

Warszawa, 27.11.2023

dr hab. inż. Tomasz Trzcíński, prof. PW  
Instytut Informatyki Politechniki Warszawskiej

**RECENZJA ROZPRAWY DOKTORSKIEJ  
DLA RADY DYSCYPLINY NAUKOWEJ INFORMATYKI  
UNIwersytetu Warszawskiego**

**Tytuł rozprawy:** Variational Inference - Applications in Deep Learning

**Autor rozprawy:** Jakub Świątkowski

## 1 Analiza strony merytorycznej rozprawy

### 1.1 Obszar problemowy

Recenzowana rozprawa doktorska skupia się na połączeniu uczenia głębokiego (deep learning) i wnioskowania wariacyjnego (variational inference), które są kluczowymi obszarami w dziedzinie uczenia maszynowego i statystyki. Praca ta ma na celu pokonanie ograniczeń modeli uczenia głębokiego, zwłaszcza w zakresie kwantyfikacji niepewności i prognoz probabilistycznych, poprzez wykorzystanie zasad wnioskowania Bayesowskiego. Recenzowana rozprawa doktorska jest oparta na cyklu czterech publikacji, które ukazały się na renomowanych konferencjach informatycznych, a których Autor rozprawy jest współautorem.

Rozprawa koncentruje się na dwóch aspektach badanych zagadnień:

- **Bayesowskie Sieci Neuronowe (BNNs):** W tej części, przypadającej na opis dwóch pierwszych publikacji, rozprawa traktuje o stosowaniu wnioskowania wariacyjnego do wag sieci w BNN. Badanie to rozszerza zrozumienie i wykorzystanie BNN, odkrywając, że popularna forma wnioskowania wariacyjnego w BNNs ma strukturę niskiego rzędu, co pozwala na redukcję liczby parametrów i przyspieszenie uczenia modelu. Ponadto, autor bada paradoksalne odkrycie, że BNNs, które odchylają się od teoretycznie optymalnego wnioskowania Bayesa, często wykazują lepszą wydajność.
- **Synteza Mowy:** W tej części rozprawy Autor płynnie przechodzi do praktycznego zastosowania wnioskowania wariacyjnego w syntezie mowy, a w szczególności do tworzenia Auto-Enkoderów Wariacyjnych (VAE), które uczą się rozłącznych, interpretowalnych i kontrolowalnych reprezentacji danych. Autor proponuje nową metodę dubbingu maszynowego, która opiera się na nauczaniu zanurzeń prozodii mowy, oddzielonych od języka, mówcy i szumów kanału. Ta metoda pozwala na transfer prozodii między językami i mówcami, generując wysokiej jakości dubbingi maszynowe. Praca wykazuje, że modelowanie prozodii na poziomie frazy, a nie całej wypowiedzi, znacznie poprawia jakość dubbingu.

Co prawda recenzowana rozprawa nie zawiera jednej jasno postawionej hipotezy badawczej, jednak można się domyślić, iż Autor jest w swoich badaniach motywowany założeniem, iż wykorzystanie wnioskowania wariacyjnego poprawia jakość współczesnych metod uczenia głębokiego, w szczególności sieci Bayesowskich oraz modeli VAE. Drobny mankament braku jasno sformułowanej hipotezy jest pewien przeskok obserwowany nawet w samej strukturze rozprawy pomiędzy tematyką pierwszych dwóch publikacji, a tematyką dwóch ostatnich. Nie jest to jednak rzecz niespotykana w dzisiejszej informatyce, ponieważ dziedzina ta rozwija się niezwykle szybko, a sformułowanie hipotezy badawczej mogącej przetrwać próbę czasu trwania doktoratu staje się coraz trudniejsze.

W recenzowanej rozprawie doktorskiej, autor stawia sobie kilka kluczowych celów związanych z wykorzystaniem wnioskowania wariacyjnego w dziedzinie uczenia głębokiego. Te cele obejmują:

- **Badanie potencjału wnioskowania wariacyjnego:** Celem jest zbadanie, jak wnioskowanie wariacyjne, będące zaawansowaną ramą do przybliżonego wnioskowania Bayesowskiego, może zostać wykorzystane w uczeniu głębokim. Autor chce sprawdzić, jak te techniki mogą pomóc w rozwiązaniu istotnych problemów w uczeniu głębokim, takich jak szacowanie niepewności i modelowanie probabilistyczne.
- **Wdrażanie i skalowalność modele probabilistyczne:** Praca ma na celu zastosowanie wnioskowania wariacyjnego do efektywnego i skalowalnego modelowania probabilistycznego w systemach uczenia głębokiego. Jest to szczególnie ważne, biorąc pod uwagę, że uczenie głębokie stało się dominującym podejściem do rozwiązywania złożonych zadań w różnych dziedzinach.
- **Rozwój teorii wnioskowania wariacyjnego w kontekście głębokiego uczenia:** Autor pragnie wnieść nowe metody, spostrzeżenia teoretyczne oraz oceny empiryczne, aby poszerzyć zrozumienie korzyści, ograniczeń i przyszłych kierunków rozwoju wnioskowania wariacyjnego w kontekście uczenia głębokiego.
- **Wskazanie korzyści i ograniczeń:** Praca ma na celu nie tylko prezentację nowych metodologii, ale także dostarczenie wglądu w zalety i ograniczenia wykorzystania wnioskowania wariacyjnego w uczeniu głębokim, co ma być podstawą do rozważań nad przyszłymi kierunkami rozwoju w tej dziedzinie.

Autor z powodzeniem realizuje powyższe cele, choć należy wspomnieć, że w moim odczuciu przedstawione w rozprawie metody i rozwiązania nie zostały w pełni zweryfikowane pod kątem skalowalności czy wydajności. Nie zmienia to faktu, że praca ta przyczynia się do otwarcia nowych możliwości w dziedzinach Bayesowskich sieci neuronowych i neuralnej syntezy mowy, dostarczając nowych metod i spostrzeżeń do dalszych badań.

Rozprawa składa się z siedmiu rozdziałów oraz dwóch dodatkowych załączników, w których autor rozszerza przedstawione publikacje o dowody przedstawianych twierdzeń oraz szczegółowe wyniki i opisy implementacyjne.

Pierwszy rozdział rozprawy to wstęp, który opisuje motywację podjęcia pracy badawczej w określonym wyżej kierunku, krótko omawia aktualnie istniejące rozwiązania oraz ich ograniczenia. W rozdziale tym Autor opisuje również zidentyfikowane luki badawcze, które prowadzą autora w kierunku proponowanych celów badawczych, a także przedstawia listę publikacji, na których opiera się recenzowana praca. Są to:

- [P1] J. Świątkowski, K. Roth, B. Veeling, L. Tran, J. Dillon, J. Snoek, S. Mandt, T. Salimans, R. Jenatton, S. Nowozin, *The k-tied Normal Distribution: A Compact Parameterization of Gaussian Mean Field Posteriors in Bayesian Neural Networks*, *International Conference on Machine Learning (ICML)*, 2020.
- [P2] F. Wenzel, K. Roth, B. Veeling, J. Świątkowski, L. Tran, S. Mandt, J. Snoek, T. Salimans, R. Jenatton, S. Nowozin, *How Good is the Bayes Posterior in Deep Neural Networks Really?*, *International Conference on Machine Learning (ICML)*, 2020.
- [P3] J. Świątkowski, D. Wang, M. Babianski, P. Tobing, R. Vipperla, V. Pollet, *Cross-lingual Prosody Transfer for Expressive Machine Dubbing*, *INTERSPEECH*, 2023.
- [P4] J. Świątkowski, D. Wang, M. Babianski, G. Coccia, P. Tobing, R. Vipperla, V. Klimkov, V. Pollet, *Expressive Machine Dubbing Through Phrase-level Cross-lingual Prosody Transfer*, *INTERSPEECH*, 2023.

W rozdziale drugim Autor przedstawia bardziej szczegółowo zagadnienia związane z Bayesowskimi sieciami neuronowymi, niezbędne do zrozumienia wkładu w rozwój nauki prezentowanego w artykułach [P1] i [P2]. Następnie przedstawione są główne aspekty tychże publikacji, wraz z kluczowymi rezultatami.

Rozdział trzeci ma podobny format do poprzedzającego rozdziału, a mianowicie składa się z bardziej rozbudowanego wstępu poświęconego neuralnej syntezie mowy oraz z opisu wkładu kolejnych publikacji z cyklu, tj. [P3] i [P4], wraz z rezultatami.

Kolejne rozdziały prezentują w całości publikacje [P1-4], a po nich Autor zamieszcza wspomniane wyżej załączniki.

Zaproponowana konstrukcja rozprawy jest prawidłowa i zgodna z praktyką pisania prac doktorskich bazujących na cyklu publikacji.

Rozprawa zawiera reprezentatywną dla poruszanej problematyki bibliografię, która w większości przypadków właściwie ilustruje omawiane zagadnienia i dokumentuje wkład własny Autora.

## 1.2 Ocena wyników oraz stopnia ich oryginalności

Oryginalne wyniki pracy badawczej doktoranta, opisane w rozprawie, osiągnięte zostały zarówno w kontekście opracowania nowych metod parametryzacji modeli czy uczenia reprezentacji, jak również identyfikacji nowych zjawisk zachodzących w głębokich sieciach neuronowych.

Zaprezentowane w rozprawie rozwiązania bazują na nowych pomysłach lub zawierają wyraźne modyfikacje w stosunku do metod przedstawionych w literaturze. Na szczególną uwagę zasługuje dogłębne zrozumienie różnorodnych aspektów wnioskowania wariacyjnego, w szczególności analiza rozkładów a posteriori w kontekście sieci Bayesowskich, oraz umiejętność zastosowania nabytej wiedzy do praktycznego rozwiązania problemu mechanicznego dubbingu w syntezie mowy.

Wśród wielu przedstawionych w rozprawie rozwiązań, dodatków technicznych i implementacyjnych oraz usprawnień istniejących metod należy wyróżnić:

1. Odkrycie, że najczęściej używana rodzina wariacyjnych rozkładów a posteriori dla wag Bayesowskich sieci neuronowych ma strukturę niskiego rzędu. [P1]
2. Zaproponowanie parametryzacji niskiego rzędu wariacyjnych rozkładów a posteriori, która redukuje liczbę parametrów o połowę i prowadzi do szybszej zbieżności treningu. [P1]
3. Postawienie i weryfikacja hipotezy iż sztuczne zmniejszenie wariancji nauczonej a posteriori w BNN poprawia wydajność predykcyjną w porównaniu z teoretycznie optymalnymi Bayesowskimi rozkładami predykcyjnymi w BNN. [P2]
4. Opracowanie nowej metody uczenia reprezentacji VAE poprawiającej wydajność wokalną z oddzielnymi cechami języka i głosu, które mogą być przenoszone między językami. [P3]
5. Zaproponowanie nowej metody regularyzacji reprezentacji na poziomie frazy opartej na jej długości, która poprawia oddzielenie wykonania wokalnego od oryginalnej treści językowej. [P4]

Należy podkreślić, że pomimo tego, iż część publikacji wchodzących w skład przedstawionej rozprawy ma kilku autorów, wkład Doktoranta jest jasno opisany i stanowi dominującą część prac, o czym świadczy również jego pierwsza pozycja na liście autorów trzech z czterech prac.

Zauważone nieliczne niedoskonałości dysertacji dotyczą raczej drobnych kwestii, takich jak brak jasno nakreślonych dalszych prac związanych z rozwijanym kierunkiem badań, w tym m.in. ograniczeń proponowanych metod, czy brak weryfikacji zwiększenia wydajności proponowanych metod w formie odpowiednich eksperymentów.

## 1.3 Zagadnienia dyskusyjne

Poniższe uwagi dotyczą ogólniejszych kwestii poruszonych w rozprawie i nie odnoszą się bezpośrednio do treści pracy. Niemniej jednak liczę na analizę tych zagadnień i odniesienie się do nich przez Doktoranta.

- *Empiryczna weryfikacja wydajności proponowanych modeli:* Pomimo wskazania na samym początku rozprawy, w rozdziale wprowadzającym, iż wnioskowanie wariacyjne pozwala na przybliżanie wnioskowania Bayesowskiego, co przekłada się na lepszą wydajność i skalowalność metod bazujących na tych technikach, Autor nie przedstawia w swojej rozprawie przekonujących dowodów, jakoby takie zwiększenie wydajności rzeczywiście miało miejsce. W szczególności rozprawa ma często charakter teoretyczny, co samo w sobie nie jest oczywiście żadną niedoskonałością. Natomiast w zestawieniu z praktycznym charakterem proponowanych metod, w tym ich zastosowaniem do syntezy mowy, brakuje mi dyskusji popartej eksperymentami, która pozwoliłaby w praktyce zweryfikować skalowalność zaproponowanych rozwiązań, w szczególności w kontekście ograniczonych zasobów obliczeniowych, którymi dysponują np. autonomiczne roboty czy telefony komórkowe. Taka analiza dopełniłaby prezentowany przez Autora obraz zalet wariacyjnego wnioskowania i uprawdopodobniłaby szerokie zastosowanie prezentowanych w rozprawie metod.

## 2 Analiza strony formalnej rozprawy

### 2.1 Ocena układu pracy i redakcji manuskryptu

Recenzowana rozprawa doktorska napisana jest w języku angielskim i obejmuje w kolejności: podziękowania, streszczenie w języku angielskim, streszczenie w języku polskim, siedem rozdziałów zasadniczych, w tym w czterech z nich przedruki artykułów, a także bibliografię. Dodatkowo praca zawiera dwa załączniki. Łącznie praca liczy 175 stron.

Bibliografia liczy ponad 200 uporządkowanych alfabetycznie pozycji. Spośród znajdujących się w bibliografii prac większość ukazała się w przeciągu ostatnich 10 lat co świadczy o dobrym umiejscowieniu tematyki rozprawy w aktualnym w skali światowej nurcie badań. Bibliografia rozprawy, po włączeniu w nią utworów cytowanych w ramach załączonych artykułów, nie budzi zastrzeżeń od strony merytorycznej, a jej redakcja jest staranna i nie dostrzegłem w niej żadnych błędów.

Układ rozprawy jest prawidłowy, jest ona także starannie opracowana pod względem edytorskim, w szczególności rozdziały pierwszy, drugi i trzeci są odpowiednio dopracowane i nie widać w nich żadnych błędów w sztuce. Użyta terminologia jest właściwa dla obszaru problemowego rozprawy, zarówno w zakresie robotyki, informatyki jak i uczenia maszynowego.

### 2.2 Uwagi szczególne

Tekst rozprawy jest poprawny pod względem językowym, stylistycznym i merytorycznym. Podczas lektury nie zauważyłem błędów redakcyjnych czy pomyłek wymagających korekt. Jedynym minimalnym błędem jaki dostrzegłem było użycie przecinka zamiast kropki przy imieniu Autora w cytowaniu pracy [P2] na stronie 12 rozprawy. Poza tym oceniam jakość redakcji rozprawy, w tym poprawność językową, niezwykle wysoko.

## 3 Konkluzja

Uważam, że recenzowana dysertacja Pana mgr. inż. Jakuba Świątkowskiego **spełnia wymagania** stawiane rozprawom doktorskim przez *Ustawę z 20 lipca 2018 roku (Dz. U. 2018 poz. 1668) Prawo o szkolnictwie wyższym i nauce*, ponieważ zawiera oryginalne koncepcje i rozwiązania istotnych problemów naukowych w **dyscyplinie informatyka**, jak również ich oryginalne zastosowania. Rezultaty nawiązują do aktualnego stanu wiedzy i mają znaczenie praktyczne, czego potwierdzeniem są liczne pierwszo-autorskie publikacje na renomowanych konferencjach (ICML - CORE A\*, 200 pkt. MEiN, INTERSPEECH - CORE A, 140 pkt. MEiN). Uzyskane wyniki eksperymentalne w wystarczającym stopniu dokumentują poprawność proponowanych koncepcji oraz skuteczność działania ich implementacji. Uwagi krytyczne sformułowane w treści recenzji, po części mające charakter dyskusyjny, nie wpływają istotnie na moją bezapelacyjnie pozytywną ocenę oryginalności i poziomu merytorycznego przedstawionej rozprawy.

Sformułowanie problemów badawczych, zaproponowane oryginalne metody ich rozwiązania, sposób przeprowadzenia badań oraz zademonstrowane umiejętności formułowania wniosków świadczą o przygotowaniu Doktoranta do pracy naukowej. Na tej podstawie **wnioskuję o dopuszczenie mgr. inż. Jakuba Świątkowskiego do dalszych kroków procedury uzyskania stopnia doktora w dziedzinie nauk ścisłych i przyrodniczych w dyscyplinie informatyka.**

Tomasz Trzciniński

