

Prof. dr hab. Stanisław Matwin

Instytut Podstaw Informatyki PAN

oraz

Faculty of Computer Science, Dalhousie University, Kanada

Halifax, 25 sierpnia 2018

Recenzja

Przedmiotem recenzji jest rozprawa doktorska Pana **mgr inż. Piotra Wójcika „Random Projection in Deep Neural Networks”**. Promotorem rozprawy jest Pan prof. zw. Witold Dzwiniel. Recenzję opracowałem na wniosek Rady Wydziału Informatyki, Elektroniki i Telekomunikacji Akademii Górniczo-Hutniczej.

[Uwaga: terminologia polska dotycząca sieci neuronowych, użyta w niniejszej recenzji, pochodzi z tłumaczenia monografii Y. Bengio , A. Courville , I. Goodfellow, „Deep Learning - Systemy uczące się”, Wydawnictwo Naukowe PWN, Wwa 2018. W kilku miejscach, tam gdzie termin polski nie wydaje się być powszechnie używany, podano też termin angielski]

1. Dziedziny nauki związane z tematyką rozprawy

Tematyka rozprawy doktorskiej, która jest przedmiotem tej recenzji, dotyczy dyscypliny Informatyka. Zgodnie z taksonomią dyscyplin naukowych wykorzystywaną przez Narodowe Centrum Nauki (NCN) – odpowiada ona panelowi dziedzinowemu ST6 (Informatyka i Technologie Informacyjne), w szczególności pod-tematom ST6_7 (Sztuczna inteligencja, systemy inteligentne i wieloagentowe) oraz ST6_11 (Uczenie maszynowe, statystyczne przetwarzanie danych i zastosowanie w przetwarzaniu sygnałów).

2. Istotność podejmowanej tematyki, jakość postawionych tez, nowatorstwo i inne walory pracy

2.1. Istotność

Tematyka jest istotna ze względu na skalowalność metod uczących się, zwłaszcza sieci neuronowych, dla zadań wielowymiarowych. Jak praca przekonywująco argumentuje, zmniejszenie wymiarowości zadań uczenia się jest kluczowe dla zastosowań, zwłaszcza przy braku wyraźnej struktury danych, eksploatowalnej przez algorytm uczący się. Autor proponuje dwa typy rzutowań losowych, które dokonują projekcji pierwotnej macierzy danych definiującej

zadanie uczące. Autor przedstawia i dyskutuje wyniki dogłębnej analizy doświadczalnej skuteczności i jakości proponowanej metody na różnorodnych zbiorach danych. Przedstawione w pracy badania zaowocowały kilkoma pracami, opublikowanymi w wysoko notowanych pismach i konferencjach, np. Pattern Analysis and Applications i ESANN. Wydaje się, że dzięki wynikom zaprezentowanym w pracy rzutowanie losowe - ze względu na swą skuteczność i niski koszt obliczeniowy - może stać się jednym z często używanych narzędzi analizy danych typu Big Data.

2.2. Teza Pracy

Główna hipoteza pracy postuluje, że rzutowanie losowe jest skuteczną metodą ułatwiającą - czy wręcz umożliwiającą - stosowanie głębokich sieci neuronowych na masywnych zbiorach danych, charakteryzujących się bardzo wysoką liczbą atrybutów (co najmniej rzędu 10^5), brakiem wewnętrznej struktury i niską gęstością niezerowych wartości tych atrybutów. Problemy tego typu, często spotykane w praktyce, cierpią na tzw. *klątwę wielowymiarowości* – “curse of dimensionality”.

Druga hipoteza pracy stanowi że inicjalizacja wag takich sieci za pomocą elementów macierzy rzutowania przypadkowego poprawia jakość wyników pewnego typu sieci.

2.3. Nowatorstwo

Nowatorstwo pracy polega na wbudowaniu rzutowania losowego w strukturę głębokiej sieci neuronowej. Dodatkowo, wyniki dotyczące inicjalizacji wag są nowatorskie. Po pierwszej lekturze pracy wydawało się, że propozycja adaptacji wartości macierzy rzutowania, zawarta w rozdz. 4.2, jest też istotną innowacją pracy, ale zapewne tak nie jest (p. poniżej).

2.4. Inne walory pracy

Praca jest jasno i klarownie napisana. Lektura pracy przedstawia Autora jako wysokiej klasy eksperta w dziedzinie sieci neuronowych, w dziedzinie technik uczenia się typu „Big Data”, a więc działających na masywnych zbiorach danych, oraz w metodologii prowadzenia i analizy wyników eksperymentów na wielką skalę w dziedzinie Uczenia się Maszynowego. Praca jest poprawna metodologicznie i ciekawa w warstwie przedstawiającej intuicję i opinie Autora. Wreszcie angielszczyzna pracy jest biegła i wartka, bez zarzutu: nie ma właściwie żadnych uwag formalnych dotyczących prezentacji czy jakości tekstu.

3. Elementy pracy, które mogłyby zostać ulepszone

3.1. Istotne pozycje literatury.

Jak stwierdziłem powyżej, adaptacja losowych wartości macierzy do danych, przedstawiona w rozdz. 4.2 i następnie badana doświadczalnie w rozdz. 5, wydaje się być oryginalną kontrybucją pracy (jak stwierdza się w rozdz. 1.2). Uwadze Autora ubiegło niestety kilka wysoce relewantnych prac, w których zaproponowano podobne rozwiązania:

1. R. Zhao i K. Mao „Semi-random Projection fir Dimensionality Reduction and Extreme Learning Machine in High-dimensional Space”, IEEE Computational Intelligence Magazine, Aug. 2016, pp. 30-41.
2. H. Zou, T. Hastie, and R. Tibshirani, “Sparse Principal Component Analysis”, Journal of Computational and Graphical Statistics_Volume 15, 2006 - Issue 2.
3. X. Chen, Y. Qi, B. Bai, Q. Lin, J. G. Carbonell, “Sparse Latent Semantic Analysis”, Proceedings of the 2011 SIAM International Conference on Data Mining.

Powyższe prace przedstawiają i realizują pomysł adaptacji wartości macierzy rzutowania losowego do danych, leżący u podstaw rozdz. 4.2 rozprawy. Należy zauważyć, że rozwiązanie zaproponowane przez Autora istotnie różni się od powyższych prac: [1] używa metody dyskryminacji liniowej, [2] używa LASSO, a [3] – optymalizacji kwadratowej, podczas gdy Autor optymalizuje niejako „wewnątrz” zadania uczenia się, przy pomocy propagacji wstecznej z użyciem SGD. Tak więc wydaje się że w odróżnieniu od podejścia przedstawionego w rozdz. 4.2, skalowalność metod [1], [2] i [3] jest niepewna, a wyniki doświadczalne prezentowane w tych pracach są otrzymane na danych o wiele szerszych, niż te używane przez Autora (jest to normalne, biorąc pod uwagę wiek większości powyższych artykułów). Jednak powinny one być cytowane i dyskutowane w końcowej wersji pracy, jako że są bezpośrednio związane z tematyką badań będącą przedmiotem rozprawy.

3.2. Szersza perspektywa.

Praca skorzystałaby, w mojej opinii, z nieco szerszego osadzenia prezentacji w tematyce Uczenia się Maszynowego. Chodzi mi o powołanie kontekstu pojęć, technik i modeli tej dyscypliny w kilku miejscach rozprawy, gdzie byłoby to celowe i chyba ciekawe. Tak na przykład sam pomysł adaptacji losowych elementów macierzy do danych w procesie selekcji atrybutów, szczególnie w ujęciu Autora gdzie rzutowanie jest wbudowane w sieć dokonującą klasyfikacji, jest rodzajem metody typu „wrapper” – klasycznej idei pochodzącej od R. Kohavi’ego z lat 90. Technika „wrappera” należy do klasycznego repertuaru algorytmów selekcji atrybutów. Jeśliby

widzieć rzutowanie losowe z adaptacją jako wrapper, możnaby być może w sposób elegancki i uzasadniony porównać to podejście z innymi metodami selekcji atrybutów.

Użyteczne byłyby też odwołania do teorii Uczenia się Maszynowego, tam gdzie jest to sensowne: np. czołowe w pracy zagadnienie „klątwy wielowymiarowości”, stanowiące jej podstawową motywację, jest nie tylko obserwacją empiryczną, ale wynika bezpośrednio ze sformułowania zadania uczenia się w modelu PAC (Probably Approximately Correct).

Charakteryzacja metody SVM z liniowym jądrem (kernel) jako metody liniowej (np. str. 3 i str. 48) jest chyba nieprawidłowa, ponieważ algorytmicznie uczenie się SVM wykonuje kwadratowe zadanie optymalizacyjne i wiadomo, że jego koszt jest $O(n^2)$, a więc bardziej właściwie byłoby używać pełnej nazwy.

4. Dyskusja

Wdaje się, że praca sugeruje pewne pytania, które być może mogłyby zostać umieszczone na liście przyszłych badań, stanowiących kontynuację pracy. Oto kilka z nich:

Ciekawa byłaby refleksja, stwiająca pytanie, w którym momencie różnorakie metody projekcji proponowane w literaturze odchodzą zbyt daleko od twierdzenia Johnsona-Lindenstraussa, aby wciąż działać? Szczególnie gdy rzutowanie losowe jest zastąpione przez rzutowanie zależne od konkretnych danych? Jest to chyba uzasadnione pytanie, szczególnie w świetle pracy Beyer et al., „When is Nearest Neighbor Meaningful”, ICDT'99, pp. 217-23.

Inicjalizacja wag na sposób losowy wydaje się z jednej strony związana z techniką odrzucania wag („dropout”), a z drugiej strony jest przeciwieństwem techniki inicjalizacji wag na sposób eksplorujący abstrakcyjną strukturę problemu, owocnej w popularnych obecnie metodach transferu poznawania („transfer learning”). Dyskusja tych nasuwających się związków mogłaby być interesująca, dowiadując rzutowanie losowe do dobrze znanych metod stanowiących niejako kościec głębokich neuronowych.

Wreszcie - na koniec – naturalne pytanie, które postawi niejeden czytelnik szerzej zainteresowany Učeniem się Maszynowym: czy metoda projekcji liniowej mogłaby być użyteczna autonomicznie, jako część wstępnej obróbki danych, przed użyciem narzędzi klasyfikacyjnych, choćby regresji czy klasyfikatora Bayesowskiego, albo nawet, w wersji ze stałą macierzą rzutowania, w wielowymiarowych zadaniach grupowania danych?

5. Dorobek naukowy kandydata

Kandydat posiada istotny dorobek naukowy. W repozytorium Google Scholar znajduje się 8 prac w których Kandydat jest współautorem, z których cztery były opublikowane w pismach z Listy A MNiSW. Podany indeks h Kandydata wynosi 6. Obie te liczby potwierdzają, że dorobek Kandydata jest, jak na tak wczesne stadium kariery naukowej, znaczący i przewyższa średnią, nawet według norm światowych.

6. Podsumowanie

Podsumowując przedstawianą opinię, stwierdzam ostatecznie, że praca **mgr inż. Piotra Iwo Wójcika** spełnia wymagania przewidziane dla rozpraw doktorskich w aktualnie obowiązującej ustawie (*Ustawa o stopniach naukowych i o tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dnia 14 marca 2003 roku, Dziennik Ustaw Nr 65, poz. 595*). Dlatego, stawiam wniosek o **przyjęcie tej pracy jako rozprawy doktorskiej i o dopuszczenie Kandydata do jej publicznej obrony.**

Prof. dr hab. Stanisław Matwin, prof. zw. IPI PAN

A handwritten signature in black ink, appearing to read 'S. Matwin', with a stylized flourish at the end.