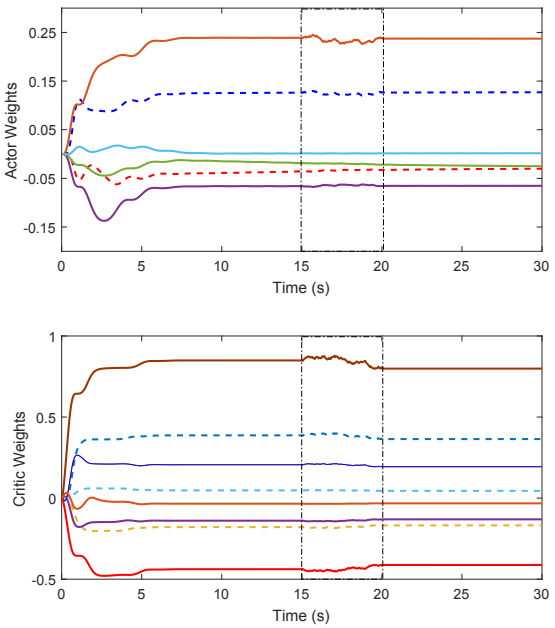


Fig. 2. Convergence of the actor network weights and critic network weights on 2-DOF simulation platform.

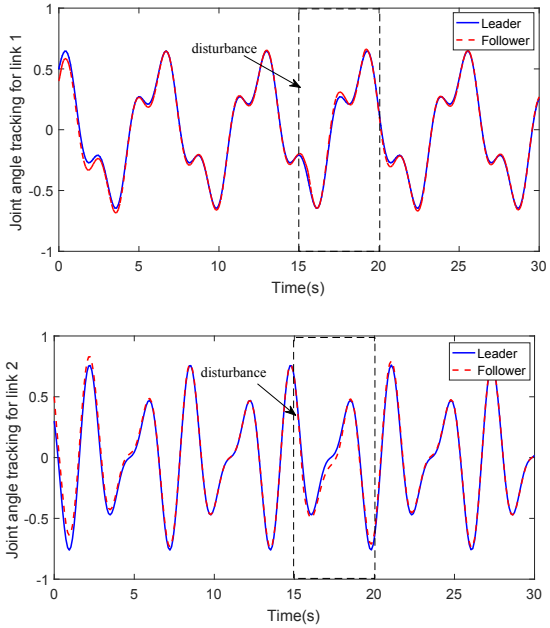


Fig. 3. The trajectories tracking performance of joint angle of Follower 1 on 2-DOF simulation platform.

The Leader system (desired trajectories) is expressed by

$$\theta_r = \begin{bmatrix} \theta_{1r} \\ \theta_{2r} \end{bmatrix} = \begin{bmatrix} 0.5cos(t)+0.2sin(3t) \\ 0.3cos(3t)-0.5sin(2t) \end{bmatrix}. \quad (19)$$

The structure of the ACNN is chosen as 3-layers back propagation (BP) NN. The initial critic NN weights and actor NN weights are set to be zero, and setting the value of the computation precision as $\delta = 10^{-5}$. The weight learning rates of the actor network and the critic network are $\kappa_{ai} = 0.03$, $\kappa_{ci} = 0.06$. Let activation functions ψ_{ai} and ψ_{ci} be the hyperbolic tangent functions, i.e., $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

2) Numerical Simulation Results and Analysis: From Fig. 2, we can see that through 6 seconds learning process, the critic NN weights and the actor NN weights are convergent. Thus, the approximate optimal controller can be obtained in Eq. (16). Therefore, based on the optimal controller, the joint angles $\theta_1 = (\theta_{11}, \theta_{12})^\top$ of Follower 1 achieve a good performance of tracking the Leader at 6 s $< t <$ 15 s, which is shown in Fig. 3.

Further, we add some disturbance signal (white noise) to the system at $t \in [15, 18]$ s to verify the robustness of our method. From Figs. 2, the ACNN weights are trained again adaptively until converge from $t = 15$ s to $t = 20$ s, and the optimal controller has been modified correspondingly. From Figs. 3, it is seen that joint angle trajectories of two links of Follower 1 are synchronized with the Leader again quickly after $t = 20$ s. These simulation results shown the better performance of the proposed DDRL algorithm, which has ability to respond to disturbances online in the system operation. It is proved that our proposed control method has good robustness against uncertainties.

B. Walking Assistance Experiments on LLE System

1) Experimental Setup: To demonstrate the adaptability of the proposed control strategy, a real LLE system, namely AIDER, which shown in Fig. 4, is designed for walking assistance case with different hemiplegia. A distributed control system is embedded in AIDER consisting of a main controller and four node controllers. The main controller is placed on the backpack to compute the control algorithm. Node controllers are fixed near by the corresponding active joints position of LLE robot which aims at receiving sensor data and executing control commands from main controller.

During the experiments on the AIDER system, three healthy subjects (1, 2, 3) with different heights (165 cm, 176 cm, 180 cm) are selected to operate the wearable LLE system. All wearers are simulated as hemiplegic patients, and the right leg of the subject are simulated as the affected leg. In the walking assistance task for all wearers, every wearer is asked to walk for 80 seconds by using the AIDER system. All the pilot's walking speed is varying from $0.1m/s$ to $0.4m/s$. Further, the accelerometer and the wearable sensory system are utilized to measure system data.

For the implementation of the proposed data-driven control strategy on the AIDER system. Note that the proposed data-driven control strategy DDRL has a learning process using the online system data at the beginning to adapt different subjects until learning the optimal control policies for the LLE system with pilots. Choosing the ACNN as 3-layer Back propagation (BP) neural networks structure [41], [42], that is, input layer, hidden layer and output layer. The initial values of weights $\hat{W}_{ci}$ and $\hat{W}_{ai}$ of the critic and actor are all set to be zero, and the activation functions ψ_{ai} and ψ_{ci} are chosen as hyperbolic tangent functions $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$. The learning rates are the same as in the simulation platform, that is $\kappa_{ai} =$

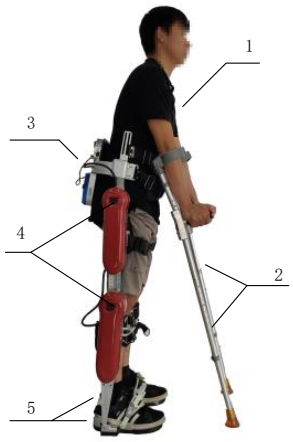


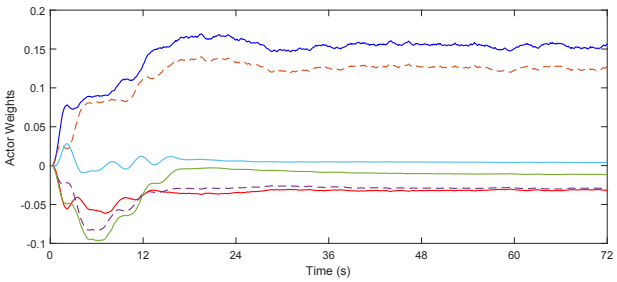
Fig. 4. A LLE system called AIDER for hemiplegic patient. 1. The subject/pilot; 2. Crutches; 3. The load backpack with main controller and power unit; 4. Active joints with node controllers (hip joints and knee joints); 5. Smart shoes with plantar sensors.

0.03, $\kappa_{ci} = 0.06$.

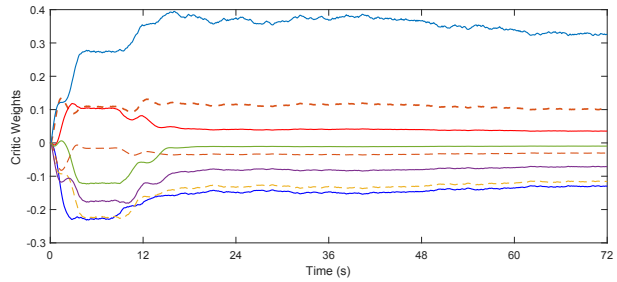
2) Results and Discussions: For subject 1, from Fig. 5 (a) and (b), we can see that, after about 15 seconds training, the weights of ACNN are bounded convergent (Uniformly Ultimately Bounded) because of the disturbances always exist in LLE system. The tracking performance of the hip joint and knee joint for the LLE system with wearer 1 is shown in Fig. 6, which states that with the aid of the optimal control policies, the two followers can achieve synchronization with the Leader's motion trajectories. Further, we can find that there are different walking motions in the process of walking, which means our proposed method has capability of adapting different gait patterns. It should be noted that the affected side of wearer with exoskeleton's side (Follower 2) has a half gait cycle delay to the side which has motion ability (Leader), which is marked with blue dashed line as shown in Fig. 6. The experiment results illustrate the effectiveness of the proposed DDRL approach for walking assistance of the exoskeleton with different pilots.

VI. Conclusion and Future Work

This paper has proposed a DDRL control strategy of a lower exoskeleton system to assist hemiplegic patient walking. A LFMAS structure has been proposed to model the interaction relation among LLE system and hemiplegic individual. The PI algorithm has been introduced to obtain optimal controller. The ACNN framework has been presented based on PI algorithm for the implementation of the proposed approach in an online-learning manner to adapt different patients. It highlights that the controller design only depends on the measured system data, without using the accurate system model and system identification. We have successfully validated the proposed method on two situations: *2-DOF manipulator* in simulation platform and *walking assistance* experiment on a real LLE system called AIDER. Experimental results have confirmed the effectiveness of the proposed control method. In the future, we will consider the RL based control for exoskeleton system with actuator fault, input time-delay.

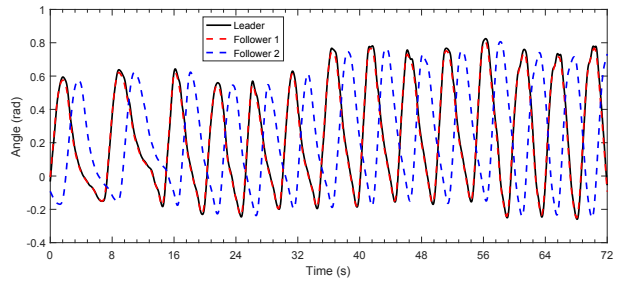


(a) Actor weights

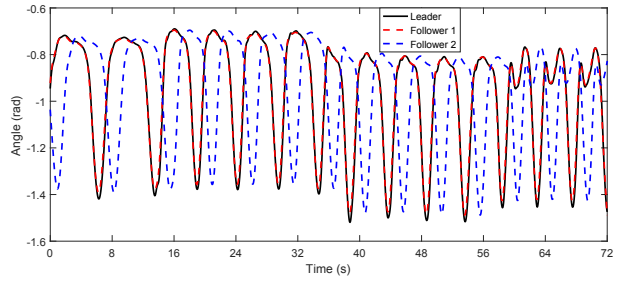


(b) Critic weights

Fig. 5. The weights convergence of the actor network and the critic network on AIDER with subject 1 in the experiment: (a) Actor weights; (b) Critic weights.



(a) Hip joint



(b) Knee joint

Fig. 6. The tracking control performance performance of the proposed DDRL strategy on AIDER with subject 1 in the experiment: (a) hip joint's angle; (b) knee joint's angle.

Acknowledgment

This work was made possible by support from National Key Research and Development Program of China (No. 2017YFB1302300), National Natural Science Foundation of China (NSFC) (No. 6150020696, 61503060) and Sichuan Science and Technology Program (No. 20SYSX0276).

REFERENCES

[1] C. J. Walsh, D. Paluska, K. Pasch, et al., "Development of a lightweight, under-actuated exoskeleton for load-carrying augmentation," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2006, pp. 3485-3491.

[2] R. Huang, H. Cheng, H. Guo, Q. Chen, X. Lin, X., "Hierarchical interactive learning for a human-powered augmentation lower exoskeleton," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 257-263.

[3] R. Huang, H. Cheng, Q. Chen, et al., "Interative learning for sensitivity factors of a human-powered augmentation lower exoskeleton," in *Proc. of IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 6409-6415.

[4] R. Huang, "Learning physical human-robot interaction with coupled cooperative primitives for a lower exoskeleton," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 4, pp. 1566-1574, 2019.

[5] J. Zhang, P. Fiers, K. A. Witte, R. W. Jackson, K. L. Poggensee, C. G. Atkeson, S. H. Collins, "Human-in-the-loop optimization of exoskeleton assistance during walking," *Science*, vol. 356, pp. 1280-1284, 2017.

[6] A. Esquenazi, M. Talaty, A. Jayaraman, "Powered exoskeletons for walking assistance in persons with central nervous system injuries: a narrative review," *PM&R*, vol. 9, no. 1, pp. 46-62, 2017.

[7] W. Huo, S. Mohammed, J. C. Moreno, and Y. Amirat, "Lower limb wearable robots for assistance and rehabilitation: A state of the art," *IEEE Systems Journal*, vol. 10, no. 3, pp. 1068-1081, 2014.

[8] Y. Sankai, "HAL: hybrid assistive limb based on cybernics," *Robotics Research*, vol. 66, pp. 25-34, 2010.

[9] A. M. Dollar, H. Herr, "Lower extremity exoskeletons and active orthoses: Challenge and state-of-art," *IEEE Transaction on Robotics*, vol. 24, no. 1, pp. 144-158, 2008.

[10] K. A. Strausser, H. Kazerooni, "The development and testing of a human machine interface for a mobile medical exoskeleton," in *Proc. of IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 4911-4916.

[11] J. Iqbal, K. Baizid, "Stroke rehabilitation using exoskeleton-based robotic exercisers: mini review," *Biomedical Research*, vol. 26, no. 1, pp. 197-201, 2015.

[12] N. S. K. Ho, K. Y. Tong, X. L. Hu, K. L. Fung, X. J. Wei, W. Rong, E. A. Susanto, "An EMG-driven exoskeleton hand robotic training device on chronic stroke subjects: task training system for stroke rehabilitation," in *Proc. of IEEE International Conference on Rehabilitation Robotics*, 2011, pp. 1-5.

[13] M. Bortole, A. Venkatakrishnan, F. Zhu, J. C. Moreno, G. E. Francisco, J. L. Pons, J. L. Contreras-Vidal, "The H2 robotic exoskeleton for gait rehabilitation after stroke: early findings from a clinical study," *Journal of Neuroengineering and Rehabilitation*, vol. 12, no. 54, pp. 1-14, 2015.

[14] D. R. Louie, J. J. Eng, "Powered robotic exoskeletons in post-stroke rehabilitation of gait: a scoping review," *Journal of Neuroengineering and Rehabilitation*, vol. 13, no. 53, pp. 1-10, 2016.

[15] S. F. Tyson, H. A. Thornton, "The effect of a hinged ankle foot orthosis on hemiplegic gait: objective measures and users' opinions," *Clinical Rehabilitation*, vol. 15, no. 1, pp. 53-58, 2001.

[16] S. Fatone, S. A. Gard, B. S. Malas, "Effect of ankle-foot orthosis alignment and foot-plate length on the gait of adults with poststroke hemiplegia," *Archives of Physical Medicine and Rehabilitation*, vol. 90, no. 5, pp. 810-818, 2009.

[17] J. A. Blaya, H. Herr, "Adaptive control of a variable impedance ankle-foot orthosis to assist drop-foot gait," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 12, no. 1, pp. 24-31, 2004.

[18] J. S. Sulzer, R. A. Roiz, M. A. Peshkin, et al., "A highly backdrivable, lightweight knee actuator for investigating gait in stroke," *IEEE Transactions on Robotics*, vol. 25, no. 3, pp. 539-548, 2009.

[19] M. Hassan, H. Kadone, T. Ueno, Y. Hada, Y. Sankai, K. Suzuki, "Feasibility of synergy-based exoskeleton robot control in hemiplegia," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 6, pp. 1233-1242, 2018.

[20] P. Maciejasz, J. Eschweiler, K. Gerlach-Hahn, A. Jansen-Troy, S. Leonhardt, "A survey on robotic devices for upper limb rehabilitation," *Journal of Neuroengineering and Rehabilitation*, vol. 11, no. 3, pp. 1-29, 2014.

[21] H. Kawamoto, T. Hayashi, T. Sakurai, et al., "Development of a single leg version of HAL for hemiplegia," in *Proc. of IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, 2009, pp. 5038-5043.

[22] H. Kawamoto, H. Kandone, T. Sakurai, et al., "Development of an assist controller with robot suit HAL for hemiplegia patients using motion data on the unaffected side," in *Proc. of IEEE Annual International Conference on Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 3077-3080.

[23] S. Fisher, L. Lucas, T. A. Thrasher, "Robot-assisted gait training for patients with hemiparesis due to stroke," *Topics in Stroke Rehabilitation*, vol. 18, no. 3, pp. 269-276, 2011.

[24] S. A. Murray, K. H. Ha, C. Hartigan, M. Goldfarb, "An assistive control approach for a lower-limb exoskeleton to facilitate recovery of walking following stroke," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 3, pp. 441-449, 2015.

[25] R. Huang, Z. Peng, H. Cheng, J. Hu, et al., "Learning-based walking assistance control strategy for a lower limb exoskeleton with hemiplegia patients," in *Proc. of IEEE International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2280-2285.

[26] J. Hu, G. Feng, "Distributed tracking control of leader-follower multi-agent systems under noisy measurement," *Automatica*, vol. 46, no. 8, pp. 1382-1387, 2010.

[27] F. L.Lewis, D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits and Systems Magazine*, vol. 9, no. 3, pp. 32-50, 2009.

[28] K. G. Vamvoudakis, F. L. Lewis, "Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878-888, 2010.

[29] R. S. Sutton, A. G. Barto, "Reinforcement Learning: An Introduction," *MIT Press*, 1998.

[30] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, et al., "Human-level control through deep reinforcement learning," it Nature, vol. 518, no. 7540, pp. 529-533, 2015.

[31] V. Mnih, A. P. Badia, M. Mirza, A. Graves, A., T. Harley, T. P. Lillicrap, D. Silver, K. Kavukcuoglu, "Asynchronous methods for deep reinforcement learning," in *Proc. of International Conference on Machine Learning (ICML)*, 2016, pp. 1928-1937.

[32] D. Wang, D. Liu, H. Li, "Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems," *IEEE Transactions on Automation Science and Engineering*, vol. 11, no. 2, pp. 627-632, 2014.

[33] D. Liu, Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 621-634, 2014.

[34] Z. Peng, Y. Zhao, J. Hu, et al., "Data-driven optimal tracking control of discrete-time multi-agent systems with two-stage policy iteration algorithm," *Information Sciences*, vol. 481, pp. 189-202, 2019.

[35] Z. Peng, J. Hu, K. Shi, R. Luo, R. Huang, B.K. Bijoy, and J. Huang, "A novel optimal bipartite consensus control scheme for unknown multi-agent systems via model-free reinforcement learning," *Applied Mathematics and Computation*, to be published, doi: 0.1016/j.amc.2019.124821.

[36] Z. Peng, J. Hu, B.K. Ghosh, "Data-driven containment control of discrete-time multi-agent systems via value iteration," *SCIENCE CHINA Information Sciences*, doi: 10.1007/s11432-018-9671-2.

[37] J. Ghan, R. Steger, H. Kazerooni, "Control and system identification for the Berkeley lower extremity exoskeleton (BLEEX)," *Advanced Robotics*, vol. 20, no. 9, pp. 989-1014, 2006.

[38] J. Ghan, H. Kazerooni, "System identification for the Berkeley lower extremity exoskeleton (BLEEX)," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2006, pp. 3477-3484.

[39] J. Si, Y. Wang, "Online learning control by association and reinforcement," *IEEE Transactions on Neural networks*, vol. 12, no. 2, pp. 264-276, 2001.

[40] F. L. Lewis, C. T. Abdallah, D. M. Dawson, "Control of Robot Manipulators," New York, NY, USA: Macmillan, 1993.

[41] A.T. Goh, "Back-propagation neural networks for modeling complex systems," *Artificial Intelligence in Engineering*, vol. 9, no. 3, pp. 143-151, 1995.

[42] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proc. of International Joint Conference on Neural Networks*, 1989, pp. 593-605.