

Gliwice, 2.05.2023

Prof. dr hab. inż. Marcin Woźniak  
Wydział Matematyki Stosowanej  
Politechnika Śląska

Sz. P.  
Prof. dr hab. inż. Robert Nowicki  
Przewodniczący  
Rady Dyscypliny Naukowej  
Informatyka Techniczna i Telekomunikacja  
Politechnika Częstochowska

**Recenzja rozprawy doktorskiej mgr inż. Michała Wróbla pt.: „Inteligentne metody rozpoznawania pisma odręcznego z wykorzystaniem ekstrakcji kresek” opracowanej pod kierunkiem dr hab. inż. Janusza Starczewskiego, Prof. P.Cz.**

W odpowiedzi na pismo R-WIMil-510-18/19 z dnia 13 lutego 2023 przedkładam niniejszą recenzję.

Przedstawiona do recenzji praca jest 117 stronicowym opracowaniem podzielonym na siedem tematycznych rozdziałów. Rozprawa rozpoczyna się wprowadzeniem w tematykę podjętych badań, a zakończona jest bibliografią, w której doktorant przedstawiła 54 źródła tematyczne wykorzystane podczas pracy badawczej. We wstępie doktorant przedstawia motywację swojej pracy naukowej oraz zastosowanie modelu kresek do ekstrakcji liter jako rozwiązanie dla problemu prawidłowej separacji poszczególnych znaków w piśmie odręcznym.

Rozprawa dotyczy rozpoznawania pisma odręcznego ze statycznego obrazu cyfrowego (rozpoznawanie offline) z zastosowaniem ekstrakcji kresek, czyli elementarnych fragmentów liter, możliwych do utworzenia pojedynczym ruchem ręki. Przyjętym celem pracy było opracowanie autorskiego algorytmu odczytującego tekst z obrazu zawierającego pojedyncze

słowo, przy czym jego szczególnym zastosowaniem ma być odczytywanie słów pochodzących z dokumentów historycznych. W związku z tym przyjęto trzy dodatkowe założenia: algorytm ma nie korzystać ze zdefiniowanego słownika, nie potrzebować dużego zbioru danych uczących, a wyniki pośrednie mają być proste do prześledzenia przez analityka (model white box).

Praca składa się z siedmiu rozdziałów, bibliografii, spisu rysunków i spisu tablic. Przeglądowi literatury poświęcony jest rozdział drugi, pierwsze podrozdziały rozdziałów trzeciego i czwartego, a także znaczne fragmenty wstępu. Pozostała część rozprawy opisuje pracę własną oraz przeprowadzone badania i opracowany algorytm.

Pierwszym rozdziałem jest wstęp. Na początku krótko omówiono w nim główne zagadnienia z obszaru automatycznego rozpoznawania pisma, takie jak HWR, OCR, rozpoznawanie online i offline. Następnie zarysowano kierunki współcześnie prowadzonych badań oraz wprowadzono pojęcie kreski. Po tym ogólnym wprowadzeniu przedstawiono założenia rozprawy, a następnie wyszczególniono główne etapy opracowanego algorytmu. Zdefiniowano także na potrzeby rozprawy pojęcia „hipoteza” oraz „wariant”.

Rozdział drugi stanowi przegląd literatury. W pierwszym podrozdziale wypunktowano główne problemy związane z rozpoznawaniem pisma odręcznego. Wspomniano także o wybranych metodach stosowanych do przewyżniania tych trudności. Kolejne podrozdziały omawiają podstawowe pojęcia z różnych dziedzin informatyki, które były wykorzystywane w dalszej części pracy. Podrozdziały te dotyczą kolejno: teorii grafów, przetwarzania obrazów, algorytmów heurystycznych (ze szczególnym uwzględnieniem zachłannych), logiki rozmytej i sieci neuronowych.

Trzeci rozdział poświęcony jest ekstrakcji kreski. W pierwszym podrozdziale zawiera on krótki przegląd metod spotykanych w literaturze, a w kolejnych podrozdziałach przedstawia zaproponowany przez Doktoranta algorytm przetwarzania obrazu wejściowego w zbiór wyekstrahowanych kreski. Algorytm ten zawiera następujące części składowe: przetwarzanie wstępne obrazu, wypełnienie obszaru liter kołami wpisanymi (wraz z selekcją tych kół), wyznaczenie połączeń (wiązań) między środkami kół, przekształcenie zbioru wiązań w zbiór kreski rozumianych jako ciągi środków kół i podział zbyt zawitych kreski. Metody te oparte są głównie na technikach przetwarzania obrazu binarnego oraz na algorytmach zachłannych. Ostatni podrozdział zawiera szczegółowe rozważania na temat tych etapów algorytmu, których opracowanie wymagało dokładniejszych badań.

W czwartym rozdziale skupiono się na rozpoznawaniu wzorców pojedynczych liter (lub krótkich zbitek) w zbiorze wyekstrahowanych kreski. Układ tego rozdziału jest zbliżony do poprzedniego: pierwszy podrozdział zawiera przegląd technik z literatury, a kolejne podrozdziały opis autorskich rozwiązań. Etap algorytmu opisany w tym rozdziale obejmuje: porównywanie pojedynczych kreski (pochodzącej z obrazu wejściowego i ze wzorca), porównywanie grup kreski oraz tworzenie zbioru hipotez rozumianych jako przypuszczenie, że dana grupa kreski z rozpoznawanego obrazu pasuje do grupy kreski z wzorca.

Zaproponowane algorytmy wykorzystują pewne aspekty teorii zbiorów rozmytych. Opracowany algorytm ma kilka wariantów, które można przełączać między sobą w trakcie badań. Oprócz samego opisu algorytmu rozdział zawiera także informacje na temat badań, których rezultatem było przyjęcie danych rozwiązań. W ostatnim podrozdziale zasygnalizowano możliwość zastąpienia tego etapu algorytmu rekurencyjną siecią neuronową.

Rozdział piąty opisuje końcowe etapy zaproponowanego algorytmu. Najpierw zbiór hipotez przekształcany jest w zbiór wariantów tekstu. Następnie ze zbioru wariantów wybierany jest napis będący rezultatem zwracany przez cały algorytm. Decyzja na temat wyboru wariantu podejmowana jest na podstawie podobieństw poszczególnych grup kresek do wzorców oraz danych statystycznych dotyczących występowania po sobie poszczególnych par kolejnych liter w danym języku.

W rozdziale szóstym opisano przeprowadzone badania eksperymentalne. Pierwszy podrozdział zawiera informacje na temat implementacji algorytmu, takie jak język programowania, zastosowane biblioteki i ogólna organizacja kodu źródłowego. Drugi podrozdział zawiera opis badania polegającego na rozpoznawaniu wybranych, przygotowanych samodzielnie obrazów z pojedynczymi słowami przy zastosowaniu niewielkiego zbioru wzorców. W ramach tego etapu badań przyglądano się listom wygenerowanych wariantów rozpoznanego tekstu dla różnych wariantów algorytmu przedstawionego w rozdziale czwartym. W trzecim podrozdziale opisano kolejny etap badań, w którym algorytm testowany był na trzech zbiorach obrazów (zawierających od 28 do 38 napisów) pochodzących z bazy IAM. Liczba wariantów algorytmu z poprzedniego kroku została tutaj zawężona do dwóch najbardziej obiecujących. Podczas badań opisanych w czwartym podrozdziale autorski algorytm porównywany był z systemem AttentionHTR opartym na głębokich sieciach neuronowych.

Ostatni, siódmy rozdział przedstawia podsumowanie i wnioski. Skomentowano tam wyniki zaprezentowane w rozdziale szóstym i wypunktowano największe, zdaniem Doktoranta, osiągnięcia uzyskane w trakcie pracy nad rozprawą. Przedstawiono także kilka pomysłów na ewentualne dalsze badania.

Zgodnie z treścią podsumowania, w ramach pracy udało się uzyskać następujące osiągnięcia. Przy uczeniu systemu za pomocą bardzo małego zbioru danych treningowych, potrafi on zwrócić lepsze wyniki, niż użyty do porównania system AttentionHTR. Wyniki pośrednie między poszczególnymi etapami algorytmu są możliwe do interpretacji przez użytkownika, co spełnia wymagania modelu white box. Etap ekstrakcji kresek zwraca dobrze wyglądające wyniki i może potencjalnie znaleźć zastosowanie także w innych obszarach, niż rozpoznawanie pisma odręcznego. Etap selekcji wariantów z wykorzystywaniem analizy sąsiednich liter stanowi kompromis między metodami słownikowymi i bezsłownikowymi: umożliwia rozpoznawanie nazw własnych przy jednoczesnym odrzucaniu ciągów znaków, które nie mają sensu w danym języku. Opracowany algorytm jest zaprojektowany tak, że

jego poszczególne etapy składowe są łatwe do ewentualnej wymiany, co może stanowić punkt wyjścia do dalszych badań.

### **Ocena merytoryczna przedstawionej rozprawy doktorskiej:**

W trakcie przeprowadzonych badań doktorant zbudował modele oraz przedstawił analizę uzyskanych wyników wraz z wnioskami płynącymi z badań naukowych. Po przeczytaniu przedstawionej rozprawy doktorskiej nasuwają się następujące pytania i sugestie:

- i. W podrozdziale 4.9 doktorant zaproponował zastąpienie etapu tworzenia wariantów tekstu (w pracy jest to realizowane za pomocą drzewa decyzyjnego) za pomocą rekurencyjnej sieci neuronowej. Uważam, że należy szczegółowo rozwinąć, jak połączyć taką sieć z pozostałymi etapami algorytmu.
- ii. Czy postać wielomianowa modelowych kresek nadaje się jako wejście sieci? Jaka postać ewentualnie mogłaby nadawać się lepiej do reprezentacji danych?
- iii. Jak przekształcić dane na wyjściu sieci do zbioru wariantów, który mógłby stanowić wejście dla kolejnego etapu algorytmu?
- iv. Dostępne zbiory treningowe zawierają ciąg znaków oczekiwany na wyjściu, ale bez informacji, który dokładnie fragment obrazu wejściowego przedstawia odpowiadającą mu literę. Zatem należy wyjaśnić jak przygotować dane do uczenia sieci.
- v. Etap wyboru wariantu tekstu (podrozdział 5.3) wymaga danych statystycznych dotyczących częstotliwości występowania liter w danym języku. Doktorant przedstawia, że użyto dane dla języka angielskiego z publikacji. Czy można użyć podobnych danych dla algorytmu w innym języku?
- vi. Pomimo, że pomysł tworzenia hipotez z ciągów kresek w podejściu bezsłownikowym wydaje się być skuteczny, system nie działa w pełni poprawnie. Niektóre słowa nie pojawiają się w ramach zestawu hipotez, stąd trudno o wybór właściwej operacji w kolejnym etapie algorytmu. Na którym etapie i z jakiej przyczyny zdaniem autora gubione są poprawne hipotezy? W jaki sposób autor mógłby to poprawić?
- vii. We wstępie pracy (str. 7) doktorant poruszył problem słabej odtwarzalności wyników prac. Jak poprawić ew. odtwarzalność danych?

### **Ocena formy, taksonomii, języka i edycji pracy:**

Przedstawiona rozprawa doktorska jest napisana poprawnym językiem. Edycja i forma dokumentu są poprawne.

### **Ocena przedstawionego dorobku naukowego doktoranta:**

W przedstawionej pracy doktorant zawarł referencje do czterech prac naukowych, które stanowią wkład w rozwój dyscypliny naukowej. Są to artykuły opublikowane w czasopiśmie tematycznych dla przedstawionego zagadnienia oraz wolumenach konferencyjnych.

Przedstawiony dorobek naukowy doktoranta pokazuje zaangażowanie w prace naukowe, które zostały uznane w środowisku naukowym. Moim zdaniem zaprezentowany dorobek pozwala stwierdzić obeznanie doktoranta z dyscypliną prowadzenia badań naukowych, analizą wyników i podejmowaniem wniosków z prowadzonych badań naukowych.

### **Podsumowanie:**

Przedstawiona do recenzji rozprawa doktorska porusza temat zastosowania modeli rekonstrukcji znaków pisma odręcznego dla uczenia maszynowego. Uczenie maszynowe, a zwłaszcza powiązanie go z różnymi modelami ekstrakcji jest tematem wielu aktualnych opracowań naukowych. Przedstawione rozwiązanie ma możliwość dalszego rozwoju, również w kierunku wdrożenia.

Zaproponowany w rozprawie doktorskiej model pokazuje wkład doktoranta w badania. Przedstawione idea pozwalają oceniać obiekty i na takiej podstawie budować model decyzyjny. Jest to temat ważny z punktu widzenia rozwoju informatyki.

Na podstawie przedstawionej powyżej oceny stwierdzam, iż rozprawa doktorska mgr inż. Michała Wróbla pt.: „Inteligentne metody rozpoznawania pisma odręcznego z wykorzystaniem ekstrakcji kresek” spełnia warunki określone w art. 13 ust. 1 Ustawy o stopniach i tytule naukowym. **Wnoszę o przyjęcie w/w rozprawy doktorskiej i dopuszczenie jej do publicznej obrony w dyscyplinie informatyka.**