

Streszczenie popularnonaukowe

Wyobraź sobie, że jest ostatni dzień przed bardzo ważnym egzaminem, do którego rzetelnie przygotowujesz się od dłuższego czasu. Czytasz ostatnią stronę w książce po czym zamykasz ją i kładziesz się spać. A teraz wyobraź sobie swoje zdziwienie kiedy w trakcie egzaminu uświadamiasz sobie, że jedyną rzeczą, którą jesteś w stanie sobie przypomnieć jest treść ostatniej strony, którą czytałeś dzień wcześniej powtarzając do egzaminu.

Okazuje się, że Twoja wiedza uległa *katastroficznemu zapominaniu* -- zjawisku, które trapi sztuczne sieci neuronowe od kilkudziesięciu lat. Katastroficzne zapominanie nie pozwala sieciom neuronowym na stopniowe przyswajanie wiedzy ponieważ, powoduje ono, że cała poprzednio zdobyta wiedza zawarta w sieci neuronowej jest zastępowana przez informacje, których sieć neuronowa uczy się w danym momencie.

Sposób, w jaki naukowcy uczący sieci neuronowe poradzi sobie z tym problemem, pozostawia wiele do życzenia i nastęcza wiele trudności. Obecnie, aby skutecznie nauczyć sieć neuronową treści zawartych w jednej książce, musielibyśmy wyrwać wszystkie kartki z tej książki, a następnie losowo, nie zważając na numerację stron, uczyć sieć informacji znajdujących się na poszczególnych kartkach. Jeśli tę czynność powtórzymy wielokrotnie, zapewniając, że każda strona zostanie przeczytana przez sieć co najmniej kilkadziesiąt razy, możemy liczyć na to, że sieć opanuje materiał zawarty w danej książce.

Istnieją trzy główne wady takiego podejścia. Po pierwsze, właśnie zniszczyliśmy książkę. Po drugie, w dowolnym momencie procesu uczenia sieci neuronowej potrzebujemy dostępu do całej książki, a książki potrafią być ciężkie. Po trzecie, ponieważ kartki z książki zostały wyrwane i losowo pomieszane, może zdarzyć się tak, że sieć neuronowa będzie próbowała zrozumieć materiał z samego końca książki, nie rozumiejąc jeszcze podstaw przedstawionych na samym początku.

Każdy z nas jest w stanie sobie wyobrazić jak bardzo nieefektywny jest obecnie wykorzystywany proces uczenia sieci neuronowych. Niestety, pomimo wielu prób, do tej pory naukowcom nie udało się określić przyczyny, ani tym bardziej rozwiązania zjawiska katastroficznego zapominania. W tym projekcie przeanalizujemy wpływ zadania, które sieć neuronowa rozwiązuje na zjawisko zapominania. Jak pokazują nasze wcześniejsze prace, zjawisko to jest silniejsze w przypadku gdy sieć neuronowa uczy się np. rozróżniania poszczególnych obiektów, natomiast, co zaskakujące, zjawisko zapominania jest zdecydowanie słabsze gdy sieć uczy się te same obiekty przerysowywać. Nasze badania wskazują na to, że sieć, która uczy się przerysowywać obiekty tworzy wewnętrzną reprezentację przerysowywanego obiektu, która jest zdecydowanie trwalsza i bardziej odporna na zapominanie niż reprezentacja obiektów, które powstają gdy sieć neuronowa ma za zadanie je rozróżnić.

W dalszych pracach planujemy przeanalizować wpływ różnego rodzaju zadań na tworzone przez sieć neuronową wewnętrzne reprezentacje i ich trwałość oraz odporność na zjawisko zapominania. Wyniki naszych badań pozwolą na lepsze zrozumienie zjawiska zapominania oraz na stworzenie metod pozwalających to zjawisko ograniczać.

Mamy nadzieję, że w niedalekiej przyszłości będziemy mogli bez żadnych obaw pożyczać swoje książki sieciom neuronowym i dostawać z powrotem nienaruszony egzemplarz.