

# **Recenzja rozprawy doktorskiej dla Rady Dyscypliny Naukowej Informatyki Technicznej i Telekomunikacji Politechniki Warszawskiej**

**Tytuł rozprawy:** Data representations in generative modelling

**Autor rozprawy:** Kamil Deja

## **1 Analiza strony merytorycznej rozprawy**

Rozprawę doktorską stanowi zbiór łącznie pięciu: czterech opublikowanych oraz jednego oczekującego na druk (zaakceptowanego), powiązanych tematycznie artykułów naukowych.<sup>1</sup>

### **1.1 Obszar problemowy**

Praca doktorska skupia się na zadaniu uczenia reprezentacji danych w przestrzeniach posiadających strukturę użyteczną do wykorzystania w zadaniach docelowych (np. klasyfikacja, predykcja) z wykorzystaniem modeli generatywnych. Swą podstawę rozważań na ten temat wywodzi z możliwości stojących za architekturami głębokich sieci neuronowych takich jak klasyczny autokoder, wariacyjny autokoder, generatywnymi sieciami współzawodniczącymi (ang. Generative Adversarial Networks (GANs)) oraz dyfuzyjnymi modelami generatywnymi (ang. Diffusion-based Deep Generative Models (DDGMs)).

W rozprawie skupiono się na badaniu wewnętrznych mechanizmów różnych modeli generatywnych, starając się odpowiedzieć na trzy zasadnicze pytania badawcze. Pierwsze z nich dotyczy sposobu kodowania przykładów danych w reprezentacjach przydatnych dla modelowania generatywnego, drugie dotyczy możliwości wykorzystania reprezentacji danych nauczonej w sposób nienadzorowany przez modele generatywne poza zadaniem modelowania generatywnego, a ostatnie dotyczy wpływu retrenowania (dotrenowywania) modeli z wykorzystaniem dodatkowych danych na zmiany reprezentacji oraz możliwości łączenia wiedzy wprowadzanej do modelu generatywnego w oddzielnych zadaniach poprzez dopasowywanie reprezentacji (ang. alignment). W ogólności, przez dokładne zbadanie tych trzech pytań badawczych, rozprawa dąży do głębszego zrozumienia mechanizmów działania modeli generatywnych, ich ograniczeń oraz możliwości ich zastosowania w różnorodnych dziedzinach. Niewielką wadą tych pytań badawczych jest ich szeroki zakres i duże spektrum możliwości, co do rodzaju badań, które można na ich podstawie przeprowadzić. Chociaż każde z pytań jest niewątpliwie interesujące, nie dają one jednoznacznego obrazu problemu, który planuje się rozwiązać. Na przykład,

---

<sup>1</sup>jedna z publikacji jest w druku w ramach konferencji ECML 2023, która odbywa się w terminie 18-22.09.2023

---

trudno jest określić, jakie kryteria powinna spełniać "użyteczna reprezentacja" (np. takie jak inwariancje czy ekwiwariancje) lub co dokładnie oznacza "użyteczność" w kontekście innych zadań niż generatywne. Mimo to, zastosowane podejście do sformułowania pytań badawczych w pracy jest zgodne z obecnym trendem, zwłaszcza gdy badania mają charakter empiryczny, a rezultaty powstają w ramach materialnych artefaktów naukowego procesu twórczego.

Pierwszy rozdział pracy stanowi kompleksowe wprowadzenie do głównych pytań i obszarów badawczych rozprawy, ukazując je w kontekście istniejących metod i potencjalnych zastosowań w uczeniu maszynowym i sztucznej inteligencji. Obejmuje on również omówienie pięciu opublikowanych wyników naukowych i wskazuje na dodatkowe publikacje, które nie zostały włączone do treści rozprawy.

Rozdział drugi wprowadza podstawowe koncepcje związane z poruszaną tematyką i przedstawia trzy różne rodzaje modeli generatywnych: (1) generatywne autokodery, jako szczególny przypadek autokoderów, posiadających pewne ograniczenia w kontekście modelowania generatywnego, i ich rozszerzenia do wariacyjnych autokoderów, (2) generatywne sieci współzawodniczące (GAN) z omówieniem ich zalet i wad, w tym braku zdolności uczenia reprezentacji dla konkretnych danych wejściowych, oraz (3) dyfuzyjne modele generatywne (DDGMs), które są postrzegane jako bardzo głębokie wariacyjne autoenkodery z określonym rodzajem posteriorów wariacyjnych. W rozdziale zawarte są wybrane detale techniczne związane z trenowaniem i modelowaniem tych sieci, a rozdział dostarcza analizę podstawowych podejść do generatywnego modelowania danych.

Rozdział trzeci przedstawia pogłębioną analizę podejść związanych już z konkretnymi propozycjami przedstawionych w doktoracie rezultatów badawczych, w tym omówiono wybrane metody, które koncentrują się na reprezentacjach ukrytych z generatywnych autoenkoderów, generatywnych modeli dyfuzyjnych i ich zastosowań do różnych zadań, w tym zadań fizyki wielkich energii. Omówiono różne rozszerzenia i modyfikacje modeli generatywnych opartych na dyfuzji (DDGMs), które zaproponowane zostały w literaturze, w tym dla wielu modalności danych rozważanych łącznie. Podkreślono, że generatywne modele odgrywają kluczową rolę w bardzo różnych dziedzinach, od syntezy mowy po analizy biomedyczne. Ogólnie rzecz biorąc, rozdział zawiera dogłębną analizę i różnorodne zastosowania modeli generatywnych w różnych dziedzinach nauki.

Kolejne trzy rozdziały zawierają teksty opublikowanych, bądź zaakceptowanych do publikacji prac, które stanowią pierwszą część zbioru publikacji, t.j.,

rozdział czwarty - Deja, K., Dubiński, J., Nowak, P., Wenzel, S., Spurek, P., & Trzcinski, T. (2020). End-to-end sinkhorn autoencoder with noise generator. *IEEE Access*, 9, 7211-7219.

rozdział piąty - Deja, K., Kuzina, A., Trzcinski, T., & Tomczak, J. (2022). On analyzing generative and denoising capabilities of diffusion-based deep generative models. *Advances in Neural Information Processing Systems*, 35, 26218-26229.

rozdział szósty - Deja, K., Trzcinski, T., & Tomczak, J. M. (2023). Learning Data Representations with Joint Diffusion Models. zaakceptowany na European Conference on Machine Learning.

Kolejny rozdział, siódmy, wprowadza nową perspektywę na rozważane dotąd problemy – uczenie reprezentacji w dynamicznych warunkach napływu danych (niestacjonarności danych). W przeciwieństwie do tradycyjnych modeli generatywnych, które są trenowane tylko raz na dostępnych danych, istnieje bowiem potrzeba ich dostosowywania się do nowych danych bez utraty wcześniej nabytej wiedzy. Poruszone jest zjawisko znane jako "katastrofalne zapomnienie", gdzie model traci zdolność do wydajnego działania na starych danych, gdy jest przeuczony na nowych. W rozdziale omawia się taksonomię składającą się z (1) metod uczenia ciągłego (continual learning) wyodrębniając podejścia

---

bazujące na regularyzacji, sprowadzających się do penalizowania nauki i spowalniania procesu nadpisywania wcześniej nauczonej wiedzy, (2) metody bazujących na dynamicznych architekturach tworzących różne wersje modelu (głównie przyrostowe) dla różnych zadań, aby poprawić jego wydajność w scenariuszu wielozadaniowym oraz (3) metody bazujące na powtarzaniu zakładające, że zapominanie można pokonać jedynie przez stałe powtarzanie wcześniejszych przykładów danych. Rozdział prezentuje również w skrócie obecny stan wiedzy na temat metod uczenia ciągłego w modelowaniu generatywnym koncentrując się na zachowaniu zdolności do generowania z tej części rozkładu, która została wyestymowana ze starszych danych w strumieniu i wymieniając wszystkie główne prace stosujące różnorodne techniki, od regularyzacji wag po dynamiczne architektury i buforowanie danych.

Kolejne dwa rozdziały zawierają teksty opublikowanych prac, które stanowią drugą część zbioru publikacji, t.j.,

rozdział ósmy - Deja, K., Wawrzyński, P., Marczak, D., Masarczyk, W., & Trzciniński, T. (2021). Binplay: A binary latent autoencoder for generative replay continual learning. In 2021 International Joint Conference on Neural Networks (IJCNN) (pp. 1-8). IEEE.

rozdział dziewiąty - Deja, K., Wawrzyński, P., Masarczyk, W., Marczak, D., & Trzciniński, T. (2022). Multiband VAE: Latent Space Alignment for Knowledge Consolidation in Continual Learning. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI-22), (pp. 2902-2908).

Rozdział dziesiąty zawiera zwięzłe podsumowanie oraz dyskusję na temat przyszłości badań nad modelami generatywnymi i wpływu, jaki mogą one mieć na inne badania w dziedzinie uczenia maszynowego. Doktorant formułuje cztery bardzo interesujące pytania badawcze skupiające się na takich aspektach jak: czy model generatywny może pomóc w uczeniu reprezentacji w sposób użyteczny dla innych zadań uczenia ciągłego, czy reprezentacje danych wygenerowane przez generatywne modele dyfuzyjne mogą być korzystne w scenariuszach uczenia ciągłego, jakie zastosowanie mogą mieć reprezentacje uzyskane z generatywnych modeli dyfuzyjnych oraz w jaki sposób zaadoptować procesy dyfuzyjne by było możliwe uczenie reprezentacji na zasadzie autokodera. Rozdział zawiera interesujące propozycje dalszych badań umożliwiających poszukiwanie odpowiedzi na te pytania.

## 1.2 Ocena wyników oraz stopnia ich oryginalności

Oryginalne osiągnięcia badawcze przedstawione w doktoracie dotyczą zarówno opracowania nowych metod generatywnych umożliwiających otrzymywanie reprezentacji danych, jak i ich zastosowaniu w scenariuszu uczenia ciągłego.

Rozwiązania zaproponowane w pracy opierają się na innowacyjnych koncepcjach lub istotnych modyfikacjach w odniesieniu do metod znanych z literatury naukowej. Warto podkreślić dążenie do tworzenia rozwiązań, które charakteryzują się zarówno wysoką jakością, jak i efektywnością. W pracy można znaleźć wiele propozycji nowych rozwiązań, usprawnień, dodatków technicznych i implementacyjnych, które są wynikiem zgłębionej analizy i konsekwentnej derywacji naukowej. Wśród uzyskanych rezultatów należy wymienić:

- wprowadzenie nowej architektury autokodera generatywnego nazwanego End-to-End Sinkhorn Autoencoder, w której proces kodowania próbek danych do ukrytych reprezentacji został oddzielony od procesu próbkowania nowych instancji, dodatkowa sieć neuronowa mapuje ustalony rozkład Gaussa na swobodnie kodowaną ukrytą reprezentację autokodera, a jako element regularyzacyjny użyta została aproksymacja Sinkhorna odległości Wassersteina, która wyrównuje zakodowane przykłady danych ze zmapowanym szumem,

- zaproponowanie hybrydowego, dwuczęściowego generatywnego modelu dyfuzyjnego, składającego się z odszumiacza i generatora, gdzie generuje się nowe cechy danych podobne do obserwowanych w danych treningowych (generator) i usuwa pozostały szum w sposób niezależny od danych (denoiser - deterministyczny autokoder),
- propozycję łącznego modelu dyfuzyjnego, w którym reprezentacje użyte są jako dane wejściowe dla klasyfikatora, a cały model jest optymalizowany łącznie na addytywnej funkcji kosztu (równanie 6.8) przy jednej parametryzacji do modelowania zarówno prawdopodobieństwa przykładu  $p(x)$ , jak i prawdopodobieństwa brzegowego  $p(y|x)$ ,
- wprowadzenie binarnego autokodera jako metody efektywnego uczenia reprezentacji (BinPlay), inspirowane mechanizmem kompresji danych, w ramach uczenia ciągłego, gdzie modele generatywne służą jako źródło przykładów powtórzeniowych dla ciągle trenowanego klasyfikatora,
- wprowadzenie metody do ciągłego dopasowania wiedzy reprezentowanej w przestrzeni ukrytej autokodera wariacyjnego, gdzie aktualizacja ciągle trenowanego VAE jest podzielona na dwie części: (1) użycie nowego modelu do nauki osadzania nowych danych i (2) dopasowania (ang. alignment) wcześniejszych i nowych danych za pomocą dodatkowego projektora MLP działającego w ukrytej przestrzeni VAE.

Rezultaty te zostały opublikowane na bardzo dobrych konferencjach o renomie międzynarodowej (NeurIPS, IJCNN, IJCAI, ECML<sup>2</sup>) oraz w przyzwoitym czasopiśmie IEEE Access. Należy zaznaczyć, że publikacje wchodzące w skład rozprawy zostały napisane przez wielu autorów (kolejne publikacje miały 6, 4, 3, 5, 5 autorów). W tekście doktoratu została zawarta informacja na temat roli i udziału doktoranta w procesie powstania tych rezultatów, ale w żadnym z oryginalnych manuskryptów publikacji nie odnalazłem tych informacji. Doktorant, będący pierwszym autorem wszystkich prac, miał znaczący udział w planowaniu, konceptualizacji, implementacji, testowaniu i ocenie metod. Praca ta była realizowana na wartościowych stażach i we współpracy z bardzo dobrymi naukowcami. Całość potwierdza, że rola doktoranta jest wyraźna i stanowi główną siłę napędową powstania prac.

### 1.3 Zagadnienia dyskusyjne

Zawarte poniżej uwagi nie mają na celu krytykowania prezentowanej rozprawy, a stanowią zaproszenie do dyskusji i dalszej refleksji nad przedstawionymi w pracy zagadnieniami.

- *Pytania badawcze.* W rozdziale pierwszym pracy zostały zaprezentowane trzy pytania badawcze, na które odpowiedzi poszukuje przedstawiony doktorat. Pojawiają się w nich określenia nieprecyzyjnie stawiające oczekiwania wobec metod uczenia reprezentacji. Proszę by Doktorant przeanalizował i odniósł się do tego, jakie kryteria powinna spełniać "użyteczna reprezentacja". Przykład dyskusji na ten temat jest zawarty w *Bronstein, M. M., Bruna, J., Cohen, T., & Veličković, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. arXiv preprint arXiv:2104.13478.*
- *Dopasowanie (ang. alignment) za pomocą MLP.* W publikacji o Multiband VAE zaproponowana jest sieć tłumacząca, która uczy się wspólnego dopasowania oddzielnych przestrzeni ukrytych w zależności od identyfikatora zadania i od pojedynczej zmiennej ukrytej  $Z$ , w której wszystkie przykłady są reprezentowane niezależnie od swojego źródłowego zadania. Czy w związku z tym mechanizm translatora dokonuje projekcji jedynie przekształcającej przestrzeń czy raczej wyucza się dopasowania (przekształcenia) przestrzeni osadzeń jednego zadania względem drugiego? Jaka jest motywacja uczenia przekształcenia złożenia warstw liniowych z leaky relu?

<sup>2</sup>w druku

---

## 2 Analiza strony formalnej rozprawy

### 2.1 Ocena układu pracy i redakcji manuskryptu

Rozprawa doktorska napisana jest w języku angielskim i pod względem edytorskim jest opracowana starannie. Zawiera podziękowania, streszczenie w językach angielskim i polskim, spis treści, dwa rozdziały wprowadzające do pierwszej części zagadnień, jeden rozdział z przeglądem literatury, kolejne trzy rozdziały będące przedrukami artykułów, rozdział wprowadzający do tematyki drugiej części zagadnień, kolejne dwa rozdziały będące przedrukami artykułów oraz rozdział podsumowujący i prezentujący możliwe dalsze badania. Bibliografia zawiera 255 wymienionych i uporządkowanych alfabetycznie, wolnych od błędów pozycji bibliograficznych. Większość prac w bibliografii została opublikowana w ciągu ostatnich kilku lat, co świadczy o aktualności tematyki rozprawy na arenie międzynarodowej.

Układ rozprawy jest właściwy, aczkolwiek dostrzeżone niewielkie niedociągnięcia w rozprawie odnoszą się do organizacji rozdziałów pracy, w której rozdziały merytoryczne są przemieszane z wprowadzającymi tło (rozdział siódmy), a rozdział drugi i trzeci stanowią niepotrzebnie rozdzieloną zawartość, o czym mogą nawet świadczyć w zasadzie identyczne nazwy ich podrozdziałów. Są to jednakże nieistotne szczegóły, w znikomym stopniu wpływające na nawigację w pracy.

Użyta w pracy terminologia jest właściwa dla analizowanego obszaru badań, obejmującego uczenie maszynowe, modele generatywne, modele głębokie. Praca posiada 181 numerowanych stron.

### 2.2 Uwagi szczegółowe

Większość treści rozprawy jest poprawna pod względem językowym, stylistycznym i merytorycznym. W trakcie czytania zauważyłem jednak kilka drobnych błędów redakcyjnych i fragmentów, które w toku dodatkowej weryfikacji możnaby poprawić. Wiele razy pominięte zostały przecinki, w innych ich brakuje. Inne pomniejsze błędy:

- p.14 In this work, we focus on those inner-workings of different generative → In this work, we focus on the inner-workings of different generative
- p.14 continual-learning → continual learning
- p.16 noise → a noise
- p.22 forgetting of past data → forgetting past data

## 3 Konkluzja

Przedstawiona praca doktorska, autorstwa mgr. inż. Kamila Deji, nie tylko spełnia, ale wręcz przewyższa kryteria określone przez Ustawę z dnia 20 lipca 2018 r. - Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668). Praca doktorska ukazuje głęboką znajomość dziedziny generatywnych technik uczenia maszynowego i pozwala wywnioskować o wysokich umiejętnościach samodzielnego prowadzenia pracy naukowej przez doktoranta. Praca koncentruje się na innowacyjnych odpowiedziach na kilka naukowych zagadnień, które łączy wspólny cel: identyfikacja efektywnych sposobów reprezentacji danych dla modeli generatywnych. Jest to praca mająca realny wpływ na dziedzinę, co można zaobserwować w licznych publikacjach autora w uznanych czasopismach oraz na międzynarodowych konferencjach z dziedziny informatyki. Przedstawione w pracy wyniki bardzo dobrze potwierdzają słuszność zgłoszonych przez autora koncepcji oraz efektywność ich zastosowania w praktyce. Wszelkie uwagi krytyczne, które pojawiły się w recenzji, nie wpływają na moją jednoznaczną i bardzo pozytywną ocenę tej dysertacji zarówno pod kątem oryginalności, jak i wartości merytorycznej. Określenie

---

zadań, proponowane innowacyjne metody ich rozwiązania, realizacja badań i kompetencje w formułowaniu wniosków dowodzą, że autor jest świetnie przygotowany do dalszej działalności naukowej. Bazując na tych przesłankach wnioskuję by mgr inż. Kamil Deja był dopuszczony do kolejnych etapów postępowania. Dodatkowo, ze względu na wysoką jakość publikacji i prestiżowych miejsc ich zamieszczenia, proponuję wyróżnienie jego pracy doktorskiej.

*Kojdalcem*