



UNIWERSYTET
WARSZAWSKI

Wydział Matematyki, Informatyki i Mechaniki
Instytut Informatyki

Recenzja rozprawy doktorskiej magistra Wojciecha Masarczyka pod tytułem *Data representations in non-stationary optimization*

Artykuły na których bazuje rozprawa

[P1] Wojciech Masarczyk and Ivona Tautkute. *Reducing catastrophic forgetting with learning on synthetic data*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2020.

[P2] Wojciech Masarczyk, Kamil Deja, and Tomasz Trzciński. *On robustness of generative representations against catastrophic forgetting*. International Conference on Neural Information Processing, 2021.

[P3] Wojciech Masarczyk, Tomasz Trzciński, and Mateusz Ostaszewski. *On consequences of finetuning on data with highly discriminative features*. NeurIPS 2023: UniReps Workshop, 2023.

[P4] Wojciech Masarczyk, Mateusz Ostaszewski, Ehsan Imani, Razvan Pascanu, Piotr Miłoś, and Tomasz Trzciński. *The tunnel effect: Building data representations in deep neural networks*. NeurIPS, 2023.

[P5] Wojciech Masarczyk, Mateusz Ostaszewski, Tin Sum Cheng, Tomasz Trzciński, Aurelien Lucchi, and Razvan Pascanu. *Unpacking softmax: How temperature drives representation collapse, compression and generalization*. Under review at NeurIPS 2025, 2025.

Tematem rozprawy są badania na temat reprezentacji danych w modelach opartych o sieci neuronowe w przypadku zmieniającego się zbioru danych. Jednym z głównych ograniczeń trenowanych sieci neuronowych jest powszechnie obserwowane zjawisko katastrofalnego zapomnienia, gdzie optymalizacja modelu pod kątem nowego zadania powoduje załamanie skuteczności na wcześniej opanowanych zadaniach.

Zanim przejdę do omówienia i oceny wyników chciałbym zwrócić uwagę na podkreślenie przez autora stawianych hipotez, konkretnych pytań, które są weryfikowane za pomocą przedstawianych podejść. Jestem gorącym zwolennikiem takiego podejścia do prowadzenia i przedstawiania wyników prac naukowych.

W pierwszych dwóch pracach [P1], [P2] przedmiotem badań jest problem ciągłego uczenia, gdzie do rozwiązania jest kilka problemów klasyfikacyjnych o rozłącznych klasach. W artykule [P1] przedstawione zostało nowe podejście polegające na generowaniu sztucznych danych za pomocą generatora G w taki sposób, aby za pomocą meta-uczenia otrzymać model, który jest mniej narażony na zjawisko katastrofalnego zapomnienia. Przy okazji dyskusji wyników pracy [P1] zabrakło mi wyjaśnienia ile razy ostatecznie wytrenowany model jest ewaluowany na prawdziwych danych testowych i czy istnieje w tym wypadku ryzyko przeuczenia. Mam nadzieję, że autor odniesie się do tego zagadnienia w trakcie obrony.

Praca [P2] stanowi naturalne rozszerzenie pracy [P1] gdzie dokonywane jest porównanie reprezentacji w modelach dyskryminujących i generatywnych. Porównanie standardowych klasyfikatorów oraz modeli takich jak autokodery (w wersji wariacyjnej i standardowej) pozwala na wyciągnięcie bardzo interesujących wniosków na temat różnicy w możliwościach uogólniania przez reprezentacje uzyskane w te dwa różne sposoby. Przedstawione wyniki jednoznacznie wskazują na przewagę modeli generatywnych w tak zdefiniowanym zadaniu.

Praca [P3] skupia się na aspekcie wprowadzenia do zbioru danych bardzo dyskryminującej cechy w postaci pojedynczego piksela doskonale skorelowanego z docelową klasą. Autor najpierw trenuje model na pierwotnym zbiorze danych a następnie po wprowadzeniu owego specjalnego piksela trenuje model ponownie. Przedstawione wyniki wskazują, że reprezentacja modelu ulega załamaniu i model skupia się praktycznie wyłącznie na wartości pojedynczego piksela, kompletnie tracąc swoją plastyczność, czyli zdolność reprezentacji do generalizacji do nowych zadań.

W kolejnych dwóch pracach [P4], [P5] tematyka badań dotyczy tego co dzieje się z wewnętrzną reprezentacją sieci neuronowych na przestrzeni jej kolejnych warstw. W pracy [P4] główna obserwacja pokazuje bardzo ciekawą relację pomiędzy tym jak skutecznie cechy danego poziomu reprezentacji pozwalają rozróżniać docelowe klasy a tym jaki jest rząd reprezentacji. W pierwszych warstwach sieci bardzo szybko rośnie skuteczność reprezentacji, ale w momencie gdy zostanie ona wysycona następuje bardzo szybki spadek rzędu. To pozwoliło na wydzielenie części sieci nazwanej tunelem, gdzie reprezentacja ulega szybkiej

kompresji mierząc ją rzędem. Praca [P5] jest tematycznie powiązana, gdyż analizie poddana jest konkretna warstwa typu softmax a dokładnie temperatura użyta w tej warstwie. Warstwa typu softmax jest nieodzownym elementem mechanizmu uwagi, dlatego też zrozumienie jej działania oraz przełożenia na możliwości uogólniania i katastrofального zapominania są ciekawym aspektem badawczym.

W ostatnich latach obserwujemy trend coraz częściej występujących modeli fundamentalnych, które trenowane są na tak zróżnicowanych zbiorach danych, że zjawisko katastrofального zapominania zostało zminimalizowane. Przykładem takich modeli w dziedzinie przetwarzania języka naturalnego są praktycznie wszystkie modele opracowywane przez OpenAI, Anthropic, Gemini, etc. Podobnie w przypadku modeli do wizji maszynowej bardzo skuteczne są modele takie jak Segment Anything lub DINO. Jest to na tyle istotny trend, że w rozprawie zabrakło mi odniesienia się do tego zagadnienia.

Przy okazji przedstawiania swojej opinii na temat rozprawy, chciałbym zwrócić uwagę na potrzebę wypracowania przez środowisko naukowe formatu rozpraw w przypadku, w którym format zbioru publikacji (tzw. zszywka) nie jest formalnie możliwy. Wyjściem pośrednim jest formalny format monograficzny rozpoczynający się od wprowadzenia do zagadnień doktoratu a następnie przedstawieniu kolejnych publikacji jako osobnych rozdziałów, co ma miejsce w omawianej rozprawie. Moim zdaniem o jakości całego doktoratu świadczy zbiór artykułów, które się na niego składają, dlatego też uważam że jako środowisko nie powinniśmy kurczowo trzymać się sztywnych zwyczajowych reguł stawianych monografiom, co czasami ma miejsce w recenzjach prac doktorskich w dziedzinach gdzie nowe wyniki pojawiają się bardzo dużym tempie, co charakteryzuje dziedziny uczenia maszynowego i sztucznej inteligencji.

Ocena rozprawy

Podsumowując, uważam, że przedstawione wyniki dotyczą ważnych i aktualnych zagadnień a artykuły w których zostały one opublikowane zostały przedstawione na rozpoznawalnych międzynarodowych warsztatach i konferencjach. Na szczególną uwagę zasługuje spójność tematyczna przedstawionych prac, które autor ujął jedną spójną tematycznie narracją. W mojej ocenie praca spełnia wszystkie wymagania ustawowe i zwyczajowe i rekomenduję przyznanie stopnia doktora na podstawie przedstawionej rozprawy.

Marek Cygan