

Jak uczyć się szybciej: w kierunku lepszej adaptacji w procesie meta-uczenia.

Streszczenie popularnonaukowe

W ciągu ostatnich lat, a szczególnie w ostatniej dekadzie, ludzkość była świadkiem niezwykłego rozwoju urządzeń, umożliwiającego wykonywanie coraz bardziej obliczeniowo kosztownych eksperymentów. Ten wzrost mocy obliczeniowej w połączeniu ze zbieraniem olbrzymiej ilości danych, całkowicie zmienił rozwój uczenia głębokiego, który stał się podstawowym wyborem dla badań nad sztuczną inteligencją.

Zwykle siła typowych metod uczenia głębokiego pochodzi z połączenia trzech czynników: złożonych modeli, wielkich zbiorów danych i wreszcie bardzo konkretnego, ograniczonego problemu do rozwiązania. Wspomniane czynniki mają wspólne podłoże - zbierane dane mają dużą, a w szczególności wystarczającą liczbę przykładów dla każdej z klas. Niestety, ten warunek jest nieprawdziwy dla większości prawdziwych problemów i jako taki nie odzwierciedla ludzkiego doświadczenia dotyczącego procesu uczenia. Ponadto, współczesne modele uczenia maszynowego zwykle są tworzone i uczone rozwiązywania tylko jednego, konkretnego problemu. Pomimo tego, że właśnie taki sposób postępowania w uczeniu głębokim zaowocował wieloma niekwestionowanymi sukcesami w wielu dziedzinach, ma też powszechne ograniczenia - np. uczone modele wymagają kosztownych obliczeń, czy ogromnych ilości ustrukturyzowanych danych. Modele są zwykle wrażliwe na rozkłady danych o długich ogonach występujące w rzeczywistych problemach, takich jak medycyna, gdzie duże ilości zbalansowanych danych należą do rzadkości. Dodatkowo, uzależnienie najnowocześniejszych modeli uczenia głębokiego od olbrzymich zasobów obliczeniowych powoduje, że tylko niewiele grup badawczych na świecie może je rozwijać.

Meta-uczenie wprowadza paradygmat „uczenia jak się uczyć” - alternatywnej strategii, w której model uczy się adaptować do licznych powiązanych zadań i zdobywa doświadczenie w trakcie wiele różnych epizodów uczenia. Następnie taki model używa zdobytego doświadczenia, by poprawić swoją wydajność na kolejnym zadaniu. Co najważniejsze, paradygmat uczenia się procesu uczenia jest analogią do procesu zdobywania doświadczenia przez ludzi i zwierzęta, a jako taki jest krokiem w kierunku zbliżenia uczenia biologicznego oraz maszynowego. Co więcej, ten paradygmat można z łatwością przystosować do rzeczywistych problemów, dla których nie mamy wielkich zbiorów danych lub ich celem jest rozwiązywanie wielu nieznacznie różniących się zadań. Wreszcie, meta-uczenie pozwala używać mniejszych ilości danych i znacznie zmniejsza wymaganą liczbę wykonywanych obliczeń, co prowadzi do demokratyzacji badań nad sztuczną inteligencją oraz korzystnie wpływa na środowisko.

Najbardziej typowym problemem meta-uczenia jest „uczenie z kilku przykładów”, gdzie dostajemy zbiór małych zadań. Każde z nich składa się ze zbioru pomocniczego zawierającego opisane przykłady jak również zbioru przykładów, które musimy sklasyfikować. Celem jest takie wytrenowanie modelu, by mógł się szybko dostosowywać do różnych zadań. W czasie ewaluacji modelu, przykłady do sklasyfikowania mogą pochodzić z klas, które nigdy nie pojawiły się w czasie samego uczenia. Próbuąc rozwiązać ten typowy przykład meta-uczenia, natrafiamy na powszechne problemy elastycznej adaptacji zarówno do nowego zadania jak i do danych pochodzących spoza rozkładu.

Te przeszkody są źródłem dla dwóch głównych gałęzi badań w meta-uczeniu - szybkiej adaptacji i szerszej generalizacji. W tym projekcie będziemy rozważać pierwszy z nich, zwiększając szybkość i elastyczność modeli w adaptacji do różnych zadań. Zainspirowani słynnym podejściem MAML i zdumiewającymi wynikami uczenia się transferu wiedzy, skoncentrujemy się na podejściach dostrajania. Ostatnie wyniki sugerują, że transfer wiedzy poprzez dostrajanie z wielkich modeli może znacząco przebić współczesne modele meta-uczenia, jednak ich największym problemem jest konieczność intensywnego przetrenowania modeli, wymagająca olbrzymich ilości danych i wielkiej mocy obliczeniowej. Jednakże, poza szeroko stosowanym modelem MAML, podejście dostrajania jest wciąż niezbadane w meta-uczeniu, w szczególności nie powstało ogólne studium nad efektywnymi sposobami lepszej adaptacji poprzez dostrajanie.

W części eksperymentalnej projektu będziemy używać szerokiego spektrum współczesnych metod uczenia głębokiego, takich jak: podejścia Bayesowskie (Bayesowskie sieci neuronowe, procesy Gaussowskie), warunkowe modele przepływowe, hipersieci, transformery i mechanizmu uwagi. Zaprezentujemy także zastosowania stworzonych metod w problemach uczenia z kilku przykładów i ciągłego meta-uczenia. Ponadto, zweryfikujemy hipotezę, czy dostrajanie jest przydatne w katastrofalnym zapominaniu przy uczeniu ciągłym.