Nr rejestracyjny: 2020/37/N/ST6/04113; Kierownik projektu: mgr Piotr Józef Kozakowski

Wybiegając Myślą Naprzód: Długofalowe planowanie z użyciem głębokiego uczenia ze wzmocnieniem bazującego na modelu

Piotr Kozakowski

Uczenie ze wzmocnieniem jest gałęzią Sztucznej Inteligencji zajmującą się uczeniem niezależnych agentów poprzez interakcję. Tacy agenci działają w danym środowisku, zdobywając nowe doświadczenie i ucząc się z niego, aby poprawić swoje zachowanie. Prostszym terminem mogłoby być "uczenie metodą prób i błędów". Co ważne, agenci uczeni przez wzmocnienie automatycznie odkrywają przydatne reprezentacje, aby wnioskować na temat świata.

Szeroki zakres problemów może być opisany w języku agentów wchodzących w interakcję ze środowiskami, co daje nadzieję na szerokie zastosowanie metod uczenia ze wzmocnieniem w praktycznych problemach. Nowe badania w tej dziedzinie stale przesuwają granicę pomiędzy problemami w których przodują komputery, a tymi w których ludzie nadal są lepsi. Spektakularne przykłady obejmują sterowanie robotami, złożone gry strategiczne, lub współczesne gry komputerowe. Niedawne badania studiowały nawet zastosowanie takich agentów w sterowaniu ekonomią, lub redukowaniu skutków pandemii.

Jednakże uczenie ze wzmocnieniem nie radzi sobie z zadaniami obejmującymi dalekie horyzonty czasowe, które są często spotykane w prawdziwym świecie. Rozwiązywanie takich problemów tymi metodami jest mozliwe, ale wymaga ogromnych nakładów ludzkiej pracy i zasobów obliczeniowych. Przykłady obejmują OpenAI Five, system który pokonał mistrzów świata w grze Dota 2, oraz AlphaStar, system który zwycięża z światowej klasy graczami w StarCraft II.

Innym sposobem rozwiązywania takich problemów jest *planowanie*. Metody planowania nie uczą się, a zamiast tego używają sprytnych algorytmów do znajdywania odpowiednich strategii, być może obejmujących dalekie horyzonty czasowe. Jednakże systemy planowania wymagają dobrych reprezentacji - musimy dobrze zdefiniować co jest stanem środowiska, oraz jakie są możliwe wybory agenta. Planowanie ze złą reprezentacją może być bardzo wymagające obliczeniowo, a znajdowanie dobrych reprezentacji wymaga dużego wysiłku.

Tak więc z jednej strony mamy metody uczenia ze wzmocnieniem, które uczą się własnych reprezentacji ale mają problemy z długimi horyzontami, a z drugiej strony mamy algorytmy planowania, które dobrze radzą sobie z długimi horyzontami, ale wymagają reprezentacji zdefiniowanych z zewnątrz. Na szczęście, istnieją sposoby na połączenie tych dwóch podejść, zachowując zalety obydwu.

Uczenie ze wzmocnieniem bazujące na modelu zakłada dostęp do *modelu* środowiska, który może być wykorzystywany do przewidywania rezultatów podejmowanych decyzji i do formułowania planów. Systemy bazujące na modelu mogą łączyć w sobie planowanie i uczenie, osiągając najlepsze z obu światów. AlphaZero jest przykładem systemu bazującego na modelu, osiągającego nadludzkie wyniki w różnorodnych, złożonych grach strategicznych. AlphaZero uczy sieci neuronowe przewidywania obiecujących ruchów, a następnie wykorzystuje je do wspomagania algorytmu planującego, aby alokował swoje zasoby obliczeniowe znacznie bardziej efektywnie.

W trakcie tego projektu zbadamy różnorodne metody łączące planowanie i uczenie, aby otrzymać bardziej efektywnych inteligentnych agentów. Będziemy badać synergie pomiędzy tymi dwoma podejściami, oraz projektować systemy, które (a) planują w sposób czyniący uczenie bardziej wydajnym, (b) uczą się w sposób czyniący planowanie bardziej wydajnym, oraz (c) uczą się własnych modeli środowiska aby osiągnąć pełną autonomię.

Oczekujemy, że metody które rozwiniemy w tym projekcie pozwolą nam na budowanie lepszych agentów Sztucznej Inteligencji, wnioskujących na dalsze horyzonty czasowe, uczących się szybciej oraz rozwiązujących bardziej złozone problemy.