



UNIwersytet  
Warszawski

Wydział Matematyki, Informatyki i Techniki  
Instytut Informatyki

Warszawa, 22.02.2019

Recenzja rozprawy doktorskiej magistra Stanisława Jastrzębskiego pod tytułem *Deep Networks Generalization and Optimization Trajectory*

Artykuły na których bazuje rozprawa

Rozprawa doktorska oparta jest na sześciu następujących artykułach:

1. Jastrzębski, Kenton, Arpit, Ballas, Fischer, Bengio, Storkey, *Width of Minima Reached by Stochastic Gradient Descent is Influenced by Learning Rate to Batch Size Ratio*, *Artificial Neural Networks and Machine Learning - ICANN 2018. Lecture Notes in Computer Science* 392-402, 2018.
2. Arpit, Jastrzębski, Ballas, Krueger, Bengio, Kanwal, Maharaj, Fischer, Courville, Bengio, Lacoste-Julien: *A Closer Look at Memorization in Deep Networks*, *ICML 2017*: 233-242
3. Jastrzębski, Kenton, Ballas, Fisher, Bengio, and Storkey: *DNN's Sharpest Directions Along the SGD Trajectory*, *Modern Trends in Nonconvex Optimization for Machine Learning Workshop at International Conference on Machine Learning*, 2018.
4. Jastrzębski, Arpit, Ballas, Verma, Che, Bengio: *Residual Connections Encourage Iterative Inference*, *International Conference on Learning Algorithms (Conference Track)*, 2017.
5. Jastrzębski, Bahdanau, Hosseini, Nourkhovitch, Bengio, Cheung: *Commonsense mining as knowledge base completion? A study on the impact of novelty*, *Proceedings of the Workshop on Generalization in the Age of Deep Learning*, 8-16, 2018.
6. Jastrzębski, Leśniak, Czarnecki: *Learning to SMILE(S)*, *International Conference on Learning Algorithms (Workshop Track)*, 2016.

M. G

## Tematyka rozprawy

W ostatnich latach miało miejsce wiele przełomowych wyników wykorzystujących głębokie sieci neuronowe w dziedzinach takich jak rozpoznawanie obrazów, automatyczne tłumaczenie czy też gra w Go. Niestety jednak wciąż nie jest jasne dlaczego modele oparte o sieci neuronowe mają tak dobre własności uogólniania ze zbioru treningowego na zbiór testowy. Jest to zaskakujące, gdyż często liczba parametrów sieci neuronowych jest większa niż liczba elementów w zbiorze treningowym, co oznacza, że a priori może łatwo dojść do przeuczenia. Podejrzewa się, że istotną rolę odgrywa tu sam proces uczenia, który oparty jest zwykle na jednym z wariantów metody gradientu (ang. SGD). Głównym celem rozprawy jest odpowiedź na pytanie jaki wpływ na proces uczenia ma wybór kroku uczenia oraz rozmiaru próbki.

## Główne wyniki rozprawy

Główna część rozprawy, rozdział trzeci, zawiera wyniki z artykułów (1)-(3). Tematem przewodnim tych prac jest wpływ kroku uczenia i rozmiaru próbki na trajektorię wyznaczoną przez algorytm SGD pozwalającą na uniknięcie trywialnych minimów funkcji kosztu w których model zapamiętuje jedynie dane treningowe. Trajektorja ta w oczywisty sposób ma bezpośredni wpływ na końcowe możliwości wytrenowanego modelu.

Najważniejszym przedstawionym wynikiem jest znaczenie ilorazu kroku uczenia i rozmiaru próbki. W pracach (1) i (2) autor wykazuje na gruncie formalnym, że przy odpowiednich założeniach, stochastyczny szum w procesie SGD jest liniowo zależny od wspomnianego ilorazu. Iloraz ten jest kluczowym czynnikiem wpływającym na dynamikę uczenia i końcowe możliwości uogólniania wyuczonego modelu. Ponadto, w pracy (3) autor przeprowadza eksperymentalną analizę krzywizny funkcji kosztu wzdłuż trajektorii wyznaczonej przez proces SGD pokazując, że krzywizna ta jest zależna od kroku uczenia i rozmiaru próbki. Co więcej, autor przywiązuje dużą wagę do początkowej części procesu uczenia i jej wpływu na krzywiznę funkcji kosztu.

## Pozostałe wyniki

Druga część rozprawy, mianowicie rozdział czwarty, oparty na pracach (4)-(6), zawiera dodatkowe wyniki, gdzie celem jest wyjaśnienie wpływu architektury sieci neuronowych na możliwości uogólniania. W artykule (4) przedmiotem badań jest blok residualny, który jest składową sieci typu ResNet. Autor przedstawia blok residualny jako element dający przedstawić proces uczenia jako iterowane przybliżanie.

M. G.

W pracy (5) autor stara się wyjaśnić w jaki sposób wybór architektury sieci neuronowej wpływa na możliwości wyuczenia czegoś. co w ludzkim kontekście można nazwać zdrowym rozsądkiem, czyli umiejętnością odpowiadania na pytania nie w oparciu o wiedzę bazującą na faktach lecz na definicjach danych znaczeń i standardowych okolicznościach świata rzeczywistego. W szczególności w pracy (5) autor pokazuje eksperymentalnie, że jedna z wcześniej proponowanych metod prowadzi jedynie do zapamiętywania zbioru treningowego.

Tematem pracy (6) jest technika ang. *virtual screening* stosowana w obszarze poszukiwania nowych leków. W obszarze tym uogólnianie i wnioskowanie jest bardzo istotne, dlatego też autor proponuje aby na tym gruncie badać możliwości uogólniania sieci neuronowych.

## Ocena rozprawy

Należy zaznaczyć, że w dziedzinie sieci neuronowych dominują badania doświadczalne, a brakuje wyników takich jak te przedstawione przez autora, gdzie na bazie matematyki możliwe jest uzasadnień zjawisk, które możemy obserwować empirycznie.

W mojej ocenie rozprawa ma bardzo wysoki poziom merytoryczny. Z całą pewnością badania dotyczą aktualnego i ważkiego tematu, który interesuje wielu uczonych z całego świata (o czym świadczy również fakt, że prace autora są wielokrotnie cytowane). Chciałbym zaznaczyć, że najmocniejszą, według mnie, częścią rozprawy są prace (1)-(4), natomiast prace (5)-(6) stanowią dodatek do wcześniejszych wyników, co w żadnym stopniu nie powinno umniejszać rozprawie, gdyż to artykuły (1)-(4) stanowią o jej sile. Rozprawa jest tematycznie spójna i zgodna z tytułem, a jakość redakcyjną tekstu oceniam bardzo dobrze. Warto zaznaczyć, że kolejność autorów w przedstawionych pracach nie jest alfabetyczna, a autor rozprawy jest wiodącym autorem we wszystkich publikacjach.

## Podsumowanie

W mojej ocenie przedstawiona rozprawa doktorska i dorobek naukowy magistra Stanisława Jastrzębskiego bez wątplenia spełnia wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim, a co za tym idzie rozprawa ta może stanowić podstawę do nadania tytułu doktora. Co więcej, rekomenduję wyróżnienie rozprawy.

  
Marek Cygan

