

RECENZJA

osiągnięcia naukowego i istotnej aktywności naukowej dr. inż. Wojciecha Kwedlo w postępowaniu habilitacyjnym w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja.

1. Podstawa opracowania recenzji

Niniejszą recenzję wykonano na zlecenie Politechniki Białostockiej na podstawie przekazanej dokumentacji, która zawierała: autoreferat, wykaz osiągnięć naukowych, kopie publikacji stanowiących osiągnięcie, kopię oświadczenia współautora jednej z publikacji stanowiącej osiągnięcie, aktywność naukową w więcej niż jednej uczelni i inne informacje. Recenzję wykonano zgodnie z art. 219 (Dz. U. z 2021 r.) Prawo o szkolnictwie wyższym i nauce oraz na podstawie zaleceń Rady Doskonałości Naukowej zawartych w poradniku Postępowania dotyczące nadawania stopnia doktora habilitowanego (ostatnia aktualizacja: 9 października 2023 r.).

2. Ogólna charakterystyka Habilitanta

2.1. Wykształcenie

dr inż. Wojciech Kwedlo w 2004 roku uzyskał na Wydziale Informatyki Politechniki Białostockiej stopień naukowy doktora nauk technicznych w dyscyplinie naukowej informatyka. Tytuł rozprawy: „Algorytmy ewolucyjne jako narzędzia pozyskiwania wiedzy z baz danych” (rozprawa obroniona z wyróżnieniem). Studia magisterskie ukończył w 1996 roku również na Wydziale Informatyki Politechniki Białostockiej, kierunek Informatyka.

2.2. Zatrudnienie

Przebieg zatrudnienia dr. inż. Wojciecha Kwedlo jest następujący:

1.10.1996 r. - 30.09.2004, asystent, Zakład Oprogramowania (następnie Katedra Oprogramowania), Instytut Informatyki (następnie Wydział Informatyki), Politechnika Białostocka

1.10.2004 r. - obecnie, adiunkt, Katedra Oprogramowania, Wydział Informatyki, Politechnika Białostocka

3. Ocena osiągnięcia naukowego

3.1. Opis osiągnięcia naukowego

Jako osiągnięcie naukowe Habilitant przedłożył cykl pięciu ściśle powiązanych tematycznie publikacji naukowych pod wspólnym tytułem „**Algorytmy grupowania danych wykorzystujące prototypy grup**”. Do cyklu publikacji Habilitant zaliczył:

- [C1] W. Kwedło. „A clustering method combining differential evolution with the *K*-means algorithm”. W: Pattern Recognition Letters 32.12 (2011), s. 1613–1621. DOI: 10.1016/j.patrec.2011.05.010. IF=1,0341, 100 pkt. MEiN2, 91 cytowań według WoS (Scopus: 119, Google Scholar: 155).
- [C2] W. Kwedło. „A parallel EM algorithm for Gaussian mixture models implemented on a NUMA system using OpenMP”. W: Proceedings of the 22nd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing PDP 2014. IEEE CPS, 2014, s. 292–298. DOI: 10.1109/PDP.2014.77. Konferencja indeksowana w bazie WoS, 20 pkt., 11 cytowań według WoS (Scopus: 12, Google Scholar: 19).
- [C3] W. Kwedło. „A new random approach for initialization of the multiple restart EM algorithm for Gaussian model-based clustering”. W: Pattern Analysis and Applications 18.4 (2015), s. 757–770. DOI: 10.1007/s10044-014-0441-3. IF=1,104, 70 pkt., 12 cytowań według WoS (Scopus: 15, Google Scholar: 22).
- [C4] W. Kwedło i P. J. Czochoński. „A hybrid MPI/OpenMP parallelization of K-means algorithms accelerated using the triangle inequality”. W: IEEE Access 7 (2019), s. 42280–42297. DOI: 10.1109/ACCESS.2019.2907885. IF=3,745, 100 pkt., wkład własny 80%, 17 cytowań według WoS (Scopus: 20, Google Scholar: 27).
- [C5] W. Kwedło. „A hybrid steady-state evolutionary algorithm using random swaps for Gaussian model-based clustering”. W: Expert Systems with Applications 208 (2022), s. 118159. DOI: 10.1016/j.eswa.2022.118159. IF2021=8,665, 140 pkt.

Prace zaliczane do cyklu powstały po uzyskaniu stopnia doktora w latach 2011–2022. Opublikowane zostały w punktowanych czasopismach (4 publikacje) oraz jedna w recenzowanych materiałach konferencji międzynarodowych. Wg. aktualnego wykazu sporządzonego przez Ministerstwo publikacje te mają (miałyby): 1 publikacja za 140 pkt, 2 za 100 pkt, 1 za 70 pkt, 1 za 20 pkt. Sumaryczny wskaźnik IF tych prac wynosi 14,548 i plasuje się na średnim poziomie. Liczbę cytowań wskazano na 131 wg. WoS, 166 wg. Scopus i 223 wg. Google Scholar. Szczególnie praca [C1] ma już 119 cytowań w bazie Scopus, co świadczy o dużym zainteresowaniu środowiska naukowego tą pracą. Także stosunkowo świeża praca [C4] ma już 20 cytowań w Scopus (25 w trakcie pisania recenzji), co także jest dobrym wynikiem.

Liczba prac w cyklu jest stosunkowo niewielka, na podkreślenie zasługuje jednak fakt, że cztery spośród pięciu prac są samodzielne. W przypadku jednej pracy współautorskiej wkład drugiego z autorów wg. jego oświadczenia był niewielki i wynosił 20%. Przedłożony cykl publikacji jest więc wynikiem głównie pracy Habilitanta. Na podstawie bazy Scopus Habilitant sumarycznie opublikował 21 prac, o liczbie cytowań 347, h-index = 10, co można ocenić jako wynik dobry.

3.2. Omówienie i ocena merytoryczna publikacji wchodzących w skład osiągnięcia

Jak już wspominałem tematyka cyklu publikacji jest ściśle powiązana i dotyczy zastosowania algorytmów optymalizacji do problemu grupowania danych oraz zrównoleglenia tych algorytmów. Problem grupowania danych jest ważnym zagadnieniem z zakresu uczenia

nienadzorowanego i jednocześnie NP-trudnym problemem optymalizacji kombinatorycznej. Tematyka prac leży więc na pograniczu maszynowego uczenia i algorytmów optymalizacji, i niewątpliwie znajduje się w dyscyplinie informatyka techniczna i telekomunikacja.

W pracy [C1] Habilitant zaproponował hybrydowy algorytm ewolucyjny dla problemu grupowania danych z funkcją celu zdefiniowaną jako sumaryczny błąd kwadratowy. Konkretnie habilitant połączył ewolucję różnicową z metodą K -średnich i algorytmem Lloyda jako algorytmem lokalnej optymalizacji. Jako operator rekombinacji wykorzystano krzyżowanie dwumianowe. W pracy zaproponowano także nową metodę generowania rozwiązań początkowych. Hybrydowe algorytmy ewolucyjne odniosły sukces w przypadku wielu problemów optymalizacji kombinatorycznej. Także zaproponowany przez habilitanta algorytm DE-KM uzyskał bardzo dobre wyniki w eksperymencie obliczeniowym obejmującym zarówno uproszczone wersje tego algorytmu, jak i znane z literatury heurystyki deterministyczne WKM i GKM. W eksperymencie wykorzystano 13 zbiorów danych. Pewną wadą tej pracy jest brak analizy istotności statystycznej uzyskanych wyników. Jak już wspomniałem, praca ta wzbudziła spore zainteresowanie środowiska naukowego o czym świadczy spora liczba cytowań.

W pracy [C2] Habilitant zaprezentował równoległą implementację algorytmu EM (expectation-maximization) dla maszyn z zachowującą spójność pamięci podręcznych architekturą NUMA, w której czas dostępu do pamięci może się znacząco różnić. Algorytm EM wykorzystuje model mieszanych wielowymiarowych rozkładów normalnych i może być potraktowany jako uogólnienie algorytmu K -średnich. W algorytmie wykorzystano równoległe przetwarzanie wierszy macierzy przez poszczególne wątki. Implementację przetestowano na systemie 64-rdzeniowym wyposażonym w cztery procesory tworzące architekturę NUMA o 8 węzłach wykorzystując dwa różne zbiory danych. Uzyskana wydajność była bliska optymalnej (98%) przy wykorzystaniu do 32 rdzeni, następnie malała do 82% ze względu na pewne ograniczenia użytych procesorów.

W pracy [C3] Habilitant zaproponował nową metodę inicjalizacji losowej przeznaczoną dla algorytmu EM z wielokrotnym startem (MREM). Ponieważ algorytm EM jest algorytmem lokalnego przeszukiwania, a ponadto może być w pewnych warunkach rozbieżny, często rozszerza się go poprzez wykorzystanie różnych losowych punktów startowych. Zaproponowany algorytm rnd-maxmin jest motywowany algorytmem inicjalizacji zaproponowanym w pracy [C1] dla metody K -średnich. Zaproponowana metoda przyjmuje równe prawdopodobieństwa a priori i losowe macierze kowariancji. Wektor oczekiwany nowego komponentu jest wybierany tak, aby zmaksymalizować minimalny kwadrat odległości Mahalanobisa od już zainicjalizowanych komponentów. Metodę rnd-maxmin porównano eksperymentalnie do 3 innych metod inicjalizacji. W eksperymencie wykorzystano zarówno dane rzeczywiste, jak i dane wygenerowane generatorem MixSim pozwalającym kontrolować pewne cechy grupowanych danych, np. współczynnik nakładania się komponentów. Zaproponowana metoda uzyskała statystycznie istotnie lepsze wyniki dla zbiorów danych o bardzo wysokiej separacji grup. Dla zbiorów danych o niskiej separacji wszystkie testowane metody uzyskały podobne wyniki. Także dla danych rzeczywistych metoda rnd-maxmin okazała się lepsza na pewnych zbiorach i nie gorsza na innych.

W pracy [C4] Habilitant wraz ze współautorem zaproponowali hierarchiczną metodę zrównoleglenia algorytmów grupowania K -średnich przyspieszonych z wykorzystaniem nierówności trójkąta, która pozwala na zmniejszenie złożoności obliczeniowej pojedynczej iteracji. W pracy wykorzystano cztery znane z literatury algorytmy przyspieszające działanie algorytmu K -średnich, tj. algorytmy Drake'a, Elkana, Annulus i Yinyang. Celem pracy było zrównoleglenie tych algorytmów dla współczesnych klastrów obliczeniowych składających się z węzłów obliczeniowych połączonych siecią. Konkretnie wykorzystano klaster Tryton z centrum TASK. W zrównolegleniu algorytmów wykorzystano własność kroku przypisania, pozwalającą na odnalezienie najbliższego centroidu niezależnie dla poszczególnych wektorów. W eksperymencie przebadano efektywność różnych strategii zrównoleglenia i uzyskano wydajność powyżej 80%. Zaproponowane podejście porównano też z równoległymi implementacjami dostępnymi w dwóch publicznie dostępnych bibliotekach wykazując przewagę zaproponowanej implementacji. Ta stosunkowo świeża praca (2019) ma już obiecującą liczbę cytowań - 17 według WoS.

Zastosowany w pracy [C3] algorytm EM z wielokrotnym startem (MREM) jest bardzo prostym, lecz niekoniecznie najbardziej skutecznym rozszerzeniem lokalnego algorytmu EM. Dlatego w pracy [C5] habilitant połączył algorytm EM z optymalizacją ewolucyjną w hybrydowym algorytmie ewolucyjnym EARSEM (ang. evolutionary algorithm with random swaps and expectation maximization). Jest to algorytm z ustalonym stanem (ang. steady state), w którym wykorzystano autorski operator rekombinacji. Rozwiązanie uzyskane w wyniku rekombinacji jest punktem startowym dla algorytmu EM. Zaproponowany algorytm porównano z algorytmem wielokrotnego startu MREM, algorytmem RSEM oraz innym hybrydowym algorytmem ewolucyjnym ESM-EM (evolutionary split and merge with expectation maximization). Każdy z testowanych algorytmów wykonywał tę samą liczbę iteracji algorytmu EM. W eksperymencie z wykorzystaniem trzydziestu zbiorów danych zaproponowany algorytm uzyskał najczęściej (choć nie zawsze) najlepsze wyniki. W niektórych wypadkach zaobserwowano też różny ranking wyników w zależności od dostępnego budżetu obliczeniowego.

Jaki widać z powyższego opisu, tematyka cyklu publikacji jest bardzo spójna, a przedstawione wyniki można z pewnością uznać za znaczące. Habilitant koncentrował się na poprawie jakości i efektywności algorytmów grupowania wykorzystując zarówno techniki algorytmiczne, jak i obliczenia równoległe na współczesnych architekturach obliczeniowych.

Niezależnie od ogólnie pozytywnej oceny chciałbym jednak zwrócić uwagę na pewne słabsze strony.

Tylko jedna praca [C5] została opublikowana w czasopiśmie na najwyższym poziomie - Expert Systems with Applications (140 pkt). Pozostałe czasopisma i konferencje można uznać za dobre lub nawet przeciętne. Nie chodzi tu jedynie o liczenie punktów ale, po pierwsze, publikacje w słabszych czasopismach są z reguły słabiej zauważane przez środowisko naukowe, po drugie, najlepsze czasopisma i konferencje poddają nadesłane artykuły wnikliwej recenzji przez najwyższej klasy ekspertów co zwiększa pewność co do poprawności prezentowanych wyników. W szczególności sądzę, że ponieważ prace Habilitanta dotyczą także ciekawych zastosowań algorytmów ewolucyjnych, można było podjąć próbę publikacji

części z nich w dobrych czasopismach i konferencjach ze środowiska algorytmów ewolucyjnych (np. IEEE TEVC, GECCO, PPSN).

Oceniany cykl publikacji powstał w dość długim okresie, od 2011 roku. Wskazuje to, że choć prace Habilitanta postępowały systematycznie, to jednak w niezbyt szybkim tempie.

Podsumowując ten punkt, biorąc pod uwagę ocenę merytoryczną oceniam pozytywnie osiągnięcie naukowe Habilitanta zgłoszone jako cykl publikacji. Stwierdzam, że wnosi ono istotny wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja. W związku z tym przesłanka dotycząca pozytywnej oceny osiągnięcia naukowego warunkująca nadanie stopnia doktora habilitowanego moim zdaniem została spełniona.

4. Ocena istotnej aktywności naukowej w więcej niż jednej uczelni, instytucji naukowej lub instytucji kultury, w szczególności zagranicznej.

Po obronie pracy magisterskiej, w 1996 roku, Habilitant pracował przez 2 miesiące w Centrum voor Wiskunde et Informatica (CWI) w Amsterdamie gdzie zajmowałem się realizacją systemów wieloagentowych opartych o modele automatów. Niestety, z tego okresu nie powstała żadna publikacja.

W 1999 roku Habilitant odbył pięciomiesięczny staż w Niemczech w ramach stypendium badawcze Niemieckiej Centrali Wymiany Akademickiej (Deutscher Akademischer Ausaschdienst - DAAD) dla młodszych pracowników naukowych. Wynikiem tego stażu była publikacja, wraz z pracownikami firmy Daimler-Chrysler na konferencji międzynarodowej Principles of Data Mining and Knowledge Discovery (PKDD) (obecnie 140 pkt. MEiN). Wydaje się jednak, że była to jedyna praca Habilitanta ze współautorami z zagranicy.

Podsumowanie

Biorąc pod uwagę zalecenia RDN dotyczące oceny istotnej aktywności naukowej w więcej niż jednej uczelni, Habilitant spełnia przesłankę warunkującą nadanie stopnia doktora habilitowanego dotyczącą tego punktu jednak tylko w sposób dostateczny.

5. Inne aktywności habilitanta.

Poza ocenianym cyklem publikacji Habilitant prowadził też inne badania dotyczące np. zrównoleglania algorytmów ewolucyjnych w eksploracji danych, algorytmów hybrydowych wykorzystujących ewolucję baldwinowską w grupowaniu danych, hybrydowych algorytmów ewolucji różnicowej w uczeniu sieci neuronowych, modyfikacji algorytmu Yinyang, gromadzenia i analizy wielkoskalowych danych medycznych, czy wykorzystania kd-drzew w równoległych algorytmach K -średnich.

Habilitant angażował się też w dydaktykę, w tym w wspólne publikacje z magistrantami i opiekę nad młodymi pracownikami naukowymi. Angażował się też w prace organizacyjne, np. budowę i utrzymanie kilku klastrów obliczeniowych na Wydziale Informatyki Politechniki

Białostockiej. Brał też udział w realizacji projektu Krajowy Klaster Linuxowy (ang. National Cluster of Linux Systems – Clusterix) kierując pracami po stronie Politechniki Białostockiej. Brał też udział w pracach komitetu organizacyjnego międzynarodowej konferencji naukowej 13th International Conference on Parallel Processing and Applied Mathematics odbywającej się 8-11.09.2019 w Białymstoku.

Zgodnie z przyjętymi zwyczajowymi zasadami oceny całokształtu dorobku naukowego, dydaktycznego i organizacyjnego, stawiane osobom ubiegającym się o nadanie stopnia doktora habilitowanego, moim zdaniem dokonania dr. inż. Wojciecha Kwedło kwalifikują go jako przyszłego samodzielnego pracownika naukowego.

Konkluzja końcowa

Oceniając osiągnięcie naukowe i istotne aktywności naukowe dr. inż. Wojciecha Kwedło w postępowaniu habilitacyjnym w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja stwierdzam, że zostały spełnione wymagania stawiane do uzyskania stopnia naukowego doktora habilitowanego (określone w Ustawie dnia z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce, z późniejszymi zmianami). Pozytywnie oceniam więc wniosek dr. inż. Wojciecha Kwedło o nadanie stopnia doktora habilitowanego nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.

A. Juchacz