

Streszczenie popularno-naukowe

Sztuczna inteligencja (AI) po cichu zrewolucjonizowała nasze codzienne życie poprzez sieci neuronowe (NN) - od zaawansowanych specjalistycznych silników rozpoznawania wzorców w medycynie i nauce, po automatyzację zadań i usprawnianie podejmowania decyzji przez asystentów osobistych.

Sieci neuronowe, częściowo zainspirowane sposobem przetwarzania informacji przez nasze mózgi, to systemy cyfrowe z tysiącami, a nawet bilionami uczących się parametrów - węzłów zwanych sztucznymi neuronami, które uczą się poprzez dostosowywanie siły swoich połączeń na podstawie dostarczonych przykładów (uczenie).

Pomimo ogromnej fascynacji najnowszymi osiągnięciami AI, to czego społeczeństwo rzadko dostrzega, to astronomiczne zasoby wymagane do treningu tych systemów. Stało się to ogromną przeszkodą, ponieważ ograniczona liczba organizacji jest w stanie przeprowadzić taki trening, nie tylko ze względu na posiadanie danych, ale także z powodu wymagań obliczeniowych.

To spowalnia postęp naukowy i ogranicza naszą zdolność do ogólnego rozwoju systemów AI. Ponadto, zwiększenie efektywności treningu sieci neuronowych przyniosłoby również znaczący wpływ na redukcję negatywnego oddziaływania na środowisko.

Proces uczenia, zwany także optymalizacją lub dopasowywaniem modelu, polega na ciągłym dostosowywaniu tych miliardów parametrów w celu minimalizacji błędów, zmuszając bazową sieć do uzyskiwania pożądanych rezultatów. Często obecne paradygmaty uczenia traktują ten proces jako jednolity przez cały czas treningu, stosując te same techniki optymalizacji od początku do końca.

Tu leży sedno problemu: większość paradygmatów optymalizacji traktuje trening sieci neuronowych jako jednolity, stały proces lub wprowadza wrażliwe harmonogramowanie, które wymaga oddzielnego dostrajania. Parametr uczący - prawdopodobnie najkrytyczniejszy parametr - jest zazwyczaj wybierany poprzez wyczerpujące i obliczeniowo kosztowne przeszukiwanie siatki bez zasadniczego rozumienia tego, jak te wybory wpływają na proces.

Proponujemy fundamentalną zmianę: dążymy do prawidłowego zrozumienia tego, co dzieje się w pierwszym etapie trajektorii optymalizacji. To niezwykle trudne zadanie, ponieważ występuje dużo stochastyczności i ogromna liczba powiązanych ze sobą podprocesów w masywnej przestrzeni parametrów.

Wczesne etapy treningu są kluczowe, ponieważ ustanawiają fundament dla trajektorii uczenia model. Analiza tych etapów może dostarczyć wglądów w to, jak sieci neuronowe zaczynają formować reprezentacje danych i jak warunki początkowe wpływają na ostateczną wydajność.

To podejście obiecuje rozwiązanie fundamentalnych wyzwań nękających rozwój AI: redukcję czasów treningu, minimalizację zużycia energii, poprawę ostatecznej wydajności oraz uczynienie rozwoju AI bardziej dostępnym i zrównoważonym.

Na koniec chcielibyśmy podkreślić, że sugerowana praca dobrze wpisuje się w nowoczesną linię badań. Coraz więcej prac skupia się mniej na większej liczbie parametrów, a raczej na lepszym ich ponownym wykorzystaniu. Przyszłość sztucznej inteligencji zależy nie tylko od budowania większych modeli, ale od budowania ich mądrzej.