

# Ćwiczenie laboratoryjne

## 11. Uczenie ze wzmocnieniem: zastosowanie modeli uczenia ze wzmocnieniem w optymalizacji procedur medycznych

### 1 Uczenie ze wzmocnieniem w optymalizacji procedur medycznych

#### 1.1 Cel ćwiczenia

Celem ćwiczenia jest zapoznanie studentów z metodami **uczenia ze wzmocnieniem** (Reinforcement Learning, RL) oraz ich zastosowaniem w optymalizacji procedur medycznych. Studenci poznają:

- formalny model procesu decyzyjnego (MDP),
- algorytm Q-learning,
- Deep Reinforcement Learning (DQN),
- metody wyjaśniania decyzji modeli RL (Explainable RL).

Ćwiczenie ma charakter **symulacyjny i dydaktyczny** i nie dotyczy rzeczywistych decyzji klinicznych.

#### 1.2 Podstawy teoretyczne uczenia ze wzmocnieniem

Problem RL formalizowany jest jako **proces decyzyjny Markowa** (MDP):

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \gamma \rangle,$$

gdzie:

- $\mathcal{S}$  – zbiór stanów (np. stany kliniczne pacjenta),
- $\mathcal{A}$  – zbiór akcji (np. decyzje terapeutyczne),
- $P(s'|s, a)$  – prawdopodobieństwo przejścia,
- $R(s, a)$  – funkcja nagrody,
- $\gamma \in [0, 1)$  – współczynnik dyskontowy.

Celem agenta jest maksymalizacja zdyskontowanej sumy nagród:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}.$$

### 1.3 Interpretacja elementów RL w kontekście medycznym

Element RL	Znaczenie medyczne
Stan ( $s$ )	stan zdrowia pacjenta
Akcja ( $a$ )	decyzja terapeutyczna
Nagroda ( $r$ )	poprawa lub pogorszenie zdrowia
Polityka ( $\pi$ )	strategia leczenia
Epizod	przebieg terapii

### 1.4 Algorytm Q-learning

Q-learning jest algorytmem bezmodelowym, który uczy się funkcji wartości:

$$Q(s, a) = \mathbb{E}[G_t \mid S_t = s, A_t = a].$$

Reguła aktualizacji:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left( r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right),$$

gdzie  $\alpha$  jest współczynnikiem uczenia.

### 1.5 Model symulacyjny procedury medycznej

W ćwiczeniu stosowany jest uproszczony model:

- Stany: `Healthy`, `Stable`, `Critical`,
- Akcje: `NoTreatment`, `Standard`, `Aggressive`,
- Nagrody: dodatnie za poprawę stanu, ujemne za pogorszenie i koszt terapii.

Model umożliwia analizę sekwencyjnych decyzji terapeutycznych.

### 1.6 Deep Reinforcement Learning (DQN)

Dla problemów o większej przestrzeni stanów tablicowy Q-learning staje się niepraktyczny. Wówczas funkcja  $Q(s, a)$  jest aproksymowana siecią neuronową:

$$Q(s, a; \theta) \approx Q^*(s, a).$$

Deep Q-Network (DQN) wykorzystuje:

- sieć neuronową do aproksymacji  $Q$ ,
- bufor doświadczeń (Replay Buffer),
- sieć docelową (Target Network).

Cel uczenia DQN:

$$y = r + \gamma \max_{a'} Q_{\text{target}}(s', a').$$

## 1.7 Implementacja DQN w TensorFlow

Sieć neuronowa:

- wejście: wektor stanu pacjenta,
- wyjście: wartości  $Q(s, a)$  dla każdej akcji.

Kod wykorzystuje bibliotekę `TensorFlow/Keras` oraz optymalizator Adam. Replay Buffer redukuje korelację próbek, a Target Network stabilizuje proces uczenia.

## 1.8 Explainable Reinforcement Learning

Uczenie ze wzmocnieniem, szczególnie Deep RL, charakteryzuje się niską interpretowalnością. W medycynie wyjaśnialność decyzji jest kluczowa.

W ćwiczeniu stosowane są następujące metody Explainable RL:

### Analiza wartości Q

Dla danego stanu  $s$  obliczane są wartości  $Q(s, a)$  dla wszystkich akcji. Wybrana akcja jest tą, dla której:

$$a^* = \arg \max_a Q(s, a).$$

### Wyjaśnienia kontrfaktyczne

Porównywane są wartości  $Q$  dla najlepszej i alternatywnej akcji:

$$\Delta Q = Q(s, a^*) - Q(s, a').$$

Mała wartość  $\Delta Q$  wskazuje na niepewność decyzji.

### Analiza wrażliwości na nagrodę

Zmiana funkcji nagrody (np. koszt leczenia agresywnego) pozwala ocenić, jak zmienia się polityka agenta, co umożliwia globalną interpretację zachowania modelu.

## 1.9 Znaczenie Explainable RL w medycynie

Explainable RL:

- zwiększa zaufanie do systemów AI,
- umożliwia audyt decyzji,
- wspiera podejście *human-in-the-loop*.

System RL nie zastępuje lekarza, lecz stanowi narzędzie wspomagające decyzje.

## 1.10 Przebieg ćwiczenia laboratoryjnego

1. Zdefiniuj środowisko MDP.
2. Zaimplementuj Q-learning.
3. Zastąp tablicę Q siecią neuronową (DQN).
4. Przeanalizuj politykę i krzywą uczenia.
5. Zastosuj metody Explainable RL.

## 1.11 Warianty zadań dla studentów

1. Zmodyfikuj funkcję nagrody i porównaj polityki.
2. Dodaj nowe stany kliniczne.
3. Porównaj Q-learning i DQN.
4. Zaimplementuj próg niepewności decyzji.
5. Zaproponuj dodatkowe metody wyjaśniania decyzji.

## 1.12 Ograniczenia i aspekty etyczne

Modele RL w medycynie:

- wymagają walidacji offline,
- nie mogą działać autonomicznie,
- muszą spełniać wymogi etyczne i regulacyjne.

## 1.13 Podsumowanie

Uczenie ze wzmocnieniem oraz Deep Reinforcement Learning stanowią obiecujące narzędzia do optymalizacji procedur medycznych w środowiskach symulacyjnych. Połączenie ich z metodami Explainable RL jest kluczowe dla bezpiecznego i odpowiedzialnego zastosowania w medycynie.

## 2 Warianty zadań RL z prawdopodobieństwami przejść (MDP)

W tej sekcji przedstawiono 15 wariantów zadań RL w postaci procesów decyzyjnych Markowa:

$$\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, R, \gamma \rangle,$$

gdzie  $P(s' | s, a)$  to prawdopodobieństwo przejścia do stanu  $s'$  po wykonaniu akcji  $a$  w stanie  $s$ . Wszystkie warianty mają charakter **symulacyjny i dydaktyczny**.

**Konwencja zapisu przejść.** Dla każdego wariantu definiujemy tablice przejść w formie list:

$$(s, a) \rightarrow \{(s'_1, p_1), (s'_2, p_2), \dots\}, \quad \sum_i p_i = 1.$$

Jeśli przejście nie jest wymienione, przyjmujemy prawdopodobieństwo 0.

### Wariant 1, 6, 11: Leczenie nadciśnienia tętniczego

**Stany  $\mathcal{S}$ :** {Normotension (N), MildHypertension (M), SevereHypertension (S)}

**Akcje  $\mathcal{A}$ :** {NoTreatment (NT), Lifestyle (LS), Drug (DR)}

**Dyskontowanie:**  $\gamma = 0.95$

**Przejścia  $P(s'|s, a)$ :**

- M, NT  $\rightarrow \{(M, 0.70), (S, 0.25), (N, 0.05)\}$
- M, LS  $\rightarrow \{(N, 0.35), (M, 0.55), (S, 0.10)\}$
- M, DR  $\rightarrow \{(N, 0.60), (M, 0.35), (S, 0.05)\}$
- S, NT  $\rightarrow \{(S, 0.75), (M, 0.20), (N, 0.05)\}$
- S, LS  $\rightarrow \{(M, 0.55), (S, 0.40), (N, 0.05)\}$
- S, DR  $\rightarrow \{(M, 0.55), (N, 0.35), (S, 0.10)\}$
- N, NT  $\rightarrow \{(N, 0.85), (M, 0.14), (S, 0.01)\}$
- N, LS  $\rightarrow \{(N, 0.90), (M, 0.10)\}$
- N, DR  $\rightarrow \{(N, 0.92), (M, 0.08)\}$

**Nagrody  $R(s, a)$  (przykład):**

- +10 za osiągnięcie stanu N (na przejściu do N),
- -5 za przejście do S,
- koszt: NT 0, LS -1, DR -2 (odejmowany zawsze).

### Wariant 2, 7, 12: Kontrola glikemii w cukrzycy

**Stany:** {Normoglycemia (N), Hyperglycemia (H), SevereHyperglycemia (S)}

**Akcje:** {NoTreatment (NT), OralMed (OM), Insulin (IN)}

$\gamma = 0.95$

**Przejścia:**

- H, NT  $\rightarrow \{(H, 0.70), (S, 0.25), (N, 0.05)\}$
- H, OM  $\rightarrow \{(N, 0.40), (H, 0.50), (S, 0.10)\}$
- H, IN  $\rightarrow \{(N, 0.65), (H, 0.30), (S, 0.05)\}$
- S, NT  $\rightarrow \{(S, 0.75), (H, 0.20), (N, 0.05)\}$

- $S, OM \rightarrow \{(H, 0.60), (S, 0.35), (N, 0.05)\}$
- $S, IN \rightarrow \{(H, 0.50), (N, 0.40), (S, 0.10)\}$
- $N, NT \rightarrow \{(N, 0.85), (H, 0.14), (S, 0.01)\}$
- $N, OM \rightarrow \{(N, 0.88), (H, 0.12)\}$
- $N, IN \rightarrow \{(N, 0.86), (H, 0.13), (S, 0.01)\}$

**Nagrody:**

- +10 za przejście do N,
- −6 za przejście do S,
- koszty: NT 0, OM −1.5, IN −3.

### Wariant 3, 8, 13: Leczenie infekcji

**Stany:** {Healthy (H), MildInfection (M), SevereInfection (S)}

**Akcje:** {Observe (OB), Antibiotic (AB), Hospitalize (HO)}

$\gamma = 0.95$

**Przejścia:**

- $M, OB \rightarrow \{(M, 0.60), (S, 0.30), (H, 0.10)\}$
- $M, AB \rightarrow \{(H, 0.55), (M, 0.40), (S, 0.05)\}$
- $M, HO \rightarrow \{(H, 0.70), (M, 0.25), (S, 0.05)\}$
- $S, OB \rightarrow \{(S, 0.75), (M, 0.20), (H, 0.05)\}$
- $S, AB \rightarrow \{(M, 0.55), (H, 0.25), (S, 0.20)\}$
- $S, HO \rightarrow \{(M, 0.55), (H, 0.35), (S, 0.10)\}$
- $H, OB \rightarrow \{(H, 0.90), (M, 0.10)\}$
- $H, AB \rightarrow \{(H, 0.92), (M, 0.08)\}$
- $H, HO \rightarrow \{(H, 0.93), (M, 0.07)\}$

**Nagrody:**

- +9 za przejście do H,
- −7 za przejście do S,
- koszty: OB 0, AB −2, HO −4.

## Wariant 4, 9, 14: Rehabilitacja poudarowa

**Stany:** {SevereDisability (S), ModerateDisability (M), Independent (I)}

**Akcje:** {NoRehab (NR), StandardRehab (SR), IntensiveRehab (IR)}

$\gamma = 0.95$

### Przejścia:

- S, NR  $\rightarrow \{(S, 0.75), (M, 0.22), (I, 0.03)\}$
- S, SR  $\rightarrow \{(M, 0.55), (S, 0.40), (I, 0.05)\}$
- S, IR  $\rightarrow \{(M, 0.60), (I, 0.20), (S, 0.20)\}$
- M, NR  $\rightarrow \{(M, 0.70), (S, 0.10), (I, 0.20)\}$
- M, SR  $\rightarrow \{(I, 0.40), (M, 0.55), (S, 0.05)\}$
- M, IR  $\rightarrow \{(I, 0.55), (M, 0.40), (S, 0.05)\}$
- I, NR  $\rightarrow \{(I, 0.85), (M, 0.13), (S, 0.02)\}$
- I, SR  $\rightarrow \{(I, 0.88), (M, 0.11), (S, 0.01)\}$
- I, IR  $\rightarrow \{(I, 0.90), (M, 0.09), (S, 0.01)\}$

### Nagrody:

- +10 za przejście do I,
- -5 za przejście do S,
- koszty: NR 0, SR -2, IR -4.

## Wariant 5, 10, 15: Zarządzanie bólem

**Stany:** {NoPain (N), ModeratePain (M), SeverePain (S)}

**Akcje:** {NoAnalgesia (NA), NSAIDs (NS), Opioids (OP)}

$\gamma = 0.95$

### Przejścia:

- S, NA  $\rightarrow \{(S, 0.75), (M, 0.20), (N, 0.05)\}$
- S, NS  $\rightarrow \{(M, 0.55), (S, 0.35), (N, 0.10)\}$
- S, OP  $\rightarrow \{(N, 0.55), (M, 0.35), (S, 0.10)\}$
- M, NA  $\rightarrow \{(M, 0.65), (S, 0.20), (N, 0.15)\}$
- M, NS  $\rightarrow \{(N, 0.45), (M, 0.50), (S, 0.05)\}$
- M, OP  $\rightarrow \{(N, 0.60), (M, 0.35), (S, 0.05)\}$
- N, NA  $\rightarrow \{(N, 0.90), (M, 0.10)\}$
- N, NS  $\rightarrow \{(N, 0.88), (M, 0.12)\}$

- $N, OP \rightarrow \{(N, 0.80), (M, 0.18), (S, 0.02)\}$

**Nagrody:**

- $+7$  za przejście do N,
- $-6$  za przejście do S,
- koszty: NA  $0$ , NS  $-1$ , OP  $-4$ .

**Uwaga dydaktyczna.** W raporcie student powinien uzasadnić dobór prawdopodobieństw przejść (np. na podstawie intuicji klinicznej: intensywne leczenie zwiększa szansę poprawy, ale niesie większy koszt/ryzyko).