

Czeladź, 18 marca 2024

Recenzja pracy doktorskiej Pana Jakuba Świątkowskiego
„*Variational Inference Applications in Deep Learning*”

Rozprawa dotyczy ważnego problemu połączenia wnioskowania wariacyjnego z nauczaniem głębokim. Jest to istotny problem zarówno z punktu widzenia teoretycznego jak i potencjalnych zastosowań w których istotna jest możliwość szacowania niepewności. Rozprawa bazuje na czterech pracach P1-P4 z których pierwsze dwie ukazały zostały przyjęte na konferencji ICML, jednej z paru najlepszych konferencji z nauczania maszynowego. Autor rozprawy jest głównym autorem prac P1, P3, P4. Należy także dodać, że prace P1 i P2 znalazły one duży oddźwięk w środowisku naukowym, o czym świadczy między innymi ich bardzo duża liczba cytowań (za Google Scholar: praca P1 ma 55, zaś praca P2 ponad 300). Prace P1 i P2 mają charakter głównie teoretyczny, zaś P3 i P4 mają nastawienie aplikacyjne.

Rozprawa zaczyna się od czytelnego i klarownie napisanego wprowadzenia do rozważanej tematyki. Rozdział 2 zawiera ogólne wprowadzenie do wariacyjnego wnioskowania w Bayesowskich sieciach neuronowych, co stanowi wprowadzenie do rozdziałów 4 i 5 (odp. prace P1 i P2). Rozdział 3 zawiera wprowadzenie do wariacyjnego wnioskowania w generowaniu mowy, wprowadzenie do rozdziałów 6 i 7 (odp. prace P3 i P4). Pracę kończy bogaty appendix zawierający dodatkowe eksperymenty, analizy i dowody które nie zmieściły się w oryginalnych pracach.

W rozdziale 3, autor rozprawy pokazuje, że Bayesowskie sieci neuronowe, trenowane za pomocą wariacyjnego wnioskowania Gaussian Mean-Field (GMFVI), uczą się macierzy odchyłeń standardowych a posteriori, które mogą być przybliżone z minimalną stratą informacji poprzez dekompozycje niskiego rzędu korzystające z rozkładu SVD (nazywane „k-tied normal posteriors”). Inaczej mówiąc, GMFVI uczy się macierzy odchylenia standardowego posteriora, która w sposób naturalny ma strukturę niskowymiarową. Choć podejście GMFVI jest raczej bazowe (najprostszy model gaussowski), w rzeczywistości jest ono bardzo często stosowane, gdyż dobrze się skaluje

i implementuje, w związku z tym badanie własności GMFVI jest kluczowe w teorii Bayesowskich sieci neuronowych. Wyniki zawarte w P3 sugerują, że bogatsze parametryzacje posteriora wariacyjnego nie zawsze są potrzebne, a parametryzacje przy pomocy znacząco mniejszej liczby parametrów potencjalnie działają równie dobrze. W konsekwencji autorzy proponują prostą, ale skuteczną parametryzację posteriora wariacyjnego, która przyspiesza uczenie i zmniejsza liczbę parametrów wariacyjnych bez pogarszania jakości predykcji. Praca P3 ma istotne znaczenie dla zastosowań sieci Bayesowskich, gdyż zmniejszenie ilości trenowalnych parametrów w sieciach Bayesowskich pozwala na ich szersze zastosowanie. W szczególności z mojej subiektywnej perspektywy widzę potencjalne zastosowania tego podejścia w modelach Bayesowskich dla problemów meta-learning (continual learning czy few-shot learning).

Praca P2 (rozdział 4) jest moim zdaniem najbardziej interesującą z prac rozprawy. Autorzy pokazują za pomocą próbkowania MCMC, że predykcje posteriora indukowane przez posterior Bayesa dają systematycznie gorsze predykcje w porównaniu z prostymi metodami, w tym oszacowaniami punktowymi uzyskanymi z SGD. Co jest szczególnie interesujące, a zupełnie niezgodne z intuicją, okazuje się, że jakość predykcji znacznie wzrasta przez użycie „zimnego posteriora”. Jest to model odbiegający od paradygmatu bayesowskiego, ale powszechnie stosowany jako heurystyka w pracach nad bayesowskim uczeniem głębokim, który polega na sztucznym zmniejszeniu parametru temperatury („wyostrzenie” gęstości) w rozkładzie a posteriori. Autorzy pracy w sposób systematyczny badają koncept zimnych posteriorów, i oceniają te hipotezy poprzez eksperymenty. Praca P2 kwestionuje cel dokładnych przybliżeń posteriorów w bayesowskim uczeniu głębokim: mianowicie jeśli prawdziwy posterior Bayesa jest słaby, nie ma sensu tworzenie bardziej dokładnych jego przybliżeń. Praca P2 jest dla mnie w związku z tym fundamentalne znaczenie w rozwoju teorii Bayesowskich sieci neuronowych. Chociaż autor rozprawy nie jest głównym autorem pracy P2, uważam za bardzo wartościowy z punktu widzenia jego rozwoju naukowego fakt, iż współpracował przy jej powstawaniu z licznym gronem doskonałych naukowców o międzynarodowej renomie.

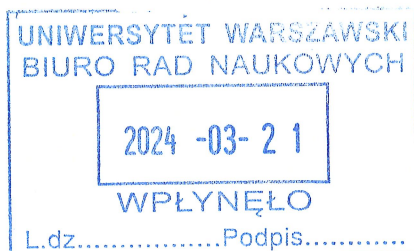
Rozdziały 5 i 6 (prace P3, P4) dotyczą bardzo istotnego z punktu widzenia zastosowań problemu wykonywania takiego dubbingu który zachowuje oryginalne charakterystyki wypowiedzi (intonacja, akcent czy rytm) czyli prozodię. Mówiąc dokładniej, prozodia to cechy wypowiedzi, które nadają jej brzmieniowy charakter, w szczególności melodia, głośność, tempo mówienia,

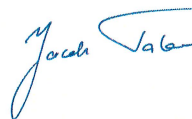
akcent, siła dynamiczna, rytmiczność. W rozdziale 5 zaprezentowano nowatorskie rozwiązanie bazujące na modelu VITS (VAE połączone z uczeniem adversarialnym), które pokazuje efektywny model transferu prozodii. Stworzony model potrafi generować mowę dubbingowaną z prozodią dopasowaną do kontekstu. Dodatkowo, autor proponuje dwie metody radzenia sobie z wyzwaniem stawianymi przez hałas w danych multimedialnych. Po pierwsze, wprowadzono moduł modelowania hałasu, który oddziela hałas od prozodii, gdzie wykorzystuje się sygnał pozbawiony hałasu wyekstrahowany z referencyjnego dźwięku. Po drugie, uzupełniono dane z hałasem o czyste dane treningowe, aby poprawić zdolność modelu do mapowania pozbawionego hałasu dźwięku referencyjnego na czystą mowę. Model zbudowany w pracy P3 pokazuje bardzo dobre praktyczne wyniki, przebijając istniejące SOTA. W rozdziale 6-tym (praca P4) zaproponowano nowe rozwiązanie, które umożliwia transfer prozodii międzyjęzykowej na poziomie fraz oraz między mówcami. Pozwala to nauczyć się modelowania informacji prozodyjnej na poziomie fraz i transferować osadzenia prozodii frazowej z języka źródłowego na język docelowy dla tłumaczonego tekstu. Prace P3 i P4 mają bardzo mocny charakter aplikacyjny i mogą być efektywnie wykorzystane w dubbingu tworzonym przez AI.

Autor rozprawy w zamieszczonych pracach pokazał, że potrafi używać i rozbudowywać szeroką klasę narzędzi z nauczania maszynowego. Co więcej, jego prace zawierają zarówno głębokie analizy teoretyczne (prace P1, P2), jak i ciekawe i wartościowe modele (prace P3, P4) które dotyczą praktycznych problemów (problem prozodii przy automatycznym dubbingu). Rozważaną rozprawę czytałem z dużym zainteresowaniem. Pomimo tego, że nie jestem specjalistą stricte z modeli Bayesowskich, prace P1 i P2 były dla mnie inspirowane. Z mojego punktu widzenia nawet same prace P1 i P2 stanowią na tyle istotny wkład naukowy w Deep Learning, że stanowią wystarczający dorobek na doktorat. Z pewnych minusów rozprawy jest jej ideologiczna niespójność, a mianowicie prace P3 i P4 można traktować jedynie jako luźno powiązane z pracami P1 i P2, nie wpływa to jednak dla mnie w zasadniczy sposób na wysoką ocenę prezentowanego dorobku.

Konkludując, rozprawa spełnia wszelkie zwyczajowe i ustawowe kryteria stawiane rozprawom doktorskim, w związku z czym wnioskuję o nadanie Panu Jakubowi Świątkowskiemu stopnia naukowego doktora w dziedzinie nauk ścisłych i przyrodniczych, w dyscyplinie informatyka.

Równocześnie uważam, że wyniki uzyskane przez doktoranta są bardzo wartościowe, w związku z czym wnioskuję o wyróżnienie rozprawy.



 Podpisany elektronicznie przez
Jacek Tabor
21.03.2024
14:42:07 +01'00'