Gdynia, 15 września 2024 r.

Prof. dr hab. inż. Ireneusz Czarnowski
Wydział Informatyki
Uniwersytet Morski w Gdyni
ul. Morska 83, 81-225 Gdynia

# RECENZJA

**rozprawy doktorskiej mgr. Macieja ŻELASZCZYKA**

pt.: *„Deep Representation Learning in Varied Settings”.*

Recenzję przygotowałem zgodnie z uchwałą Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej z dnia 21 maja 2024 roku w sprawie wyznaczenia recenzentów rozprawy doktorskiej w postępowaniu w sprawie nadania stopnia doktora mgr. Maciejowi Żelaszczykowi, w której wskazano moją osobę jako recenzenta rozprawy.

## 1. Problematyka naukowa rozprawy

Rozprawa doktorska zatytułowana „*Deep Representation Learning in Varied Settings*” została przygotowana w dyscyplinie informatyka techniczna i telekomunikacja. Tematyka rozprawy mieści się w obszarze teorii sztucznej inteligencji, dotyczy badań związanych z uczeniem maszynowym, a problem badawczy nad którym skupił się Doktorant dotyczy uczenia się reprezentacji.

Problem uczenia się reprezentacji Doktorant rozważa w oparciu o analizę głębokiego uczenia oraz wybrane zadania/problemy klasyfikacji związane z przetwarzaniem danych o zróżnicowanej i złożonej strukturze, w tym w szczególności obrazu oraz danych audio. W oparciu o te zadania Doktorant podjął się oceny możliwości uchwycenia istotnych informacji w danych wejściowych, w tym znalezienia właściwej reprezentacji danych w kontekście procesu uczenia się, które z założenia powinno być autonomiczne (mając na uwadze autonomiczność inteligentnego systemu, ale również autonomiczność w sensie niezależności od struktury i postaci danych wejściowych). Ocenę tą Doktorant prowadzi również w kontekście uczenie adwersarialnego. Doktorant główną hipotezę pracy doktorskiej sformułował w kontekście uczenia się reprezentacji jako głównego czynnika pozwalającego na uczenie się systemu niezależnie od rozwiązywanego problemu oraz obszaru jego implementacji.

Doktorant swoją dysertację oparł na wynikach badań, które zostały wcześniej zawarte w następujących artykułach naukowych (opublikowanych w ramach materiałów pokonferencyjnych):

(1) Maciej Żelaszczyk and Jacek Mańdziuk, Adversarial Defenses via a Mixture of Generators. Proceedings of the International Conference on Neural Information Processing, ICONIP 2021, CCIS 1516, 566-574 (70 pkt.)
(2) Maciej Żelaszczyk and Jacek Mańdziuk, Audio-to-Image Cross-Modal Generation. 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1-8, doi: 10.1109/IJCNN55064.2022.9892863 (70 pkt.)

Doktorant wskazał także dwa inne artykuły, które są na etapie przygotowania lub recenzji. O wszystkich wspomnianych artykułach jest mowa w podrozdziale 1.2 (Research Hypothesis), a następnie w podrozdziale 1.3 (Outline), gdzie Doktorant krótko je charakteryzuje, gdy pełne wyniki uzyskanych badań (w tym proponowane rozwiązana algorytmiczne) objęte tymi artykułami, prezentuje dalej w kolejnych rozdziałach pracy (Rozdziały 3-6).

Recenzowana rozprawa wpisuje się w aktualny nurt badawczy. Tematyka rozprawy jest aktualna, a sam problem badawczy został sformułowany prawidłowo.

## 2. Treść rozprawy

Rozprawa doktorska została przygotowane w języku angielskim. Opracowanie łącznie składa się z 144 stron, w tym:

- wstępu, w którym Doktorant przedstawił tło problemu, sformułował hipotezę badawczą oraz odniósł się do poszczególnych rozdziałów rozprawy,
- rozdziału tzw. wprowadzającego,
- czterech kolejnych rozdziałów, w których Doktorant przedstawił proponowane rozwiązania,
- zakończenia, obejmującego zwarte podsumowanie uzyskanych wyników badań,
- wykazu bibliografii, w którym wskazano 110 pozycji literaturowych.

Pracę opatrzono również w spis algorytmów i rysunków.

Rozdział 2 rozprawy, zatytułowany „*Preliminaries*", jest rozdziałem wprowadzającym. Doktorant w rozdziale tym wprowadza definicje, przedstawia ideę działania modelu autoencodera oraz architekturę GAN. Drugi wątek podjęty przez Doktoranta w rozdziale 2 dotyczy aspektu ataków adwersarialnych.
W tym kontekście Doktorant wprowadza do problemu, a następnie odwołuje się do metod generycznych związanych z generowaniem danych adwersarialnych oraz prowadzi dyskusję związaną z typami ataków adwersarialnych. Ostatni aspekt, który Doktorant podejmuje w tym rozdziale, dotyczy procesu przekształcania danych i mapowania ich na wektory o niższej liczbie wymiarów, co w literaturze anglojęzycznej nosi nazwę embeddingu (osadzania).

Rozdział 3 rozprawy zatytułowany „*Adversarial Defenses via a Mixture of Generators*" prezentuje wyniki badań opierających się na założeniu, iż istnieje możliwość odzyskania reprezentacji danych uszkodzonych w wyniku ataku adwersarialnego, w kontekście docelowej klasyfikacji danych, poprzez zastosowanie systemu przetwarzania danych opartego na komitecie konkurujących ekspertów (w rozumieniu agentów programowych). W tym podejściu każdy z ekspertów ma zadanie usunięcie efekt błędu wprowadzonego atakiem adwersarialnym. Innymi słowy każdy z ekspertów jest „odwrotną aproksymacją" ataku wprowadzającego uszkodzenie w reprezentacji danych. Tym samym, podejście zaproponowane przez Doktoranta opiera się na założeniu o możliwości trenowania systemu ekspertów w taki sposób, aby docelowo możliwe było odzyskanie właściwej reprezentacji danych istotnej dla dalszego procesu klasyfikacji, uszkodzonej w wyniku zaistniałych ataków destrukcyjnych. Doktorant analizę tego przypadku badawczego oparł na przetwarzaniu obrazów oraz przyjął również założenie, że ogólna struktura ataku jest znana i nie ulega ona zmianie. W konsekwencji udało się zbadać Doktorantowi możliwość zastosowania modelu GAN do przetwarzania obrazów uszkodzonych celem przywrócenia ich oryginalnych reprezentacji, a tym samym oryginalnych etykiet klas przypisanych do tych obrazów. Doktorant badał również możliwość odzyskiwania reprezentacji danych uszkodzonych kilkoma generowanymi atakami, czyli rekonstrukcję uszkodzonych danych obrazowych w tym

przypadku w oparciu o zmodyfikowany model GAN. Doktorant przedstawił działanie proponowanego systemu oraz analizę możliwości odzyskiwania reprezentacji danych w oparciu o serię eksperymentów, dla których założenia oraz uzyskane wyniki omówił szeroko w podrozdziale 3.5.

Rozdział 4 rozprawy został zatytułowany „*Audio-to-image cross-modal generation*" i obejmuje wyniki badań Doktoranta nad problemem mapowania reprezentacji danych audio na reprezentację obrazową. Mapowanie to zostało przez Doktoranta zaproponowane poprzez użycie współdzielonej reprezentacji opartej na tak zwanym generowaniu międzymodalnym reprezentacji danych. Tym samym Doktorant swoje badania oparł na łączeniu reprezentacji danych audio i reprezentacji danych obrazu w docelowe dane treningowe oraz rekonstrukcji danych obrazowych z wykorzystaniem dedykowanego systemu. System ten poprzez współdzielone reprezentacje umożliwia generowanie obrazów nie występujących w danych wejściowych, ale odpowiednich dla klasy przypisanej do próbki danych audio podanych na wejście systemu. W szczególności Doktorant poddał analizie zastosowanie dwóch modeli generatywnych, tj. modelu autoencodera VAEs oraz modelu GAN, oraz ich kombinacji. Modele te zostały wykorzystane do ekstrakcji cech z danych audio, a następnie do generowania danych obrazowych. Zaproponowane przez Doktorant systemowe podejście zostało poddane walidacji na drodze eksperymentów obliczeniowych z wykorzystaniem przygotowanych danych oraz zaproponowanych procedur i scenariuszy oceny jakości mapowania, ocenianej poprzez pryzmat klasyfikacji danych obrazowych. Istotnym dla tego aspektu badawczego było również przeanalizowanie dopasowania generowanych próbek danych audio do danych obrazowych, w tym w oparciu o tzw. mapowania „wiele do jednego" i „jeden do wielu".

Rozdział piąty rozprawy dotyczy aspektu interpretowalności w oparciu o współdzielone reprezentacje danych. Doktorant w tym przypadku za cel przyjął wykazanie, iż wprowadzenie zestawu współdzielonych reprezentacji pozwala na tworzenie interpretowalnych reprezentacji zmiennych tabelarycznych, co może być przydatne przy rozwiązywaniu problemu uczenia wielozadaniowego (ang. *multi-task learning*). Doktorant w tym rozdziale stara się wykazać, iż dzięki współdzielonym reprezentacjom zmiennych możliwe jest ograniczenie ich formy lub formy ich kombinacji. W konsekwencji takie ograniczenie skutkuje różnymi poziomami interpretowalności, a także poprawą dokładności klasyfikacji. Niemniej jednak Doktorant podkreśla, że ostatecznie istotny jest kompromis pomiędzy dokładnością klasyfikacji a poziomem interpretowalności. Ten aspekt badawczy Doktorant rozważa również w kontekście przekształcania danych opartego na embeddingu. Doktorant przedstawił koncepcję tego przetwarzania danych, wprowadzając przy tym tak zwane współdzielone osadzanie zmiennych. Architektura przetwarzania zmiennych osadzonych została zaimplementowana do rozwiązania zadań klasyfikacji oraz poddana ocenie na drodze eksperymentów obliczeniowych. Ocena ta został przeprowadzona również pod kątem interpretowalności zmiennych.

Kolejny rozdział - zatytułowany „*Restricting representations*", związany jest z aspektem ograniczania reprezentacji danych dla docelowego modelu predykcyjnego poprzez uczenia się relacji między tak zwanymi komponentami zawartymi w reprezentacji danych. Założeniem proponowanym przez Doktoranta rozwiązania jest włączenie tych zidentyfikowanych relacji do procedury uczenia adwersarialnego, w konsekwencji czego, zaproponował on stosowną procedurę. Stopień ograniczenia reprezentacji danych został przebadany poprzez pryzmat oceny korelacji między komponentami reprezentacji oraz oceny ich redundancji. Stosowne wyniki eksperymentów obliczeniowych zostały przedstawione w rozprawie wraz z ich omówieniem.

Ostatni rozdział stanowi podsumowanie oraz formułuje najważniejsze wnioski z przeprowadzonych badań. Doktorant w rozdziale tym podkreśla również rolę uczenia się reprezentacji, szczególnie w kontekście systemów autonomicznych.

Literatura została dobrana właściwie i jest on aktualna w kontekście omawianych i przywoływanych problemów.

## 3. Najważniejsze wyniki uzyskane w pracy

Do oryginalnych osiągnięć oraz wartościowych wyników badań Doktoranta zaliczam:

- zbadanie możliwości wykorzystania modelu generatywnego – GAN, do przetwarzania i rekonstrukcji obrazów uszkodzonych w wyniku ataku adwersarialnego,
- zaproponowanie rozwiązania do generowania obrazów z dźwięku poprzez wyodrębnianie cech audio i przetwarzanie ich w celu uzyskania podzbioru cech audiowizualnych — wspólnych cech istotnych dla przetwarzania powiązanych danych wizualnych i słuchowych,
- zbadanie stopnia dopasowania zestawu danych do generowanych próbek danych audio, z mapowaniem wiele do jednego i jeden do jednego,
- zaproponowanie mechanizmu współdzielonego osadzania zmiennych dla potrzeb rozwiązywania problemu uczenia wielozadaniowego,
- zaproponowanie podejścia do ograniczania reprezentacji danych dla potrzeb uczenia adwersarialnego.

## 4. Pytania i uwagi do recenzowanej pracy

Recenzowana rozprawa doktorska podejmuje bardzo aktualny problem związany z uczeniem maszynowym. W szczególności prezentuje wybrane aspekty związane z uczeniem się reprezentacji. Zaproponowane rozwiązania algorytmiczne nie budzą wątpliwości. Zostały one zaprezentowane w sposób czytelny oraz podane ocenie na drodze eksperymentów obliczeniowych, których zakres należy uznać za wystarczający. Sam sposób prezentacji wyników badań wraz z ich analizą oceniam również wysoko. Nie budzi również uwag ich wartość merytoryczna. Niemniej jednak, w kontekście każdego zaproponowanego rozwiązania algorytmicznego nasuwają się pytanie dotyczące jego złożoności obliczeniowej. Jaka zatem jest złożoność obliczeniowa proponowanych algorytmów? Na ile zaproponowane rozwiązania mogą mieć wpływ na i tak kosztowne procesy głębokiego uczenia? Czy uczenie się reprezentacji może przyczynić się do zmniejszenia kosztów obliczeń w kontekście modeli generatywnych? Czy i w jaki sposób proponowane rozwiązania można ukierunkować implementacyjnie tak, aby koszty obliczeń nie wzrosły istotnie? Ponadto, czy proponowane rozwiązania uczenia się reprezentacji są niezależne od modeli generatywnych, które możemy użyć? Czy jednak istnieją pewne ograniczenia lub warunki które trzeba spełnić? Odpowiedzi na te pytania Doktorant mógłby udzielić podczas publicznej obrony.

Poza tym, dokonując oceny i zrozumienia zaprezentowanych algorytmów nasuwa się uwaga związana z ich zapisem. Dobrze byłoby, aby algorytmy te miały jasno wskazane zmienne wejściowe. Niekiedy brakuje również wskazania (zdefiniowania) zastosowanych oznaczeń, jak np. $G_m$ i $D$ w Algorytmie 3.1. Podobnie, nie definiowano czym jest *audio-image test set samples* oraz *audio-image test set* w Algorytmach 4.3 lub 4.4, lub jak zdefiniowano „numer of batches …" w Algorytmie 4.4. Są to oczywiście wybrane przykłady związane z niezbyt precyzyjnym opisem algorytmów, a uwaga ta ma charakter generalny w stosunku do wszystkich prezentowanych w pracy algorytmów.

## 5. Ocena redakcji i przygotowania rozprawy

Organizacja pracy jest przejrzysta. Praca został przygotowana w sposób staranny. Prace wyposażono w stosowny spis algorytmów oraz rysunków. Dla przejrzystości, można było również wprowadzić spis najważniejszych użytych w rozprawie oznaczeń oraz akronimów, czego jednak nie uczyniono.

W pracy występują błędy gramatyczne, w tym błędy użycia przedimków, niemniej jednak nie umniejszają one ogólnej ocenie przygotowania rozprawy, którą oceniam pozytywnie.

Pomimo moich powyższych uwag, redakcję rozprawy oceniam wysoko.

## 6. Konkluzja

Doktorant w rozprawie podejmuje bardzo aktualny problem badawczy związany z uczeniem maszynowym, a uzyskane wyniki badań oceniam jako bardzo wartościowe.

Rozprawa prezentuje i potwierdza ogólną wiedzę teoretyczną odpowiednio dla osoby ubiegającej się o nadanie stopnia doktora. Pomimo mich sformułowanych w recenzji pytań i uwag, uważam, że wyniki badań zostały zaprezentowane w sposób wystarczający dla oceny ich oryginalności i ważności dla dyscypliny informatyka techniczna i telekomunikacja. Stwierdzam, że Doktorant wykazał się umiejętnością samodzielnego rozwiązywania problemów badawczych, w tym doborem odpowiednich metod, potwierdził też, że posiada umiejętności związane z metodyką i metodologią prowadzenia badań naukowych.

Podsumowując, uważam, że rozprawa doktorska mgr. Macieja Żelaszczyka pt. ”*Deep Representation Learning in Varied Settings*” spełnia wymogi stawiane przy ubieganiu się o nadanie stopnia doktora w dyscyplinie informatyka techniczna i telekomunikacja.

Wnioskuję o dopuszczenie rozprawy doktorskiej mgr. Macieja Żelaszczyka do publicznej obrony, wnioskuję również o wyróżnienie rozprawy.

Ireneusz Czarnowski

Dokument podpisany przez Ireneusz Czarnowski
Data: 2024.09.19 14:18:03 CEST