Title:
Trajectory Planning of Rehabilitation Exercises using an Integrated Reward Function in Reinforcement Learning

Authors:
Yanlin Shi, shiy3418@myumanitoba.ca, University of Manitoba
Qingjin Peng, qingjin.peng@umanitoba.ca, University of Manitoba
Jian Zhang, jianzhang@stu.edu.cn, Shantou University

Keywords:
Reinforcement learning, Reward function, Rehabilitation, Trajectory planning

DOI: 10.14733/cadconfP.2021.187-191

Introduction:
Rehabilitation devices help patients to recover injured body parts such as elbow and knee joints [3]. Trajectory planning of rehabilitation exercises determines a suitable moving path to guide patients in daily recovery activities for body parts based on injured levels and joints [4]. It is expected that the rehabilitation process is smooth and comfortable. The existing trajectory planning are mainly manual methods that require physicians to plan the rehabilitation exercise trajectory [7], which is inefficient and inaccurate [1].

Reinforcement learning (RL) uses intelligent agents to plan actions in environments for maximum rewards [5]. Using RL, a rehabilitation device can autonomously learn and plan a trajectory for required exercise actions in different conditions. Based on the range of rotation angles and movement speed required in the rehabilitation of patients, a reward function can generate the optimal trajectory for patients to approach the target position in rehabilitation exercises efficiently and accurately [6].

An integrated reward function is proposed in this paper to plan the trajectory of rehabilitation exercises. Based on injured joints of a patient recorded by motion sensors, the range of rotation angles and movement speeds are restricted and planed for the patient using RL. The rotation angles and movement speeds are reset for injured joints based on the daily progress of the patient recovery to improve performance of the rehabilitation.

Main Idea:
Azimuth reward function for rehabilitation exercise trajectory:
Based on conditions of injured joints of a patient recorded by motion sensors, the maximum rotation angle and movement speed of injured parts can be defined. An Azimuth reward function is proposed to restrict the rotation angle of the exercise trajectory as follows.

$$R_a^i = \begin{cases} 1 - \dfrac{0.8\theta_i^{\max} - \theta_i}{0.8\theta_i^{\max}} & \theta_i \le \dfrac{4}{5}\theta_i^{\max} \\ \dfrac{\theta_i^{\max} - \theta_i}{\theta_i^{\max}} & \dfrac{4}{5}\theta_i^{\max} \le \theta_i \le \theta_i^{\max} \\ -\dfrac{\theta_i^{\max} - \theta_i}{\theta_i^{\max}} & \theta_i > \theta_i^{\max} \end{cases} \quad (1)$$

where θ_i is the i_{th} rotation angle of rehabilitation device. $\theta_i^{\max}$ is the maximum rotation angle.

Speed reward function for rehabilitation exercise trajectory:

The speed is defined by balancing efficiency and comfort of the rehabilitation exercise based on conditions of injured joints of the patient using Eqn. (2).

$$R_s = \begin{cases} 0 & V > V_m \\ \dfrac{V_m - V}{V_m} & V_a < V < V_m \\ 1 - \dfrac{V_a - V}{V_a} & 0 < V < V_a \end{cases} \tag{2}$$

where V is the speed at an end point of the rehabilitation device. V_m is the maximum acceptable speed of injured joints for a patient. V_a is the average movement speed of a joint for healthy people. When the maximum acceptable speed of injured joints for a patient is lower than the average movement speed of a joint for healthy people, the speed reward function is defined as follows.

$$R_s = \begin{cases} 0 & V > V_m \\ 1 - \dfrac{V_m - V}{V_m} & V < V_m \end{cases} \tag{3}$$

Position reward function for rehabilitation exercise trajectory:

A position reward is defined to guide the device to approach the target point using Eqn. (4).

$$R_p = \frac{\max[d_{et}] - d_{et}}{\max[d_{et}]} \tag{4}$$

where d_{et} is the distance between an end point of the rehabilitation device and the target point.

Trajectory planning using defined reward functions:

By combining the Azimuth reward, position reward and speed reward, an integrated reward function for the rehabilitation exercise is defined in Eqn. (5). Where i is the total number of rotation angles of the rehabilitation device.

$$R_{total} = \frac{1}{3}\left(\frac{1}{i}\sum_{i=1}^{i} R_a^i + R_s + R_p\right) \tag{5}$$

A flowchart of the training process is shown in Fig. 1. After training the RL model based on Eqn. (5), a trajectory can be generated to guide patients in daily recovery exercises. Based on the progress of the patient recovery, the rotation angle and movement speed are reset to update the exercise trajectory using the trained model in RL.

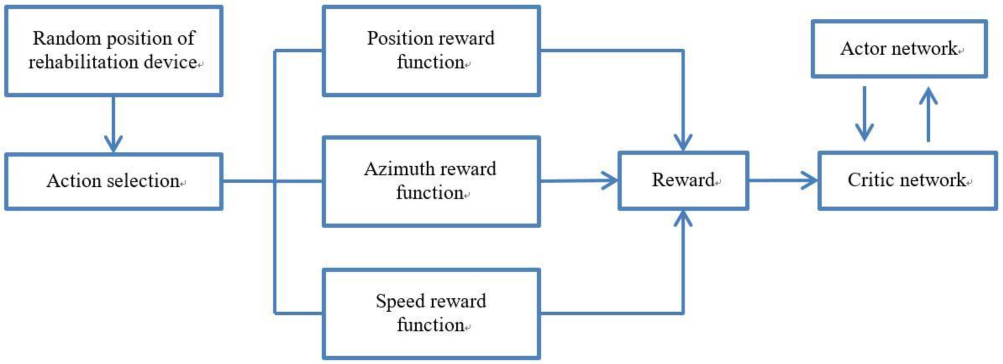
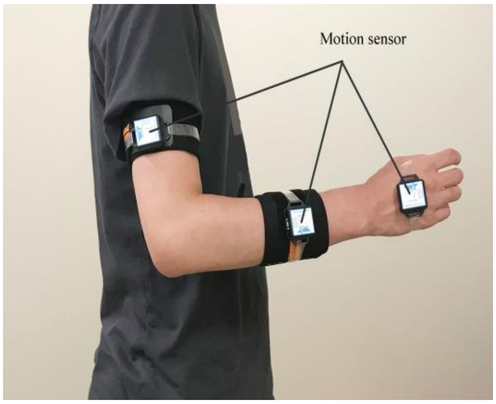


Fig. 1: The training process based on the reward function.



a. Rotation angles test

b. Rotation angles in the exercise

Fig. 2: Rotation angles of the patient.

No.	*Rotation angles of arms*	*Maximum value*
1	shoulder flexion angle θ_1	90°
2	shoulder abduction angle θ_2	120°
3	elbow flexion angle θ_3	90°
4	forearm pronation angle θ_4	70°
5	wrist flexion angle θ_5	35°
6	wrist radial deviation angle θ_6	50°

Tab. 1: Maximum rotation angles of patient's arm.

Case study:

A case study of the upper limb rehabilitation exercise for a patient with the injury arm is used to verify the proposed RL method for trajectory planning of the rehabilitation exercise. Injured levels and joints of a patient are recorded by motion sensors to determine the rotation angles and movement speed of the patient as shown in Fig. 2a. Maximum rotation angles of patient's arm are defined as shown in Tab. 1. There are six rotation angles including shoulder flexion angle θ_1, shoulder abduction angle θ_2, elbow flexion angle θ_3, forearm pronation angle θ_4, wrist flexion angle θ_5 and wrist radial deviation angle θ_6. The maximum acceptable speed V_p of the arm movement is 0.45 m/s based on the injured level of the patient. The average movement speed of the arm movement for healthy people is 0.6 m/s in daily activities.

Based on maximum rotation angles of patient's arm in Tab. 1 and maximum acceptable speed of the arm movement, a reward function is defined using Eqn. (5). Unity ML-Agents are used for the sysem implementation of upper limb rehabilitaiton planning as shown in Fig. 3 [2]. The model of upper limb rehabilitaiton device is trained in Unity. After training the model, the upper limb rehabilitaiton device can guide patient's injured arm to complete daily revovery exercise.

Using the reward function in training RL agents, the convergence of the system is reached as shown in Fig. 4. The RL model convergence is achieved after 40000 episodes. The training process took approximately 6 hours. The trained agent is then used to generate the exercise trajectory as shown in Fig. 5.

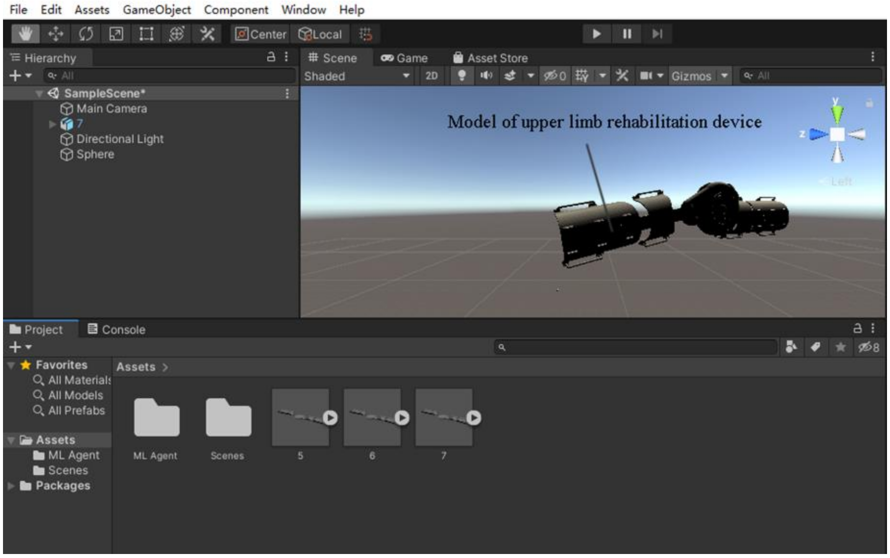
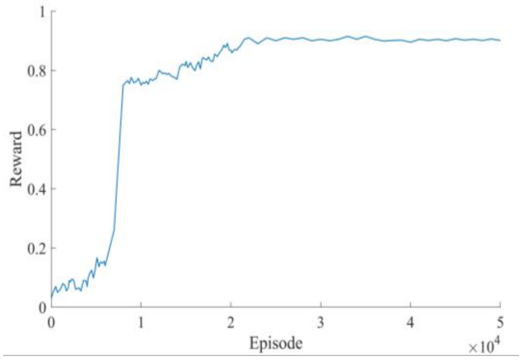


Fig. 3: Simulation environment of the rehabilitation device using Unity ML-Agents.



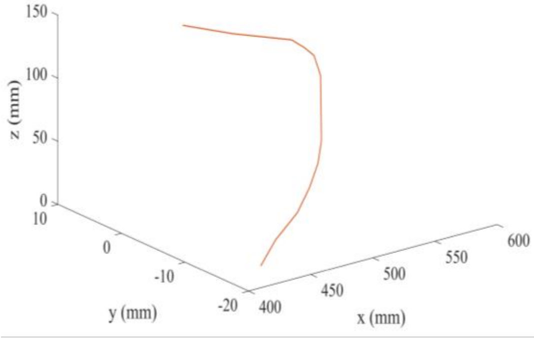


Fig. 4: Convergence of the RL model. Fig. 5: Recovery trajectory of end point of the arm.

The trajectory is used by the device to guide the patient to complete daily recovery exercise. Based on the recovery progress monitored by motion sensors, the trajectory is regenerated to update the rotation angle and movement speed as shown in Fig. 6. The maximum elbow rotation angle in the first week is 90 degree. The maximum elbow rotation angles in the second and third week are 96 and 101 degrees, respectively.

Conclusion:

This paper proposed an integrated reward function in the RL model for trajectory planning of rehabilitation exercises. The reward function was proposed to restrict and optimize the range of rotation angles and movement speed for patients with different injured levels and joints. A trajectory is generated automatically for rehabilitation exercises of patients to improve the daily performance of the recovery. By resetting rotation angles and movement speed of injured joints of patients based on the recovery progress, the restriction angles and speed are updated to generate new trajectory using the trained RL model. The performance of the rehabilitation is improved.

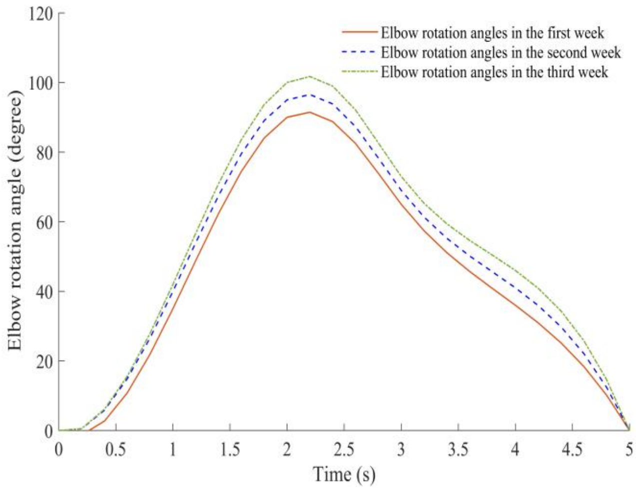


Fig. 6: Elbow rotation angles of the patient.

Acknowledgements:
The authors wish to acknowledge that this research has been supported by the Discovery Grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada, University of Manitoba Graduate Fellowships (UMGF) and the Graduate Enhancement of Tri-Council Stipends (GETS) program from the University of Manitoba.

References:
[1] Fan, T.; Long, P.; Liu, W.; Pan, J.: Distributed multi-robot collision avoidance via deep reinforcement learning for navigation in complex scenarios, The International Journal of Robotics Research, 39(7), 2020, 856-892. https://doi.org/10.1177/0278364920916531
[2] Juliani, A.; Berges, V. P.; Teng, E., Cohen, A.; Harper, J.; Elion, C.; Lange, D.: Unity: A general platform for intelligent agents. arXiv preprint arXiv:1809.02627, 2018.
[3] Maciejasz, P.; Eschweiler, J.; Gerlach-Hahn, K.; Jansen-Troy, A.; Leonhardt, S.: A survey on robotic devices for upper limb rehabilitation, Journal of neuroengineering and rehabilitation, 11(1), 2014, 1-29. https://doi.org/10.1186/1743-0003-11-3
[4] Miao, Q.; Zhang, M.; Cao, J.; Xie, S. Q.: Reviewing high-level control techniques on robot-assisted upper-limb rehabilitation, Advanced Robotics, 32(24), 2018, 1253-1268. https://doi.org/10.1080/01691864.2018.1546617
[5] Niroui, F.; Zhang, K.; Kashino, Z.; Nejat, G.: Deep reinforcement learning robot for search and rescue applications: Exploration in unknown cluttered environments, IEEE Robotics and Automation Letters, 4(2), 2019, 610-617. https://doi.org/10.1109/LRA.2019.2891991
[6] You, C.; Lu, J.; Filev, D.; Tsiotras, P.: Advanced planning for autonomous vehicles using reinforcement learning and deep inverse reinforcement learning, Robotics and Autonomous Systems, 114, 2019, 1-18. https://doi.org/10.1016/j.robot.2019.01.003
[7] Zhang, S.; Guo, S.; Fu, Y.; Boulardot, L.; Huang, Q.; Hirata, H.; Ishihara, H.: Integrating compliant actuator and torque limiter mechanism for safe home-based upper-limb rehabilitation device design, Journal of Medical and Biological Engineering, 37(3), 2017, 357-364. https://doi.org/10.1007/s40846-017-0228-2