

dr hab. Piotr Miłoś, prof IMPAN
Instytut Matematyczny Polskiej Akademii Nauk

RECENZJA ROZPRAWY DOKTORSKIEJ mgr Szymona Knopa

Tytuł rozprawy: Deep Generative Models in Image Processing

Autor rozprawy: Szymon Knop

Promotor rozprawy: dr hab. Igor T. Podolak

Promotor pomocniczy rozprawy: dr hab. Przemysław Spurek

1 Analiza strony merytorycznej rozprawy

1.1 Obszar problemowy

Recenzowana rozprawa dotyczy zagadnień budowy modeli generatywnych w oparciu o głębokie uczenie maszynowe oraz ich zastosowań w przetwarzaniu obrazów. Głównym tematem rozprawy jest analiza metod opartych o przestrzeń ukrytą w generatywnych modelach. Autor skupia się na badaniu różnych aspektów związanych z przestrzenią ukrytą i jej wpływem na jakość generowanych próbek oraz możliwości modeli generatywnych. Tematyka modeli generatywnych, w tym operujących na modalności wizyjnej jest bardzo aktualna. Jedyne zastrzeżenie do tematyki rozprawy to ewaluacja na raczej prostych zbiorach danych (MNIST, CelebA).

Autor stawia ogólnie postawione pytanie naukowe: "Jakie są sposoby strukturyzowania przestrzeni ukrytej i dopasowywania jej rozkładu w celu poprawy wydajności generatywnych modeli?" W rozprawie analizowane są różne metody i techniki mające na celu strukturyzowanie przestrzeni ukrytej w generatywnych modelach oraz kształtowania jej rozkładu. Autor bada, jak te czynniki wpływają na jakość modeli do generowania nowych wiarygodnych obrazów. Uważam, że pytanie naukowe jest dobrze postawione i wyjście od tak ogólnie postawionego problemu jest dobrym punktem wyjścia do badania konkretnych zagadnień w pracach P_1 - P_6 . Tematyka takowa mieści w sposób oczywisty w dziedzinie informatyki.

Pierwszy rozdział przedstawia opis tła i formalizm jakim posługuje się obszar modeli generatywnych badanych w rozprawie, tj. autoencoderów i GANów. Rozdział jest poprawny, choć moim zdaniem zbyt zwięzły. Autor mógłby przedstawić (nieco) szerszy zarys dziedziny i spróbować umiejscowić w nim swoje badania. Jest to szczególnie istotne ze względu na szybki postęp, którego doświadcza dziedzina modeli generatywnych. Autor też nie odnosi się do tego w konkluzjach (ostatnia sekcja rozprawy). Jednym z wielu narzucających się pytań jest to, czy uzyskane wyniki mogą być wykorzystane w modelach multi-modalnych, np. CLIP, gdzie przestrzeń ukryta jest wspólna dla obrazów i tekstów.

Kolejny rozdział zarysowuje strukturę po czym następuje omówienie prac P_1 - P_6 . W rozdziale tym Autor przedstawia również swoje wkłady w rozwój nauki.

Praca P_1 przedstawia i bada autoencodery Cramera-Wolda. Techniczne głównym pomysłem jest użycie odległości Cramera-Wolda jako człon regularyzacyjny dla przestrzeni ukrytej.

Uzyskano w ten sposób dwie nowe architektury: Cramer-Wold Autoencoder i Carmer-Wold Generator Istotną zaproponowanych rozwiązań jest istnienie 'zamkniętych' wzorów, co poprawia możliwości optymalizacji. Architektury zostały przetestowane empirycznie uzyskując rozsądne wyniki.

Praca P_2 jest kontynuacją i rozszerzeniem P_1 . Głównym rezultatem jest konstrukcja kilku nowatorskich autoenkoderów. W artykule zaproponowano i oceniono kilka modeli: SCvMAE, SKSAE, SCFWAE i SCWAE. Wyniki artykułu potwierdzają istotność wprowadzonych wcześniej koncepcji w pracy P_1 . Uzyskane wyniki pokazują pewne poprawy wyników, acz wydają się one zbyt skromne, żeby uzasadnić ich użycie w praktyce

W artykule P_3 zaproponowano analizę wpływu rozmiaru 'batcha' w autoencoderach generatywnych. Autor wykazał, że zbyt małe jego małe rozmiary degeneracji przestrzeni ukrytej, ale lepiej sprawdzają się w zadaniu rekonstrukcji. Efektem pracy jest konceptualizacja tego zjawiska ('the reconstruction-distribution trade-off') oraz zaproponowanie rozwiązania nowej procedury treningowej, która pozwala go uniknąć. Procedura ta, w uproszczeniu, proponuje, aby używać innych rozmiarów batcha dla członu rekonstrukcyjnego i regularyzacyjnego. Autorzy przetestowali swoje rozwiązanie na kilku zbiorach danych i uzyskali poprawę jakości generowanych próbek. Istotną zaletą praktyczną jest to, że rozwiązanie może być zastosowane w treningach używających małej ilości VRAMu.

Publikacja P_4 kontynuuje powyższy cykl badań. Autorzy proponują dwa usprawnienia do architektury autoencondera, mianowicie jest to użycie metryki Carmera-Wolda w części rekonstrukcyjnej funkcji straty i dodatkowego generatora dla przestrzeni ukrytej ('latent trick'). Ten drugi element jest, w moim przekonaniu, szczególnie ciekawy, gdyż pozwala na ominięcie części problemów związanych z faktem iż struktura przestrzeni ukrytej jest zwykle nieznana (choć w założeniu powinna być bliska przejętemu rozkładowi P_Z). Zaznaczyć należy, że 'latent trick' nie jest specyficzny dla przyjętej architektury i może być zastosowany do innych modeli autoencoderowych. W pracy wykazano empirycznie, że zaproponowane rozwiązania poprawiają jakość generowanych próbek, które dorównują jakością próbkom uzyskanym z GANów.

Kolejna zaprezentowana praca P_5 dotyczy GANów. Autor zauważa, iż w przeciwieństwie do autoencoderów, w GANach nie istnieje 'ustalona' struktura przestrzeni ukrytej. Autor proponuje, aby nadać strukturę, która odpowiada strukturze obrazka. Mianowicie każdy 'piksel' przestrzeni ukrytej zawiera lokalną informację, która modeluje lokalne cechy obrazka. Wymuszenie lokalności odbywa się poprzez narzucenie architektury konwolucyjnej (w pracy dodatkowo przyjmuje się istnienie globalnego komponentu przestrzeni ukrytej odpowiedzialnego za wysokopoziomowe, nielocalne cechy obrazka). Interesującą konsekwencją takiego podejścia jest to, że generacja i trening są możliwa w trybie lokalnym. Interesujące byłoby przetestowanie zaproponowanego rozwiązanie w do uzupełniania części obrazka (inpainting).

Ostatnia z zaprezentowanych prac P_6 przedstawia zastosowanie uzyskanych wcześniej wyników (w pracach P_1 - P_4) do problemu do paradygmatu uczenia ciągłego (continual learning). Mianowicie człon regularyzacyjny jest użyty do tego aby zapobiegać zapomnianiu. Rozwiązanie zostało przetestowane na standardowych zadaniach Split MNIST, Split CIFAR-10 and Permuted MNIST.

1.2 Ocena wyników oraz stopnia ich oryginalności

Rozprawa dotyczy problematyki architektur generatywnych w głębokim uczeniu maszynowym. Autor skupia się na analizie przestrzeni ukrytej i jej modelowaniu w celu poprawy jakości generowanych próbek. W przedstawionych pracach P_1 - P_6 uzyskano szereg oryginalnych wyników (omówionych poniżej). Pewien niepokój budzi raczej mała liczba cytowań (pomijając pracę P_1), co być może po części wynika z faktu, że 4 z prezentowanych prac ukazały się w ubiegłym roku. Podobnie, niepokojący jest fakt, że jedynie jedna praca została opublikowana w czasopiśmie z maksymalną liczbą punktów według listy MNiSW, a praca P_2 być może powinna być wyłączona z dorobku. Brakuje też publikacji na którejś z wiodących konferencji z dziedziny uczenia

maszynowego (np. NeurIPS).

Pozytywnie należy ocenić wykład Doktoranta w wyżej wymienione prace. Wszystkie one są w szerokiej współpracy wieloautorskiej (co jest normą w dziedzinie). W pracach P_2 - P_4 Doktorant miał dominującą funkcję wśród autorów. W pracy P_1 był jednym z głównych wykonawców, wreszcie w pracach P_5 - P_6 miał wpływ na konceptualizację i eksperymenty (i jak przypuszczam mentorowanie w przypadku P_6).

Uzyskano szereg oryginalnych wyników, które można podzielić na trzy (główne) kategorie:

Nowe architektury generatywne W pracach P_1 , P_2 , P_4 , P_5 zaproponowano nowe architektury do problemu generacji obrazków. Przetestowano je, używając najczęściej metryki FID i zbiorów MNIST, CelebA. Oceniam tę część dorobku jako istotną i ważną. Pewnym mankamentem może być to, że w żadnym przypadku nie przebito znacząco 'state-of-the-art', choć w dziedzinie o tak dynamicznym rozwoju nie można tego traktować jako krytyczny zarzut.

Nowe procedury treningowe W pracach P_3 , P_4 badano i przedstawiono nowe procedury treningowe. Ta część wydaje mi się szczególnie istotna ze względu na rozwój intelektualny Doktoranta. Zadanie to wymagało konceptualizacji istniejących problemów i oryginalnego podejścia do ich rozwiązania. Warto zauważyć, że w obu przypadkach uzyskano znaczące poprawy w stosunku do istniejących metod.

Badanie struktury przestrzeni ukrytej W pracach P_3 , P_4 , P_4 badano strukturę przestrzeni ukrytej. Podobnie jak w poprzednim punkcie, wydaje mi się, że jest to zdanie analityczne, które wymaga właściwej dojrzałości badacza. Jednocześnie jest ono ważne dla dziedziny, ponieważ pozwala lepiej zrozumieć działanie modeli generatywnych.

Praca P_6 odbiega nieco od głównego nurtu rozprawy, ponieważ dotyczy problemu uczenia ciągłego. Zauważył warto, że pokazuje jak zastosować uzyskane uprzednio wyniki do nowego problemu.

1.3 Zagadnienia dyskusyjne

Poniższa sekcja zawiera uwagi krytyczne dotyczące pracy.

Poziom dorobku Jak już wspomniano, pewne zastrzeżenia budzi poziom dorobku. Jeśli wyjąć pracę P_6 (w której, jak przypuszczam, Doktorant brał udział głównie jako mentor) i pracę P_2 (w czasopiśmie z niskim IF), to pozostają jedynie cztery prace. Niepokój budzi też brak pracy na którejś z wiodących konferencji. Po stronie pozytywnej należy tutaj wspomnieć, że prace dobrze łączą się w cykl tematyczny i dotyczą bardzo aktywnego obszaru badań.

Brak opisu dziedziny Rozprawa bardzo pobieżnie opisuje dziedzinę, opis obejmuje w zasadzie jedynie formalizm używany w rozprawie, natomiast pomija zupełnie opis problematyki i istniejących rozwiązań. Symptomatyczne jest to, że bibliografia zawiera jedynie 22 pozycje. Nie pozwala to ocenić, w jakim stopniu Doktorant jest zaznajomiony z dziedziną.

Brak pogłębionej analizy teoretycznej Prace P_1 i P_2 zawierają istotny komponent teoretyczny. Z przedstawionego opisu nie jest jasne jaka jest ich motywacja. Być może da się przedstawić argumenty, które nie są li tylko empiryczne.

Dalsza analiza 'reconstruction-distribution tradeoff' Przedstawiony 'reconstruction-distribution tradeoff' jest bardzo interesującym zjawiskiem. Moim zdaniem Autor powinien zaznaczyć, czy to jest jego obserwacja, czy też została ona już opisana w literaturze (być może praca [9]). W pierwszym przypadku wydaje się, iż warto byłoby przeprowadzić bardziej pogłębioną analizę tego

zjawiska. Być może zdanie 'This is attributed to a mismatch between intrinsic data and latent space dimensionalities' na str 11 jest krokiem w tę stronę, acz jest ono dla mnie trudne do zrozumienia.

Detale LocoGAN Niektóre detale metody opracowanej w P_5 są niejasne. Na przykład w jaki sposób przekazywana jest informacja globalna'. Czy to wymaga oddzielnej procedury treningowej, kopiowania jej między 'pikselami'. Nawiasem mówiąc wydawałoby mi się naturalniejsze powiedzenie, iż przestrzeń ukryta ma strukturę $2D$, podobnie jak obrazki mają strukturę $2D$ i kanały.

2 Analiza strony formalnej rozprawy

Rozprawa doktorska, która została poddana recenzji, została napisana w języku angielskim i składa się z 9 rozdziałów (włączając Bibliografię). Po tych rozdziałach wklejone są prace składające się na rozprawę. Układ ten jest zgodny z zaleceniami. Rozprawa jest przygotowana bardzo dobrze od strony edytorskiej. Także warstwa tekstowa nie budzi zastrzeżeń. Uznałbym nawet ją za mocną stronę rozprawy. Opisy w większości przypadków są zwięzłe i klarowne. Pozwalają one łatwo zorientować się w tematyce rozprawy i jej wynikach. Dodatkowe są z sobą powiązane logicznie. Użyta terminologia jest właściwa dla obszaru problemowego rozprawy w zakresie informatyki. Autor wyraźnie przedstawia zakres przeprowadzonych przez siebie prac, co ułatwia ocenę rozprawy. Wstęp, mimo, że dosyć krótki, zawiera jawnie sformułowane pytanie badawcze, które pozwala spiąć klamrą opisane w rozprawie wyniki.

Wyjątkami od tej pozytywnej oceny jest brak szerszego opisu dziedziny (patrz wyżej) i raczej niejasny opis metody w publikacji P_6 . Sugerowałbym poprawienie tego rozdziału.

3 Konkluzja

Pomimo wskazanych powyżej uwag i zastrzeżeń, uważam, że recenzowana dysertacja Pana mgr. Szymona Knopa **spełnia w sposób dostateczny** stawiane rozprawom doktorskim przez *Ustawę z 20 lipca 2018 roku (Dz. U. 2018 poz. 1668) Prawo o szkolnictwie wyższym i nauce*. Rozprawa zawiera oryginalne koncepcje z zakresu architektury i trenowania modeli generatywnych i tym samym adresuje istotny problem w dziedzinie informatyki. Uzyskane wyniki eksperymentalne w wystarczającym stopniu dokumentują poprawność proponowanej koncepcji oraz skuteczność działania jej implementacji. Na tej podstawie **wnioskuję o dopuszczenie mgr. Szymona Knopa do dalszych kroków procedury uzyskania stopnia doktora nauk ścisłych i przyrodniczych (w dyscyplinie informatyka)**..



Piotr Miłoś