目录

[控制 2](#_Toc46492185)

[1. "Reinforcement Learning Control of a Flexible Two-Link Manipulator: An Experimental Investigation," 2](#_Toc46492186)

[2. "Learning Physical Human–Robot Interaction With Coupled Cooperative Primitives for a Lower Exoskeleton," 2](#_Toc46492187)

[3. "DMP-Based Motion Generation for a Walking Exoskeleton Robot Using Reinforcement Learning," 3](#_Toc46492188)

[4. Reinforcement learning based compliance control of a robotic walk assist device 4](#_Toc46492189)

[5. "Does Reinforcement Learning outperform PID in the control of FES-induced elbow flex-extension?," 5](#_Toc46492190)

[6. Position/force control of robot manipulators using reinforcement learning 6](#_Toc46492191)

[7. Improving Input-Output Linearizing Controllers for Bipedal Robots via Reinforcement Learning 7](#_Toc46492192)

[8. Reinforcement learning control of a single-link flexible robotic manipulator 8](#_Toc46492193)

[步态评价 调节参数 8](#_Toc46492194)

[1. "Learning task-parametrized assistive strategies for exoskeleton robots by multi-task reinforcement learning," 9](#_Toc46492195)

[2. Learning assistive strategies for exoskeleton robots from user-robot physical interaction 10](#_Toc46492196)

[3. Wearer-Prosthesis Interaction for Symmetrical Gait: A Study Enabled by Reinforcement Learning Prosthesis Control 11](#_Toc46492197)

[4. Online Reinforcement Learning Control for the Personalization of a Robotic Knee Prosthesis 12](#_Toc46492198)

[5. Teaching a humanoid robot to walk faster through Safe Reinforcement Learning 13](#_Toc46492199)

[6. Offline Policy Iteration Based Reinforcement Learning Controller for Online Robotic Knee Prosthesis Parameter Tuning 14](#_Toc46492200)

[相关领域论文： 15](#_Toc46492201)

[1. 论文题目：Deep Predictive Policy Training using Reinforcement Learning 15](#_Toc46492202)

[2. 论文题目：Learning Agile and Dynamic Motor Skills for Legged Robots 16](#_Toc46492203)

# 控制

## 1. "Reinforcement Learning Control of a Flexible Two-Link Manipulator: An Experimental Investigation,"

W. He, H. Gao, C. Zhou, C. Yang and Z. Li, "**Reinforcement Learning Control of a Flexible Two-Link Manipulator: An Experimental Investigation**," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, doi: 10.1109/TSMC.2020.2975232.

|  |
| --- |
| This article discusses the control design and experiment validation of a flexible two-link manipulator (FTLM) system represented by ordinary differential equations (ODEs). A reinforcement learning (RL) control strategy is developed that is based on actor-critic structure to enable vibration suppression while retaining trajectory tracking. Subsequently, the closed-loop system with the proposed RL control algorithm is proved to be semi-global uniform ultimate bounded (SGUUB) by Lyapunov's direct method. In the simulations, the control approach presented has been tested on the discretized ODE dynamic model and the analytical claims have been justified under the existence of uncertainty. Eventually, a series of experiments in a Quanser laboratory platform are investigated to demonstrate the effectiveness of the presented control and its application effect is compared with PD control.  本文论述了一个灵活的控制设计和实验验证两连杆机械手(FTLM)系统由常微分方程(ode)表示。强化学习(RL)控制策略是**基于actor-critic结构发达,使振动抑制,同时保留轨迹跟踪**。随后,提出了**RL的闭环系统控制算法**被证明是semi-global一致终极有界(SGUUB)李雅普诺夫直接法。模拟、离散的控制方法提出了测试颂歌动态模型和分析要求合理的下不确定性的存在。最终,一系列的实验在Quanser实验室平台调查证明了控制的有效性及其应用效果与PD控制相比较。 |

## 2. "Learning Physical Human–Robot Interaction With Coupled Cooperative Primitives for a Lower Exoskeleton,"

R. Huang, H. Cheng, J. Qiu and J. Zhang, "**Learning Physical Human–Robot Interaction With Coupled Cooperative Primitives for a Lower Exoskeleton**," in IEEE Transactions on Automation Science and Engineering, vol. 16, no. 4, pp. 1566-1574, Oct. 2019, doi: 10.1109/TASE.2018.2886376.

|  |
| --- |
| Human-powered lower exoskeletons have received considerable interests from both academia and industry over the past decades, and encountered increasing applications in human locomotion assistance and strength augmentation. One of the most important aspects in those applications is to achieve robust control of lower exoskeletons, which, in the first place, requires the proactive modeling of human movement trajectories through physical human-robot interaction (pHRI). As a powerful representative tool for motion trajectories, dynamic movement primitives (DMP) have been used to model human movement trajectories. However, canonical DMP only offers a general representation of human movement trajectory and may neglects the interactive term, therefore it cannot be directly applied to lower exoskeletons which need to track human joint trajectories online, because different pilots have different trajectories and even same pilot might change his/her motion during walking. This paper presents a novel coupled cooperative primitive (CCP) strategy, which aims at modeling the motion trajectories online. Besides maintaining canonical motion primitives, we model the interaction term between the pilot and exoskeletons through impedance models, and propose a reinforcement learning method based on policy improvement and path integrals (PI 2 ) to learn the parameters online. Experimental results on both a single degree-of-freedom platform and a HUman-powered Augmentation Lower EXoskeleton (HUALEX) system demonstrate the advantages of our proposed CCP scheme.  在过去的几十年中，人力驱动的下骨骼受到了学术界和工业界的极大关注，并且在人体运动辅助和力量增强中得到了越来越多的应用。 这些应用程序中最重要的方面之一是实现对下骨骼的鲁棒控制，这首先​​要求通过物理人机交互（pHRI）对人类运动轨迹进行主动建模。 作为运动轨迹的有力代表工具，动态运动原语（DMP）已用于对人类运动轨迹进行建模。 但是，规范的DMP仅提供了人体运动轨迹的一般表示，并且可能忽略了交互项，因此它不能直接应用于需要在线跟踪人体关节轨迹的下骨骼，因为不同的飞行员具有不同的轨迹，甚至同一位飞行员可能会改变 他/她在行走过程中的动作。 本文提出了一种新颖的耦合合作原语（CCP）策略，该策略旨在在线建模运动轨迹。 除了维护经典运动原语外，我们还通过阻抗模型对飞行员与外骨骼之间的相互作用项进行建模，并提出一种基于策略改进和路径积分（PI 2）的强化学习方法以在线学习参数。 在单自由度平台和人力增强下外骨骼（HUALEX）系统上的实验结果都证明了我们提出的CCP方案的优势。 |

## 3. "DMP-Based Motion Generation for a Walking Exoskeleton Robot Using Reinforcement Learning,"

Y. Yuan, Z. Li, T. Zhao and D. Gan, "**DMP-Based Motion Generation for a Walking Exoskeleton Robot Using Reinforcement Learning**," in IEEE Transactions on Industrial Electronics, vol. 67, no. 5, pp. 3830-3839, May 2020, doi: 10.1109/TIE.2019.2916396.

|  |
| --- |
| **Abstract:**  For the purpose of the assistance for human walking, this paper describes a novel coupled movement sequences planning and motion adaption based on dynamic movement primitives (DMPs) for a walking exoskeleton robot. The developed exoskeleton robot has eight degrees of freedom (DOFs). The hip and knee of each artificial leg can provide two electric-powered DOFs to flexion or extension, two passive-installed DOFs of the ankle are to achieve the motion of inversion/eversion and plantarflexion/dorsiflexion, and two passive DOFs of the hip are to achieve the motion of roll or yaw. A novel trajectory-learning scheme based on reinforcement learning (RL) combined with DMPs is presented for a lower limb exoskeleton robot, aiming to give assistance to human walking. In the proposed strategy, a two-level planning is designed. In the first level, the inverted pendulum approximation under the consideration of the locomotion parameters is utilized to guarantee the zero-moment point within the ankle joint of the support leg in the phase of single support. In the second level, the joint trajectories are modeled and learned by DMPs. Meanwhile, the RL is adopted to learn the trajectories for eliminating the effects of uncertainties in joint space. The experiment involving four subjects based on a lower limb exoskeleton robot demonstrates that the proposed scheme can effectively suppress the disturbances and uncertainties.  为了辅助人类步行，本文描述了一种新颖的基于运动外骨骼机器人的基于动态运动原语（DMP）的运动序列规划和运动适应。 开发的外骨骼机器人具有八个自由度（DOF）。 每条人造腿的髋部和膝盖可提供两个电动自由度来屈曲或伸展，踝部的两个被动安装的自由度可实现内翻/外翻和plant屈/背屈的运动，而髋部的两个被动性自由度可实现 实现横摇或横摇的运动。 提出了一种基于强化学习（RL）结合DMP的新型轨迹学习方案，旨在为下肢外骨骼机器人提供帮助。 在提出的策略中，设计了一个两级计划。 在第一级中，考虑运动参数的倒立摆近似用于确保在单次支撑阶段支撑腿的踝关节内的零矩点。 在第二级中，联合轨迹由DMP建模和学习。 同时，采用RL来学习消除关节空间不确定性影响的轨迹。 基于下肢外骨骼机器人的四个对象的实验表明，该方案可以有效地抑制干扰和不确定性。 |

## 4. Reinforcement learning based compliance control of a robotic walk assist device

Khan S G, Tufail M, Shah S H, et al. **Reinforcement learning based compliance control of a robotic walk assist device**[J]. Advanced Robotics, 2019, 33(24): 1281-1292.

|  |
| --- |
| **ABSTRACT**  Millions of people around the globe have to deal with walking disability. Robotic walk assist devices can help people with walking disabilities, especially those with weak legs. However, safety, cost, efficiency and user friendliness are some of the key challenges. For robotic walk assist devices, light weight structure and energy efficient design as well as optimal control are vitally important. In addition, compliance control can help to improve the safety of such devices as well as contribute to their user friendliness. In this paper, an optimal adaptive compliance control is proposed for a Robotic walk assist device. The suggested scheme is based on bio-inspired reinforcement learning. It is completely dynamic-model-free scheme and employs joint position and velocity feedback as well as sensed joint torque (applied by user during walk) for compliance control. The efficiency of the controller is tested in simulation on a robotic walk assisting device model.    KEYWORDS: [Robotics walk assist device](https://www.tandfonline.com/keyword/Robotics+Walk+Assist+Device), [reinforcement learning](https://www.tandfonline.com/keyword/Reinforcement+Learning), [compliance control](https://www.tandfonline.com/keyword/Compliance+Control), [model reference](https://www.tandfonline.com/keyword/Model+Reference), [lower extremity exoskeleton](https://www.tandfonline.com/keyword/Lower+Extremity+Exoskeleton)  全球数以百万计的人不得不应对步行障碍。 机器人步行辅助设备可以帮助有步行障碍的人，尤其是腿弱的人。 但是，安全性，成本，效率和用户友好性是一些关键挑战。 对于机器人步行辅助设备，轻巧的结构和节能设计以及最佳控制至关重要。 另外，顺从性控制可以帮助提高此类设备的安全性，并有助于提高其用户友好性。 在本文中，提出了机器人步行辅助设备的最优自适应顺应性控制。 建议的方案基于生物启发的强化学习。 它是完全无动态模型的方案，并采用关节位置和速度反馈以及感应到的关节扭矩（在步行过程中由用户施加）来进行顺从性控制。 在机器人步行辅助设备模型的仿真中测试了控制器的效率。  关键词：机器人行走辅助装置，强化学习，合规控制，模型参考，下肢外骨骼 |

## 5. "Does Reinforcement Learning outperform PID in the control of FES-induced elbow flex-extension?,"

D. D. Febbo et al., "**Does Reinforcement Learning outperform PID in the control of FES-induced elbow flex-extension?**," 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, 2018, pp. 1-6, doi: 10.1109/MeMeA.2018.8438800.

|  |
| --- |
| **Abstract:**  Functional electrical stimulation (FES) is an effective technology in post-stroke rehabilitation of the upper limbs. Because of the complexity of the system, traditional linear controllers are still far to drive accurate and natural movements. In this work, we apply reinforcement learning (RL) to design a nonlinear controller for an upper limb FES system combined with a passive exoskeleton. RL methods learn by interacting with the environment and, to efficiently use the collected data, we simulated large numbers of experience episodes through artificial neural network (ANN) models of the electrically stimulated arm muscles. The performance of the novel control solution was compared to a PID controller on five healthy subjects during planar reaching tasks. Both controllers correctly drove the arm at the target position, with a mean absolute error <; 1°. The RL control significantly outperformed the PID in terms of setting time, position accuracy and smoothness. Future trials are needed to confirm these promising results.  功能性电刺激（FES）是上肢中风后康复的有效技术。 由于系统的复杂性，传统的线性控制器仍然无法驱动准确自然的运动。 在这项工作中，我们应用强化学习（RL）来设计与被动外骨骼相结合的上肢FES系统的非线性控制器。 RL方法是通过与环境交互来学习的，为了有效地使用收集到的数据，我们通过电刺激手臂肌肉的人工神经网络（ANN）模型来模拟大量的体验事件。 将新型控制解决方案的性能与PID控制器在平面伸手任务期间对五个健康受试者的性能进行了比较。 两个控制器正确地将手臂驱动到目标位置，平均绝对误差<; 1°。 在设置时间，位置精度和平滑度方面，RL控制明显优于PID。 需要进一步的试验来确认这些有希望的结果。 |

## 6. Position/force control of robot manipulators using reinforcement learning

Perrusquia A, Yu W, Soria A, et al. **Position/force control of robot manipulators using reinforcement learning**[J]. Industrial Robot-an International Journal, 2019, 46(2): 267-280.

|  |
| --- |
| Abstract  Purpose  The position/force control of the robot needs the parameters of the impedance model and generates the desired position from the contact force in the environment. When the environment is unknown, learning algorithms are needed to estimate both the desired force and the parameters of the impedance model.  Design/methodology/approach  In this paper, the authors use reinforcement learning to learn only the desired force, then they use proportional-integral-derivative admittance control to generate the desired position. The results of the experiment are presented to verify their approach.  Findings  The position error is minimized without knowing the environment or the impedance parameters. Another advantage of this simplified position/force control is that the transformation of the Cartesian space to the joint space by inverse kinematics is avoided by the feedback control mechanism. The stability of the closed-loop system is proven.  Originality/value  The position error is minimized without knowing the environment or the impedance parameters. The stability of the closed-loop system is proven.  **目的**：机器人的位置/力控制需要阻抗模型的参数并生成所需的位置接触力的环境。当环境是未知的,需要学习算法来估计所需的力和阻抗的参数模型  **设计/方法/方法**  在本文中，作者使用强化学习仅学习所需的力，然后他们使用比例积分微分导纳控制来生成所需的位置。 提出了实验结果以验证其方法。  **发现**  在不知道环境或阻抗参数的情况下将位置误差最小化。 这种简化的位置/力控制的另一个优点是，反馈控制机制避免了通过逆运动学将笛卡尔空间转换为关节空间。 闭环系统的稳定性得到了证明。  **创意/价值**  在不知道环境或阻抗参数的情况下将位置误差最小化。 闭环系统的稳定性得到了证明。 |

## 7. Improving Input-Output Linearizing Controllers for Bipedal Robots via Reinforcement Learning

|  |
| --- |
| The main drawbacks of input-output linearizing controllers are the need for precise dynamics models and not being able to account for input constraints. Model uncertainty is common in almost every robotic application and input saturation is present in every real world system. In this paper, we address both challenges for the specific case of bipedal robot control by the use of reinforcement learning techniques. Taking the structure of a standard input-output linearizing controller, we use an additive learned term that compensates for model uncertainty. Moreover, by adding constraints to the learning problem we manage to boost the performance of the final controller when input limits are present. We demonstrate the effectiveness of the designed framework for different levels of uncertainty on the five-link planar walking robot RABBIT.  输入输出线性化控制器的主要缺点是**需要精确的动力学模型**，并且**无法考虑输入约束**。 在几乎所有机器人应用中，模型不确定性都很常见，并且在每个现实世界系统中都存在输入饱和。 在本文中，我们将通过使用强化学习技术来解决两足机器人控制特定情况下的两个挑战。 采用标准输入输出线性化控制器的结构，我们使用了可加性学习项来补偿模型的不确定性。 此外，通过为学习问题添加约束，当输入限制出现时，我们设法提高了最终控制器的性能。 我们在五连杆平面行走机器人RABBIT上证明了所设计框架针对不同级别不确定性的有效性。 |

## 8. Reinforcement learning control of a single-link flexible robotic manipulator

|  |
| --- |
| In this study, the authors focus on the reinforcement learning control of a single-link flexible manipulator and attempt to suppress the vibration due to its flexibility and lightweight structure. The assumed mode method and the Lagrange's equation are adopted in modelling to enhance the satisfaction of precision. Two radial basis function neural networks (NNs) are employed in the designed control algorithm, actor NN for generating a policy and critic NN for evaluating the cost-to-go. Rigorous stability of the system has been proven via Lyapunov's direct method. Through Matlab simulation and experiment on the Quanser flexible link platform, the superiority and feasibility of the reinforcement learning control are verified.  在这项研究中，作者专注于单连杆柔性机械手的强化学习控制，并试图抑制由于其柔性和轻巧结构而引起的振动。 在建模中采用假设模式方法和拉格朗日方程，以提高精度的满意度。 在设计的控制算法中，使用了两个径向基函数神经网络（NN），即参与者NN用于生成策略，评论家NN用于评估运行成本。 系统的严格稳定性已通过Lyapunov的直接方法得到证明。 通过在Matser柔性链接平台上的Matlab仿真和实验，验证了强化学习控制的优越性和可行性。 |

# 步态评价 调节参数

## 1. "Learning task-parametrized assistive strategies for exoskeleton robots by multi-task reinforcement learning,"

M. Hamaya, T. Matsubara, T. Noda, T. Teramae and J. Morimoto, "**Learning task-parametrized assistive strategies for exoskeleton robots by multi-task reinforcement learning**," 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, pp. 5907-5912, doi: 10.1109/ICRA.2017.7989695.

|  |
| --- |
| Recent studies suggest that reinforcement learning has great potential for generating assistive strategies in exoskeletons through physical interactions between a user and a robot. Previous methods focused on a task-specific assistive strategy, where for every single task (situation/context), the user needs to interact with a robot to learn an appropriate assistive strategy. Therefore, the learned strategies cannot be generalized for a new task. Since the sampling cost is expensive for such human-in-the-loop systems as exoskeletons, generalization must be enabled. In this paper, we propose to learn task-parametrized assistive strategies for exoskeleton robots. Our method employs an assistive strategy, which depends on the task parameter and the state variable, that can be learned from multiple sets of human-robot interaction data across different tasks and generalized even for an unseen task, given the task parameter without additional learning. To alleviate the user's burden in the learning process across multiple tasks, we exploit a data-efficient multi-task reinforcement learning framework. To verify the effectiveness of our method, we developed an experimental platform with an exoskeleton robot. We conducted a series of experiments whose experimental results show that our method can learn such a task-parametrized assistive strategy and be generalized for unseen tasks to reduce the user's electromyography signals (EMGs) during tasks. |
| 最近的研究表明，通过用户与机器人之间的物理交互，强化学习具有在外骨骼中生成辅助策略的巨大潜力。 先前的方法专注于特定于任务的辅助策略，其中对于每个任务（情况/上下文），用户都需要与机器人进行交互以学习适当的辅助策略。 因此，学到的策略不能推广到新任务上。 由于采样成本对于诸如外骨骼这样的“人在环”系统来说是昂贵的，因此必须实现通用化。 在本文中，我们建议学习外骨骼机器人的任务参数化辅助策略。 我们的方法采用了一种辅助策略，该策略取决于任务参数和状态变量，可以从跨不同任务的多组人机交互数据中获知，并且即使对于看不见的任务也可以将其推广，只要任务参数无需额外学习即可。 为了减轻跨多个任务的学习过程中用户的负担，我们开发了一种数据有效的多任务强化学习框架。 为了验证我们方法的有效性，我们开发了带有外骨骼机器人的实验平台。 我们进行了一系列实验，实验结果表明，我们的方法可以学习这种任务参数化的辅助策略，并且可以针对看不见的任务进行推广，以减少任务期间用户的肌电信号（EMG）。 |

## 2. Learning assistive strategies for exoskeleton robots from user-robot physical interaction

Hamaya M, Matsubara T, Noda T, et al. **Learning assistive strategies for exoskeleton robots from user-robot physical interaction**[J]. Pattern Recognition Letters, 2017: 67-76.

|  |
| --- |
| **Keywords:**  Exoskeleton robot; Human-robot physical interaction; Human-in-the- loop; Reinforcement learning  **Abstract**  Social demand for exoskeleton robots that physically assist humans has been increasing in various situations due to the demographic trends of aging populations. With exoskeleton robots, an assistive strategy is a key ingredient. Since interactions between users and exoskeleton robots are bidirectional, the assistive strategy design problem is complex and challenging. In this paper, we explore a data-driven learning approach for designing assistive strategies for exoskeletons from user-robot physical interaction. We formulate the learning problem of assistive strategies as a policy search problem and exploit a data-efficient model-based reinforcement learning framework. Instead of explicitly providing the desired trajectories in the cost function, our cost function only considers the user’s muscular effort measured by electromyography signals (EMGs) to learn the assistive strategies. The key underlying assumption is that the user is instructed to perform the task by his/her own intended movements. Since the EMGs are observed when the intended movements are achieved by the user’s own muscle efforts rather than the robot’s assistance, EMGs can be interpreted as the “cost” of the current assistance. We applied our method to a 1-DoF exoskeleton robot and conducted a series of experiments with human subjects. Our experimental results demonstrated that our method learned proper assistive strategies that explicitly considered the bidirectional interactions between a user and a robot with only 60 seconds of interaction. We also showed that our proposed method can cope with changes in both the robot dynamics and movement trajectories. |
| 由于人口老龄化的趋势，在各种情况下，对物理上辅助人类的外骨骼机器人的社会需求一直在增长。 对于外骨骼机器人，辅助策略是关键要素。 由于用户与外骨骼机器人之间的交互是双向的，因此辅助策略设计问题既复杂又具有挑战性。 在本文中，我们探索了一种数据驱动的学习方法，用于从用户与机器人的物理交互中设计外骨骼的辅助策略。 我们将辅助策略的学习问题表述为策略搜索问题，并利用基于数据有效模型的强化学习框架。 我们的成本函数并未考虑在成本函数中明确提供所需的轨迹，而是仅考虑通过肌电图信号（EMG）测量的用户的肌肉力量来学习辅助策略。 关键的基本假设是通过用户自己的预期动作指示用户执行任务。 由于在通过用户自己的肌肉力量而不是机器人的帮助实现预期运动时会观察到EMG，因此EMG可以解释为当前帮助的“成本”。 我们将我们的方法应用于1-DoF外骨骼机器人，并针对人类受试者进行了一系列实验。 我们的实验结果表明，我们的方法学习了正确的辅助策略，该策略明确考虑了用户和机器人之间的双向交互（交互仅需60秒）。 我们还表明，我们提出的方法可以应对机器人动力学和运动轨迹的变化。 |

## 3. Wearer-Prosthesis Interaction for Symmetrical Gait: A Study Enabled by Reinforcement Learning Prosthesis Control

Y. Wen, M. Li, J. Si and H. Huang, "**Wearer-Prosthesis Interaction for Symmetrical Gait: A Study Enabled by Reinforcement Learning Prosthesis Control**," in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 4, pp. 904-913, April 2020, doi: 10.1109/TNSRE.2020.2979033.

|  |
| --- |
| **Abstract:**  With advances in robotic prostheses, rese-archers attempt to improve amputee's gait performance (e.g., gait symmetry) beyond restoring normative knee kinematics/kinetics. Yet, little is known about how the prosthesis mechanics/control influence wearer-prosthesis' gait performance, such as gait symmetry, stability, etc. This study aimed to investigate the influence of robotic transfemoral prosthesis mechanics on human wearers' gait symmetry. The investigation was enabled by our previously designed reinforcement learning (RL) supplementary control, which simultaneously tuned 12 control parameters that determined the prosthesis mechanics throughout a gait cycle. The RL control design facilitated safe explorations of prosthesis mechanics with the human in the loop. Subjects were recruited and walked with a robotic transfemoral prosthesis on a treadmill while the RL controller tuned the control parameters. Stance time symmetry, step length symmetry, and bilateral anteroposterior (AP) impulses were measured. The data analysis showed that changes in robotic knee mechanics led to movement variations in both lower limbs and therefore gait temporal-spatial symmetry measures. Consistent across all the subjects, inter-limb AP impulse measurements explained gait symmetry: the stance time symmetry was significantly correlated with the net inter-limb AP impulse, and the step length symmetry was significantly correlated with braking and propulsive impulse symmetry. The results suggest that it is possible to personalize transfemoral prosthesis control for improved temporal-spatial gait symmetry. However, adjusting prosthesis mechanics alone was insufficient to maximize the gait symmetry. Rather, achieving gait symmetry may require coordination between the wearer's motor control of the intact limb and adaptive control of the prosthetic joints. The results also indicated that the RL-based prosthesis tuning system was a potential tool for studying wearer-prosthesis interactions. |
| 随着机器人假肢的进步，重新弓箭手试图改善被截肢者的步态表现（例如，步态对称性），而不是恢复正常的膝盖运动学/运动学。 然而，关于假体力学/控制如何影响穿戴者假体的步态性能（如步态对称性，稳定性等）的了解甚少。本研究旨在研究机器人股骨假体力学对人类穿戴者步态对称性的影响。 我们先前设计的强化学习（RL）辅助控制使这项研究得以实现，该辅助控制同时调整了12个控制参数，这些参数确定了整个步态周期中的假体力学。 RL控制设计有助于在人为干预的情况下安全地探索假体力学。 招募受试者并用机器人经股假体在跑步机上行走，而RL控制器调整控制参数。 测量姿态时间对称性，步长对称性和双侧前后（AP）脉冲。 数据分析表明，机器人膝关节力学的变化导致下肢的运动发生变化，从而导致步态时空对称性测量。 在所有受试者中，肢间AP脉冲测量结果与步态对称性一致：站立时间对称性与净肢间AP脉冲显着相关，步长对称性与制动和推进脉冲对称性显着相关。 结果表明，可以个性化经股假体控制，以改善时空步态的对称性。 然而，仅调整假体力学不足以使步态对称最大化。 相反，要实现步态对称，可能需要在穿戴者对完整肢体的运动控制与义肢关节的自适应控制之间进行协调。 结果还表明，基于RL的假体调整系统是研究佩戴者与假体相互作用的潜在工具。 |

## 4. Online Reinforcement Learning Control for the Personalization of a Robotic Knee Prosthesis

Y. Wen, J. Si, A. Brandt, X. Gao and H. H. Huang, **"Online Reinforcement Learning Control for the Personalization of a Robotic Knee Prosthesis**," in IEEE Transactions on Cybernetics, vol. 50, no. 6, pp. 2346-2356, June 2020, doi: 10.1109/TCYB.2019.2890974.

|  |
| --- |
| **Abstract:**  Robotic prostheses deliver greater function than passive prostheses, but we face the challenge of tuning a large number of control parameters in order to personalize the device for individual amputee users. This problem is not easily solved by traditional control designs or the latest robotic technology. Reinforcement learning (RL) is naturally appealing. The recent, unprecedented success of AlphaZero demonstrated RL as a feasible, large-scale problem solver. However, the prosthesis-tuning problem is associated with several unaddressed issues such as that it does not have a known and stable model, the continuous states and controls of the problem may result in a curse of dimensionality, and the human-prosthesis system is constantly subject to measurement noise, environmental change and human-body-caused variations. In this paper, we demonstrated the feasibility of direct heuristic dynamic programming, an approximate dynamic programming (ADP) approach, to automatically tune the 12 robotic knee prosthesis parameters to meet individual human users' needs. We tested the ADP-tuner on two subjects (one able-bodied subject and one amputee subject) walking at a fixed speed on a treadmill. The ADP-tuner learned to reach target gait kinematics in an average of 300 gait cycles or 10 min of walking. We observed improved ADP tuning performance when we transferred a previously learned ADP controller to a new learning session with the same subject. To the best of our knowledge, our approach to personalize robotic prostheses is the first implementation of online ADP learning control to a clinical problem involving human subjects.  机器人假肢比被动假肢具有更大的功能，但是我们面临着调整大量控制参数以使截肢者个性化设备的挑战。 传统的控制设计或最新的机器人技术不容易解决此问题。 强化学习（RL）自然很有吸引力。 AlphaZero最近取得的空前成功证明了RL是一种可行的大规模问题解决方案。 但是，假体调整问题与一些未解决的问题相关，例如，它没有已知的稳定模型，问题的连续状态和控制可能导致维数的诅咒，并且人工假体系统不断 可能会受到测量噪声，环境变化和人为因素的影响。 在本文中，我们演示了直接启发式动态编程（一种近似动态编程（ADP）方法）自动调整12个机器人膝关节假体参数以满足个人用户需求的可行性。 我们在跑步机上以固定速度行走的两个对象（一个身体健全的对象和一个被截肢的对象）测试了ADP调谐器。 ADP调谐器学会了**平均以300个步态周期或10分钟的步行达到目标步态运动学**。 当我们将先前学习的ADP控制器转移到具有相同主题的新学习会话时，我们观察到ADP调整性能得到了改善。 据我们所知，我们使机器人假体个性化的方法是对涉及人类受试者的临床问题的在线ADP学习控制的首次实施。 |

## 5. Teaching a humanoid robot to walk faster through Safe Reinforcement Learning

Garcia J, Shafie D. **Teaching a humanoid robot to walk faster through Safe Reinforcement Learning[J]. Engineering Applications of Artificial Intelligence**, 2020.

|  |
| --- |
| Abstract  Teaching a humanoid robot to walk is an open and challenging problem. Classical walking behaviors usually require the tuning of many control parameters (e.g., step size, speed). To find an initial or basic configuration of such parameters could not be so hard, but optimizing them for some goal (for instance, to walk faster) is not easy because, when defined incorrectly, may produce the fall of the humanoid, and the consequent damages. In this paper we propose the use of Safe Reinforcement Learning for improving the walking behavior of a humanoid that permits the robot to walk faster than with a pre-defined configuration. Safe Reinforcement Learning assumes the existence of a safe baseline policy that permits the humanoid to walk, and probabilistically reuse such a policy to learn a better one, which is represented following a case based approach. The proposed algorithm has been evaluated in a real humanoid robot proving that it drastically increases the learning speed while reduces the number of falls during learning when compared with state-of-the-art algorithms.  教人形机器人走路是一个开放且具有挑战性的问题。 经典的步行行为通常需要调整许多控制参数（例如步长，速度）。 找到这些参数的初始或基本配置并不难，但是针对某个目标（例如，走得更快）优化它们并不容易，因为如果定义不正确，可能会导致人形生物坠落，并因此而跌倒。 损害赔偿。 在本文中，我们建议使用安全强化学习来改善类人动物的行走行为，从而使机器人的行走速度比预先定义的配置更快。 安全强化学习假定存在允许人形生物行走的安全基线策略，并概率性地重用这种策略以学习更好的策略，该策略是基于案例的方法表示的。 所提出的算法已在真实的人形机器人中进行了评估，证明与最新算法相比，该算法可大大提高学习速度，同时减少学习过程中的跌倒次数。 |

## 6. Offline Policy Iteration Based Reinforcement Learning Controller for Online Robotic Knee Prosthesis Parameter Tuning

|  |
| --- |
| **Abstract:**  This paper aims to develop an optimal controller that can automatically provide personalized control of robotic knee prosthesis in order to best support gait of individual prosthesis wearers. We introduced a new reinforcement learning (RL) controller for this purpose based on the promising ability of RL controllers to solve optimal control problems through interactions with the environment without requiring an explicit system model. However, collecting data from a human-prosthesis system is expensive and thus the design of a RL controller has to take into account data and time efficiency. We therefore propose an offline policy iteration based reinforcement learning approach. Our solution is built on the finite state machine (FSM) impedance control framework, which is the most used prosthesis control method in commercial and prototypic robotic prosthesis. Under such a framework, we designed an approximate policy iteration algorithm to devise impedance parameter update rules for 12 prosthesis control parameters in order to meet individual users' needs. The goal of the reinforcement learning-based control was to reproduce near-normal knee kinematics during gait. We tested the RL controller obtained from offline learning in real time experiment involving the same able-bodied human subject wearing a robotic lower limb prosthesis. Our results showed that the RL control resulted in good convergent behavior in kinematic states, and the offline learning control policy successfully adjusted the prosthesis control parameters to produce near-normal knee kinematics in 10 updates of the impedance control parameters.  本文旨在开发一个机器人的最优控制器能自动提供个性化控制膝关节假体,以最好地支持单个假肢穿戴者的步态。我们引入一个新的强化学习(RL)为此控制器基于承诺RL控制器来解决最优控制问题的能力通过与环境的交互,而不需要显式的系统模型。然而,从human-prosthesis系统收集数据是昂贵的,因此RL控制器的设计必须考虑数据和时间效率。因此,我们提出基于离线策略迭代的强化学习方法。我们的解决方案是建立在有限状态机(FSM)阻抗控制框架,这是最常用的假肢控制方法在商业和被机器人假肢。在这样的一个框架,我们设计了一个近似策略迭代算法设计阻抗参数更新规则12假肢控制参数以满足个人用户的需求。基于强化学习的目标控制生育正常的膝盖在步态运动学。我们测试了RL控制器实时获得离线学习实验涉及相同的健全的人类穿着机器人下肢假肢。我们的研究结果表明,RL控制导致了运动状态,良好的收敛行为和离线学习控制策略成功地调整了假肢控制参数产生接近于正常的膝关节运动学在10阻抗控制参数的更新。 |

# 相关领域论文：

## 论文题目：Deep Predictive Policy Training using Reinforcement Learning

中文题目：深度预测策略的强化学习训练方法

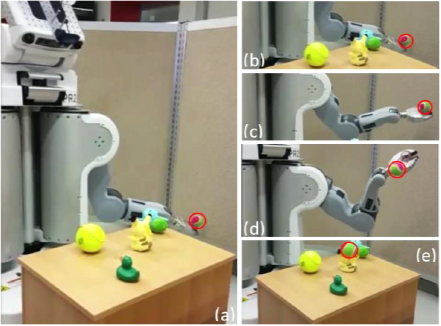
论文作者：Ali Ghadirzadeh, Atsuto Maki, Danica Kragic and Marten Bjorkman.

论文出处：Robotics: Science and Systems，2019

论文地址：https://arxiv.org/pdf/1903.11239.pdf

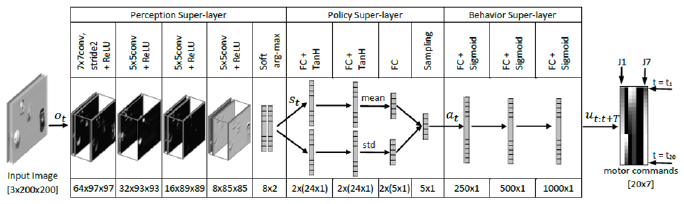
**研究问题：**

由于感知运动过程的固有延迟，机器人任务学习最好通过**预测动作策略**来实现。然而，训练这样的预测策略是具有挑战性的，因为它涉及到在整个动作过程中找到运动激活的轨迹。本文中，提出一个基于深度神经网络的**数据高效深度预测策略训练（DPPT）框架**，将图像观测映射到一系列的运动激活。该体系结构由三个子网络组成，分别称为感知层、策略层和行为层。感知层和行为层迫使我们对视觉和行为进行抽象分别用合成训练样本和模拟训练样本训练数据。策略层是一个较小的子网络，具有较少的参数来映射抽象流形之间的数据。使用策略搜索强化学习的方法对每个任务进行训练。通过在PR2 机器人上训练熟练抓取和投球的预测策略。下图表示在机器人抛掷ball 过程的瞬间图。



**研究方法：**

由感知层、策略层和行为层组成的深度预测策略体系结构如下图所示。作为网络输入，给出了一个中心RGB 图像。感知层将图像数据抽象为与任务相关的对象对应的若干空间位置。策略层将抽象状态随机映射到操作流形中的一个点。最后，针对给定的采样动作，行为层生成一长轨迹的电机指令，并应用于机器人连续T 个时间步长。



**研究结果：**

文章证明了所提出的结构和学习框架的适用性。该方法的有效性通过以下事实得到了证明:这些任务仅使用180 次真正的机器人进行训练，并提供定性的最终奖励。

## 2. 论文题目：Learning Agile and Dynamic Motor Skills for Legged Robots

**中文题目**：面向腿式机器人的敏捷动态特性的技能学习

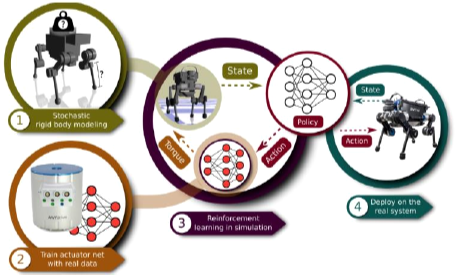
**论文作者**：Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Hoonho Lee, Vassilios Tsounis, Vladlen Koltun and Marco Hutter.

论文出处：Science Robotics, 2019

论文地址：https://arxiv.org/pdf/1901.08652.pdf

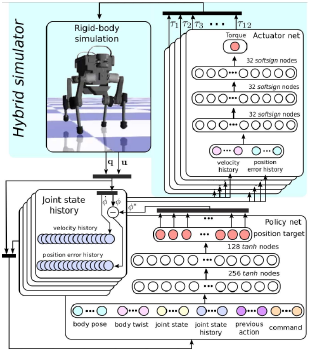
研究问题：

近年来，腿式机器人是机器人技术中最大的挑战之一。**动物的动态和敏捷的动作**是现有的方法无法模仿的，而这些方法是由人类精心设计的。一个令人信服的替代方案是**强化学习**，它需要**最少的技能并促使控制策略的自然演化更新**。然而，到目前为止，对腿式机器人的强化学习研究主要**局限于仿真，**在**实际系统中应用比较简单的例子较少**。主要原因是，使用真正的机器人进行训练，尤其是使用动态平衡系统，既复杂又昂贵。在本论文中，我们**提供了一种新的方法，在模拟中训练一个神经网络策略，并将其迁移到一个最先进的腿系统，因此我们利用快速、自动化和经济有效的数据生成方案**。



研究方法：

对于腿式机器人的敏捷动态性技能学习的过程，首先是**系统建模**，针对于四足机器人的物理参数的辨识以及确定机器人动态参数的不确定性指标，这个过程可能需要环境参数估计，物理动态性能估计等；其次是**训练驱动神经网络**，这个过程一般通过**构建机器人状态到机器人电机控制的映射函数实现**，随着深度神经网络的广泛认可，这样的非线性映射函数现大多采用深度神经网络拟合；然后在**仿真中完成基于强化学习的驱动神经网络的学习过程**，最后将训练好的驱动神经网络拟合的控制策略**应用在实际的系统中**。



整个系统的控制网络由三部分构成，首先是**策略网络**，用于将当前的观测量和之前的关节状态量映射到目标关节量（下一时刻关节控制量），然后是**驱动网络**，用于在刚体关节控制中将**历史关节状态**映射到**关节力矩控制量**上，机器人状态量有各关节的位置信息q 与速度信息u。

**研究结果：**

应用于一个复杂的中型犬大小的四足系统ANYmal 机器人，使得在模拟中训练的四足机器人的运动策略超越了以前的方法，ANYmal 能够精确和高效地遵循高水平的身体速度指令，比以前跑得更快，甚至在复杂的配置中也能从跌倒中恢复过来。