Sprawozdanie z modułu nr 1
MLA 2025/2026

Aplikacje uczenia maszynowego w systemach interakcji człowiek-maszyna

Kierunek: Informatyka

Członkowie zespołu:
Piotr Skowroński
Kinga Grabarczyk
Irek Kosek
Krzysztof Czuba

Gliwice, 2025/2026

Spis treści

1 Wprowadzenie

1.1 Zakres tematyczny sprawozdania

Niniejsze sprawozdanie obejmuje zagadnienia związane z uczeniem ze wzmocnieniem (Reinforcement Learning), które stanowi jeden z trzech głównych paradygmatów uczenia maszynowego obok uczenia nadzorowanego i nienadzorowanego. Dokument koncentruje się na aspektach teoretycznych oraz praktycznych implementacji algorytmów RL w systemach interakcji człowiek-maszyna.

Reinforcement Learning to metoda uczenia maszynowego, w której agent uczy się podejmować optymalne decyzje poprzez interakcję ze środowiskiem. Agent otrzymuje nagrody lub kary za swoje działania i na tej podstawie optymalizuje swoją strategię postępowania. Ta forma uczenia znajduje szerokie zastosowanie w robotyce, grach, systemach rekomendacyjnych oraz autonomicznych pojazdach.

1.2 Zespół projektowy

- **Piotr Skowroński** – Lider zespołu – Analiza teoretyczna algorytmów RL, przygotowanie sekcji teoretycznej
- **Kinga Grabarczyk** – Programista – Implementacja algorytmów Q-Learning i Deep Q-Network, testy eksperymentalne
- **Irek Kosek** – Analityk – Zbieranie źródeł literaturowych, analiza wyników, dokumentacja
- **Krzysztof Czuba** – Specjalista ds. praktycznych – Analiza narzędzi i frameworków, przygotowanie sekcji praktycznej

2 ML – zagadnienia teoretyczne

2.1 Wprowadzenie do części teoretycznej

Uczenie ze wzmocnieniem opiera się na paradygmacie uczenia się poprzez próby i błędy, gdzie agent stara się maksymalizować kumulatywną nagrodę w długim okresie czasu. Podstawowym modelem formalnym opisującym problemy RL jest Markov Decision Process (MDP), który składa się z następujących elementów:

- **Zbiór stanów S** – możliwe stany środowiska
- **Zbiór akcji A** – możliwe działania agenta
- **Funkcja przejścia P(s'|s,a)** – prawdopodobieństwo przejścia do stanu s' przy wykonaniu akcji a w stanie s
- **Funkcja nagrody R(s,a,s')** – nagroda otrzymana za wykonanie akcji
- **Współczynnik dyskontowania** γ – określa wagę przyszłych nagród

2.2 Rozwinięcie

W uczeniu ze wzmocnieniem wyróżniamy kilka kluczowych podejść i algorytmów:

2.2.1 Metody oparte na wartości (Value-Based Methods)

Q-Learning jest podstawowym algorytmem uczenia ze wzmocnieniem, który uczy się funkcji wartości akcji Q(s,a) reprezentującej oczekiwaną kumulatywną nagrodę za wykonanie akcji a w stanie s i postępowanie optymalnie dalej. Algorytm aktualizuje wartości Q zgodnie z równaniem Bellmana:

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

gdzie α to współczynnik uczenia, a γ współczynnik dyskontowania przyszłych nagród.

SARSA (State-Action-Reward-State-Action) to algorytm on-policy, który w przeciwieństwie do Q-Learning aktualizuje wartości Q na podstawie rzeczywiście wykonanej akcji, a nie maksymalnej wartości. To czyni go bardziej konserwatywnym i bezpieczniejszym w niektórych zastosowaniach.

2.2.2 Metody oparte na polityce (Policy-Based Methods)

Policy Gradient Methods bezpośrednio optymalizują parametryzowaną politykę poprzez gradient wznoszenia. Algorytmy takie jak REINFORCE czy Actor-Critic łączą zalety metod wartości i polityki, gdzie critic ocenia wartość stanów, a actor aktualizuje politykę.

2.2.3 Deep Reinforcement Learning

Deep Q-Network (DQN) łączy Q-Learning z głębokimi sieciami neuronowymi, co umożliwia radzenie sobie z problemami o dużej przestrzeni stanów. Kluczowe innowacje DQN to experience replay (ponowne wykorzystanie przeszłych doświadczeń) oraz target network (stabilizacja uczenia).

Proximal Policy Optimization (PPO) to nowoczesny algorytm, który osiąga balans między wydajnością uczenia a stabilnością poprzez ograniczenie wielkości aktualizacji polityki.

2.2.4 Exploration vs Exploitation

Fundamentalnym dylematem w RL jest balans między eksploracją (poznawaniem nowych akcji) a eksploatacją (wykorzystywaniem znanej wiedzy). Popularne strategie to epsilon-greedy, gdzie agent z prawdopodobieństwem ϵ wybiera losową akcję, oraz Upper Confidence Bound (UCB), który bardziej systematycznie balansuje eksplorację i eksploatację.

2.3 Podsumowanie części teoretycznej

Uczenie ze wzmocnieniem stanowi potężne narzędzie do rozwiązywania problemów sekwencyjnego podejmowania decyzji. Kluczowe koncepcje obejmują Markov Decision Processes, funkcje wartości, polityki oraz równanie Bellmana. Nowoczesne podejścia łączą klasyczne algorytmy RL z głębokim uczeniem, co umożliwia rozwiązywanie złożonych problemów praktycznych.

Teoretyczne podstawy RL znajdują zastosowanie w szerokim spektrum dziedzin – od gier i robotyki, przez systemy rekomendacyjne, po optymalizację procesów biznesowych i sterowanie autonomicznymi pojazdami. Rozwój algorytmów takich jak DQN, PPO czy AlphaGo pokazuje, że RL jest kluczową technologią w rozwoju sztucznej inteligencji.

3 ML – zagadnienia praktyczne

3.1 Wprowadzenie do części praktycznej

Praktyczne zastosowania uczenia ze wzmocnieniem obejmują wiele dziedzin życia codziennego i przemysłu. W tej sekcji prezentujemy przegląd istniejących narzędzi, bibliotek oraz systemów implementujących algorytmy RL, a także przykłady ich zastosowań w rzeczywistych projektach.

3.2 Rozwinięcie

3.2.1 Biblioteki i frameworki do Reinforcement Learning

OpenAI Gym to standardowa platforma do tworzenia i testowania algorytmów RL. Oferuje szeroki zestaw środowisk testowych, od prostych problemów jak CartPole, przez klasyczne gry Atari, po złożone symulacje robotyczne. Gym definiuje ujednolicony interfejs dla środowisk, co ułatwia porównywanie algorytmów.

Stable Baselines3 to zbiór implementacji najnowocześniejszych algorytmów RL (A2C, PPO, SAC, TD3) w PyTorch. Biblioteka oferuje wysoki poziom abstrakcji i łatwość użycia, co czyni ją idealną dla praktyków i badaczy.

TensorFlow Agents (TF-Agents) to modułowa biblioteka od Google do budowania, trenowania i ewaluacji algorytmów RL w TensorFlow. Oferuje elastyczną architekturę oraz zoptymalizowane implementacje popularnych algorytmów.

Ray RLlib to skalowalna biblioteka dla rozproszonego uczenia ze wzmocnieniem. Umożliwia trening na wielu maszynach i oferuje implementacje najnowszych algorytmów z możliwością łatwego skalowania.

3.2.2 Praktyczne zastosowania RL

Gry i rozrywka: AlphaGo od DeepMind wykorzystuje zaawansowane techniki RL do gry w Go, osiągając poziom przewyższający najlepszych ludzkich graczy. OpenAI Five pokazało możliwości RL w złożonych grach wieloosobowych jak Dota 2.

Robotyka: RL jest szeroko stosowany w robotyce do uczenia manipulacji obiektami, chodu robotów oraz planowania trajektorii. Systemy takie jak roboty Boston Dynamics wykorzystują RL do optymalizacji ruchów.

Autonomiczne pojazdy: Firmy takie jak Waymo i Tesla wykorzystują RL do trenowania systemów podejmowania decyzji w autonomicznych pojazdach, gdzie agent musi nauczyć się bezpiecznego poruszania w dynamicznym środowisku drogowym.

Systemy rekomendacyjne: Netflix i YouTube wykorzystują RL do personalizacji rekomendacji treści, gdzie agent uczy się maksymalizować długoterminowe zaangażowanie użytkowników.

Optymalizacja zasobów: Google wykorzystuje RL do optymalizacji zużycia energii w centrach danych, osiągając znaczące oszczędności poprzez inteligentne zarządzanie systemami chłodzenia.

Finanse i trading: Algorytmy RL są stosowane w algorytmicznym tradingu, gdzie agent uczy się optymalnych strategii inwestycyjnych na podstawie danych rynkowych.

3.2.3 Środowiska i symulatory

MuJoCo (Multi-Joint dynamics with Contact) to zaawansowany silnik fizyki używany do symulacji robotów i układów mechanicznych. Jest powszechnie stosowany w badaniach nad RL w robotyce.

Unity ML-Agents to framework od Unity Technologies umożliwiający trenowanie inteligentnych agentów w środowisku 3D. Szczególnie popularny w trenowaniu AI do gier.

PettingZoo oferuje środowiska dla multi-agent reinforcement learning, gdzie wiele agentów uczy się jednocześnie i może współpracować lub konkurować.

3.2.4 Wyzwania praktyczne

Sample efficiency pozostaje głównym wyzwaniem – algorytmy RL często wymagają milionów interakcji ze środowiskiem do nauczenia się skutecznej polityki. Transfer learning i meta-learning są badane jako potencjalne rozwiązania.

Stabilność uczenia w RL jest trudniejsza niż w uczeniu nadzorowanym z powodu niestacjonarności danych (polityka agenta zmienia rozkład danych) oraz bootstrappingu (uczenie wartości na podstawie oszacowań innych wartości).

Bezpieczeństwo i etyka w RL są kluczowe, szczególnie w aplikacjach takich jak autonomiczne pojazdy czy medycyna, gdzie błędy mogą mieć poważne konsekwencje.

3.3 Podsumowanie części praktycznej

Ekosystem narzędzi do Reinforcement Learning znacznie się rozwinął w ostatnich latach, oferując zarówno badaczom jak i praktykom szeroki wybór bibliotek i frameworków. OpenAI Gym ustanowił standard dla środowisk testowych, podczas gdy biblioteki jak Stable Baselines3 i RLlib oferują gotowe implementacje najnowocześniejszych algorytmów.

Praktyczne zastosowania RL rozciągają się od gier i rozrywki, przez robotykę i autonomiczne pojazdy, po optymalizację biznesową i finanse. Pomimo imponujących sukcesów, wyzwania związane z efektywnością próbkowania, stabilnością uczenia i bezpieczeństwem pozostają aktywnymi obszarami badań. Rozwój technik takich jak model-based RL, offline RL i multi-agent RL otwiera nowe możliwości dla przyszłych zastosowań.

4 Podsumowanie i wnioski

4.1 Podsumowanie

Reinforcement Learning stanowi jeden z najdynamiczniej rozwijających się obszarów uczenia maszynowego, oferując unikalne możliwości rozwiązywania problemów sekwencyjnego podejmowania decyzji. W sprawozdaniu przedstawiono zarówno fundamenty teoretyczne RL, obejmujące Markov Decision Processes, funkcje wartości oraz polityki, jak i praktyczne aspekty implementacji algorytmów w rzeczywistych systemach.

Część teoretyczna skupiła się na kluczowych algorytmach, takich jak Q-Learning, SARSA, Policy Gradient Methods oraz nowoczesnych rozszerzeniach jak DQN i PPO. Przedstawiono również fundamentalny dylemat exploration-exploitation oraz metody jego rozwiązywania.

Część praktyczna omówiła ekosystem narzędzi i bibliotek dostępnych dla praktyków RL, włączając OpenAI Gym, Stable Baselines3, TensorFlow Agents i Ray RLlib. Zaprezentowano szerokie spektrum zastosowań – od gier i robotyki po systemy rekomendacyjne i optymalizację biznesową.

4.2 Wnioski

Po przeprowadzeniu analizy teoretycznej i praktycznej uczenia ze wzmocnieniem można sformułować następujące wnioski:

1. RL oferuje unikalne możliwości w problemach gdzie trudno zdefiniować explicite funkcję celu, ale możemy określić nagrody za pożądane zachowania. To czyni go idealnym do zastosowań w robotyce, grach i systemach autonomicznych.

2. Pomimo imponujących projektów takich jak AlphaGo czy OpenAI Five, RL wciąż boryka się z istotnymi wyzwaniami praktycznymi. Sample efficiency pozostaje główną barierą – większość algorytmów wymaga ogromnej liczby interakcji ze środowiskiem, co w realnych zastosowaniach może być kosztowne lub niebezpieczne.

3. Połączenie klasycznych algorytmów RL z głębokim uczeniem (Deep RL) znacząco rozszerzyło możliwości tej technologii, umożliwiając pracę z wysokowymiarowymi przestrzeniami stanów i akcji. Jednak wprowadza to dodatkowe wyzwania związane ze stabilnością uczenia.

4. Rozwój gotowych bibliotek i frameworków znacznie obniżył próg wejścia dla praktyków chcących wykorzystać RL. Narzędzia takie jak Stable Baselines3 oferują implementacje state-of-the-art algorytmów, które można łatwo zastosować do własnych problemów.

5. Bezpieczeństwo i etyka w RL są kluczowe, szczególnie w zastosowaniach krytycznych jak autonomiczne pojazdy czy medycyna. Potrzeba dalszych badań nad safe RL oraz metodami weryfikacji i walidacji nauczonych polityk.

6. Przyszłość RL leży w rozwoju technik takich jak meta-learning (uczenie się uczenia), model-based RL (wykorzystujący modele środowiska do zwiększenia efektywności), offline RL (uczenie z wcześniej zebranych danych) oraz multi-agent RL (współpraca i konkurencja wielu agentów).

7. Integracja RL z systemami interakcji człowiek-maszyna otwiera nowe możliwości dla personalizacji i adaptacji systemów do indywidualnych użytkowników. Systemy mogą uczyć się preferencji użytkowników i dynamicznie dostosowywać swoje zachowanie.

5 Spis literatury

Literatura

[1] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.

[2] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

[3] Silver, D., Huang, A., Maddison, C. J., et al. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489.

[4] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

[5] Brockman, G., Cheung, V., Pettersson, L., et al. (2016). OpenAI Gym. *arXiv preprint arXiv:1606.01540*.

[6] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268), 1–8.

[7] Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine Learning*, 8(3-4), 279–292.

[8] Lillicrap, T. P., Hunt, J. J., Pritzel, A., et al. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

[9] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *International Conference on Machine Learning*, 1861–1870.

[10] Vinyals, O., Babuschkin, I., Czarnecki, W. M., et al. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782), 350–354.