Optimizing Patient Scheduling Using Reinforcement Learning: Prioritizing Urgent Cases

By

Letaru Prudence Butele

MSDS

S25M19/007

B34977

Background

Public health facilities in Uganda often struggle with long queues, inefficient patient flow, and manual scheduling.

Patients arrive with varying urgency levels categorised in to two in this project:

- Urgent (Red) patients that should be served soon.
- Non-Urgent (Yellow) patients who can wait.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Problem statement

- Approaches that are commonly used in health facilities such as first-come-first-served ordering often result in;
- Delays for urgent patients
- Overcrowding
- Inefficient use of medical staff

Reinforcement Learning (RL) offers a dynamic solution to this problem.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Main Objective

To develop a Reinforcement Learning-based patient scheduling system that:

- Dynamically selects which patient to serve next
- Minimizes overall waiting times
- Prioritizes Red cases appropriately
- Adapts to changing arrival patterns

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Hypothesis

❑Null Hypothesis (H_0):

The reinforcement learning agent **does not significantly reduce** average waiting times compared to a baseline policy (such as random or first-come-first-served assignment of patients).

❑Alternative Hypothesis (H_1):

The reinforcement learning agent **significantly reduces** average waiting times compared to a baseline policy.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Markov Decision Process (MDP) Problem Formulation

❑State Space (S)

The State includes the current triage categories, accumulated waiting times, queue size as well as the number of available clinicians and their ongoing service durations.

❑Action Space (A)

The agent’s action is to select which patient to serve next from the queue.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

MDP Problem Formulation Cont’d

Reward Function (R)

Reward encourages the agent to prioritize urgent patients while minimizing waiting times for all categories.

Priority reward: Red = 40, Yellow = 30,

Threshold bonuses: If a patient is served within the waiting-time threshold (Red ≤15 min, Yellow ≤30 min), a positive reward is applied: Red = 40, Yellow = 30.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug

MDP Problem Formulation Cont’d

- **Transition Dynamics (P)**

After the agent selects a patient basing on the category;

- Service time is sampled for the selected patient and doctor.
- Simulation clock is advanced by minimum service time.
- Waiting times for of all the other patients is updated.
- Served patient is removed from queue.
- New arrivals are added based on stochastic process.
- Rewards computed.
- New observation and rewards are returned.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Algorithm Choice

Deep Q-Learning (DQN) is selected because the patient scheduling problem is a discrete-action, high-dimensional, and has a model-free environment.

- The scheduling decision of choosing which patient category to serve next forms a discrete action space, making DQN more suitable than continuous-action methods such as PPO or SAC.
- The environment involves stochastic patient arrivals, varying service times, and dynamic queue states, which cannot be modeled analytically; therefore, a model-free RL method is required.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Justification for Choosing the DQN Algorithm

- DQN's neural network function approximator allows it to handle the complex, multi-feature state space representing waiting times, urgency levels, and doctor availability.
- Additionally, DQN's value-based framework aligns with the objective of optimizing cumulative performance
- Techniques like experience replay and target networks stabilize learning
- These characteristics make DQN the most appropriate algorithm for optimizing patient flow.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Simulation & Training

Environment Setup

- The project uses a custom-built simulation environment in Python. The environment models the outpatient department as a time-driven system with the following components:
- **Dynamic queue:** containing the waiting patients, their urgency level, and waiting time.
- **Multiple doctors:** Each doctor has a service status (“free” or “busy”) and a remaining service time counter.
- **Variable service times:** Each patient category has an average service duration. The model uses approximate average service times for each category. Red ≈ 15 minutes, Yellow ≈10 minutes
- To implement randomness in service duration, these average values are modelled using a lognormal distribution.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Environment Setup

- **Random patient arrivals:** At each time step, a new patient may arrive based on a Poisson arrival rate λ. Each incoming patient is assigned a triage category:
  - Red (urgent)
  - Yellow (non urgent)

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Hyperparameter Tuning

❑ 1. Learning Rate (α)

The size of steps taken to update the neural network weights (Q values) when it learns from new experience.

Value Used:

$\alpha = 0.0005$ (5e-4)

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Hyperparameter Tuning

- **2. Discount Factor (γ)**

The discount factor determines how much the model values future rewards vs. immediate rewards.

γ close to 1 → values long-term outcomes

γ close to 0 → only cares about immediate reward

- **Value Used:**

γ = 0.95

The model should focus on long-term congestion reduction.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Hyperparameter Tuning

- 3. **Exploration Rate (ε)**

ε-greedy strategy

It controls how often the agent explores new actions instead of choosing the best-known action.

- **Strategy Used**

ε_start = 1.0 (pure exploration at beginning)

ε_end = 0.1 (Mostly exploitation)

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Hyperparameter Tuning

- 4. Batch Size

Number of experience samples used in one training update.

- Value Used:

batch_size = 64

- 5. Replay Buffer Size

This is the memory that stores past experience (state, action, reward, next state).

- Value Used:

buffer_size = 50,000

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Hyperparameter Tuning

❑6. Target Network Update Frequency

The target network update frequency determines how often the weights of the online Q-network are copied into the target network.

❑Value Used:

target_update = every 1,000 steps

Smooths out Q-value updates, prevents divergence.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Evaluation Metrics

- **1. Average Episode Reward**

Average reward per episode: 1895.70

- 2.Average wait times

(Red, Yellow): 3.90, 5.86

- 3. Percentage served within thresholds

(Red, Yellow): 100.00%, 100.00%

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Evaluation Metrics

- **4. Convergence Speed**

This refers to how quickly (after how many steps and episodes) the agent’s learning stabilizes, meaning:

Q-values stop changing wildly

episode reward becomes stable

policy stops fluctuating

- At 0 timesteps → mean reward: 1940.10
- At 10000 timesteps → mean reward: 1869.30
- At 20000 timesteps → mean reward: 1874.40

- Model converged at ~20000 timesteps.

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

links

The project includes training, inference, and deployment via a FastAPI endpoint.

For running the API locally:

http://127.0.0.1:8000/docs

For Web-based Testing:

https://rl-hospital-api.onrender.com/docs

Key libraries: numpy, gymnasium, stable-baselines3, torch, matplotlib, fastapi, pydantic, uvicorn

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Sample state vector for testing the FastApi inference API

{ "state": {   "free_doctors": 2,   "longest_wait_red": 8,   "longest_wait_yellow": 12,   "red_queue_length": 4,   "yellow_queue_length": 5,   "doctor1_busy_time": 0,   "doctor2_busy_time": 0,   "doctor3_busy_time": 2 }}

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug

State vector for testing the FastApi inference API

Edit Value | Schema


```json
{
  "state": {
    "free_doctors": 2,
    "longest_wait_red": 8,
    "longest_wait_yellow": 12,
    "red_queue_length": 4,
    "yellow_queue_length": 5,
    "doctor1_busy_time": 0,
    "doctor2_busy_time": 0,
    "doctor3_busy_time": 2
  }
}
```

Test OutPut

Details

Response body


```json
{
  "action": "Serve Red",
  "reward": 76,
  "wait_time": 8,
  "free_doctors": 2
}
```


Response headers

P.O. Box 4, Mukono, Uganda, Plot 67-173, Bishop Tucker Road, Mukono Hill
Tel: +256 (0) 312 350 800 Email: info@ucu.ac.ug Web: https://ucu.ac.ug
Founded by the Province of the Church of Uganda. Chartered by the Government of Uganda

Thank You

Uganda Christian University

P.O. Box 4 Mukono, Uganda
Tel: 256-312-350800
Email: info@ucu.ac.ug

@UgandaChristianUniversity

@UCUniversity

@ugandachristianuniversity

https://ucu.ac.ug/

A Complete Education for A Complete Person