

Streszczenie

Modele głębokiego uczenia w ostatnich latach okazały się niezwykle biegłe w wielu domenach, takich jak widzenie komputerowe, przetwarzanie języka naturalnego i chemioinformatyka. Jednocześnie dany model głębokiego uczenia jest zwykle ograniczony do jednego konkretnego zastosowania i nie może zostać łatwo zaadaptowany do nowych warunków. Z drugiej strony, w świecie rzeczywistym zwierzęta i ludzie muszą stale dostosowywać się do nowych warunków, rozwiązywać sekwencje heterogenicznych zadań i kalibrować czas reakcji w zależności od sytuacji. Nieustanne zmiany charakterystyczne dla świata rzeczywistego są jedną z głównych przeszkód na drodze do budowy bardziej uniwersalnych modeli sztucznej inteligencji, które mogą być stosowane do różnorodnych problemów. W tej rozprawie badamy różne sposoby adaptacji wcześniej wytrenowanych sieci do nowych zadań i ustawień.

Niniejsza rozprawa składa się z pięciu prac i jest podzielona na dwie części. Pierwsza część rozprawy zawiera Publikacje [I-III], które skupiają się na problemie ciągłego uczenia. Ten obszar badawczy zajmuje się budowaniem algorytmów zdolnych do trenowania na zmieniającym się rozkładzie danych przy efektywnym wykorzystaniu wiedzy z przeszłości do uczenia się nowych zadań. Jest to istotne, jako że w sieciach neuronowych występuje zjawisko katastroficznego zapominania, co oznacza, że w obliczu przesunięcia rozkładu danych model traci zdolność rozwiązywania wcześniej napotkanych zadań. Publikacja [I] wprowadza środowisko oraz protokół ewaluacji dla ciągłego uczenia ze wzmocnieniem, oparte na sekwencji zadań manipulacji robotycznej wraz z zestawem kluczowych metryk. Publikacja [II] rozszerza poprzednią pracę, skupiając się na zagadnieniu transferu wiedzy, tj. jak wykorzystać wiedzę z przeszłości, aby wydajnie uczyć się na nowych zadaniach. Identyfikujemy kluczowe aspekty architektoniczne oraz algorytmiczne transferu wiedzy oraz proponujemy zestaw wytycznych, które prowadzą do bardziej efektywnych rozwiązań. W Publikacji [III] proponujemy metodę ciągłego uczenia z gwarancjami na zapominanie, pokazując, że spadek wydajności może być ograniczony z góry przy użyciu arytmetyki interwałowej.

Druga część tej rozprawy skupia się na metodach adaptacji istniejących modeli do innych paradygmatów. W Publikacji [IV] pokazujemy, jak efektywnie przekształcić model, który wykonuje obliczenia z wykorzystaniem statycznej ścieżki obliczeniowej, tak aby dynamicznie adaptował ścieżkę obliczeniową do każdego przykładu. Dokonujemy tego, wprowadzając klasyfikatory wczesnego wyjścia, które umożliwiają zmianę czasu przetwarzania w zależności od wejścia. W Publikacji [V] wykorzystujemy technikę normalizujących przepływów do wprowadzenia warunkowania do wcześniej wytrenowanych nienadzorowanych modeli generatywnych, co pozwala na lepszą kontrolę procesu generowania.

Wszystkie powyżej wymienione Publikacje zostały zaprezentowane na topowych konferencjach rangi A* i we wszystkich jestem pierwszym autorem. Powstały w wyniku

współpracy zarówno polskiej, jak i międzynarodowej. Dodatkowo opublikowałem szereg prac dotyczących innych zagadnień głębokiego uczenia, brałem udział w stażach naukowych w renomowanych instytucjach, byłem recenzentem dla czasopism i konferencji najwyższej rangi, a także organizowałem warsztaty i szkoły letnie z zakresu głębokiego uczenia.

Słowa kluczowe: głębokie uczenie, adaptacja, uczenie ciągłe, uczenie ze wzmocnieniem