

Streszczenie popularnonaukowe

Najnowsze badania nad sztuczną inteligencją zaowocowały ogromną liczbą metod, które możemy wyszkolić tak by naśladowały ludzkie zachowanie w wielu zadaniach. Mamy już autonomiczne samochody, maszyny które mogą zastąpić operatorów w prostych zadaniach takich jak kontrola jakości produktów, czy nawet bardziej złożone systemy w których komputery oparte o sztuczną inteligencję są w stanie obliczać np. stawkę ubezpieczenia dokładniej niż jakikolwiek ludzki ekspert. Ukrytym założeniem, kryjącym się za tymi sukcesami jest ogromna ilość zebranych danych treningowych, skrupulatnie oznaczonych przez ludzi. O ile założenie to znacznie upraszcza problem nauki modeli AI do prostego rozpoznawania wzorców w danych, niekoniecznie prowadzi to do uniwersalnych metod uczenia maszynowego. Ten tak zwany nadzorowany paradygmat uczenia maszynowego jest poważnym ograniczeniem w budowaniu faktycznie inteligentnych modeli sztucznej inteligencji.

Jako ludzie już we wczesnym dzieciństwie uczymy się nie tylko w oparciu o wskazówki naszych opiekunów. Często eksplorujemy otaczający nas świat w sposób nienadzorowany i uczymy się poprzez samodzielne doświadczanie otaczającego nas świata (często w oparciu o własne błędy). W uczeniu maszynowym metody, które koncepcyjnie próbują naśladować takie zachowanie nazywamy samonadzorowanymi metodami uczenia.

Niestety, w przeciwieństwie do dzieci, wszystkie sztuczne sieci neuronowe cierpią z powodu tzw. "katasroficznego zapominania" - sytuacji, w której wydajność modelu spada drastycznie za każdym razem, gdy jest on ponownie trenowany z wykorzystaniem nowych informacji. Przykładowo, jeśli sieć wcześniej wytrenowana do wykrywania infekcji wirusowych jest dotrenowana na danych opisujących nowy, niedawno odkryty szczep, precyzja diagnostyczna na wszystkich poprzednich wirusach spadnie w sposób drastyczny. W kategoriach ludzkich to tak jakbyśmy tracili wszelką pamięć na temat poprzednich doświadczeń za każdym razem, gdy przyuczamy się do wykonywania nowego zadania.

W tym projekcie stawiamy hipotezę, że samonadzorowane uczenie maszynowe może być kluczowym składnikiem w rozwoju metod starających się przeciwdziałać zapomnianiu. Wiemy już, że odkrywanie wiedzy w systemach samonadzorowanych jest bardziej zbliżone do sposobu w jaki uczymy się jako ludzie. W związku z tym, na podstawie tej obserwacji, naszych dotychczasowych badań i najnowszych eksperymentów stawiamy tezę badawczą, że różnica ta może być również obserwowana w sposobie, w jaki systemy uczenia maszynowego zapominają. W związku z tym, planujemy dokładnie zbadać i zrozumieć istotę zapominania w sztucznych systemach samonadzorowanych, po to by docelowo zaproponować nowatorskie rozwiązania, które bez nadzoru będą nieustannie gromadzić wiedzę na temat otaczającego nas świata.