

prof. dr hab. inż. Tomasz Trzcíński
Instytut Informatyki Politechniki Warszawskiej

**RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY DYSCYPLINY NAUKOWEJ
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA
INSTYTUTU PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK**

Tytuł rozprawy: Continual Learning with Experience Replay

Autor rozprawy: mgr inż. Andrii Krutsylo

1. Analiza strony merytorycznej rozprawy

1.1. Obszar problemowy

Recenzowana rozprawa doktorska dotyczy zagadnień związanych z uczeniem ciągłym (continual learning, CL), czyli obszarem badań, który zyskał szczególnie dużą dynamikę rozwojową w ostatnich latach. Modele uczone sekwencyjnie, w warunkach zmieniających się danych lub zmieniających się zadań, muszą radzić sobie z problemem *catastrophic forgetting*: utraty wiedzy nabytej we wcześniejszych etapach. W kontekście praktycznych zastosowań – takich jak robotyka, sterowanie autonomiczne, systemy personalizacji oraz systemy rekomendacyjne – zagadnienie to ma fundamentalne znaczenie.

Autor słusznie zauważa, że współczesne systemy uczenia maszynowego, szczególnie te pracujące w środowiskach dynamicznych, nie mogą być stale rekonstruowane „od zera”, dlatego efektywne techniki uczenia ciągłego są koniecznością. W rozprawie skoncentrowano się przede wszystkim na metodach bazujących na mechanizmie *experience replay*, w których przechowuje się niewielki fragment danych z poprzednich zadań, by ułatwić modelowi zachowanie wiedzy historycznej. Paradygmat ten rzeczywiście odgrywa dziś ważną rolę, choć Autor powinien szerzej uzasadnić tezę o jego dominującej pozycji, prezentując odpowiednie odniesienia literaturowe.

Obszar problemowy został zarysowany w sposób poprawny, jednak stosunkowo skrótowy. Zabrakło między innymi:

- szerszego kontekstu współczesnych badań nad CL, obejmujących foundation models, modele multimodalne oraz architektury transformacyjne,
- omówienia skali najnowszych benchmarków, takich jak SplitCIFAR, ImageNet-R, CORe50 czy CLVision.

W efekcie rozprawa nieco zawęża kontekst badań, mimo że jej tematyka dotyczy zagadnienia szerokiego i dynamicznie rozwijającego się, co widać choćby w najnowszych publikacjach NeurIPS, ICLR czy ICML.

1.2. Przegląd literatury i jego kompletność

W porównaniu do wzorcowych rozpraw doktorskich prezentujących szerokie ujęcie stanu wiedzy (np. z bibliografią przekraczającą 300–400 pozycji), bibliografia w niniejszej pracy liczy mniej niż 100 pozycji, a ponadto znaczna część odnosi się do prac sprzed 2020 roku. W szczególności zabrakło mi omówienia aktualnych wyników badań nad replay (np. o adaptacyjnych buforach lub metodach inteligentnego próbkowania), oraz dokładniejszego omówienia głębokich modeli generatywnych używanych do replay bez bufora (np. MultibandVAE).

Co więcej, niektóre metryki, takie jak BWT czy FWT, choć zdefiniowane, nie zostały później wykorzystane w analizie eksperymentalnej, co zaburza spójność pomiędzy częścią teoretyczną a empiryczną.

1.3. Oryginalność pracy i wyniki przedstawione w rozprawie

Autor deklaruje trzy główne kontrybucje, w tym propozycja nowych metod wyboru próbek do powtórzenia, sposobu ich łączenia z danymi treningowymi oraz nowej metryki ewaluacyjnej. Choć zaproponowane metody są ciekawe i oryginalne, szczególnie zaskakujący dla mnie jest fakt umieszczenie kontrybucji dotyczącej ewaluacji jako trzeciej kontrybucji, mimo że sekcja 3.1.2 sugeruje jej fundamentalny charakter dla całej pracy. Ewaluacja jest kluczowym elementem badań nad CL i powinna być wyraźniej wyeksponowana. Co więcej, zaproponowana metryka *Knowledge Retention* jest ciekawa i z chęcią zobaczyłbym jak proponowane w rozdziałach czwartym i piątym modyfikacje wpływają na nią.

W części empirycznej można docenić rzetelność przygotowanych eksperymentów, jednak występują istotne niedociągnięcia:

- użyte w eksperymentach zbiory danych są stosunkowo niewielkie i mało aktualne,
- wyniki między metodami są nierzadko bardzo bliskie – różnice nie są statystycznie istotne (przykład: Tabela 5.1),
- brakuje informacji o liczbie powtórzeń w części tabel (np. Tabela 5.2),
- brak wariancji czasów wykonania w Tabeli 5.2: podano wartości bez zakresów niepewności, co utrudnia ich interpretację,
- brak analizy istotności statystycznej, mimo naturalnej wariancji w sieciach neuronowych.

Zwraca uwagę wykorzystanie SentenceBERT jako zamrożonego modelu w metodzie CVM, co stoi w pewnej sprzeczności z motywacją pracy dotyczącą urządzeń zasobooszczędnych. Modele SentenceBERT liczą dziesiątki milionów parametrów (od 20 do 100 milionów), a ich użycie na zasobooszczędnych platformach (np. robotach), o których rozprawa traktuje we wprowadzeniu, może być utrudnione ze względu na dostępną w nich ograniczoną pamięć czy zasoby bateryjne.

1.4. Uwagi dyskusyjne i zagadnienia wymagające doprecyzowania

Poniższe uwagi dotyczą zarówno konkretnych kwestii merytorycznych poruszonych w rozprawie, jak i bardziej ogólnych aspektów recenzowanej rozprawy. Liczę na analizę tych zagadnień i odniesienie się do nich przez Doktoranta.

- **Efektywność energetyczna.** Autor kilkakrotnie apeluje do znaczenia efektywności obliczeniowej, jednak nie przedstawia żadnych metryk energetycznych. Choć przedstawione wyniki obejmują kwestie czasu potrzebnego do obliczenia, nie do końca wiadomo jak zaproponowane przez Autora modyfikacje mogą przełożyć się na oszczędność energetyczną związaną z zastosowaniem proponowanych metod. Analiza złożoności ogranicza się więc do czasu wykonania, brakuje także danych dotyczących pamięci całego modelu, nie tylko bufora replay. Czy Doktorant mógłby skomentować ten aspekt?
- **Brak metod niereplayowych w analizach porównawczych.** Metody takie jak EWC, SI czy LwF pojawiają się dopiero w późniejszej części pracy, brakuje mi ich w wielu ewaluacjach przedstawionych w rozdziale czwartym. Czy Autor mógłby skomentować lub uzupełnić te braki?
- **Szerszy kontekst ewaluacji proponowanych metod.** Moim zdaniem przeprowadzone eksperymenty mogłyby zostać znacząco poprawione, choćby poprzez wykorzystanie wprowadzonych w rozdziale drugim i szóstym metryk ewaluacyjnych, w tym BWT/FWT czy KR. Dodatkowo, rekomendowałbym ewaluację na bardziej aktualnych i większych zbiorach danych, na których wyniki mogłyby klarowniej pokazywać relację między testowanymi metodami, np. ImageNet-R. W pracy brakuje mi również porównań metod z innymi podejściami zajmującymi się np. analizą przestrzeni ukrytej, choćby metoda MultibandVAE. Czy Autor może odnieść się do tej kwestii?

2. Analiza strony formalnej pracy

2.1. Redakcja i układ rozprawy

Praca została przedstawiona w języku angielskim i obejmuje wstęp, część teoretyczną, opis proponowanych metod, eksperymenty oraz wnioski. Struktura jest poprawna, jednak znacznie krótsza niż przeciętne rozprawy w obszarze uczenia maszynowego, co samo w sobie nie świadczy na jej niekorzyść, jednak w połączeniu z poruszonymi powyżej kwestiami wskazuje na przestrzeń do poprawy.

W szczególności, moim zdaniem, w rozprawie brakuje pełniejszego streszczenia kolejnych rozdziałów we wprowadzeniu, szerszego omówienia literatury, w szczególności metod innych niż replay-based, oraz integrującego rozdziału porównawczego, w którym proponowane metody byłyby przeanalizowane pod kątem omówionych w rozprawie metryk.

2.2. Uwagi językowe i stylistyczne

Praca jest napisana poprawną angielszczyzną, jednak zauważyłem kilka błędów:

- s. 8 - Błąd w przenoszeniu pomiędzy wersami: *Mod-els* podczas gdy powinno być *Mo-dels*.
- s. 10 - Długie zdanie, prowadzące do błędnej formy czasownika *replace*: *This dissertation examines explicit memory construction, through sample selection strategies that promote diversity, implicit construction, through feature-level representations learned during training, and memory alternatives, which replaces the dynamic buffer with a fixed, pretrained language model that provides stable embeddings as anchors for each new class.*
- s. 44 – Błąd: “Omiglot” → powinno być “Omniglot”.
- Powtarzane wielokrotnie długie (45 wyrazów lub więcej) zdania, szczególnie w sekcjach kontrybucji oraz metodologii.

3. Konkluzja

Podsumowując, recenzowana rozprawa mgr inż. Andrii Krutsylo zawiera wartościowe elementy naukowe, a przedstawione wyniki eksperymentalne i koncepcje metodyczne stanowią wkład w rozwój metod uczenia ciągłego, choć praca wymagałaby pogłębienia pod względem przeglądu literatury, jakości analiz eksperymentalnych oraz szerokości ujęcia problemu.

Stwierdzam, że rozprawa mgr inż. Andrii Krutsylo spełnia wymagania określone w Ustawie z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce. Przedstawione badania oraz sposób ich raportowania świadczą o przygotowaniu Autora do prowadzenia samodzielnej pracy naukowej.

Wnioskuje o dopuszczenie Doktoranta do dalszych etapów przewodu doktorskiego.



Tomasz Trzcinski