

Skok w rozwoju AI. Polacy przełamują impas w uczeniu ze wzmocnieniem (RL)

© Ludwika Tomala (PAP)

Nauka w Polsce, 09.12.2025

<https://naukawpolsce.pl/aktualnosci/news%2C110836%2Cskok-w-rozwoju-ai-polacy-przelamuja-impas-w-uczeniu-ze-wzmocnieniem-rl.html>

Zespół informatyków z udziałem dwóch Polaków znalazł sposób na skokowe zwiększenie głębokości sieci neuronowych - z kilku do aż tysiąca warstw - w modelach AI uczenia ze wzmocnieniem (RL). Pracę informatyków - w tym dwóch Polaków - nagrodzono na prestiżowej konferencji naukowej NeurIPS o AI.

W badaniach kierowanych przez prof. Benjamina Eysenbacha z Princeton brali udział doktorant Michał Bortkiewicz i prof. Tomasz Trzciński z Politechniki Warszawskiej. Ich praca znalazła się wśród pięciu nagrodzonych w ramach najbardziej prestiżowej konferencji naukowej dotyczącej sztucznej inteligencji, Neural Information Processing Systems (NeurIPS). W tym roku na konferencję nadesłano ponad 20 tys. prac, a zaakceptowano około 5 tys.

Uczenie ze wzmocnieniem (Reinforcement Learning, RL) to jeden z kilku głównych nurtów uczenia maszynowego - obok uczenia nadzorowanego, nienadzorowanego i samonadzorowanego. Modele RL m.in. zapewniły zwycięstwo maszyny - AlphaGo - nad człowiekiem w grze Go czy w grach wideo, takich jak Dota. Pomagają wszędzie tam, gdzie trzeba przeczesać niezliczone kombinacje rozwiązań: szukać nowych leków, projektować syntezę białek czy wspomagać decyzje ekonomiczne.

Jak wygląda np. nauka chodzenia w wykonaniu modelu RL? Dajemy komputerowemu ludzikowi nogi, ręce, oczy i wiedzę, że kiedy dotrze do środka labiryntu, to dostanie nagrodę. Ludzik nie może obserwować, jak chodzi mama, nikt nie trzyma go za rączki, żeby się nie przewrócił, nie dostaje do przeczytania instrukcji używania nóg ani buziaczków za pierwsze kroki. Dotrze do celu? Nagroda! Nie dotrze - brak nagrody. Może styl chodzenia, jaki wypracuje przywoła na myśl ministerstwo głupich kroków, a może pozazdrości mu Robert Korzeniowski? Nieważne. Byleby dotarł do celu.

Do tej pory ludzika dawało się wyposażać tylko w kilka warstw sieci neuronowych, tworząc jego „minimózg”. Zwiększanie liczby warstw takiego „minimózgu” nie prowadziło do rozwiązania problemu, a wręcz go uniemożliwiało. Teraz zaś zespół z USA i Polski pokazał zupełnie nowy pomysł: odkrył, jak zwiększyć w modelu RL liczbę warstw sieci neuronowej do nawet 1024. A skokowe zwiększenie głębokości takich sieci tworzy już wcale nie mini-, ale „makromózg”.

SKOK W ROZWOJU AI. NAD ŚCIANĄ LABIRYNTU

O ile przy paru takich warstwach sieci ludzik niezgrabnie odkrywał, do czego służą kończyny i koślawo ciągnął swoje ciało po labiryncie, to dzięki rozwiązaniu badaczy, tzw. contrastive RL (CRL) ludzik dokonał - całkiem dosłownie - skokowego postępu w nauce poruszania się. Liczba sukcesów w dotarciu do celu wzrosła ponad 50-krotnie. I tak np. przy 256 warstwach ludzik odkrył, że jeśli kucnie, a potem odbije się w górę, to uda mu się przeskoczyć nad ścianami labiryntu, żeby szybciej dotrzeć do nagrody. To się nazywa myślenie nieszablonowe – thinking outside the box!

Skoro zaś już nawet stare modele RL - o kilku warstwach - zdołały wypracować strategię wygrania z człowiekiem w grę Go i szukają coraz lepszych leków, to co będą w stanie zrobić te, które mają 1000 warstw i przeskakują przez labirynt? Pora się przekonać! Algorytm CRL jest udostępniony bezpłatnie wszystkim zainteresowanym (<https://wang-kevin3290.github.io/scaling-crl/>).

CEBULA MA WARSTWY? MODEL RL MA WARSTWY

Jeden z twórców rozwiązania, prof. Tomasz Trzciniński z PW w rozmowie z PAP tłumaczy, że warstwy to element architektury sieci neuronowej, który pozwala uzyskiwać kolejne etapy przetwarzania informacji w sieci neuronowej. Im więcej warstw, tym trudniejsze operacje mogą zachodzić pomiędzy między informacją wejściową, a końcowy rezultatem. Jeśli zwiększa się liczbę warstw, a więc głębokość sieci, model jest w stanie nauczyć się bardziej skomplikowanych koncepcji i buduje bogatszą reprezentację świata, zanim podejmie działania.

- W przypadku zadania z labiryntem, model ma więcej stopni swobody niż tylko krok w lewo czy w prawo – może skakać, zginać się, wyciągać rękę. To są dodatkowe możliwości, które pozwalają na znajdowanie nowych, kreatywnych rozwiązań - skomentował prof. Trzciniński.

Skalowanie, czyli zwiększanie liczby warstw sieci neuronowych (a więc ich głębokości), było z powodzeniem stosowane w wielu modelach, np. w LLM-ach (dużych modelach językowych), takich jak GPT. Naukowcy potrafili więc rozwijać „duże mózgi” tylko w innych architekturach stosowanych w uczeniu maszynowym.

Tymczasem modele RL pozostawały w tym zakresie daleko w tyle - wykorzystywały bardzo płytkie sieci neuronowe mające zaledwie 2-5 warstw. - Kiedy próbowano dorzucić kolejne warstwy, algorytm się gubił i model przestawał się trenować. Przyjęto się więc sądzić, że modele RL takie już są: muszą mieć płytkie sieci i nie da się tego zmienić - podsumował prof. Trzcíński.

W swoich badaniach w ramach doktoratu Michał Bortkiewicz pokazał, że kluczem do rozwoju tych kolejnych warstw - wewnętrznych mózgów modelu - jest wykorzystanie techniki znanej z innego nurtu badań nad uczeniem maszynowym - to tzw. samonadzorowane uczenie (SSL). W takiej odmianie uczenia wykorzystuje się tzw. zadania pretekstowe (ang. pretext tasks), które zmuszają model do zrozumienia wewnętrznej struktury danych i nadania im etykiet, zanim wyruszy na poszukiwanie odpowiedzi. W metodzie SSL przygotowywany jest choćby pierwszy etap - pretrening - dużych modeli językowych LLM.

Autorzy badań stwierdzają, że w nowoczesnych modelach te dwa typy uczenia: RL i SSL nie muszą być przeciwstawnymi regułami, ale mogą zostać połączone, aby uzyskać niespotykane rezultaty. - Relatywnie mała zmiana, którą odkryliśmy, powoduje tak ogromne, przełomowe osiągnięcia - zauważył prof. Trzcíński.

Co ciekawe, mimo gigantycznego wzrostu liczby warstw, model nie stał się bardziej energożerny w przeliczeniu na efekt – „myśli” wprawdzie w sposób bardziej skomplikowany, ale dzięki temu szybciej znajduje rozwiązanie, zamiast długo błądzić bez celu.

- Nasze badania pokazują, że warto kwestionować utarte szlaki i myśleć „outside the box”. Nawet w Polsce, gdzie finansowanie nauki i badań podstawowych nie jest wystarczające, czy porównywalne do innych krajów rozwiniętych, można zadawać trafne pytania i kwestionować status quo, aby zmienić świat i odkrywać rzeczy, na które nikt do tej pory nie wpadł - podsumował prof. Trzcíński.

Informatyk wyraził nadzieję, że modele CRL przyczynią się do projektowania nowych leków na śmiertelne choroby czy ułatwią proces personalizacji leków. - Chciałbym też zobaczyć, jak te metody pozwalają na opracowanie takich modeli sztucznej inteligencji, które będą w stanie samodzielnie się ulepszać, aby np. kreatywnie tworzyć nowe pomysły i prowadzić do kolejnych etapów rozwoju nauki - ocenił prof. Trzcíński.

Ludwika Tomala (PAP)