Preference Augmented Q-Learning for Patient Exercise Scheduling in a Robotic Rehabilitation Gym

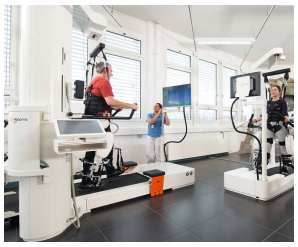
Varun R. Bharadwaj, Benjamin A. Miller, Vesna D. Novak and Chao Jiang

University of Wyoming

University of Cincinnati

Introduction

- **A Robotic gym** equipped with several rehabilitation robots enables multiple patients to exercise simultaneously.
- Optimally assigning multiple patients to a limited number of robots in real time is a complex task.
- An automated system can **learn patient-robot assignment strategies** from expert feedback and past outcomes.



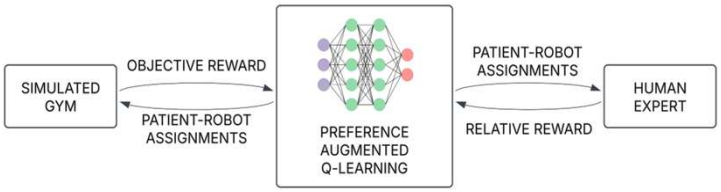
- But **real patient data** is scarce.
- **Simulate patient exercise** via a simulated robot gym.
- Patients have a **hidden skill level** for each associated exercise.
- Patients' **success rate and effort** are observable.
- Patients' **motivation and fatigue** are hidden and affect how they develop.



Preference Augmented Q-Learning

Our algorithm: Preference Augmented Q-Learning learns a neural network model encapsulating the underlying patient-robot assignment strategies, as well as infer hard-to-quantify factors like motivation and fatigue from expert feedback.

Training Procedure:



- **Data generation:** Generated 10000 simulated patient data where individual patients exercise for a 12-timestep exercise session.

Reward Design

- **Objective Rewards:** The algorithm is given rewards based on patient exercise characteristics.
  - The rate of change of patient success rate between consecutive timesteps.
  - If the patient is skilled enough to start exercising on higher difficulty level.
- **Relative Preference Reward:** Human feedback provides relative rewards.
  - Helps the algorithm learn preferences between different robot choices.
  - Helps infer about the effects of motivation and fatigue.

$$Relative\ Preference\ Reward = \left[\left(\frac{e^{R_e}}{e^{R_e} + e^{R_m}}\right) * 40\right] - 20$$

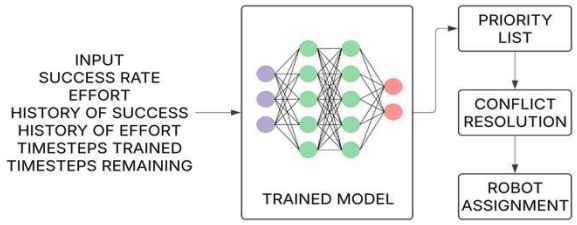
$$where$$

$$R_e = Q(s, a_e) + \sqrt{(Q(s, a_e) - Q(s, a_m))^2}$$

$$R_m = Q(s, a_m) - \sqrt{(Q(s, a_e) - Q(s, a_m))^2}$$

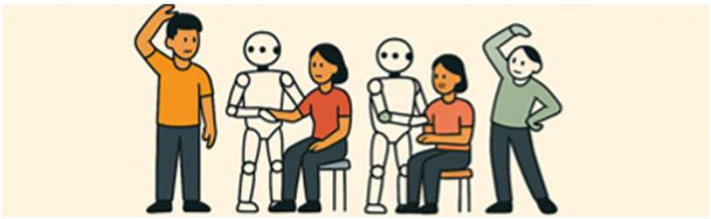
Patient-Robot Assignment Procedure

- **Trained Model:** Generates patient-robot assignment values based on patient characteristics.
- **Priority List:** Ranks prospective assignments by value.
- **Conflict Resolution:** Resolves any assignment conflicts.
- **Robot Assignment:** Patient-robot assignment based on priorities.



Evaluation Scenario

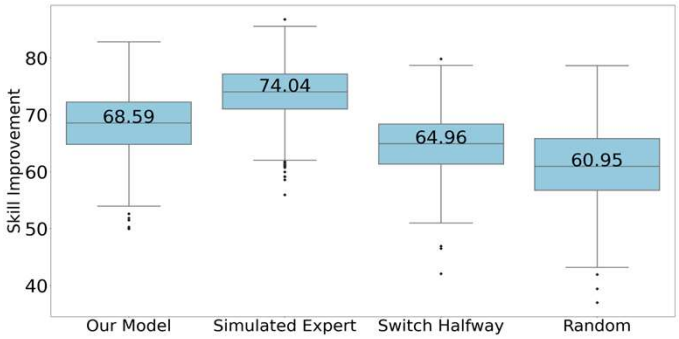
- Consisted of 5 simulated patients and robots, and lasted for 12 timesteps.
- Methodology tested using 1000 simulated patient groups.



Baselines:

- **Simulated Expert:** Patients are assigned to robots at each timestep based on their current characteristics, motivation and fatigue.
- **Switch Halfway:** In the beginning, patients are assigned to the robots expected to give the best results for one timestep, and halfway through the session they switch to their next best robot.
- **Random:** Patients are assigned to robots randomly at each timestep.

Results



- One way repeated-measures **ANOVA with Post-hoc Tukey** tests showed that all schedulers were different from each other at **$p < 0.05$**.
- Our model outperformed all the baselines except the expert, **without access to patient motivation and fatigue**. This shows the benefit of human-in-the-loop feedback.

Conclusion and Future Work

- Our algorithm implicitly incorporates hard-to-quantify factors through explicit expert patient-robot assignment feedback.
- Future work aims to reduce therapist feedback frequency and adapt to varying gym sizes, group dynamics, and stochasticity.