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

Yanlin Shi[1], Qingjin Peng[2] and Jian Zhang[3]

[1]University of Manitoba, shiy3418@myumanitoba.ca
[2]University of Manitoba, Qingjin.Peng@umanitoba.ca
[3]Shantou University, jianzhang@stu.edu.cn

Corresponding author: Qingjin Peng, Qingjin.Peng@umanitoba.ca

Abstract. Trajectory planning of rehabilitation devices determines movements in daily recovery activities of patients based on injured levels and joints. The existing trajectory planning methods are mainly manual methods based on the experience of therapists, which is inefficient and inaccurate. Reinforcement learning (RL) uses intelligent agents to plan actions in an environment for the maximum reward. Using RL, a rehabilitation device can autonomously learn to plan a trajectory for required exercise actions in different conditions of injured joints and levels of patients. An integrated reward function is proposed in this paper using a RL model to plan the trajectory of rehabilitation exercises. Based on injured joints of a patient recorded by motion sensors, ranges of rotation angles and speeds are restricted and planed for the patient using RL. Exercise actions can be easily reset for injured joints based on the daily progress of the patient recovery using the trained RL model to improve performance of the rehabilitation exercise.

Keywords: Reinforcement learning, Reward function, Rehabilitation, Trajectory planning
DOI: https://doi.org/10.14733/cadaps.2022.1042-1054

1 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. Based on injured levels of patients, therapists plan a series of rehabilitation trajectories for the rehabilitation device to guide patients in daily recovery activities. During the rehabilitation exercise, patients recover their injured body parts such as arms and legs following the planned trajectory [4]. It is expected that the rehabilitation process is progressed for rehabilitation devices to take patients' injured body part to complete daily actions in the rehabilitation exercise.

The existing method of trajectory planning is mainly a manual method by therapists to plan the rehabilitation exercise trajectory [29], which is inefficient and inaccurate [7]. Different patients

have different injured levels for rehabilitation. Therapists have to plan a specific rehabilitation exercise trajectory for each individual patient. The exercise difficulty levels may not be able to be updated on time for the rehabilitation, which impacts performance of the rehabilitation exercise. Therefore, it is necessary to have an efficient method to plan a suitable trajectory for patients in different conditions.

With the development of reinforcement learning (RL), an artificial intelligent (AI) agent can simulate a therapist to plan a suitable trajectory for rehabilitation exercises. The trajectory can be generated automatically based on injured body parts and levels of patients [11]. RL uses intelligent agents to plan actions in an environment for the maximum reward [15]. 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 speeds required in the rehabilitation, a reward function can generate the optimal trajectory for patients to approach the target position in rehabilitation exercises efficiently and accurately [28].

Machine learning agent enables the simulation of an environment in training intelligent agents for rehabilitation exercises. The model of a rehabilitation device can be applied for trajectory planning. By simulating the human therapist using the machine learning agent, a trajectory can be automatically generated.

This paper proposes an automatic trajectory planning method based on the injured level of patients using RL. The injured level of patients is examined to define the exercise objective. An integrated reward function is proposed to plan the trajectory of rehabilitation exercises. Based on injured joints of a patient recorded by motion sensors, the range of rotation angles and speeds are restricted and planed for the patient using RL. The rotation angles and speeds are reset for injured joints based on the daily progress of the patient recovery to improve performance of the rehabilitation.

Following parts of the paper are organized as follows. Literature review of the related work is discussed in Section 2. A trajectory planning method is proposed using RL in Section 3. Section 4 is a case study for trajectory planning of the upper limb rehabilitation device using the proposed method. Section 5 is the solution evaluation and discussion, followed by the research conclusion in Section 6.

2 LITERATURE REVIEW

2.1 Existing Methods for Trajectory Planning of Rehabilitation Exercise

Existing methods for trajectory planning of rehabilitation exercises including kinematic analysis, analytic hierarchy process (AHP) and multi-objective optimization.

Kinematic analysis plans a trajectory for the rehabilitation device based on coordinates of each link of the device. For example, Yang et al proposed a rehabilitation trajectory by simulating a 4-DOF upper limb rehabilitation device based on parameters of structures and connecting rods [27]. Wang et al planed the human-machine motion of a rigid-flexible hybrid lower limb rehabilitation device based on parameters of the device mechanism [23]. Liao et al proposed the ankle injury rehabilitation using the screw theory based on a mathematical model of the parallel mechanism [13].

AHP method compares different trajectories to determine the best rehabilitation trajectory for patients. For example, Wang et al combined fuzzy AHP and QFD methods to improve the design of a hand training device [22]. Vaida et al defined a high degree of universality for the exercise trajectory by modeling design characteristics and constraints for shoulder rehabilitation devices [20]. Xing et al proposed a motor-function evaluation method for improving the trajectory accuracy in rehabilitation exercise using a fuzzy AHP method and online self-correction function synthetical evaluation model [26].

Multi-objective optimization method searches for the best rehabilitation trajectory by balancing different rehabilitation requirements such as the movement speed and force applied in injured parts. Hamida et al improved movement ranges of a trajectory for upper-limb rehabilitation

exercises using a multi-objective genetic algorithm [9]. Huang et al proposed a multi-objective optimal trajectory planning model based on parameters such as the motion time, dynamic disturbance, and jerk using a multi-objective particle swarm optimization method [10]. Abe et al proposed an improved trajectory planning method to reduce the cost of energy for the vibration control of a flexible manipulator based on the maximum residual vibration amplitude and operating energy [1].

However, the trajectory planning for rehabilitation is mainly a manual method that requires therapists to plan the exercise trajectory based on injured levels of patients, which is inefficient and inaccurate. In addition, it is difficulty to update trajectory for patients in the daily recovery exercise on time.

2.2 RL for Trajectory Planning

Reinforcement learning (RL) takes actions in an environment to maximize the cumulative reward based on reward functions using intelligent agents. There are two kinds of model-free algorithms for RL including value-based and policy-based RL.

Value-based RL includes Q-learning and Deep Q-Network (DQN) methods to find the optimal value function for describing relations between states and actions of the agent. Marchesini et al proposed an approach to reduce the training time for determining the maples navigation based on Double Deep Q-Network in parallel asynchronous training [14]. Sallab et al used a DQN framework for autonomous driving with improved convergence and performance using a lane keeping assist function [17]. Du et al developed a trajectory planning method for automated parking systems using Q-learning method to improve performance of the parking agent [6].

Policy-based RL includes deep deterministic policy gradient (DDPG) and asynchronous advantage actor-critic (A3C) methods to find the optimal policy in the high-dimensional action space or continuous action spaces by optimizing parameters of policy functions. Wu et al proposed an online trajectory planning method by learning a policy that maps states to actions via trial and error in a simulation environment for a free-floating space robot using the DDPG method [24]. Guo et al built an autonomous path planning model for unmanned ships based on the continuous interaction with the environment and historical experience data of paths using the DDPG method [8]. Bouhamed et al developed an autonomous unmanned aerial vehicle (UAV) path planning framework by training the UAV to navigate through or over obstacles to reach its assigned target in the continuous action space using DDPG method [2].

Spaces of output actions yielded by value-based RL such as Q-learning and DQN are discrete, which cannot be applied for continuous action spaces like the trajectory planning of rehabilitation exercises. For the policy-based RL, the DDPG method shows the convergence efficiency in trajectory planning. Therefore, the DDPG method is selected to train the agent in this research.

2.3 Reward Functions for Trajectory Planning

For generating trajectory automatically for rehabilitation exercises using RL, there are two types of reward functions including sparse reward function and continuous reward function.

The sparse reward function defines a trajectory based on the result of each action such as catching a target point using a rehabilitation device. A self-consistent trajectory can be generated using a spare reward function in RL with trajectory embedding for several simulated tasks [5]. For example, a learning paradigm was proposed in the presence of multiple sparse reward signals based on the sparse rewards of externally defined target tasks, learning by playing solving sparse reward tasks from scratch [16]. A model-free approach was developed using RL with sparse rewards using a deep deterministic policy gradient algorithm to solve the high dimensional control problems in trajectory planning [21].

The continuous reward function can continuously change based on environment observations and actions of the rehabilitation trajectory. A deep RL model was proposed to acquire navigation skills for wheel-legged robots in complex environments by adjusting an extra weight parameter for the reward function of path planning [4]. Xie et al used an optimized continuous reward function to

increase the convergence rate for the training model and improve the efficiency of robotic trajectory planning in the unstructured working environment with obstacles [25]. Tai et al used RL for the continuous control of mobile robots in the mapless navigation based on an improved reward function [19].

Sparse reward functions slow down learning because the agent needs many actions before getting any reward in trajectory planning. Continuous reward functions use a simple network structure to improve the convergence in training for rehabilitation exercises. In addition, a continuous reward function can obtain the global optimal solution. Therefore, the continuous reward function is used in this research to plan trajectory for rehabilitation exercises.

3 PROPOSED TRAJECTORY PLANNING METHOD

For defining the best trajectory for rehabilitation exercises, a RL-based method is proposed based on injured conditions of patients as shown in Figure 1.

3.1 Exercise Target and Reward Function

Based on conditions of injured joints of a patient recorded by motion sensors, the maximum rotation angle and speed of injured parts can be defined for defining injured levels and conditions of patients. According to the dimension of rehabilitation devices and maximum rotation angles for injured parts of a patient, the coordinate for current limiting motions of the patient can be defined using Equations (1) and (2).

$$ {}^{0}_{e}\mathrm{T} = {}^{0}_{1}\mathrm{T}\,{}^{0}_{2}\mathrm{T}\ldots{}^{n-1}_{n}\mathrm{T}\ldots{}^{n}_{e}\mathrm{T} \tag{1} $$

$$ {}^{n-1}_{n}T = Rot(X, \alpha_{n-1})Trans(X, a_{n-1})Rot(Z, \theta_n)Trans(Z, b_n) \tag{2} $$

where e is the degree of freedom. α_{n-1} is the rotation angle of linkage n-1. a_{n-1} is the length of linkage n-1. b_n is the offset of the linkage. θ_n is the maximum joint angle of the patient.

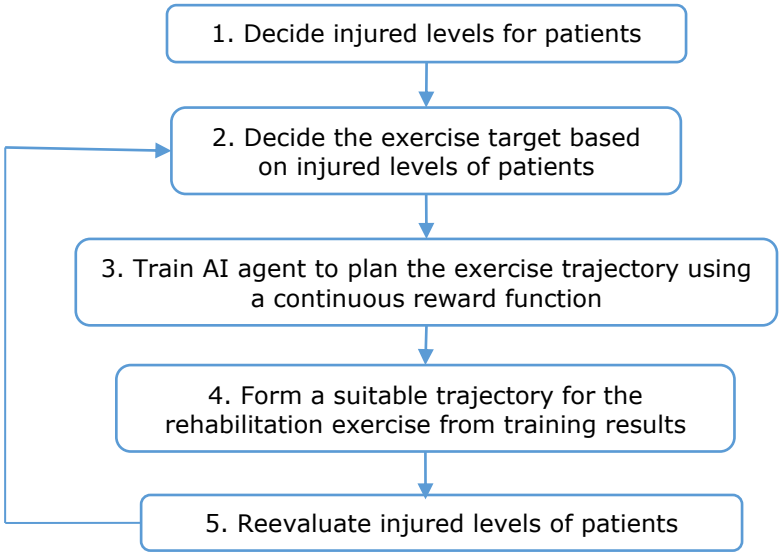


Figure 1: Proposed trajectory planning using RL.

A right-hand Cartesian coordinate system is used to represent joint angles of patient arms as follows. z is the direction from feet to head, x is the direction of the left-hand side of the patient, and y is the posterior direction of the patient. Based on the current limiting movement position for the patient, the coordinate of target points (*x, y, z*) can be defined for rotation angles in the rehabilitation exercise using Equations (3) and (4). For difficult levels of rehabilitation exercises and number of target points in a rehabilitation exercise, I is a set of integers as follows.

$$\begin{cases} (x-x_0)^2+(y-y_0)^2+(z-z_0)^2 \le r^2 \\ |x-x_0| \in I; |y-y_0| \in I; |z-z_0| \in I \end{cases} \tag{3}$$

$$r = (1/5) * l \tag{4}$$

where (x_0, y_0, z_0) is the coordinate point for current limiting movement. l is the length of injured parts such as an arm or leg.

Target point is the aim point and destination for the rehabilitation device to take patient injured parts in rehabilitation exercises. The start point is the origin of rehabilitation devices. Trajectory planning determines the path of rehabilitation devices from the starting point to target point. The difficulty of rehabilitation exercises is defined based on coordinates of the target point. For example, the height of a target point can determine shoulder rotation angles in the rehabilitation exercise.

Based on conditions of injured joints of a patient defined by Equations (1) to (4), the maximum rotation angle and speed of injured parts can be defined. An azimuth reward function is proposed to restrict rotation angles of the exercise trajectory as follows.

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

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

The speed is defined by balancing efficiency and comfort of rehabilitation exercises based on conditions of injured joints of the patient using Equation (6).

$$R_s = \begin{cases} 0 & V > V_m \\ \dfrac{V_m - V}{V_m} & V_a < V < V_m \\ 1 - \dfrac{V_a - V}{V_a} & 0 < V < V_a \end{cases} \tag{6}$$

where *V* is the speed at an end point of the rehabilitation device. V_m is the maximum 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.

$$R_s = \begin{cases} 0 & V > V_m \\ 1 - \dfrac{V_m - V}{V_m} & V < V_m \end{cases} \tag{7}$$

A position reward is defined to guide the device to approach the target point using Equation (8).

$$R_p = \frac{\max[d_{et}] - d_{et}}{\max[d_{et}]} \tag{8}$$

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

By combining the Azimuth reward, position reward and speed reward, an integrated reward function for the rehabilitation exercise is defined in Equation (9). Where i is the total number of rotation angles of a rehabilitation device.

$$R_{total} = \frac{1}{3}(\frac{1}{i}\sum_{i=1}^{i} R_a^i + R_s + R_p) \tag{9}$$

3.2 Exercise Trajectory Generation

The RL training process simulates a therapist to plan trajectory for rehabilitation exercises of patients using a rehabilitation device. The AI agent is the rehabilitation device in Figure 2. The environment is the patient and exercise space. The state is the distance between the endpoint of the rehabilitation device and target location. The action is the movement of a rehabilitation device model in the exercise space. The goal is to automatically find a suitable trajectory for the patient to exercise the injured arm with a suitable rotation angle and speed using the rehabilitation device to recover the movement function.

Steps for interactions between the agent and environment for trajectory planning are shown in Figure 2. For meeting the goal, a reward function is proposed by combining the Azimuth reward function, speed reward function and position reward function to train the agent to generate a trajectory for rehabilitation exercises. Based on the patient injuring assessment in Section 3.1, the maximum movement angle and speed of the patient are recorded to define a target point in the rehabilitation. The environment state changes based on the action. The rehabilitation device model acts based on the reward and environment state. Algorithm 1 shows the trajectory planning based on the proposed reward function using DDPG.

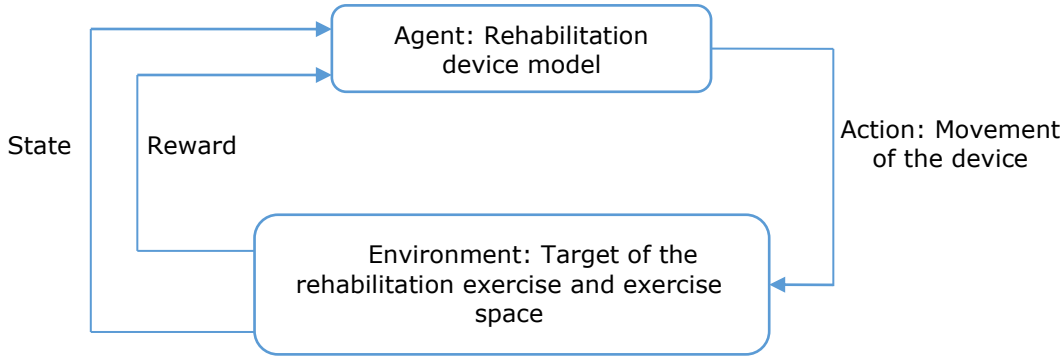


Figure 2: Interactions between AI agent and environment for trajectory planning.

Algorithm 1: Trajectory planning

Input: Environment status S ; maximum episode M; maximum training steps in each episode T.

Output: Action a

1: Initialize Actor Network $\mu(\mathrm{S}|\Theta^u)$ and Critic Network $Q(\mathrm{S}, a|\Theta^Q)$
2: for episode=1 to M do
3: for t=1 to T do
4: $a_t \leftarrow \mu(\mathrm{S}|\Theta^u)$
5: Determine value of R_t based on the proposed reward function
6: Update weight of Actor Network Θ^u
7: Update weight of Critic Network Θ^Q
8: end for
9: end for

As shown in Figure 2 and Algorithm 1, the training process of trajectory planning for rehabilitation exercises has 4 stages including the initialization of actor and critic network, action selection, reward calculation for current action and network training. At the first stage, the actor network and critic network for trajectory planning are initialized randomly. The critic network evaluates values for the action, and the actor network predicts the action to be performed. In the second stage, based on environment state S defined by relative distances between the target location and endpoint of the rehabilitation device and the value defined by the critic network, the actor network decides the action. In the third stage, a reward for the current action is computed by the proposed reward function for critic network in training and evaluation. In the last stage, weights of the network are updated considering environmental status S, action a and reward R comprehensively. By repeating the process in updating the environment state and agent action for training the agent, the rehabilitation device model can generate the best trajectory to approach the target location for the rehabilitation exercise.

After the RL model training based on the proposed reward function in Equation (9) and algorithm 1, a trajectory can be generated to guide the patient to approach the target location for daily recovery exercises. For avoiding a suddenly stop at the target point in rehabilitation exercises, the rehabilitation device moves patients' arm for a gradual stop at the end in the rehabilitation exercise.

Based on the trained agent, the daily updating trajectory can be automatically formed based on the progress of the patient recovery for the rehabilitation exercise.

3.3 Trajectory Planning for the Rehabilitation Exercise

Based on the trained model, the system can generate a trajectory to take patients' injured parts to approach the target in an accurate movement using rehabilitation devices in a rehabilitation exercise. Difficulty levels of rehabilitation exercises for patients are defined based on the distance between the target point of rehabilitation exercise and endpoint of the rehabilitation device using Equation (10).

$$d_i = \sqrt{(x_i - a)^2 + (y_i - b)^2 + (z_i - c)^2} \tag{10}$$

where (x_i, y_i, z_i) is the coordinate of the i_{th} target point. (a, b, c) is the coordinate of the endpoint of the rehabilitation device.

The endpoint of the rehabilitation device is normally fixed with extremity of injured parts such as hand or foot. The Euclidean distance between the target location and endpoint of the rehabilitation device can reflect the difficulty level for rehabilitation exercises. When the shortest distance d_1 is in between the target location p_1 and endpoint of the rehabilitation device, target point p_1 is defined as the easiest target point for the rehabilitation exercise. The first target point p_1 is defined as the easiest rehabilitation exercise and the last target point p_m is defined as the most difficult rehabilitation exercise.

The trajectory for approaching target location p_i is provided for a rehabilitation device to take patients' injured parts for recovery. After the rehabilitation exercise with the rehabilitation device, if the patient can move their injured parts such as arms and legs to repeat the trajectory movement, a more difficult target location p_j will be set to update the exercise using Equation (11). The patient can follow the device for more difficult rehabilitation exercises.

$$p_j = p_{i+1} \tag{11}$$

If the patient feels uncomfortable or cannot repeat the trajectory movement to approach target location p_i by their injured parts, an easier target location p_j will be set to update the exercise using Equation (12). The patient can follow the device for an easier rehabilitation exercise.

$$p_j = p_{i-1} \tag{12}$$

After completing all the rehabilitation exercises in this stage, the patient is reevaluated for the injured level. Based on progress of the patient recovery, the rotation angle and speed are reset to update the exercise trajectory using the trained RL model.

4 CASE STUDY

A case study of the upper limb rehabilitation exercise for a patient with an injured arm is used to verify the proposed RL method for trajectory planning of the rehabilitation exercise. Injured levels and joints of the patient are recorded by three motion sensors to determine rotation angles and speeds of the rehabilitation device as shown in Figure 3. The patient is required to follow a ball in the computer screen to complete some actions such as catching a virtual ball and touching a target point in the test of injury conditions.

Three motion sensors shown in Figure 3 record six rotation angles and speeds of attached arm parts. The first sensor is attached in the upper arm to record shoulder flexion angle θ_1 and shoulder abduction angle θ_2. The second sensor is located in the lower arm to record elbow flexion angle θ_3 and forearm pronation angle θ_4. The third sensor is fixed on the hand back to record wrist flexion angle θ_5, wrist radial deviation angle θ_6 and movement speed V_p of the hand.

Obtained maximum rotation angles of the arm are shown in Tab. 1, which shows that the patient has difficulty in four rotations: shoulder flexion angle θ_1, shoulder abduction angle θ_2, forearm pronation angle θ_4, and wrist flexion angle θ_5. Therefore, a trajectory is planned for the patient to recover the injured arm for these four rotation angles. For avoiding the risk of further injury in the rehabilitation process, the maximum acceptable speed V_p of the hand is decided as 0.45 m/s based on the record of the hand movement speed by the third motion sensor to restrict the speed in the rehabilitation exercise, compared to 0.6 m/s measured for the healthy people.

By using Equations (1) and (2), the current limiting movement positions for the patient is defined as (55, 79, 112). The coordinates for target points are defined using Equations (3) and (4). Based on the maximum rotation angles of patient's arm in Table 1 and maximum acceptable speed of the arm movement, a reward function is defined using Equations (5) to (9). Unity ML-Agents are used for the system implementation of upper limb rehabilitation planning as shown in Figure 4 [2]. The model of the upper limb rehabilitation device is trained in Unity. After training the model, the upper limb rehabilitation device can guide patient's injured arm to complete daily recovery exercises.

The proposed reward function is used in training RL agents. The model convergess after 40000 episodes as shown in Figure 5. The training time is approximately 6 hours. Using the Unity ML-Agents [11], a trajectory of the arm for the recovery exercise is formd as shown in Figure 6.

The trained system is applied to generate exercise trajectories of all target points using Equation (3). An exercise plan is then proposed using Equations (10) and (11) to decide a suitable rehabilitation exercise trajectory for the patient in recovery based on the daily progress of the exercise.

Comparison of the existing and proposed methods for trajectory planning is shown in Table 2. Using the existing method, therapists have to manually update the trajectory, which may not be able to catch progress of the patient recovery on time [20]. Our proposed method can generate exercise trajectory automatically based on data of patients to improve the efficiency of rehabilitation trajectory planning. In addition, the proposed method can update exercise difficulty levels based on the patient recovery on time, which improves performance of rehabilitation exercises and reduces time of the patient recovery.

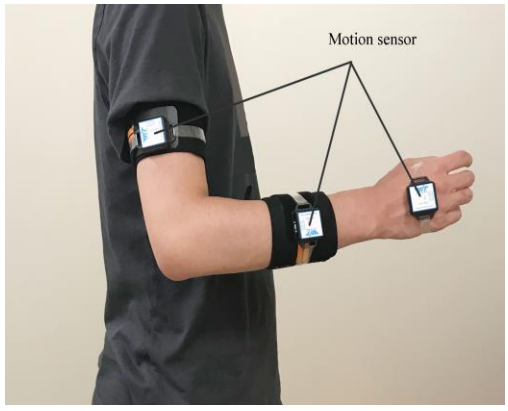


Figure 3: Rotation angles of the patient.

No.	*Rotation angles of arms*	*Maximum angles for the patient*	*Maximum angles for healthy people*
1	shoulder flexion angle θ_1	90°	180°
2	shoulder abduction angle θ_2	120°	150°
3	elbow flexion angle θ_3	90°	90°
4	forearm pronation angle θ_4	70°	90°
5	wrist flexion angle θ_5	35°	70°
6	wrist radial deviation angle θ_6	50°	50°

Table 1: Maximum rotation angles of the patient arm.

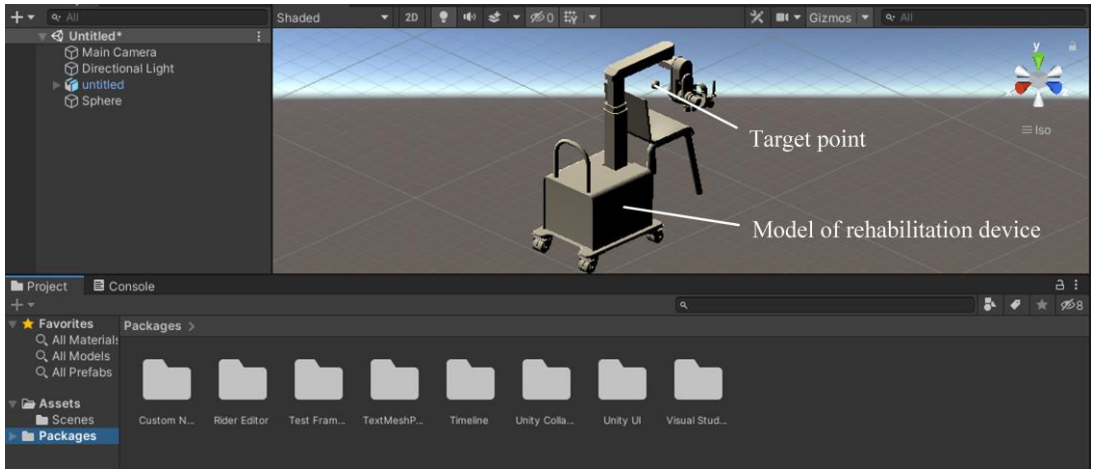
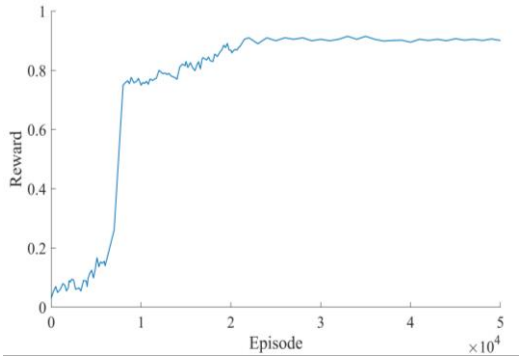


Figure 4: Simulation environment of the rehabilitation device in Unity.



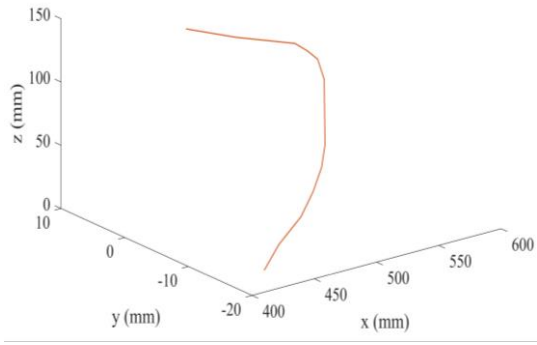


Figure 5: Convergence of the RL model. **Figure 6:** Recovery trajectory of end point of the arm.

No.	*Characters of trajectory*	*Trajectory defined by existing methods*	*Trajectory defined by the proposed method*
1	Time to plan trajectory	2-3 days	3-5 hours
2	Frequency for updating trajectory	5-7 days	15-30 minutes
3	Number of exercise levels	5-10 levels	50-80 levels
4	Generate trajectory automatically	No	Yes
5	Adjust the exercise difficulty levels during the exercise	No	Yes

Table 2: Comparison of the existing and proposed methods for trajectory planning.

5 SOLUTION EVALUATION AND DISCUSSION

For verifying the proposed trajectory planning method, the trajectory defined by the therapist manually using the kinematic analysis method [12] is compared with the trajectory formed by the proposed method as shown in Table 2.

The mannual method is time-consuming by therapists to decide rotation angles for updating the recovery trajectory. In addition, the frequency of updating trajectory cannot be always on time to meet the patient revocery progress. The exercise difficulty levels cannot be adjusted during the exercise, which affects the efficeincy of rehabilitation exercises.

The proposed method can generate rehabilitation trajectories for different difficulty levels automatically. The rehabilitation device can apply the best trajectory for a suitable exercise for patients in the recovery process. Patients can exercise their injured arms to improve the movement performance based on the daily progress. The proposed method is efficient in trajectory planning as all the rehabilitation trajectories are automatically defined by the AI agent compared to the conventional method.

6 CONCLUSIONS

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

The future work will consider the long-term performance of patient recovery in clinic applications to improve the proposed method. An optimal plan of the rehabilitation exercise will be developed for patients with different injury levels.

7 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.

Yanlin Shi, https://orcid.org/0000-0002-5537-5809
Qingjin Peng, http://orcid.org/0000-0002-9664-5326

REFERENCES

[1] Abe, A.; Komuro, K.: Minimum energy trajectory planning for vibration control of a flexible manipulator using a multi-objective optimisation approach, International Journal of Mechatronics and Automation, 2(4), 2012, 286-294. https://doi.org/10.1504/IJMA.2012.050499.

[2] Bouhamed, O.; Ghazzai, H.; Besbes, H.; Massoud, Y.: Autonomous UAV navigation: A DDPG-based deep reinforcement learning approach, In 2020 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2020, October, 1-5. https://doi.org/10.1109/ISCAS45731.2020.9181245.

[3] Buonamici, F.; Furferi, R.; Governi, L.; Lazzeri, S.; McGreevy, K.-S.; Servi, M.; Volpe, Y.: A CAD-based procedure for designing 3D printable arm-wrist-hand cast, Computer-Aided Design & Application, 16(1), 2019, 25-34. https://doi.org/10.14733/cadaps.2019.25-34.

[4] Chen, X.; Ghadirzadeh, A.; Folkesson, J.; Björkman, M.; Jensfelt, P.: Deep reinforcement learning to acquire navigation skills for wheel-legged robots in complex environments, In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) IEEE, 2018, October, 3110-3116. https://doi.org/10.1109/IROS.2018.8593702.

[5] Co-Reyes, J.; Liu, Y.; Gupta, A.; Eysenbach, B.; Abbeel, P.; Levine, S.: Self-consistent trajectory autoencoder: Hierarchical reinforcement learning with trajectory embeddings, In International Conference on Machine Learning PMLR, 2018, July, 1009-1018

[6] Du, Z.; Miao, Q.; Zong, C.: Trajectory Planning for Automated Parking Systems Using Deep Reinforcement Learning, International Journal of Automotive Technology, 21(4), 2020, 881-887. https://doi.org/10.1007/s12239-020-0085-9.

[7] 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.

[8] Guo, S.; Zhang, X.; Zheng, Y.; Du, Y.: An autonomous path planning model for unmanned ships based on deep reinforcement learning, Sensors, 20(2), 2020, 426. https://doi.org/10.3390/s20020426

[9] Hamida, I.-B.; Laribi, M.-A.; Mlika, A.; Romdhane, L.; Zeghloul, S.; Carbone, G.: Multi-Objective optimal design of a cable driven parallel robot for rehabilitation tasks, Mechanism and Machine Theory, 156, 2021, 104-141. https://doi.org/10.1016/j.mechmachtheory.2020.104141

[10] Huang, P.; Liu, G.; Yuan, J.; Xu, Y.: Multi-objective optimal trajectory planning of space robot using particle swarm optimization, In International Symposium on Neural Networks

Springer, Berlin, Heidelberg, 2008, September, 171-179. https://doi.org/10.1007/978-3-540-87734-9_20

[11] 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.

[12] Li, G., Fang, Q., Xu, T., Zhao, J., Cai, H., Zhu, Y.: Inverse kinematic analysis and trajectory planning of a modular upper limb rehabilitation exoskeleton, Technology and Health Care, 27(S1), 2019, 123-132. https://doi.org/10.3233/THC-199012.

[13] Liao, Z.; Yao, L.; Lu, Z.; Zhang, J.: Screw theory based mathematical modeling and kinematic analysis of a novel ankle rehabilitation robot with a constrained 3-PSP mechanism topology, International journal of intelligent robotics and applications, 2(3), 2018, 351-360. https://doi.org/10.1007/s41315-018-0063-9

[14] Marchesini, E., Farinelli, A.: Discrete deep reinforcement learning for map-less navigation, In 2020 IEEE International Conference on Robotics and Automation (ICRA) IEEE, 2020, May, 10688-10694. https://doi.org/10.1109/ICRA40945.2020.9196739.

[15] 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.

[16] Riedmiller, M.; Hafner, R.; Lampe, T.; Neunert, M.; Degrave, J.; Wiele, T.; Springenberg, J.-T.: Learning by playing solving sparse reward tasks from scratch, In International Conference on Machine Learning PMLR, 2018, July, 4344-4353

[17] Sallab, A.-E.; Abdou, M.; Perot, E.; Yogamani, S.: Deep reinforcement learning framework for autonomous driving, Electronic Imaging, 19, 2017, 70-76. https://doi.org/10.2352/ISSN.2470-1173.2017.19.AVM-023.

[18] Shi, Y.; Peng, Q.; Zhang, J.: An Objective Weighting Method of Function Requirements for Product Design Using Information Entropy, Computer-Aided Design & Application, 16(1), 2020, 966-978. https://doi.org/10.14733/cadaps.2020.966-978.

[19] Tai, L.; Paolo, G.; Liu, M.: Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation, In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) IEEE, 2017, September, 31-36. https://doi.org/10.1109/IROS.2017.8202134.

[20] Vaida, C.; Carbone, G.; Plitea, N.; Ulinici, I.; Pisla, D.: Preliminary design for a spherical parallel robot for shoulder rehabilitation, In New Advances in Mechanism and Machine Science, 2018, 155-164. Springer, Cham. https://doi.org/10.1007/978-3-319-79111-1_15.

[21] Vecerik, M.; Hester, T.; Scholz, J.; Wang, F.; Pietquin, O.; Piot, B.; Riedmiller, M.: Leveraging demonstrations for deep reinforcement learning on robotics problems with sparse rewards, 2017. arXiv preprint arXiv:1707.08817

[22] Wang, D.; Yu, H.; Wu, J.; Meng, Q.; Lin, Q.: Integrating fuzzy based QFD and AHP for the design and implementation of a hand training device, Journal of Intelligent & Fuzzy Systems, 36(4), 2019, 3317-3331. https://doi.org/10.3233/JIFS-181025.

[23] Wang, K.-Y.; Yin, P.-C.; Yang, H.-P.; Tang, X.-Q. The man-machine motion planning of rigid-flexible hybrid lower limb rehabilitation robot, Advances in Mechanical Engineering, 10(6), 2018. https://doi.org/10.1177/1687814018775865.

[24] Wu, Y.-H.; Yu, Z.-C.; Li, C.-Y.; He, M.-J.; Hua, B.; Chen, Z.-M.: Reinforcement learning in dual-arm trajectory planning for a free-floating space robot, Aerospace Science and Technology, 98, 2020, 105657. https://doi.org/10.1016/j.ast.2019.105657.

[25] Xie, J.; Shao, Z.; Li, Y.; Guan, Y.; Tan, J.: Deep reinforcement learning with optimized reward functions for robotic trajectory planning, IEEE Access, 7, 2019, 105669-105679. https://doi.org/10.1109/ACCESS.2019.2932257.

[26] Xing, L., Jianhui, W., Xiaoke, F. A kind of motor-function evaluation method for upper-limb rehabilitation robot, In Mechanical Engineering and Technology, Springer, Berlin, Heidelberg, 2012, 229-235. https://doi.org/10.1007/978-3-642-27329-2_32.

[27] Yang, L.-J.; Shen, L.-Y.; Ding, H.: The trajectory planning simulation of 4-DOF upper limb rehabilitation robot. Computer Simulation, 33(8), 2016, 332-337
[28] 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.
[29] 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.