

dr hab. inż. Robert Burduk
Politechnika Wroclawska
Wydział Informatyki i Telekomunikacji
Ul. Wybrzeże Stanisława Wyspiańskiego 27
50-370 Wrocław

Wrocław, dnia 24.11.2022 r.

S E K R E T A R I A T
Rady Dyscypliny AEEITK
Wpłynęło dnia..... 02. 12. 2022
Zarejestrowano pod nr
Podpis *Am*

RECENZJA

rozprawy doktorskiej mgr inż. Michała Karwatowskiego
zatytułowanej: „**Sprzętowa akceleracja wymagających obliczeniowo operacji
na potrzeby algorytmów sztucznej inteligencji w układach FPGA**”

Recenzja została sporządzona w związku z powołaniem przez Radę Dyscypliny Automatyka, Elektronika i Elektrotechnika Akademii Górniczo-Hutniczej im. Stanisława Staszica w Krakowie, piszącego niniejszą recenzję, jako recenzenta rozprawy doktorskiej mgr inż. Michała Karwatowskiego pismem z dnia 30 września 2022 r.

Kryteria oceny dysertacji wynikają z przepisów zawartych w art. 13 ust. 1 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65 r., poz. 595 z późn. zm.).

Problem badawczy i jego znaczenie

Zakres recenzowanej rozprawy dotyczy szerokiej i dynamicznie rozwijającej się problematyki akceleracji sprzętowej w układach FPGA wybranych metod uczenia maszynowego. W szczególności Doktorant koncentruje się na zagadnieniach dotyczących przetwarzania języka naturalnego, redukcji precyzji oraz wymiarowości danych klasyfikowanych za pomocą algorytmu k-NN oraz konwersji wysokopoziomowego opisu sieci neuronowych do implementacji w układach FPGA.

W rozprawie sformułowano trzy tezy badawcze, które dotyczą:

- optymalizacji energetycznej obliczeń oraz skrócenia czasu wykonywania algorytmów uczenia maszynowego akcelerowanych w układach FPGA,

- uzyskania niskich latencji przy zachowaniu wysokiej wartości metryki predykcji algorytmów uczenia maszynowego akcelеровanych w układach FPGA,
- konwersji wysokopoziomowego opisu sieci neuronowych do układów FPGA.

Przedstawione hipotezy jednocześnie definiują cele badawcze, którymi są:

- opracowanie uniwersalnego narzędzia DL2HDL pozwalającego na konwersję wysokopoziomowego opisu sieci neuronowych do układów FPGA,
- przeprowadzenie szeregu badań eksperymentalnych pozwalających ocenić akcelerację w układach FPGA wybranych metod oraz algorytmów uczenia maszynowego.

Praca ma charakter koncepcyjno-eksperymentalny, zatem postawione tezy badawcze zostały zweryfikowane na drodze eksperymentalnej. Zamierzone cele badawcze zostały również osiągnięte przy zastosowaniu eksperymentalnej metod badawczej. Schemat eksperymentu naukowego, przeprowadzonego przez Doktoranta polegał na: zdefiniowaniu zmiennych niezależnych (parametrów eksperymentów), przeprowadzeniu badań, przedstawieniu wyników badań w formie tabel oraz rysunków, przedstawieniu wniosków jak zmiana wartości zmiennych niezależnych wpływa na zmianę wartości zmiennych zależnych. Zastosowana metoda badawcza jest prawidłowa. Omówienie wyników badań zawarte po każdym przeprowadzonym eksperymencie jest również prawidłowe.

Tematyka podjęta przez mgr inż. Michała Karwatowskiego jest interesująca, w pełni uzasadniona oraz odpowiada na wyzwania współczesnej elektroniki. Wynik przedstawione w recenzowanej pracy bez wątplenia mogą znaleźć zastosowanie praktyczne. W szczególności opracowane narzędzie DL2HDL pozwala na akcelerację sprzętową modeli uczenia maszynowego uzyskiwanych przy pomocy rekurencyjnych sieci neuronowych, a wyniki eksperymentalne określają przydatność akceleracji sprzętowej w układach FPGA wybranych metod oraz algorytmów uczenia maszynowego.

Struktura pracy oraz wiedza Autora

Recenzowana praca została napisana w języku polskim i liczy 145 stron maszynopisu (bez załącznika). Składa się z czterech rozdziałów merytorycznych, wstępu, podsumowania, bibliografii, słownika terminów, załącznika, streszczenia w języku polskim oraz angielskim. Rozdział nr 1 przedstawia kolejno: motywację dotyczącą podjęcia tematyki badawczej, cele badawcze nieco zawile przedstawione w tekście, sformułowane trzy hipotezy badawcze, oraz strukturę dysertacji.

Rozdział 2 zawiera wprowadzenie dotyczące problematyki programowalnych układów logicznych oraz metod i algorytmów z zakresu sztucznej inteligencji. Kolejność merytorycznych treści zawarta w następujących po sobie punktach nie jest konsekwentna. Na przykład punkt 2.1 przedstawia wybrane zastosowania sztucznej inteligencji, opis metod uczenia maszynowego znajduje się w punkcie 2.6, a zagadnienia związane z FPGA przedstawiono w punkcie 2.2.

Rozdział 3 dysertacji przedstawia akcelerację obliczeń dotyczących wybranych aspektów przetwarzania języka naturalnego, do których należy metoda TF-IDF służąca do wyznaczeniu statystycznej wagi termów oraz metryki podobieństwa kosinusowego dokumentów. Badania eksperymentalne dotyczą oceny czasu przetwarzania oraz zużycia energii dla różnych platform sprzętowych.

Rozdział 4 recenzowanej rozprawy przedstawia wyniki badań dotyczące redukcji precyzji wartości miary kosinusowej oraz redukcji dokładności algorytmu k-NN przy zastosowaniu akceleracji sprzętowej. Dodatkowo rozdział przedstawia problematykę przeniesienia modelu sieci neuronowych na platformę FPGA dla różnych wartości przycinania oraz kwantyzacji.

Rozdział 5 przedstawia opracowane przez Doktoranta narzędzie DL2HDL służące do konwersji wysokopoziomowego opisu sieci neuronowych zaprojektowanych w języku *Python* do implementacji w układach programowalnych. Omawiany rozdział zawiera również wyniki badań eksperymentalnych wykonany na trzech zbiorach danych, których celem było określenie latencji odpowiedzi sieci neuronowej oraz określenie wartości wykorzystanych zasobów dla różnych wartości rzadkości oraz kwantyzacji.

Recenzowaną dysertację kończy rozdział 6, który jest podsumowaniem wyników pracy badawczej magistra inż. Michała Karwatowskiego. Niniejszy rozdział niestety nie zawiera propozycji dalszych prac badawczych.

Spis literatury liczy 82 pozycje. Cytowane prace dobrane są prawidłowo, są aktualne i odnoszą się do omawianych w pracy problemów.

Wkład Autora — oryginalne osiągnięcia

Wkład Autora w rozwój dyscypliny Elektronika dotyczy:

Opracowania narzędzia o nazwie *DL2HDL*, którego celem jest mapowanie modeli sieci neuronowych do układów FPGA. Wykorzystując narzędzie DL2HDL można dokonać

konwersji wysokopoziomowego opisu architektury sieci neuronowej do niskopoziomowego języka opisu sprzętu.

Przeprowadzenia badań eksperymentalnych, z wykorzystaniem trzech różnych zbiorów danych, oceniających latencję sieci neuronowych mapowanych do układu FPGA za pomocą narzędzia DL2HDL.

Zaprojektowania wraz z implementacją akceleratora sprzętowego układu FPGA do następujących metod uczenia maszynowego: algorytm k-NN, metryka podobieństwa kosinusowego.

Przeprowadzenia badań eksperymentalnych dotyczących oceny dokładności metryki podobieństwa kosinusowego oraz jakości klasyfikacji algorytmu k-NN akcelerowanych w układzie FPGA.

Przeprowadzenia badań eksperymentalnych dotyczących oceny zużycia energii w zadaniu obliczania metryki podobieństwa kosinusowego akcelerowanej w układzie FPGA.

Wymienione osiągnięcia świadczą o umiejętności samodzielnego prowadzenia pracy naukowej przez magistra inżyniera Michała Karwatowskiego oraz o oryginalnym rozwiązaniu problemu naukowego.

Uwagi krytyczne

Tematyka rozprawy dotyczy akceleracji sprzętowej algorytmów uczenia maszynowego w układach FPGA. Niestety Autor w rozprawie nie wspominał o dedykowanych układach do uczenia maszynowego (machine learning processing unit) czy też układach dedykowanych do uczenia głębokiego (deep learning processing unit).

W punkcie przedstawiającym wyniki badań eksperymentalnych (4.1.2.3) Autor dysertacji wykorzystuje pojęcie korelacji. W pracy brakuje wzoru przedstawiającego wyliczenie wartości współczynnika korelacji (używanej przez Autora), którego wartość znajdują się na osi rzędnych rysunku 4.4. Tabela 4.1 przedstawia wyniki dla wybranych szerokości bitów, ale bez uzasadnienia wyboru szerokości bitów równej 8 oraz 16.

W punkcie 4.2 Autor stwierdza, że wartość parametru k algorytmu k-NN nie ma wpływu na zużycie zasobów PPGA. Należy jednak zauważyć, że algorytm k-NN jest algorytmem, który nie tworzy modelu reprezentacji wiedzy, lecz przeszukuje zbiór uczący. Jedną z czynności, która uzależniona jest od wartości parametru k jest sortowanie wyników dotyczących odległości obiektu klasyfikowanego od wszystkich obiektów znajdujących się w zbiorze uczącym. Sortowanie wyników pozwala na znalezienie k najbliższych sąsiadów. W tym

kontekście Autor powinien szerzej omówić wykorzystanie akceleracji sprzętowej algorytmu k-NN.

Wzór 5.10 przedstawia miarę będącą różnicą wartości błędu średniokwadratowego po przycinaniu i kwantyzacji oraz wartości błędu średniokwadratowego bez przycinania i kwantyzacji pomnożoną przez wartość 100. Autor w tekście rozprawy wspominał, że jest to „specjalna miara”. Na czym polega „specjalność” miary wyrażonej worem 5.10? Wynik przedstawione w kolumnie o nazwie „Quality score” (tabele 5.3, 5.4 oraz 5.6) można podzielić przez wartość 100, wówczas będą reprezentowały wspomnianą różnicę wartości błędu średniokwadratowego.

Protokół eksperymentalny stosowany w punkcie 5.3 nie jest spójny dla przeprowadzonych eksperymentów dotyczących różnych zbiorów danych. W każdym eksperymencie wartości rzadkości oraz kwantyzacji, będące parametrami eksperymentów, są różne. Co jest powodem zastosowania różnych wartości parametrów w każdym eksperymencie?

Błędy językowe, kompozycji tekstu oraz typograficzne

W punkcie 2.6.3.5 zostały przedstawione normy, czyli funkcje określające długość wektora. Nazwa punktu sugeruje przedstawienie metryk odległości.

Algorytm obliczania miary podobieństwa kosinusowego (str. 43) został przedstawiony za pomocą opisu słownego. Znacznie czytelniejsze jest wykorzystanie do prezentacji algorytmu schematu blokowego czy też pseudokodu.

W punkcie 2.6.3 przedstawiono metryki jakim są czułość (*recall*) oraz precyzja (*precision*) z wykorzystaniem oznaczeń zawartych z macierzy pomyłek (Tab. 2.3). Metryka dokładność (*accuracy*) nie została przedstawiona z wykorzystaniem wspomnianych oznaczeń, co jest swego rodzaju niekonsekwencją przedstawiania metryk.

Punkt 5.2.3.2 nie zawiera właściwego tekstu. Znajduje się w nim notatka Autora jaka treść powinna się znaleźć we wskazanym punkcie.

W tekście rozprawy można znaleźć wiele błędów edycyjnych, którymi są:

Tzw. wiszące znaki, np. na stronie 57 dwa razy „w”.

Str. 12: „stanowi ona znaczą część”

Str. 20: „FPGA ą najlepszym”

Str. 24: „bitów na dla obecnie”

Str. 25: brak wzoru po tekście „TF-IDF to”

Str. 29: wielkość nawiasów w równaniu 2.10

- Str. 30: „wycięty i n tej podstawie”
- Str. 35: słowo „przyjazny” przed równaniem 2.21 nie jest prawidłowym sformułowaniem
- Str. 60: „Korelacja zachowana spadki”
- Str. 68: „FPGA. uzyskane” – słowo uzyskane powinno być pisane dużymi literami
- Str. 74: duże litery w nagłówku punktu 4.3.4
- Str. 79: [56] [57] [6] – kolejność pozycji literatury
- Str. 98: „za darmo”
- Str. 100: FPGA jest
- Str. 112: „obliczany jest wzór 2.18” – wzór nie jest obliczany, oblicza się zgodnie ze wzorem

Podsumowanie

Reasumując stwierdzam, iż mgr inż. Michał Karwatowski posiada ogólną wiedzę teoretyczną w zakresie projektowania oraz implementacji akceleracji sprzętowej w układach FPGA, która mieści się w dyscyplinie Elektronika. Dodatkowo posiada ogólną wiedzę teoretyczną dotyczącą metod uczenia maszynowego, co świadczy o interdyscyplinarności wiedzy Doktoranta.

Lektura dysertacji pozwala stwierdzić, że Autor zaprezentował na jej łamach umiejętność samodzielnego prowadzenia pracy naukowej w dyscyplinie Elektronika, odpowiadającej dziedzinie nauk inżyniersko-technicznych, dyscyplinie Automatyka, Elektronika i Elektrotechnika wg klasyfikacji określonej w stosownym rozporządzeniu MNiSW z dnia 20 września 2018 r. w sprawie dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych. Osiągnięte przez mgr inż. Michała Karwatowskiego rezultaty mogą mieć zastosowanie w sferze gospodarczej.

Wobec powyższego, recenzowana praca spełnia wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami). Konkludując, wnoszę o przyjęcie rozprawy oraz dopuszczenie mgr inż. Michała Karwatowskiego do publicznej obrony.

R. Purodek