

dr hab. inż. Maciej Jaworski, prof. PK
Katedra Informatyki
Wydział Informatyki i Telekomunikacji
Politechnika Krakowska
Warszawska 24, 31-155 Kraków

Kraków, 27.04.2023

Recenzja rozprawy doktorskiej mgr inż. Michała Wróbla pt. „Inteligentne metody rozpoznawania pisma odręcznego z wykorzystaniem ekstrakcji kresek”

Niniejsza recenzja została przygotowana na prośbę wystosowaną w piśmie prof. dra hab. inż. Roberta Nowickiego, Kierownika dyscypliny naukowej informatyka techniczna i telekomunikacja Wydziału Inżynierii Mechanicznej i Informatyki Politechniki Częstochowskiej, z dnia 13.02.2023, nr pisma R-WIMil-510-18/19.

Problematyka rozprawy

Tematem recenzowanej rozprawy doktorskiej jest problem rozpoznawania pisma odręcznego. Zagadnienie to wpisuje się w tematykę optycznego rozpoznawania znaków. W odróżnieniu od rozpoznawania tekstu drukarskiego, rozpoznawanie pisma odręcznego jest zadaniem znacznie trudniejszym. Wymaga przeprowadzenia kilku dodatkowych operacji na danych, m.in. separacji poszczególnych liter i oddzielenia poszczególnych wierszy tekstu. Rozpoznawanie pisma może być wykonywane zarówno w trybie online jak i offline. W niniejszej pracy skupiono się na tym drugim podejściu, opracowując algorytmy rozpoznawania pisma statycznego. Ostatnimi czasy, jak to ma miejsce w większości zagadnień w których wykorzystywane są algorytmy sztucznej inteligencji, także w przypadku rozpoznawania pisma odręcznego prym wiodą metody bazujące na sieciach neuronowych. Niemniej jednak w celu ich poprawnego wytrenowania wymagane są zwykle duże zbiory uczące. Pozyskanie takiego zbioru może okazać się czasochłonne i kosztowne. Dlatego Doktorant przyjął, jako jedno z głównych założeń przy implementacji swojego algorytmu, że system nie powinien wymagać zbyt wielu danych uczących do właściwego funkcjonowania. Drugim założeniem jest działanie bezsłownikowe, aby algorytm mógł rozpoznawać także nazwy własne i słowa archaiczne, nie pojawiające się zwykle w powszechnie używanych zbiorach danych. Kolejnymi założeniami są interpretowalność modeli oraz ich blokowa struktura, tak aby każdy fragment algorytmu mógł funkcjonować oddzielnie i w razie potrzeby zostać podmieniony przez inną metodę, jeśli pierwotne pomysły nie sprawdzałyby się w praktyce. W celu sprostania tym założeniom, opracowany algorytm bazuje na

metodach matematyczno-statystycznych. Zaproponowana metoda została porównana z opisaną w literaturze siecią neuronową AttentionHTR, wypadając w tym porównaniu korzystnie.

Zawartość pracy oraz najważniejsze wyniki

Zaproponowany w pracy doktorskiej algorytm rozpoznawania pisma odręcznego zbudowany jest z trzech głównych modułów: ekstrakcja kresek, rozpoznawanie wzorców oraz wybór wariantów tekstu. Pierwszy z modułów jest najbardziej rozbudowany i składa się z kilku następujących po sobie etapów przetwarzania danych. Wszystkie zastosowane w pracy metody mają solidne podstawy matematyczne. Należy tutaj podkreślić biegłość i swobodę Doktoranta w posługiwaniu się formalizmem matematycznym. Wszystkie kroki algorytmu zostały bardzo ściśle i jednoznacznie opisane w pracy. Czytelnik zainteresowany ewentualnym powtórzeniem eksperymentów z pewnością bez problemu będzie w stanie zaimplementować przedstawione algorytmy. Doktorant wykazał się znajomością szerokiego spektrum przeróżnych metod matematycznych i informatycznych, które zastosował przy projektowaniu swoich rozwiązań. Po wstępnym przetwarzaniu, polegającym na binaryzacji i szkieletyzacji obrazu z tekstem, następuje proces wypełniania wnętrza liter kołami. Następnie, spośród wszystkich kół, wybierane są tylko te niezbędne do odtworzenia kształtu litery. W tym celu zastosowany został z powodzeniem algorytm zachłanny. Z kolei przy łączeniu wyselekcjonowanych kół w więzy, a także przy wyodrębnianiu finalnych kresek, zaproponowane metody bazują na teorii grafów. Finalnie, kształty kresek są przybliżane wielomianami trzeciego stopnia (zarówno współrzędne x jak i y , a więc każda kreska scharakteryzowana jest ośmioma liczbami). Wszystkie etapy algorytmu ekstrakcji kresek zawierają kilka istotnych parametrów, których wartości należy w jakiś sposób ustalić. W tym celu Doktorant zastosował heurystyczną metodę zwaną symulowanym wyżarzaniem. Cały algorytm ekstrakcji kresek jest zdecydowanie najlepiej działającym elementem proponowanego systemu i w razie potrzeby może z powodzeniem być stosowany jako oddzielne narzędzie.

Kolejnym elementem algorytmu rozpoznawania pisma odręcznego jest rozpoznawanie wzorców, a więc wyszukiwanie podobieństw wyekstrahowanych kresek do kresek wzorcowych, reprezentujących litery lub zbitki dwóch liter. W tym celu zaproponowano dwa podejścia: wielomianowe i interwałowe. W pierwszym z nich porównywane są współczynniki wielomianów reprezentujących kreski (oczywiście bez wyrazów wolnych, informujących wyłącznie o położeniu danej kreski na obrazie). W podejściu interwałowym natomiast Doktorant wykazał się znajomością kolejnego istotnego zagadnienia matematycznego, a mianowicie logiki rozmytej i zbiorów rozmytych. W tym podejściu zakłada się, że dysponujemy kilkoma wersjami wzorców kresek dla danej litery. Pozwala to skonstruować reprezentację kreski w formie zbiorów rozmytych, zamiast wartości ostrych. W obu podejściach nadmiarowe hipotezy dotyczące przypasowania kresek do wzorców były usuwane ponownie z wykorzystaniem odpowiedniego algorytmu zachłannego.

Trzecim etapem zaprezentowanego systemu jest tworzenie i wybór wariantów tekstu. W pracy opisano możliwość zastosowania do tego celu algorytmu zachłanego. Algorytm taki nie daje jednak gwarancji znalezienia optymalnego rozwiązania, dlatego zdecydowano się na zastosowanie metody wykorzystującej drzewa decyzyjne. Do oceny jakości utworzonych wariantów zastosowano wskaźnik uwzględniający dwie rzeczy: zgodność poszczególnych liter ze wzorcami, a także sensowność występowania poszczególnych sekwencji liter w danym języku. Algorytm do prawidłowego funkcjonowania wymaga więc znajomości pewnych statystyk dotyczących częstości pojawiania się ciągów znaków.

Zaproponowany system rozpoznawania pisma został przetestowany zarówno na próbkach tekstów przygotowanych przez Doktoranta jak i powszechnie dostępnym zbiorze IAM. Podobieństwo tekstu wykrytego do tekstu prawdziwego zostało zmierzone za pomocą wspomnianego wcześniej wskaźnika, a także przez obliczenie odległości Levenshteina. Algorytm zostały porównany z siecią neuronową AttentionHTR. Wyjątkowo na potrzeby tego porównania wykorzystano miarę będącą procentową skutecznością algorytmów. W przypadku uczenia sieci zbiorem danych IAM, porównanie wypada na korzyść autorskiego algorytmu Doktoranta. W pracy dokonano także porównania z siecią AttentionHTR nauczoną na innym, znacznie większym zbiorze danych Imgur5K. W tym przypadku, czego należało się spodziewać, znacznie efektywniejsza okazuje się być sieć neuronowa.

Uwagi krytyczne i zagadnienia do dyskusji

Jak wspomniano wcześniej, rozprawa doktorska ogólnie jest napisana bardzo starannie i zrozumiale, a wszystkie metody zostały opisane przy użyciu bardzo precyzyjnego formalizmu matematycznego. Właściwie jedyną kwestią, która może budzić pewien niedosyt czytelnika, jest jej część eksperymentalna. Prawdą jest, że uzyskane wyniki są porównywalne, a nawet lepsze, niż wyniki uzyskane za pomocą przeznaczonej do tego zadania sieci neuronowej (przy sprawiedliwym uczeniu sieci na tym samym małym zbiorze danych). Niemniej jednak, bezwzględny poziom uzyskiwanych dokładności, rzędu 20%-30%, wydaje się być bardzo niski aby z powodzeniem móc zastosować opracowany system w praktyce. Przedstawione w tabelach przykłady wyrazów prawdziwych oraz rozpoznanych również nie napawają optymizmem. Wydaje mi się, że przyczyn tego stanu rzeczy należy szukać w modułach odpowiadających za porównywanie kresek z wzorcami oraz wybór wariantu tekstu. Część algorytmu odpowiadająca za ekstrakcję kresek działa bowiem perfekcyjnie i moim zdaniem stanowi największą wartość recenzowanej pracy. Należy także odnotować, że Doktorant jest świadomy wad wspomnianych wcześniej dwóch modułów algorytmu, proponując w podsumowaniu kilka kierunków badań i modyfikacji, których zastosowanie mogłyby w przyszłości poprawić efektywność całego algorytmu.

Czytając rozprawę doktorską nasunęły mi się jeszcze dwa pytania/zagadnienia, które pozwolę sobie sformułować poniżej. Nie mają one charakteru krytycznego, mogą ewentualnie stanowić inspirację do dalszych badań:

- Głównym założeniem pracy było skonstruowanie algorytmu nie wymagającego uczenia na dużym zbiorze treningowym, co wyklucza zastosowanie sieci neuronowych. Ostatnimi czasy jednak pojawia się coraz więcej gotowych, nauczonych sieci, przeznaczonych do różnych zadań. Z pewnością można więc znaleźć gotową sieć neuronową do rozpoznawania pisma odręcznego, a następnie wykonać na niej procedurę uczenia transferowego, dostrajając ją do naszego, niewielkiego, zbioru danych. Czy Doktorant rozważał takie podejście? I czy należałoby je traktować jako rozwiązanie wymagające czy niewymagające dużego zbioru danych uczących?

- W reprezentacji kresek poprzez parametry dopasowanych wielomianów Doktorant słusznie wykluczył wyrazy wolne, gdyż odpowiadają one jedynie za położenie przestrzenne danego wzorca. Czy jest możliwe zmodyfikowanie reprezentacji tak, aby była ona odporna na ewentualne rotacje lub różnice w nachyleniu liter, a także na różnice w ich skali?

Podsumowanie i konkluzja recenzji

Podsumowując wszystkie powyższe uwagi, należy stwierdzić, iż rozprawa doktorska Pana mgr inż. Michała Wróbla stanowi istotny wkład w rozwój tematyki inteligentnego rozpoznawania pisma odręcznego. W pracy zaprezentowano nowatorski algorytm, którego głównymi zaletami są brak wymagania dostępu do dużego zbioru danych uczących a także jego bezsłownikowość. Najbardziej wartościowym elementem zaprezentowanego algorytmu wydaje się być moduł odpowiadający za ekstrakcję kresek. Algorytm jest bardzo złożony, a jego opracowanie wymagało od Doktoranta znajomości bardzo szerokiego spektrum różnych zagadnień matematyczno-informatycznych. Warto także podkreślić, że treści przedstawione w rozprawie zostały także opublikowane w trzech artykułach konferencyjnych oraz w jednym artykule w recenzowanym czasopiśmie naukowym *Journal of Artificial Intelligence and Soft Computing Research*, któremu Ministerstwo Edukacji i Nauki przyznaje 140 punktów. We wszystkich czterech pracach mgr inż. Michał Wróbel jest pierwszym autorem.

Stwierdzam, że rozprawa doktorska mgr inż. Michała Wróbla spełnia we wszystkich możliwych aspektach wymagania nakładane przez Ustawę o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki z dn. 14.03.2003 z późn. zm. Wnoszę o dopuszczenie rozprawy do publicznej obrony.

Maciej Jaworski