



Marcin Kurdziel
Wydział Informatyki
Akademia Górniczo-Hutnicza im. St. Staszica w Krakowie
al. Mickiewicza 30, 30-059 Kraków
kurdziel@agh.edu.pl

Recenzja rozprawy doktorskiej dla Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Poznańskiej

Tytuł rozprawy: *Beyond traditional Curriculum Learning: shaping the optimization path to find better minima, faster*

Autor rozprawy: mgr inż. Izabela Krysińska

Niniejszą recenzję przygotowałem w odpowiedzi na pismo Prodziekana Wydziału Informatyki i Telekomunikacji Politechniki Poznańskiej, dra hab. inż. Szymona Wilka, prof. PP, z dnia 30 października 2025 r., w związku z powołaniem mnie przez Radę Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Poznańskiej na recenzenta w postępowaniu w sprawie nadania stopnia doktora Pani mgr inż. Izabeli Krysińskiej.

1 Obszar badań i uzyskane wyniki

Recenzowana rozprawa doktorska mieści się w obszarze informatyki badającym algorytmy tzw. uczenia według programu (ang. *Curriculum Learning*). Inspiracją stojącą za tą klasą algorytmów jest obserwacja, że człowiek przyswaja wiedzę oraz nabywa umiejętności w sposób etapowy, rozpoczynając od prostych pojęć i technik, stopniowo przechodząc do coraz trudniejszych zagadnień. Taki etapowy sposób uczenia kontrastuje z powszechnie przyjętą w uczeniu maszynowym praktyką losowego wyboru przykładów uczących. Badania w obszarze uczenia według programu poszukują więc metod szeregowania przykładów uczących w sposób, który pozwala poprawić jakość wytrenowanego modelu lub poprawić efektywność wykorzystania danych (ang. *sample efficiency*).

Pomimo intuicyjnie przekonującej motywacji stojącej za metodami uczenia według programu, algorytmy te nie znalazły dotychczas szerokiego zastosowania w praktyce uczenia maszynowego. Doktorantka celnie podkreśla w rozprawie rozdźwięk pomiędzy potencjałem uczenia według programu a obserwowanymi w praktyce, rozczarowującymi wynikami

empirycznymi. Na tym tle celem Doktorantki jest systematyczne zbadanie wpływu dwóch głównych komponentów uczenia według programu na uzyskiwane rezultaty: funkcji szacującej poziom trudności przykładu (ang. *scoring function*) i funkcji porządkującej przykłady według trudności (ang. *spacing function*).

Realizując tak postawione zadanie badawcze, Doktorantka w pierwszej kolejności porównała dwa podejścia do oceny trudności przykładu uczącego, tj. ocenę statyczną bazującą na *typowości* przykładu oraz ocenę dynamiczną, biorącą pod uwagę oszacowanie niepewności predykcji. Następnie Doktorantka porównała kilka funkcji porządkujących przykłady uczące, koncentrując się na sposobie ich kompozycji w mini-batche przetwarzane przez algorytm uczący. Na podstawie uzyskanych wyników Doktorantka argumentuje, że odpowiednio zaprojektowane metody uczenia według programu poprawiają efekty uczenia w szerokiej klasie problemów klasyfikacyjnych, począwszy od klasyfikacji danych tabelarycznych, poprzez klasyfikację tekstu i obrazów, a skończywszy na wybranych elementach systemu rozpoznawania mowy.

Rezultaty uzyskane w trakcie przygotowania rozprawy doktorskiej Doktorantka opublikowała w 4 artykułach naukowych:

- [A1] Krysińska, I., Morzy, M. i Kajdanowicz, T. (2021). *Curriculum Learning Revisited: Incremental Batch Learning with Instance Typicality Ranking*. W: *International Conference on Artificial Neural Networks*, strony 279–291.
- [A2] Kunst, J. R., Gundersen, A. B., Krysińska, I., Piasecki, J., Wójtowicz, T., Rygula, R., van der Linden, S. i Morzy, M. (2024). *Leveraging artificial intelligence to identify the psychological factors associated with conspiracy theory beliefs online*. *Nature Communications*, 15(1), 7497.
- [A3] Krysińska, I., Wójtowicz, T., Olejnik, A., Morzy, M. i Piasecki, J. (2021). *Be careful who you follow: the impact of the initial set of friends on COVID-19 vaccine tweets*. W: *Proceedings of the 2021 Workshop on Open Challenges in Online Social Networks*, strony 1–8.
- [A4] Gwiazdziński, P., Gundersen, A. B., Piksa, M., Krysińska, I., Kunst, J. R., Noworyta, K., Olejnik, A., Morzy, M., Rygula, R., Wójtowicz, T. i inni (2023). *Psychological interventions countering misinformation in social media: A scoping review*. *Frontiers in psychiatry*, 13, 974782.

Wyniki te omówię szczegółowo w dalszej części niniejszej recenzji.

Uczenie według programu

Zasadniczą część rozprawy doktorskiej rozpoczyna rozdział 2, w którym Doktorantka wprowadza podstawowe pojęcia z obszaru tematycznego pracy. W rozdziale tym strategia uczenia według programu zdefiniowana jest jako para funkcji $(\mathcal{S}, \mathcal{P})$. Funkcja \mathcal{S} odpowiada za ocenę stopnia trudności przykładów wejściowych. Funkcja \mathcal{P} komponuje natomiast sekwencję mini-batchy dla algorytmu uczącego, biorąc pod uwagę oszacowany stopień trudności przykładów i numer kroku uczenia. Doktorantka wprowadza również podział strategii uczenia według programu na strategie sterowane danymi, w których każdy przykład wejściowy

oceniany jest niezależnie, oraz strategie szeregujące zadania, w których przykłady wejściowe grupowane są w pewne podzadania o rosnącej skali trudności.

W dalszej części rozdziału Doktorantka przedstawia kluczową literaturę poświęconą obu elementom strategii uczenia według programu. Publikacje poświęcone szacowaniu trudności przykładów uczących omawia w podziale na metody statyczne i dynamiczne. Pierwsza klasa algorytmów bierze pod uwagę wyłącznie zawartość przykładu uczącego. Stopień trudności może być wówczas oceniony przez eksperta (np. radiologa oceniającego zdjęcie rentgenowskie) lub z wykorzystaniem miar heurystycznych (np. długość dokumentu w zagadnieniach klasyfikacji tekstu). Metody dynamiczne oceniają stopień trudności przykładu, biorąc pod uwagę sposób jego przetwarzania przez algorytm uczący. Oceną stopnia trudności może być wówczas funkcja kosztu w sieci neuronowej, entropia zwróconego przez sieć rozkładu prawdopodobieństwa nad klasami, czy też zmienność predykcji obserwowana w kolejnych epokach uczenia. Rozdział zamyka przedstawienie kilku statycznych funkcji porządkujących przykłady uczące (np. zwiększających stopień trudności w sposób logarytmiczny, liniowy lub wykładniczy) oraz wybranych prac proponujących porządkowanie adaptacyjne, uwzględniające skuteczność trenowanego modelu.

Formalizm wprowadzony w rozdziale 2 pozwala Doktorantce uporządkować przegląd literatury wzdłuż dwóch głównych aspektów uczenia według programu poruszanych w pracy, tj. oceny trudności przykładów uczących i strategii ich porządkowania. Znajduje on również odzwierciedlenie w układzie rozdziałów prezentujących zasadnicze wyniki badań Doktorantki. W pracy nie podjęto natomiast próby poszukiwania wyników formalnych zbudowanych na bazie wprowadzonych pojęć. Nie jest to jednak zaskakujące w kontekście wykorzystania uczenia według programu do trenowania sieci neuronowych. Formalna analiza procesu uczenia sieci neuronowych wciąż pozostaje otwartym problemem badawczym.

Ocena *typowości* przykładów uczących

Klasyczne strategie uczenia według programu szeregują przykłady na podstawie oszacowania ich trudności. W rozdziale 3 Doktorantka bada alternatywne rozwiązanie: szeregowanie przykładów uczących na podstawie oceny ich *typowości*, tj. podobieństwa do innych przykładów uczących. Wedle hipotezy badanej w tym rozdziale, szeregowanie przykładów od najbardziej do najmniej typowych może poprawić reprezentacje (Doktorantka posługuje się tu pojęciem *core concepts*) poszczególnych klas, tym samym zwiększając dokładność klasyfikacji.

W ramach weryfikacji postawionej hipotezy Doktorantka zbadała trzy miary typowości przykładu uczącego. Punktem wyjścia do każdej z tych miar jest graf, którego wierzchołkami są poszczególne przykłady uczące, zaś krawędzie łączą przykłady, których podobieństwo przekracza pewien próg. Typowość przykładu jest następnie definiowana jako jedna z trzech miar centralności wierzchołka: stopień wierzchołka, miara zbudowana na podstawie entropii empirycznego rozkładu klas w sąsiadach wierzchołka, oraz miara łącząca stopień wierzchołka i entropię empirycznego rozkładu klas. Wykorzystując tak zdefiniowane miary, Doktorantka zbadała skuteczność *uczenia według typowości* na wybranych zbiorach danych z repozytorium *UCI Machine Learning Repository*. Aby ocenić istotność statystyczną uzyskanych wyników, Doktorantka wykonała eksperymenty wielokrotnie, każdorazowo dzieląc

losowo dane na zbiór uczący i ewaluacyjny. Uzyskane wyniki pokazały, że dwie spośród badanych miar typowości dają wyniki odmienne od klasycznego SGD ($p < 0.05$). Niestety, w obu przypadkach były to wyniki słabsze od SGD. Dla trzeciej miary typowości przykładu nie uzyskano wyników istotnych statystycznie. Ostatecznie więc hipoteza badana w tym rozdziale nie znalazła potwierdzenia eksperymentalnego.

W mojej ocenie negatywny wynik uzyskany w rozdziale 3 może być pochodną zestawu danych wykorzystanych w eksperymentach. Wszystkie eksperymenty wykonano mianowicie na danych tabularycznych. Choć dla takich danych można policzyć miarę „typowości”, to jednak nie jest jasne, jak zweryfikować, czy faktycznie szacuje ona interesującą nas cechę. Ciekawe byłoby więc rozszerzenie badań na proste obrazy (przykładowo, w zadaniu rozpoznawania twarzy), gdzie można pokusić się o jakościową ocenę miary „typowości”.

Wyniki uzyskane w rozdziale 3 zostały opublikowane na konferencji ICANN 2021 [A1].

Szeregowanie przykładów na podstawie wariancji gradientu

W rozdziale 4 Doktorantka bada strategię uczenia według programu porządkującą przykłady uczące na podstawie wariancji gradientu (ang. *Variance of Gradients*, VoG). W przeciwieństwie do metody badanej w rozdziale 3, ocena stopnia trudności przykładu uczącego ma tu charakter dynamiczny. Zależy ona mianowicie od wartości wag w trenowanej sieci neuronowej. Bardziej precyzyjnie, w celu oceny trudności przykładu \mathbf{x} wyznaczany jest gradient logitu prawidłowej klasy ze względu na \mathbf{x} . Gradient ten wyznaczany jest co pewną liczbę kroków – zebrane wartości pozwalają następnie oszacować wariancję każdego elementu macierzy gradientu. Przeciętna wariancja elementów tej macierzy jest traktowana jako miara trudności przykładu wejściowego. Sam VoG nie jest oryginalną propozycją doktorantki¹. Uprzednio nie był jednak badany jako element strategii uczenia według programu.

Algorytm badany w rozdziale 4 wybiera przykłady uczące w sposób losowy, z prawdopodobieństwem proporcjonalnym do aktualnego oszacowania VoG. Co ciekawe, jest to strategia istotnie odmienna od klasycznego uczenia według programu: losowanie przykładów z prawdopodobieństwem proporcjonalnym do VoG skutkuje nadreprezentacją przykładów stanowiących wyzwanie dla aktualnych parametrów trenowanej sieci. Konstrukcja ta motywowana jest chęcią skupienia procesu uczenia na przykładach, dla których sieć nie ma jeszcze stabilnej granicy decyzyjnej.

Doktorantka oceniła efekty zaproponowanej strategii uczenia, wykorzystując cztery zbiory danych: MNIST oraz trzy zbiory tekstowe. Warto tu podkreślić, iż dwa z wykorzystanych zbiorów tekstowych zostały stworzone z udziałem Doktorantki i są obecnie publicznie dostępnymi benchmarkami. W eksperymentach wykorzystano zarówno mniejsze sieci konwolucyjne i perceptrony wielowarstwowe, jak i duży model pretrenowany (RoBERTA base). Uzyskane wyniki pokazały, że zaproponowana strategia uczenia według programu poprawia w sposób istotny statystycznie dokładność wytrenowanej sieci. Doktorantka pokazała również, że zaproponowana strategia uczenia zmniejsza końcowe wartości VoG względem klasycznego SGD oraz zmienia populację przykładów z dużą wariancją gradientu.

¹Został zaproponowany w: Agarwal, C., et al., CVPR 2022, strony 10368-10378.

Zbiory danych wykorzystane w rozdziale 4 zostały opublikowane w dwóch pracach, których współautorem jest Doktorantka: pracy w renomowanym czasopiśmie *Nature Communications* [A2] oraz pracy z warsztatów organizowanych w ramach konferencji HT 2021 [A3]. Przygotowanie tych zbiorów danych poprzedziła publikacja przeglądowa w czasopiśmie *Frontiers in psychiatry*, której współautorem jest Doktorantka [A4].

Orkiestracja mini-batchy w uczeniu według programu

Rozdział 5 poświęcony jest ocenie strategii uczenia według programu z perspektywy orkiestracji mini-batchy z wejściowymi przykładami. Pod pojęciem tym kryją się strategie, które realizują uczenie według programu poprzez adaptację rozmiaru mini-batchy, kompozycji mini-batchy oraz konstrukcji sekwencji mini-batchy. Doktorantka bada tu hipotezę, że odpowiednio zaprojektowana orkiestracja mini-batchy pozwala poprawić skuteczność trenowanej sieci.

Podobnie jak omówione dotychczas wyniki, Rozdział 5 ma charakter pracy eksperymentalnej. Punktem wyjścia do przeprowadzonych badań była luka dostrzeżona w literaturze: brak było wówczas prac, które systematycznie badałyby strategie uczenia według programu polegające na jednoczesnej adaptacji rozmiaru mini-batchy i sekwencji przykładów uczących. Wychodząc z tej przesłanki, Doktorantka zaproponowała dwie serie eksperymentów: eksperymenty łączące statyczne szeregowanie według typowości z różnymi wariantami SGD oraz eksperymenty łączące szeregowanie przykładów według wariacji gradientu z wybranymi strategiami adaptacji rozmiaru mini-batchy.

W pierwszej serii eksperymentów porównano SGD, mini-batch SGD, strategię polegającą na wielokrotnym przetwarzaniu tego samego mini-batcha (przed przejściem do kolejnego mini-batcha, z mniej typowymi przykładami) oraz modyfikację metody mini-batchy, w której i -ta iteracja przetwarza pierwsze $b + im$ przykładów (szeregowanych względem typowości), dla pewnych b i m . Rezultaty eksperymentów sugerują, że SGD z metodą mini-batchy i szeregowaniem według typowości daje słabsze wyniki niż klasyczne uczenie bez programu. Szeregowanie według typowości poprawia natomiast wyniki w połączeniu z pozostałymi wariantami orkiestracji mini-batchy. Doktorantka stawia w tym kontekście hipotezę, wedle której różnice pomiędzy kolejnymi mini-batchami (skonstruowanymi z przykładów posortowanych według typowości) wprowadzają szum utrudniający optymalizację sieci. Pojawia się tu jednak pytanie, dlaczego analogiczny problem nie jest obserwowany dla SGD (a więc wariantu z jednym przykładem per krok optymalizacji)? W mojej ocenie warto byłoby zbadać tą kwestię szerzej. Przykładowo, czy obserwowane relacje są stabilne dla różnych rozmiarów mini-batcha? Wariant z rosnącym rozmiarem mini-batcha nie odpowiada wprost na to pytanie, ponieważ nie jest to klasyczna metoda mini-batchy.

W drugiej serii eksperymentów porównano trzy schematy adaptacji rozmiaru mini-batchy w trakcie uczenia: stały rozmiar mini-batcha, przyrost liniowy oraz przyrost wykładniczy z cyklicznym powrotem. Przykłady były albo szeregowane według wariacji gradientu, albo losowane z prawdopodobieństwem proporcjonalnym do wariacji gradientu (wariant VGL). Punktem odniesienia była przypadkowa kolejność przykładów. Uzyskane wyniki nie wykazują tu jednoznacznej przewagi żadnej z ewaluowanych metod. Jedyny istotny statystycznie wynik uzyskano w eksperymencie z liniowo rosnącym rozmiarem mini-batcha,

gdzie VGL poprawił rezultaty względem losowego porządku przykładów. Wyniki te były jednak en bloc gorsze od uczenia ze stałym, dużym rozmiarem mini-batcha. Doktorantka słusznie zauważyła, że na uzyskane rezultaty może w istotny sposób wpływać założenie stałego tempa uczenia. Wykonała również analizę ablacji, w której adaptacja tempa uczenia poprawiła wyniki referencyjne. Otwartym pozostaje pytanie, jak adaptacja tempa uczenia wpłynęłaby na pozostałe wyniki w tej serii eksperymentów.

Wyniki zaprezentowane w tym rozdziale 5 powstały na podstawie publikacji [A1].

Szeregowanie zadań w Uczeniu według programu

W rozdziale 6 Doktorantka badała wariant uczenia według programu, w którym szeregowane są nie tyle poszczególne przykłady, co pewne podzadania. Wątkiem przewodnim w tych pracach było zagadnienie trenowania systemu do automatycznego rozpoznawania mowy. Doktorantka zaproponowała w tym kontekście metodę *Transitional Object Learning* (TOL), w której funkcja kosztu dekomponowana jest na średnią ważoną z funkcji kosztu dla poszczególnych podzadań. W trakcie uczenia wagi podzadań są adaptowane w taki sposób, by początkowo dominował koszt podzadań prostszych, zaś pod koniec koszt zadań trudniejszych.

Doktorantka zdefiniowała dwa podzadania dla problemu automatycznego rozpoznawania mowy: rozpoznawanie rodzajów głosek (samogłoski, spółgłoski dźwięczne, spółgłoski bezdźwięczne, pauza) oraz rozpoznawanie poszczególnych fonemów. Początkowo funkcja kosztu optymalizowała rozpoznawanie rodzajów głosek, zaś pod koniec uczenia dominował koszt związany z rozpoznawaniem fonemów. Doktorantka porównała tę strategię z popularnym algorytmem CTC (ang. *Connectionist Temporal Classification*), uzyskując istotną statystycznie poprawę wyników dla trzech ewaluowanych zbiorów danych. Zaproponowana strategia okazała się również poprawiać stabilność trenowania sieci. Co istotne, Doktorantka wykonała również test ablacyjny polegający na losowym przemieszaniu etykiet klas w zadaniu rozpoznawania rodzajów głosek. Spowodowało to istotne pogorszenie rezultatów, co pokazuje, że uzyskane pierwotnie wyniki faktycznie wynikają z dekompozycji zadania na dwa podproblemy.

Warto tu podkreślić, że efektem prac Doktorantki jest nie tylko nowy algorytm uczenia dla modeli rozpoznawania fonemów, lecz także publicznie dostępny korpus danych wraz z pretrenowanym modelem.

2 Ocena uzyskanych wyników

Wyniki uzyskane przez Doktorantkę oceniam pozytywnie. Tematyka podjęta w pracy jest wymagająca: strategie uczenia według programu obejmują kilka niezależnych elementów algorytmicznych (ocena trudności lub typowości przykładu, szeregowanie przykładów, dekompozycja na podzadania), co istotnie zwiększa obszar poszukiwania skutecznych rozwiązań. Co więcej, tempo rozwoju uczenia głębokiego w ostatnich latach powoduje szybkie wysycenie rezultatów w dostępnych zbiorach ewaluacyjnych, utrudniając tym samym porównywanie algorytmów. W tym kontekście należy docenić uporządkowanie problematyki

uczenia według programu zaproponowane w rozdziale 2. Pozwala ono Doktorantce na systematyczne badanie w kolejnych rozdziałach poszczególnych elementów konstruowanych strategii. Cennym elementem pracy jest również wyjście poza dominujący w dziedzinie paradygmat, wedle którego strategie uczenia według programu winny szeregować przykłady uczące względem rosnącego stopnia trudności. Doktorantka zbadała między innymi algorytmy, w których priorytet mają przykłady trudne (Rozdział 4), uzyskując pozytywne wyniki. Wykorzystała w tym zakresie adaptacyjną miarę trudności przykładu uczącego, która nie była uprzednio badana w kontekście uczenia według programu. Warto również podkreślić ciekawe wyniki uzyskane w obszarze uczenia modeli do rozpoznawania fonemów: prosta ale dobrze umotywowana dekompozycja problemu na dwa podzadania znacząco poprawiła tu efekty uczenia. Doktorantka każdorazowo weryfikuje istotność statystyczną uzyskiwanych wyników i jasno wskazuje badania, w których uzyskano rezultaty negatywne. W połączeniu z szerokim spektrum badanych rozwiązań, daje to dobrą intuicję odnośnie strategii mogących potencjalnie poprawić efekty uczenia. Cennym efektem prac Doktorantki są również nowe, publicznie dostępne zbiory danych i pretrenowane modele. W mojej ocenie, wyniki te stanowią sumarycznie istotny wkład w badania nad metodami uczenia według programu.

Lektura pracy doktorskiej utwierdziła mnie w przekonaniu, że Doktorantka dysponuje dobrym warsztatem badawczym w dziedzinie uczenia maszynowego. Potrafi prawidłowo projektować eksperymenty i analizować ich wyniki. Dobrze orientuje się w literaturze poświęconej zagadnieniom uczenia według programu oraz szerszym aspektom uczenia głębokiego. Wiele z prezentowanych wyników powstało w ramach współpracy w projektach z zewnętrznym finansowaniem. Wszystko to świadczy o dojrzałości naukowej Doktorantki. Sformułowane w recenzji pytania i prośby o doprecyzowanie stanowią jedynie zaproszenie do dalszej dyskusji nad uzyskanymi rezultatami.

Praca napisana jest w sposób czytelny, bez większych błędów językowych. Doktorantka zadbała również o przejrzysty skład tekstu. Jedynie w rozdziale 5 wkraśl się techniczny błąd w odniesieniach do tabeli, który został dostrzeżony już po złożeniu pracy.

3 Konkluzja

W mojej ocenie rozprawa doktorska Pani mgr inż. Izabeli Krysińskiej spełnia wymogi sformułowane w art. 187 Ustawy *Prawo o szkolnictwie wyższym i nauce* z dnia 20 lipca 2018 r., z późniejszymi zmianami. Wnioskuje o dopuszczenie Pani Izabeli Krysińskiej do dalszych etapów postępowania w sprawie nadania stopnia doktora w dyscyplinie *Informatyka Techniczna i Telekomunikacja*.

Z poważaniem,
Marcin Kurdziel