



UNIWERSYTET
WARSZAWSKI

Wydział Matematyki, Informatyki i Mechaniki
Instytut Informatyki

Warszawa, 8.06.2023

Recenzja rozprawy doktorskiej magistra Szymona Knopa pod tytułem *Deep Generative Models in Image Processing*

Artykuły na których bazuje rozprawa

- (1) S. Knop, P. Spurek, J. Tabor, I. Podolak, M. Mazur, S. Jastrzębski, "Cramer-Wold Auto-Encoder", *Journal of Machine Learning Research*, vol. 21 (2020), 1-28
- (2) S. Knop, M. Mazur, J. Tabor, I. Podolak, P. Spurek, "Sliced Generative Models", *Schedae Informaticae*, vol. 27 (2018), 69-79,
- (3) S. Knop, P. Spurek, M. Mazur, J. Tabor, I. Podolak, "Batch size reconstruction-distribution trade-off in kernel based autoencoders", *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, 3728-3732,
- (4) S. Knop, M. Mazur, P. Spurek, J. Tabor, I. Podolak, "Generative models with kernel distance in data space", *Neurocomputing*, vol. 487 (2022), 119-129,
- (5) Ł. Struski, S. Knop, P. Spurek, W. Daniec, J. Tabor "LocoGAN — locally convolutional GAN", *Computer Vision and Image Understanding*, vol. 221 (2022), 103462,
- (6) M. Mazur, Ł. Pustelnik, S. Knop, P. Pagacz, P. Spurek, "Target Layer Regularization for Continual Learning Using Cramer-Wold Distance", *Information Sciences*, vol. 609 (2022), 1369-1380,

Tematyka rozprawy

Rozważając modele i algorytmy uczenia maszynowego możemy dokonać następującego podziału na dwie duże poddziedziny. Z jednej strony mamy zagadnienia modelowania, których celem jest przewidywanie wartości zmiennych ukrytych na podstawie znanych cech, co w przypadku problemu klasyfikacji sprowadza się do wyznaczenia granicy pomiędzy obszarami klas w wielowymiarowej przestrzeni cech.

Z drugiej strony wyobraźmy sobie sytuację w której mamy dany zbiór próbek i naszym celem jest wygenerowanie nowych próbek, ale w sposób zgodny z rozkładem z którego pochodzi oryginalny zbiór. Jest to dziedzina metod generatywnych i właśnie w tej dziedzinie osadzona jest omawiana rozprawa doktorska. Bardzo naturalnym przykładem problemu typu generatywnego jest generowanie zdjęć, przykładowo mając zbiór zdjęć ludzkich twarzy celem jest wygenerowanie nowych zdjęć, które będą przedstawiać nowe twarze, których nie ma w oryginalnym zbiorze danych.

Główną wspólną ośią przedstawianych w rozprawie artykułów jest wprowadzenie i wykorzystanie na różne sposoby metryki Cramera-Wolda (która bazuje na twierdzeniu Cramera-Wolda). Użycie tej metryki pozwala na zastosowanie zwartych wzorów w miejscach, gdzie poprzednio potrzebne było próbkowanie, co pozwala na osiągnięcie nowych wyników w dziedzinie modeli generatywnych, poparte eksperymentami dla kilku powszechnie stosowanych zbiorów danych zdjęć.

Pięć spośród sześciu przedstawionych w rozprawie artykułów wykorzystuje technikę uczenia autokoderów (ang. *autoencoders*), gdzie trenowane są dwa modele: koder i dekodek. Koder pozwala na przejście z przestrzeni próbek (na przykład obrazów) do przestrzeni ukrytej, natomiast dekodek pozwala na przejście z przestrzeni ukrytej z powrotem do przestrzeni próbek. Istotnym elementem tej metody jest sposób regularyzacji przestrzeni ukrytej, który pozwala na próbkowanie w tej przestrzeni w sposób pozwalający na generowanie nowych próbek za pomocą dekodeka (w etapie generowania koder nie jest już potrzebny).

Wyniki rozprawy

W pracy (1) zdefiniowana została metryka Cramera-Wolda, która polega na definiowaniu odległości w oparciu o wiele rzutowań na jednowymiarową przestrzeń. Metoda rzutowań była poprzednio użyta w autokoderach typu ang. *Sliced-Wasserstein* (SWAE). Praca (1) zawiera zarówno matematyczne wyprowadzenie proponowanej metryki jak również zestaw wyników eksperymentalnych pokazujących, że autokodery oparte o nową metrykę potrafią osiągać lepsze wyniki niż poprzednie wersje. Warto tutaj zaznaczyć, że w dziedzinie modeli generatywnych nie jest oczywiste w jaki sposób mierzyć jakość otrzymanych próbek. Jednym z kryteriów stosowanych w analizie wyników jest miara FID (ang. *Frechet Inception Distance*). Przeprowadzone eksperymenty w oparciu o zbiory MNIST, Fashion-MNIST, CELEB A oraz CIFAR-10 pokazują, że metoda pozwala na osiągnięcie lepszych wyników niż alternatywne autokodery, a w dodatku przedstawione modele CWAE wykazują się

mniejszą wrażliwością na hiperparametry w procesie treningu. Praca (2) stanowi bezpośrednie rozszerzenie przedstawionych tematów o badania innych typów autokoderów również polegających na metrykach opartych o metody rzutowania.

Prace (3) i (4) koncentrują się na dwóch aspektach związanych z trenowaniem autokoderów: rozmiarem próbki w jednym kroku gradientu (ang. *batch size*) a także rozmiarem przestrzeni ukrytej. Optymalizowana funkcja kosztu w autokoderach posiada dwa składniki: jeden związany z rekonstrukcją a drugi z regularyzacją przestrzeni ukrytej. Praca (3) eksperymentalnie wskazuje, że te dwa składniki mają inne optymalne wielkości próbek. Żeby pozwolić na uzyskanie najlepszych wyników dla obu składników jednocześnie zaproponowano pomysłową metodę polegającą na rozszerzeniu wielkości próbki na potrzeby składnika regularyzującego o wektory przestrzeni ukrytej z poprzednich iteracji gradientu.

Tematem pracy (4) jest wymiarowość przestrzeni ukrytej w autokoderach. Z jednej strony zwiększanie wymiarowości przestrzeni ukrytej pozwala na zmniejszenie błędu rekonstrukcji (z powodu poszerzenia wąskiego gardła), jednakże utrudnia to próbkowanie w przestrzeni ukrytej. Aby pogodzić te dwa aspekty wprowadzono dwuetapowy proces próbkowania w przestrzeni ukrytej - najpierw próbkowany jest wektor w przestrzeni Z' (zgodnie z wielowymiarowym rozkładem normalnym), a następnie wektor ten jest przetwarzany przez wyuczony koder, którego celem jest dopasowanie się do rozkładu panującego w przestrzeni Z .

W pracy (6) metryka Cramera-Wolda została wykorzystana w zadaniu uczenia kontynuacyjnego (ang. *continual learning*), gdzie celem jest wyuczenie modelu dla kolejnych zadań w sposób pozwalający na uniknięcie zjawiska dramatycznego spadku skuteczności dla wcześniej postawionych zadań. Jedną z metod stosowanych w tego typu przypadkach jest regularyzacja parametrów pozwalająca zachować niewielką odległość (w przestrzeni parametrów) od poprzednio znalezionej odpowiedzi. W tym miejscu autorzy wykorzystali metrykę Cramera-Wolda, pierwotnie wprowadzoną w pracy (1).

Jedyną pracą ze zbioru, która skupia się na generatywnych modelach przeciwstawnych (ang. *GANs*), zamiast autokoderów, jest praca (5). W pracy tej pokazano w jaki sposób wprowadzić dodatkową strukturę do tensora przestrzeni ukrytej, w sposób pozwalający na rozdzielanie cech lokalnych (takich jak wartości poszczególnych pikseli w otoczeniu) oraz globalnych (takich jak płeć, kolor włosów, itd). Przedstawiona metoda LocoGAN pozwala na generowaniu obrazów o innej rozdzielczości niż miało to miejsce w procesie treningu, co pozwala między innymi na generowanie spójnych (cyklicznych) tekstur.

Ocena rozprawy

Moim zdaniem siłę rozprawy stanowi aktualność prowadzonych badań oraz spójność tematyczna przedstawionych artykułów. Mgr Szymon Knop na przestrzeni kilku lat dokonał wielu obserwacji i eksperymentów, co pozwoliło na publikację przedstawionych wyników w sposób z pewnością stanowiący wartość dodaną w bardzo szybko rozwijającej się dziedzinie nauki, badanej na całym świecie. Warto zaznaczyć, że większość artykułów została opublikowana w rozpoznawalnych czasopismach o bardzo wysokiej renomie.

Starając się szukać słabych stron rozprawy mogę jedynie wskazać na fakt, że modele oparte o autokodery nie są obecnie powszechnie stosowane w dziedzinie generowania obrazów, która to dziedzina została zdominowana przez generatywne sieci przeciwstawne, a w ostatnim czasie przez modele dyfuzyjne. Pomimo wspomnianej dominacji modeli dyfuzyjnych oraz GAN metody oparte o autokodery mają inne zalety, co powoduje, że są one wykorzystywane na przykład w robotyce do tworzenia przestrzeni bardziej skomplikowanych zachowań i umiejętności robotów.

Podsumowanie

W mojej ocenie przedstawiona rozprawa doktorska i dorobek naukowy magistra Szymona Knopa spełnia wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim, a co za tym idzie rozprawa ta z pewnością może stanowić podstawę do nadania stopnia doktora.

Marek Cygan

