Usprawnianie skalowalności i trenowalności architektur neuronowych w uczeniu ze wzmocnieniem

Uczenie głębokie umożliwiło rozwój szerokiej gamy technologii, od przetwarzania języka naturalnego i analizy obrazu po projektowanie leków i materiałów. Jednakże postępy te w dużej mierze zależały od dostępu do ogromnych zbiorów danych oraz potężnych modeli zdolnych do uczenia się wzorców na ich podstawie. W przeciwieństwie do tego agenci uczenia ze wzmocnieniem (ang. reinforcement learning, RL) często uczą się ze strumienia doświadczeń, a nie ze stałych zbiorów danych. Ten niestacjonarny rozkład danych stwarza szereg wyzwań dla stabilności treningu, zwłaszcza w przypadku wykorzystania dużych sieci neuronowych. Do powszechnych problemów należą: pogorszona plastyczność sieci neuronowych, katastroficzne zapominanie wcześniej nabytej wiedzy oraz niekontrolowany wzrost gradientów i norm wag, co w ostateczności często prowadzi do zepsucia uczonych reprezentacji oraz pogorszenia wydajności modelu. Co więcej, nawet w sytuacjach, gdy dostęp do danych jest praktycznie nieograniczony, wydajność współczesnych metod RL często wysyca się na wczesnych etapach optymalizacji. W rezultacie szybkość gromadzenia danych, jaką zapewniają symulatory fizyki wspierane przez GPU, często nie jest w pełni wykorzystywana.

Niniejsza rozprawa bada wyzwania związane ze skalowalnością oraz procesem uczenia się architektur neuronowych przeznaczonych do podejmowania decyzji. Zaproponowano w niej rozwiązania architektoniczne, które odpowiadają na te problemy i umożliwiają efektywne uczenie dużych modeli uczenia ze wzmocnieniem w robotyce. Przedstawione prace mają zastosowanie w szerokiej klasie metod aktor-krytyk, zarówno w metodach typu off-policy, jak i on-policy, ze szczególnym uwzględnieniem metod opartych na uczeniu temporal-difference oraz samonadzorowanym uczeniu ze wzmocnieniem. Choć większość przedstawionych prac ma charakter empiryczny, rozprawa dostarcza również istotnych wniosków teoretycznych dotyczących własności architektur neuronowych stosowanych w uczeniu ze wzmocnieniem. Rozprawa jest zorganizowana wokół następującego głównego pytania badawczego:*W jaki sposób należy projektować architektury neuronowe agentów RL, aby maksymalizować ich zdolność do uczenia się?*

Pierwsza część pracy bada nieefektywność uczenia się agentów RL oraz identyfikuje komponenty architektury neuronowej, które poprawiają efektywność wykorzystania danych oraz zdolności wnioskowania. Pierwszy artykuł

analizuje trzy główne problemy w treningu agentów RL: (1) przeszacowanie wartości (ang. value overestimation), czyli nadmiernie optymistyczne szacunki wartości akcji, (2) przeuczenie (overfitting) względem obserwacji treningowych, oraz (3) degradację plastyczności, czyli zmniejszoną zdolność do uczenia się na podstawie nowych danych. Kolejny artykuł porusza problem plastyczności w uczeniu ciągłym, wprowadzając regularyzację spektralną. Ogranicza ona wartości osobliwe macierzy wag sieci neuronowej, utrzymując je blisko wartości charakterystycznych dla świeżo zainicjowanej sieci, która cechuje się wysoką podatnością na uczenie. Ostatnia praca w tej grupie dowodzi, że iteracyjne udoskonalanie decyzji poprzez tzw. architektury iteracyjne poprawia generalizację wyuczonej polityki. Te trzy prace dostarczają dowodów na to, że wydajność agentów RL jest przede wszystkim ograniczona przez nieefektywną akumulację wiedzy oraz dynamikę uczenia w obrębie architektury sieci neuronowej, a nie przez realizowany algorytm uczenia.

Druga część pracy skupia się na metodach samonadzorowanego uczenia ze wzmocnieniem (ang. self-supervised reinforcement learning, SSRL) oraz ich skalowalności. Pierwsza praca w tej grupie prezentuje JaxGCRL – szybką bazę kodu i środowiska testowe (ang. benchmark) do uczenia ze wzmocnieniem warunkowanym celem w trybie online, co umożliwiło rozwój kolejnych projektów. Druga praca opiera się na JaxGCRL i wskazuje kluczowe decyzje projektowe oraz wnioski empiryczne dotyczące skalowania samonadzorowanego RL, kładąc szczególny nacisk na metody kontrastywne. Ostatnia praca bada, w jakim stopniu agenci RL, w tym modele SSRL, potrafią komponować fragmenty zachowań wyuczone podczas treningu w celu rozwiązywania zadań o dłuższym horyzoncie czasowym podczas ewaluacji. Przedstawione prace pokazują, że przy stabilnych architekturach i algorytmach uczenia możliwe jest budowanie głębokich agentów RL, których wydajność i zdolność generalizacji skalują się wraz z liczbą warstw.

Słowa kluczowe: Uczenie ze Wzmocnieniem, Uczenie Ciągłe, Samo-Nadzorowane Uczenie ze Wzmocnieniem, Plastyczność Sieci Neuronowych