

Warszawa, 11.08.2025

Prof. dr hab. Henryk Rybinski
Politechnika Warszawska
Instytut Informatyki

**Ocena dorobku naukowego oraz
podstawowego osiągnięcia naukowego, złożonych we wniosku
dra Macieja Piernika o przeprowadzenie postępowania habilitacyjnego**

Niniejsza recenzja została przygotowana w odpowiedzi na decyzję Rady Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Poznańskiej, powołującą mnie na recenzenta w postępowaniu habilitacyjnym Pana dra inż. Macieja Piernika z Politechniki Poznańskiej.

Recenzja została przygotowana w oparciu o materiały prezentujące (1) osiągnięcie naukowe dra Macieja Piernika wraz ze związanym z tym autoreferatem, oraz (2) dorobek naukowy, w szczególności wykaz opublikowanych prac naukowych, a także informacje o działalności naukowej. W recenzji ustosunkuję się najpierw do zgłoszonego osiągnięcia dra Macieja Piernika, a następnie przedstawię opinię na temat całokształtu dorobku naukowego.

Osiągnięcie naukowe

Przedmiotem oceny jest osiągnięcie naukowe w formie cyklu publikacji powiązanych ze sobą tematycznie, zatytułowane:

Metody klasyfikacji złożonych typów danych: podejścia oparte na cechach, odległościach i wzorcach

Osiągnięcie obejmuje 7 prac opublikowanych w latach 2017-2024, w którego skład wchodzi publikacje:

1. Maciej Piernik, Tadeusz Morzy, Robert Susmaga, Izabela Szczęch: On feature extraction using distances from reference points, *Foundations of Computing and Decision Sciences*, 2024, vol. 49, no. 3, s. 287-302 [MNiSW 2024: 40 pkt., IF 2024: 1.8]

2. Maciej Piernik, Dariusz Brzeziński, Paweł Zawadzki: Random Similarity Forests w: *Machine Learning and Knowledge Discovery in Databases: European Conference*, ed. Massih-Reza Amini, ECML PKDD 2022, Grenoble, France, Sept. 19–23, 2022, Proceedings, Part V, Springer, 2023 - s. 53-69 [MNiSW 2023: 140 pkt.]

3. Maciej Piernik, Dariusz Brzeziński, Paweł Sztromwasser, Klaudia Pacewicz, Weronika Majer-Burman, Michał Gniot, Dawid Sielski, Oleksii Bryzghalov, Alicja Woźna, Paweł Zawadzki: DBFE: distribution-based feature extraction from structural variants in whole-genome data, *Bioinformatics*, 2022, vol. 38(19), ss. 4466-4473 [MNiSW 2022: 200 pkt., IF 2022: 6,931]

4. Maciej Piernik, Tadeusz Morzy: A study on using data clustering for feature extraction to improve the quality of classification, *Knowledge and Information Systems* - 2021, vol. 63(7), ss. 1771-1805 [MNiSW 2021: 100 pkt., IF 2021: 2,531]

5. Maciej Piernik, Joanna Sołomiewicz, Arkadiusz Jachnik: Assessing the Effectiveness of Sequences of Treatments Using Sequential Patterns w: *Artificial Intelligence in Medicine: 17th Conference on Artificial Intelligence in Medicine*, AIME 2019, Poznan, Poland, June 26–29, 2019: Proceedings, David Riaño, Szymon Wilk (WI), Annette ten Teije (eds), Springer International Publishing, 2019 - s. 131-135 [MNiSW 2019: 70 pkt.]

6. Maciej Piernik, Tadeusz Morzy: Partial Tree-Edit Distance: A Solution to the Default Class Problem in Pattern-Based Tree Classification, w: *Advances in Knowledge Discovery and Data Mining: 21st Pacific-Asia Conference, PAKDD 2017*, Jeju, South Korea, May 23-26, 2017, Proceedings, Part II, Jinho Kim (ed.), Switzerland: Springer, 2017 - s. 208-219 [MNiSW 2017: 140 pkt.]

7. Maciej Piernik, Dariusz Brzeziński, Tadeusz Morzy, Mikołaj Morzy: Using Network Analysis to Improve Nearest Neighbor Classification of Non-network Data, w: *Foundations of Intelligent Systems: 23rd International Symposium, ISMIS 2017*, Warsaw, Poland, June 26-29, 2017: Marzena Kryszkiewicz (eds), Switzerland: Springer, 2017 - s. 105-115 [MNiSW 2017: 20 pkt.]

Przedstawiony cykl publikacji obejmuje:

1. 3 artykuły z czasopism z listy JCR ([1, 3, 4]
2. 4 publikacje w materiałach konferencyjnych konferencji z listy ministerialnej (wszystkie wysoko punktowane).

Łączna wartość publikacji zgłoszonych w osiągnięciu w punktach MEiN (wg nowej punktacji) wynosi 710 pkt. Skumulowana wartość IF publikacji z czasopism JCR wynosi 11,262 (zgodnie z rokiem opublikowania). W zgłoszonym wykazie we wszystkich pracach dr Maciej Piernik jest pierwszym autorem. Zgodnie z oświadczeniami dra Macieja Piernika oraz współautorów udział habilitanta był następujący:

1. W pracach [2,3,5,6] dr Piernik jest pomysłodawcą metody, w pracy [1] opracował pojęcia sąsiedztwa (*theta-neighborhood*) oraz neuronową interpretację punktów referencyjnych;
2. W pracach [1,2,4,6] habilitant samodzielnie zaimplementował metodę, w pozostałych pracach był odpowiedzialny za całość prac implementacyjnych;
3. W pracach [1, 4, 6] samodzielnie przeprowadził eksperymenty, w pracy [5] zaprojektował eksperymenty oraz analizował wyniki, w pracach [2, 3, 7] był współodpowiedzialny za przeprowadzenie eksperymentów;
4. We wszystkich pracach dr Piernik miał istotny udział w przygotowaniu tekstu.

Podsumowując, uważam, że udział habilitanta w pracach wskazanych w osiągnięciu był istotny i stanowi odpowiednią podstawę do analizy wartości merytorycznej osiągnięcia.

Tematyka badań

Prace składające się na osiągnięcie naukowe poświęcone są problematyce klasyfikacji. Problematyka ta od dawna jest przedmiotem zainteresowań wielu zespołów badawczych. W badaniach nad klasyfikatorami istnieje duża różnorodność podejść. W latach 80-tych ubiegłego stulecia prace nad metodami klasyfikacji stanowiły zasadniczą część dziedziny sztucznej inteligencji, w szczególności uczenia maszynowego (prekursorami tu byli Michalski, Carbonell, Mitchell). Natomiast w połowie lat 90-tych pojawiły się propozycje bazujące na nowatorskim wówczas podejściu eksploracji danych (polegającym na odkrywaniu cech w dużych zbiorach danych). Oba te podejścia są niezwykle ważne w

kontekście ogólnie rozumianej analizy danych. Eksploracja danych jest bardziej zorientowana na proces odkrywania pewnych wzorców w dużych zbiorach danych, natomiast uczenie maszynowe jest zorientowane na utworzenie modelu, którego główną cechą jest umiejętność przewidywania klasy obiektu na podstawie jego cech. Prace przedstawione w zgłoszonym osiągnięciu nawiązują do obu tych nurtów.

Zawartość merytoryczna i ocena ogólna osiągnięcia

Różnorodność metod klasyfikacji danych w dużej mierze wynika z różnorodności charakteru danych. Jednym z wyróżników jest stopień złożoności danych, mocno zależny od konkretnego zastosowania. Dość często, szczególnie w pierwszych zastosowaniach, analizowane dane to były dane tabelaryczne, czy ogólniej, dane opisywane przez pewien stały, określony zestaw cech. W wielu nowszych zastosowaniach złożoność danych wymusza indywidualizację podejścia do konkretnych przypadków. W pracach wchodzących w skład osiągnięcia habilitant zajmuje się problematyką analizy danych złożonych. W kontekście danych złożonych wyróżnia się trzy podejścia do problematyki danych złożonych:

1. bazujące na cechach
2. bazujące na odległościach
3. bazujące na wzorcach

W pracach przedstawionych w osiągnięciu habilitant podejmuje próbę łączenia tych podejść w taki sposób, aby w poszczególnych fazach analizy danych wykorzystywać najbardziej odpowiednie podejście. Poniżej omówione są osiągnięcia habilitanta w trzech wymienionych podejściach.

Podejście bazujące na cechach

Podejście to jest wynikiem badań nad metodami klasyfikacji guzów nowotworowych na podstawie danych pochodzących z sekwencjonowania całogenomowego. Efektem badań są trzy metody ekstrakcji cech z rozkładów długości wariantów genetycznych. Wyniki analizy danych pozwalają zbudować klasyfikatory ogólnego przeznaczenia. Przeprowadzono eksperymenty dla nowotworów jajnika, płuc i piersi, które potwierdziły przydatność opracowanych rozwiązań. W prezentowanym osiągnięciu wyniki badań w tym zakresie przedstawione jest w pracach [3] i [7]. Praca [7] łączy w sobie podejście bazujące na cechach z podejściem bazującym na odległościach, mianowicie zastosowano

tu przekształcenie danych do postaci sieci, przy czym w przekształceniu tym korzysta się z odległości pomiędzy przykładami.

Warto tu zaznaczyć, że efektem praktycznym badań nad podejściem bazującym na cechach jest stworzenie przez habilitanta biblioteki DBFE w Pythonie, obejmującej procedury do ekstrakcji cech.

Podejście bazujące na odległościach

Podejście bazujące na odległościach stanowiło przedmiot badań zaprezentowanych w pracach [4] z roku 2021 i [1] z roku 2024. Praca [4] była próbą pokonania ograniczeń, które wystąpiły w metodzie opisanej w pracy [7]. Podstawowym ograniczeniem metody obliczania wag przykładów z [7] był problem z wykorzystaniem ich jako cechy, ponieważ obliczenia były prowadzone dla wszystkich przykładów jednocześnie, to zaś powodowało, że pojawienie się nowych przykładów wymagało ponownego przeliczenia dla wszystkich przykładów, tym samym wykluczało możliwość uczenia inkrementalnego. Zaproponowana w [4] metoda usuwała ten problem. Podsumowanie badań bazujących na odległościach zostało przedstawione w pracy [1]. Stanowi ona istotne uogólnienie podejścia prezentowanego w [4]. Wprowadzono tu pojęcie θ -sąsiedztwa, co pozwala nawigować po grafach z wykorzystaniem odległości między przykładami. Opracowana metoda charakteryzuje się bardzo dobrymi parametrami klasyfikacji i przewyższa znane w literaturze podejścia.

W dorobku habilitanta jest klasyfikator nazwany przez autorów *Random Similarity Forest* (RSF), zaproponowany w [2]. Klasyfikator ten nie miesza trzech rodzajów metod, jest ściśle związany z podejściem bazującym na odległościach. Jego interesującą właściwością jest to, że może on działać na heterogenicznym zbiorze obiektów. Algorytm został przetestowany na 30 heterogenicznych zbiorach danych złożonych (sekwencje zbiorów, przebiegi czasowe, grafy) oraz mieszanych (w których pojawiały się zarówno cechy numeryczne jak i złożone). Wyniki eksperymentów pokazują, że RSF działa równie dobrze jak Random Forest na danych liczbowych, ale przewyższa go jakością na danych złożonych i mieszanych.

Podejście bazujące na wzorcach

W tej grupie badań mieszczą się prace [6] i [5] (odpowiednio z lat 2017 i 2018). Praca [6] łączy w sobie podejście bazujące na wzorcach z podejściem bazującym na odległościach,

mianowicie zaproponowano tu nową miarę odległości dla struktur drzewowych (nazwaną *partial tree-edit distance*, oznaczaną jako PTED), polegającą na określeniu stopnia zawierania się jednego drzewa w drugim. Jednocześnie wykorzystywane jest podejście bazujące na wzorcach, w którym miara PTED pozwala istotnie poprawić jakość klasyfikacji. Pokazano, że miara PTED daje się obliczyć w czasie wielomianowym. Przeprowadzone eksperymenty pokazały dużą przydatność opracowanej metody do budowania klasyfikatorów dla danych drzewiastych.

W pracy [5] rozważany jest podobny problem, z tym, że zamiast struktur drzewiastych analizowane są sekwencje zbiorów. Zaprezentowany tu algorytm stanowi nowe podejście do badania wzorców sekwencyjnych. Pokazano przydatność podejścia w złożonych zastosowaniach do analizy danych medycznych.

Podsumowując, stwierdzam co następuje:

1. W większości przypadków omawianych w publikacjach osiągnięcia habilitanta inspiracją do badań są konkretne praktyczne potrzeby klasyfikacji danych;
2. warsztat naukowy wykorzystywany w zbiorze prac stanowiących osiągnięcie habilitanta jest ściśle związany z tematyką eksploracji danych, oraz z metodami uczenia maszynowego i budowania klasyfikatorów;
3. całość stanowi spójny cykl prac.

Podsumowując tę część opinii, stwierdzam, że wyniki uzyskane przez dra inż. Macieja Piernika, zaprezentowane w cyklu mają duże znaczenie naukowe.

Ocena dorobku naukowego

Dr Maciej Piernik ukończył studia magisterskie w 2010 roku na Politechnice Poznańskiej na kierunku Informatyka. W roku 2015 tamże obronił doktorat w zakresie eksploracji danych w dyscyplinie Informatyka, na podstawie rozprawy p.t. *"Pattern-based clustering and classification of XML data"*.

Tematyka badań dra Macieja Piernika skupia się wokół problematyki sztucznej inteligencji – przede wszystkim w zakresie metod klasyfikacji, oraz zagadnień eksploracji danych.

Łącznie habilitant wykazał 18 prac (7 prac zgłoszonych w osiągnięciu habilitacyjnym i 11 poza osiągnięciem); 5 prac przed uzyskaniem stopnia doktora, 13 po uzyskaniu stopnia doktora). W tej liczbie znajdują się następujące kategorie prac:

1. 9 publikacji w czasopismach, w tej liczbie 9 pozycji z listy JCR (w tym 3 przed uzyskaniem stopnia doktora, 6 po uzyskaniu stopnia doktora, w tym 3 z wykazu osiągnięcia);
2. 5 publikacji w materiałach konferencyjnych (konferencje z listy ministerialnej), w tej liczbie 4 w osiągnięciu, ;
3. 4 abstrakty zamieszczone w czasopismach z listy JCR

Sześć prac (dwa artykuły i 4 abstrakty) dotyczyły zagadnień medycznych (były one opublikowane w czasopismach medycznych) – w pracach tych wkład habilitanta związany był z wykorzystaniem metod uczenia maszynowego w celu opracowania wyników eksperymentów medycznych.

Przybliżona wartość skumulowana IF w czasopismach z listy JCR (liczona dla artykułów z pominięciem abstraktów) wynosi ok. 19,699 .

Według bazy Scopus, dorobek habilitanta obejmuje 16 publikacji, cytowanych przez 96 prac, indeks Hirscha wynosi 6. Według WoS dorobek obejmuje 17 dokumentów, liczba cytowań – 61 (w tym jedno cytowanie w dokumencie patentowym), indeks H=4. Według GS liczba prac (w tym rozprawa doktorska) wynosi 29, zaś liczba cytowań wynosi 182 i H=6. Niewątpliwie na uwagę zasługuje dynamika cytowań, zarówno w bazie Google Scholar jak też w bazie Scopus. W GS tylko za okres ostatnich 5 lat H=5.

Zarówno z punktu widzenia parametrów bibliometrycznych, jak też wartości merytorycznych moja ocena dorobku publikacyjnego habilitanta jest pozytywna.

Habilitant brał udział w projektach naukowo badawczych: dwa projekty OPUS (NCN), dwa projekty MNiSW, w tym, w jednym projekcie MNiSW habilitant był kierownikiem).

Wysoko oceniam dorobek dra inż. Macieja Piernika w zakresie współpracy z otoczeniem społecznym i gospodarczym. W obszarze działalności wdrożeniowej habilitant jest współtwórcą czterech wdrożeń przemysłowych w zakresie zastosowań metod sztucznej inteligencji, jest też autorem modułów zintegrowanego systemu informatycznego dla służby zdrowia. Ponadto od roku 2020 współpracuje z firmą MNM Diagnostics nad rozwojem technologii wspomagającej diagnostykę pacjentów nowotworowych.

Potwierdzeniem międzynarodowej pozycji naukowej habilitanta jest fakt, że wielokrotnie był zapraszany do recenzowania prac w wysoko punktowanych czasopismach. Baza WoS potwierdza wykonanie 11 recenzji.

Podsumowując tę część opinii, pozytywnie oceniam dorobek naukowy dra inż. Macieja Piernika. Są to ważne osiągnięcia na styku dziedzin sztucznej inteligencji (przede wszystkim uczenia maszynowego), oraz metod eksploracji danych. Wysoko oceniam aktywność w obszarze działalności wdrożeniowej, łączącej wiedzę teoretyczną z zastosowaniami w praktycznych rozwiązaniach – uzyskane wyniki świadczą o dojrzałości i samodzielności naukowej habilitanta.

Podsumowanie

Mając na uwadze moją pozytywną ocenę osiągnięcia naukowego dra inż. Macieja Piernika, zatytułowanego „Metody klasyfikacji złożonych typów danych: podejścia oparte na cechach, odległościach i wzorcach” oraz pozytywną ocenę dorobku naukowego, stwierdzam, że dr inż. Maciej Piernik spełnia wymagania określone w odnośnej ustawie o stopniach i tytule naukowym. Na tej podstawie popieram wniosek dra inż. Macieja Piernika o nadanie mu stopnia naukowego doktora habilitowanego w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.