

Porównanie metod nauczanie przez wzmacnianie na podstawie gry wideo

Kacper Majkowski

22 maja 2025

1 Wstęp

1.1 Motywacja

W dzisiejszych czasach niewiele jest dziedzin nauki które rozwijają się tak szybko jak sztuczna inteligencja. Technologie, które jeszcze dekadę czy dwie temu byłyby uznane za science fiction, używane są dzisiaj na porządku dziennym. Rozwój takich narzędzi jak ChatGPT czy Copilot całkowicie odmieniły sposób w jaki ludzie rozwiązują problemy, czy to w pracy czy poza nią. Jedną z dziedzin nauczania maszynowego jest Nauczanie przez Wzmacnianie (eng. Reinforcement Learning)

1.2 Historia nauczania przez wzmacnianie

Współczesne nauczanie przez wzmacnianie jest wynikiem połączenia kilku dziedzin nauki i ich aplikacji dla programów komputerowych. Jedną z tych dziedzin jest psychologia, a konkretnie badania mające ustalić czy zwierzęta potrafią się uczyć, a jeżeli tak to na podstawie jakiego mechanizmu. Już na początku XX wieku naukowcy znajdowali dowody na to że zwierzęta faktycznie uczą się na podstawie doświadczeń

“Z kilku odpowiedzi na tę samą sytuację te, które są połączone z satysfakcją zwierzęcia lub następują bezpośrednio po niej, będą, przy innych równych warunkach, silniej powiązane z sytuacją, tak że gdy sytuacja się powtórzy, będą się częściej powtarzać; te, które są połączone z dyskomfortem zwierzęcia lub następują bezpośrednio po nim, będą, przy innych równych warunkach, słabiej powiązane z tą sytuacją, tak że gdy sytuacja się powtórzy, będą się rzadziej powtarzać. Im większa satysfakcja lub dyskomfort, tym silniejsze lub słabsze jest połączenie.” - (Thorndike, 1911, p. 244) [?]

Kolejną dziedziną nauki która przyczyniła się do rozwoju nauczania przez wzmacnianie jest Teoria Sterowania Optymalnego (eng. Optimal Control Theory). Jednym z naukowców zajmujących się tym tematem w latach 50-tch XX wieku był Richard Bellman. Jego wkład przyczynił się do rozwoju programowania dynamicznego, zdefiniowania Równania Bellmana oraz w końcu na ich podstawie do rozwoju nauczania przez wzmacnianie.

1.3 Nawiązanie do natury

Wspomniane było wcześniej że nauczanie przez wzmacnianie jest powiązane z tym jak zwierzęta uczą się na podstawie doświadczeń. Możemy to również zaobserwować u ludzi. Jeżeli dziecko dodknie raz gorącej kuchenki i się oparzy, to najprawdopodobniej więcej już tego nie zrobi. Innymi słowy, podjęło ono pewną akcję, za którą zostało "ukarane", więc w przyszłości będzie tego unikać. Z drugiej strony jeżeli dziecko znajdzie szafkę ze słodyczami, to będzie chodzić tam częściej, wiedząc że za każdym razem czekają tam na nie słodycze. Mówiąc inaczej, wykonało ono pewną akcję i dostało za to "nagrodę", więc w przyszłości będzie próbowało powtórzyć tę akcję.

Analogicznie możemy spojrzeć na to jak zachowują się zwierzęta. Wyobraźmy sobie mysz która wkrada się do domu, szukając pożywienia. Jeżeli natrafi ona na groźnego kota, to od tej pory będzie już wiedzieć aby do tego domu nie chodzić. Ale czy na pewno? Może kot strzeże wartościowej spiżarni? Może nagroda byłaby wystarczająca aby mysz podjęła się tego ryzyka?

Są to życiowe, uproszczone przykłady, ale dobrze pokazują sposób działania nauczania przez wzmacnianie oraz problemy jakimi się ono zajmuje.

1.4 Intuicja

Możemy więc zdefiniować pewną intuicję, która pozwoli zrozumieć nam na czym polega nauczanie ze wzmacnieniem. Podmiot (nazywany później "agentem") znajduje się w pewnej sytuacji (stanie) i ma do podjęcia decyzję (akcję). Za tę akcję dostaje następnie nagrodę lub karę i ją zapamiętuje. Na podstawie tego może w przyszłości podjąć decyzję, kiedy znajdzie się w tym stanie ponownie w przyszłości to czy powinien podjąć tę samą decyzję. W późniejszych rozdziałach przedstawiona zostanie bardziej dokładna i rygorystyczna definicja, jednakże ta intuicja pozwoli nam lepiej zrozumieć esencję nauczania ze wzmacnianiem, zanim przejdziemy do bardziej matematycznych, statystycznych i programistycznych rozważań.

Literatura

- [1] Thorndike, E.L. (1911) Animal Intelligence: Experimental Studies. MacMillan, New York.