

Częstochowa, 24 stycznia 2024 r.

prof. dr hab. inż. Robert Nowicki
Katedra Inteligentnych Systemów Informatycznych
Wydział Inżynierii Mechanicznej i Informatyki
Politechnika Częstochowska
ul. Armii Krajowej 36
42-200 Częstochowa

**Ocena osiągnięć naukowych dra Andrzeja Janusza ubiegającego się o nadanie
stopnia doktora habilitowanego
opracowana w ramach postępowania habilitacyjnego toczącego się
w Instytucie Badań Systemowych Polskiej Akademii Nauk**

Niniejsza opinia została przygotowana na prośbę wystosowaną w piśmie dra hab. inż. Jana Owsieńskiego, Zastępcy Dyrektora ds. Naukowych Instytutu Badań Systemowych Polskiej Akademii Nauk z dnia 14 listopada 2023 r., nr syg. BA3-0005-4/23.

Opinia ta została opracowana w ramach postępowania habilitacyjnego dra Andrzeja Janusza prowadzonego przed Radą Instytutu Badań Systemowych Polskiej Akademii Nauk na podstawie decyzji Rady Doskonałości Naukowej.

Ocena głównego osiągnięcia naukowego

Dorobek naukowy przedstawiony przez Kandydata jako osiągnięcie naukowe będące podstawą ubiegania się przez Niego o nadanie stopnia doktora habilitowanego jest cyklem jedenastu artykułów naukowych przedstawionych we wniosku pod zbiorczym tytułem „Wybrane metody reprezentacji obiektów i pojęć w danych dla algorytmów uczenia maszynowego”. Wszystkie publikacje należące do cyklu są współautorskie, przy czym tylko w dwóch Kandydat nie jest pierwszym autorem. Cztery z nich to prace wydane w recenzowanych materiałach z konferencji międzynarodowych, sześć zostało opublikowane w czasopismach naukowych ujętych w wykazie sporządzonym zgodnie z przepisami wydanymi na podstawie art. 267 ust. 2 pkt 2 lit. b ustawy Prawo o szkolnictwie wyższym i nauce. W pięciu przypadkach czasopismom lista ta przypisuje 140 i 200 punktów. Cykl uzupełnia rozdział z monografii.

[O1] Andrzej Janusz, Sebastian Stawicki, Hung Son Nguyen: Adaptive Learning for Improving Semantic Tagging of Scientific Articles. in FedCSIS 2014: 27-34

[O2] Andrzej Janusz, Dominik Ślęzak: Rough Set Methods for Attribute Clustering and Selection. Appl. Artif. Intell. 28(3): 220-242 (2014)

[O3] Andrzej Janusz, Dominik Ślęzak: Computation of Approximate Re ducts with Dynamically Adjusted Approximation Threshold. ISMIS 2015: 19-28

[O4] Andrzej Janusz, Marek Grzegorowski, Marcin Michalak, Łukasz Wróbel, Marek Sikora, Dominik Ślęzak: Predicting seismic events in coal mines based on underground sensor measurements. *Eng. Appl. Artif. Intell.* 64: 83-94 (2017)

[O5] Sebastian Stawicki, Dominik Ślęzak, Andrzej Janusz, Sebastian Widz: Decision bireducts and decision reducts- a comparison. *Int. J. Approx. Reason.* 84: 75-109 (2017)

[O6] Andrzej Janusz, Dominik Ślęzak: Investigating Similarity between He arthstone Cards: Text Embeddings and Interchangeability Approaches. *SMC 2018*: 3421-3426

[O7] Mateusz Przyborowski, Tomasz Tajmajer, Łukasz Grad, Andrzej Janusz, Piotr Biczuk, Dominik Ślęzak: Toward Machine Learning on Granulated Data- a Case of Compact Autoencoder-based Representations of Satellite Images. *IEEE BigData 2018*: 2657-2662

[O8] Dominik Ślęzak, Marek Grzegorowski, Andrzej Janusz, Michał Kozłowski, Sinh Hoa Nguyen, Marek Sikora, Sebastian Stawicki, Łukasz Wróbel: A framework for learning and embedding multi-sensor forecasting models into a decision support system: A case study of methane concentration in coal mines. *Inf. Sci.* 451-452: 112-133 (2018)

[O9] Andrzej Janusz, Dominik Ślęzak: Analytics over Multi-sensor Time Series Data- A Case-Study on Prediction of Mining Hazards. *Information Technology in Geo-Engineering. ICITG 2019*: 815-830

[O10] Andrzej Janusz, Daniel Kałuża, Maciej Matraszek, Łukasz Grad, Maciej Świechowski, Dominik Ślęzak: Learning multimodal entity representations and their ensembles, with applications in a data-driven advisory framework for video game players. *Inf. Sci., Volume 617*: 193-210 (2022)

[O11] Andrzej Janusz, Dominik Ślęzak, Sebastian Stawicki, Krzysztof Stencel: A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings using Decision Bireducts. *Inf. Sci., In print* (2023)

Zgodnie z art. 219 ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* cykl artykułów naukowych będących podstawą do nadania stopnia doktora habilitowanego powinien składać się z artykułów naukowych powiązanych tematycznie. Definiując tytuł osiągnięcia, Kandydat wskazał jako temat wiążący publikacje cyklu metody reprezentacji obiektów i pojęć w danych. Rzeczywiście, istotnym elementem działań przedstawionych w publikacjach należących do cyklu jest przygotowanie, swego rodzaju preprocesing, danych pozyskanych ze źródeł rzeczywistych, na potrzeby uczenia maszynowego. Prace te są wiążą się także z publikacjami spoza cyklu korzystając z przedstawionych tam wyników. Przykładem jest metoda estymacji prawdopodobieństwa braku relacji pomiędzy elementami zaproponowana w

[D1] Janusz, A., Ślęzak, D.: Random Probes in Computation and Assessment of Approximate Reducts. In M. Kryszkiewicz et al., ed.: *Proceedings of RSEISP 2014. Volume 8537 of Lecture Notes in Computer Science.*, Springer (2014) 53–64

a rozwijana i wykorzystywana w [O3], [O5]. Działalność naukową Kandydata wiąże bez wątpienia sprawne posługiwanie się teorią zbiorów przybliżonych (ang. *rough set theory*) prof. Pawłaka i wkład w jej rozwój. W pracach należących do cyklu jest wykorzystywany algorytm obliczania reduktów aproksymacyjnych z dynamicznie określonym współczynnikiem aproksymacji (ang. *dynamically adjusted approximate reducts*, DAAR) opublikowany w pracy [O3] oraz koncepcja bireduktów [O5], [O11]. Skuteczność i poprawność algorytmu została potwierdzona w zastosowaniach przedstawionych

w innych publikacjach. O jej przydatności świadczy też deklarowana przez Kandydata liczba pobrań biblioteki dla języka R, zawierającej implementację algorytmu. Mój sprzeciw budzi jednak formalny zapis algorytmu przedstawiony w postaci Algorytmów 1 i 2 zamieszczonych w publikacji [O3] i przeniesionych do autoreferatu z tą samą numeracją. W linii 5. Algorytmu 1 losowany jest podzbiór $A' \subseteq A \setminus AR$ o określonej liczebności, z czego można wnioskować, że $A' \cap AR$ jest zbiorem pustym, a więc $A' \setminus AR = A'$. Dlaczego zatem w linii następnej otwierana jest pętla (foreach) dla elementów $a \in A' \setminus AR$, nie zaś dla $a \in A'$? Niezrozumiałe bez zapoznania się z Algorytmem 2 jest dla czytelnika pojawienie się w linii 13. Algorytmu 1 niezdefiniowanej zmiennej \hat{a}_{best} , która będzie reprezentowała element zdefiniowany i losowany wewnątrz Algorytmu 2. Nie wiadomo też co oznacza indeks *best*. W przypadku Algorytmu 2 nie wyjaśniono znaczenia wartości *nProbes* dla przypadku *globalPerms = TRUE*, w którym wartość *nProbes* nie ma wpływu na przebieg algorytmu, a ma wpływ na wynik.

Znaczna część prac należących do zgłoszonego cyklu publikacji dotyczy wykorzystania algorytmu DAAR w projektowaniu systemów inteligencji obliczeniowej pracujących na danych rzeczywistych. Część z nich to niezwykle perspektywiczne koncepcje związane z bezpieczeństwem i wczesnym ostrzeganiem o zagrożeniu w miejscu pracy. Przedstawione wyniki potwierdzają skuteczność i przydatność algorytmu DAAR, a oryginalnym elementem prac jest dostosowanie struktury danych do potrzeb projektowania systemu. Zgodnie z tytułem cyklu. Praca [O4] dotyczy przetwarzania danych mających przewidzieć zagrożenia sejsmiczne w kopalniach węgla kamiennego. Praca [O8], również zajmuje się zagrożeniami w górnictwie węgla kamiennego, korzystając z podobnych metod, opisuje propozycję przygotowania danych do uczenia systemów przewidujących poziomy koncentracji metanu w eksploatowanym złożu. Siostrzaną publikacją jest pozycja [O9] dotycząca również przetwarzania danych z czujników koncentracji metanu, tym razem w reprezentacji wielowymiarowej. Również w tej pracy wykorzystana została biblioteka *RoughSets* i zawarty w niej algorytm DAAR.

Kandydat wykazuje się również dużą aktywnością w zakresie popularyzacji zagadnień opisanych danymi, które przetwarza, oraz stworzonej biblioteki. Jedną z dróg są liczne konkursy, które organizuje w ramach konferencji naukowych. Dają one także możliwość bezpośredniego porównania skuteczności metod przetwarzania danych, proponowanych przez zespoły z różnych ośrodków badawczych z własnymi propozycjami. W publikacja [O9] zamieszczono listy rankingowe dwóch takich konkursów w zestawieniu z wynikami osiągniętymi z zastosowaniem modelu opracowanego przez Kandydata.

W pracy [O9] wykorzystano także koncepcję wymienialności atrybutów warunkowych, której załączek został przedstawiony w publikacji [O2]. Rozważano zagadnienie rozróżnialności i wymienialności (zastępowalności) atrybutów warunkowych oparte na heurystycznej koncepcji uwzględnienia częstości współwystępowania poszczególnych atrybutach w wygenerowanych reduktach. W sposób arbitralny wykluczono z rozważań współistnienie w reduktach atrybutów z samym sobą, przypisując tej częstości wartość 0, podczas gdy wartość wynikająca bezpośrednio z koncepcji częstości współwystępowania wyniosłaby 1 – atrybut zawsze występuje w redukcje, w którym ten atrybut występuje. Nie rozważano skutków pozostawienia wartości naturalnej (1), które być może uzasadniałyby zastosowany zabieg.

Ciekawym wątkiem badawczym jest wykorzystanie zaproponowanych metod przetwarzania danych w analizie m.in. wirtualnej gry karcianej *HearthStone* [O6] oraz gier video [O10]. W pracy [O1] rozważany było rozwiązanie z zakresu uczenia adaptacyjnego jako rozwinięcie algorytmu *Explicit Semantic Analysis*. Celem było indeksowanie tematyczne (nadawanie słów kluczowych) tekstom naukowym. Badania zostały przeprowadzone na obszernych bazach rzeczywistych publikacji naukowych. Podobny cel przyświecał przetwarzaniu ofert pracy metodami omówionymi w publikacji [O11]. W badaniach

omówionych w publikacji [O7] wykorzystano autoenkodery splotowe do uzyskania skompresowanej reprezentacji obrazów satelitarnych wykorzystywanej w uczeniu maszynowym.

W publikacjach należących do cyklu pojawia się wiele potencjalnych zastosowań uczenia maszynowego (przewidywanie zagrożeń w górnictwie węgla kamiennego w postaci podwyższonego stężenia metanu oraz zdarzeń sejsmicznych, przetwarzania języka naturalnego w postaci tekstów naukowych oraz ofert pracy, przetwarzanie zdjęć satelitarnych oraz analizy gier karcianych i gier video). W każdym przypadku zostaje zaproponowana metoda przetworzenia danych do reprezentacji wykorzystywanej następnie w uczeniu maszynowym. Przedstawione wyniki wskazują, że zastosowanie odpowiednich metod reprezentacji danych jest istotne dla uczenia maszynowego, a wybór metod jest słuszny i pozytywnie wpływa na proces uczenia. Publikacje i zawarte w nich wyniki potwierdzają deklarowany w autoreferacie wkład Kandydata w rozwój dyscypliny naukowej Informatyka Techniczna i Telekomunikacja w postaci:

- nowych algorytmów selekcji cech (redukty aproksymacyjne, mRMR), wykorzystujących jako kryterium stopu statystyczny test permutacyjny,
- algorytmów generowania reduktów decyzyjnych przy wykorzystaniu metod grupowania atrybutów w tablicach decyzyjnych,
- metody mierzenia stopnia wymienialności atrybutów w kontekście zadanego problemu predykcyjnego,
- nowych, efektywnych obliczeniowo algorytmów konstruowania bireduktów decyzyjnych,
- metod skalowalnej ekstrakcji cech szeregów czasowych zaimplementowane w modelu obliczeń Map-Reduce,
- metody nauki zanurzeń dokumentów tekstowych w przestrzeni powiązań z pojęciami z zadanej taksonomii,
- algorytmów uczenia się zanurzeń obiektów opisywanych różnymi modalnościami danych, inspirowanymi technikami wywodzącymi się z dziedziny przetwarzania języka naturalnego,

a wkład ten należy uznać za znaczny. Dostarczona dokumentacja jednoznacznie określa merytoryczny wkład Kandydata w powstanie publikacji wchodzących w skład osiągnięcia.

Ocena innych osiągnięć badawczych

Ponieważ art. 219 ustawy Prawo o szkolnictwie wyższym i nauce stanowi, że „stopień doktora habilitowanego nadaje się osobie, która [...] posiada w dorobku osiągnięcia naukowe [...] stanowiące znaczny wkład w rozwój określonej dyscypliny...” używając liczby mnogiej. Również poradnik „Postępowania dotyczące nadawania stopnia doktora habilitowanego” wydany przez Radę Doskonałości Naukowej wskazuje, że wymaga się wykazania co najmniej dwóch osiągnięć stanowiących znaczny wkład w rozwój określonej dyscypliny. We wniosku, ani w autoreferacie Kandydat nie wskazał wprost co najmniej drugiego osiągnięcia naukowego, aczkolwiek nazwał ocenione powyżej osiągnięcie w postaci cyklu publikacji osiągnięciem głównym, a także wskazał „pozostałe kierunki badań” oraz „najważniejsze publikacje spoza [głównego] osiągnięcia naukowego” wraz z ich omówieniem. Pozwala to domniemywać, że aktywność Kandydata nie ogranicza się do działań, które doprowadziły do uzyskania głównego osiągnięcia naukowego, a pośród pozostałych kierunków badań znajduje się i inne osiągnięcie naukowe. Znaczna część wymienionych publikacji przedstawia pracę badawczą poprzedzającą główne osiągnięcie naukowe Kandydata lub wpisujące się w ten sam wątek badań – rozwijanie metod reprezentacji obiektów i pojęć w danych dla algorytmów

uczenia maszynowego. Osiągnięciem badawczym nie jest też zorganizowanie konkursów eksploracji danych, pomimo że jest to inicjatywa niezwykle pracowita i cenna dla środowiska naukowego.

Bazując na materiale ograniczonym, w tym zakresie, do autoreferatu za osiągnięcie można uznać adaptację dostępnych i opracowanych modeli i metod uczenia maszynowego do rzeczywistych i nietrywialnych zadań praktycznych. Zadania te dotyczyły w głównej mierze monitorowania bezpieczeństwa i były elementem realizowanych projektów badawczych. Modele te zostały także wykorzystane w ocenie rozwiązań wchodzących w skład głównego osiągnięcia naukowego oraz w organizowanych przez Kandydata konkursach eksploracji danych.

Ocena działalności poza ośrodkiem macierzystym

Kolejną przesłanką do nadania stopnia doktora habilitowanego jest wykazanie przez Kandydata istotnej aktywności naukowej realizowanej w więcej niż jednej uczelni lub instytucji naukowej. Wskazując jej spełnienie Kandydat przedstawia w autoreferacie listę projektów badawczych, w których uczestniczył oraz omawia współpracę z innymi niż macierzysty Uniwersytet Warszawski instytucjami naukowymi. O ile udział w projekcie realizowanym we współpracy z inną niż macierzysta instytucja badawcza nie wyczerpuje przesłanki, w autoreferacie wymieniono także cztery projekty, z opisu których wynika, że Kandydat wykonywał zadania w innych podmiotach niż macierzysty Uniwersytet Warszawski. Są to:

- Grail: Opracowanie modułowego narzędzia do implementacji i optymalizacji zaawansowanej sztucznej inteligencji szerokiego zastosowania w grach wideo. Numer projektu: POIR.01.02.00-00-0150/16, finansowany przez NCBR. Projekt realizowany przez firmę Silver Bullet Labs Sp. z o.o. Habilitant pełnił rolę wykonawcy badań i był członkiem kadry kluczowej projektu (2017 – 2018).
- SENSEI: System wspomagający rozwój umiejętności graczy. Numer projektu: POIR.01.02.00-00-0184/17-00, finansowany przez NCBR. Projekt realizowany przez firmę ESENSEI Sp. z o.o. Habilitant pełnił rolę kierownika naukowego części badawczej i był członkiem kadry kluczowej projektu (2018 – 2020).
- LitL: Labelling in the Loop. Opracowanie opartego na uczeniu maszynowym systemu automatycznego i półautomatycznego oznaczania danych w dużych zbiorach danych. Numer projektu: POIR.01.01.01-00-0213/19, finansowany przez NCBR. Projekt realizowany przez firmę QED Software. Habilitant pełni rolę wykonawcy badań i był członkiem kadry kluczowej projektu (2020 – 2022).
- BrightBox: Narzędzia klasy explainable AI służącego do poprawy interpretowalności i przewidywalności działania metod uczących oraz diagnostyki poprawności działania wyuczonych modeli AI/ML. Numer projektu: MAZOWSZE/0198/19, finansowany przez NCBR. Projekt realizowany przez firmę QED Software. Habilitant pełni rolę kierownika badań i był członkiem kadry kluczowej projektu (2020 – 2022).

Funkcje pełnione przez Kandydata oraz tematyka projektów wskazują, że można tą aktywność naukową uznać za istotną.

Ocena osiągnięć dydaktycznych, organizacyjnych oraz popularyzujących naukę

Obowiązujące przepisy dotyczące postępowania o nadanie stopnia doktora habilitowanego nie odnoszą się w żadnym zakresie do aktywności innej niż oceniona powyżej. Kandydat umieścił jednak w autoreferacie pewne informacje na temat swej aktywności na Wydziale Matematyki Informatyki i Mechaniki, Uniwersytetu Warszawskiego. Trudno się więc do tych danych nie odnieść. Wykazane zostały dwa zakresy działalności. Zgodnie z informacjami podanymi w autoreferacie, w okresie od uzyskania stopnia doktora w roku 2014 pełnił funkcję promotora 11 prac magisterskich i licencjackich, co odnosząc się do realiów w miejscu pracy recenzenta, jest liczbą nader skromną. Tematyka prac jest bardzo zbliżona do tematyki badawczej Kandydata. Wymienione są także trzy doktoraty, w których Kandydat był, bądź jest promotorem pomocniczym. W Autoreferacie wielokrotnie przywoływana jest aktywność w zakresie organizacji konkursów eksploracji danych. Przytoczone dane wskazują, że konkursy te cieszą się dużym zainteresowaniem międzynarodowym, co zapewne jest wynikiem ich rzetelnego przygotowania.

Wszystkie aktywności Kandydata, wykazane w autoreferacie i dołączonych dokumentach przenikają się wzajemnie, a łączy je tematyka badawcza sięgająca co najmniej doktoratu.

Podsumowanie

Ustawa z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* określa bardzo konkretne przesłanki do nadania stopnia doktora habilitowanego. Są to: posiadanie stopnia doktora, osiągnięcia naukowe stanowiące znaczny wkład w rozwój dyscypliny oraz istotną aktywność naukową w więcej niż jednym ośrodku naukowym. Skan dyplomu doktora został dołączony do dokumentacji. Kandydat, niewątpliwie znając treść ustawy i ogólnie dostępnych interpretacji i komentarzy, wykazał wprost jedynie jedno osiągnięcie naukowe. Oceniam je bardzo pozytywnie. Istnienia pozostałych przesłanek zostało wykazane nie wprost poprzez przedstawienie listy pozostałego dorobku, omówienie kierunków badań prowadzonych poza głównym osiągnięciem naukowym oraz pracy w projektach. Przesłanki te są tym materiale widoczne, chociaż słabo udokumentowane.

Powyzsza ocena pozwala mi stwierdzić, że działalność badawcza dra Andrzeja Janusza wyczerpuje wszystkie przesłanki niezbędne do nadania stopnia doktora habilitowanego, określone w ustawie z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz. U. 2018 poz. 1668 z późniejszymi zmianami). Wnoszę zatem o dalsze procedowanie postępowania o nadanie drowi Andrzejowi Januszowi stopnia doktora habilitowanego w dyscyplinie naukowej Informatyka Techniczna i Telekomunikacja.


prof. dr hab. inż. Robert Nowicki