



Politechnika
Wrocławska

Prof. dr hab. inż. Halina Kwaśnicka
Katedra Sztucznej Inteligencji
Wydział Informatyki i Telekomunikacji
Politechnika Wrocławska
Wyb. Wyspiańskiego 27, 50 370 Wrocław
E-mail: halina.kwasnicka@pwr.edu.pl

15 stycznia 2024r.

Ocena dorobku naukowo-dydaktycznego dr Andrzeja Janusza

w związku z prowadzonym postępowaniem habilitacyjnym przez Instytut Badań Systemowych PAN i powołaniem mnie przez Radę Doskonałości Naukowej na recenzenta, pismo DRKN.Z2.400.126.2023

1. Informacje ogólne o Habilitancie

Dr. Andrzej Janusz jest magistrem matematyki, dyplom magistra uzyskał 12 grudnia 2007 roku na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego. Praca magisterska pod tytułem „Relacja podobieństwa w zagadnieniach uczenia się z danych” łączyła się z maszynowym uczeniem. W tym też obszarze Habilitant kontynuował badania w ramach pracy doktorskiej, obronił ją 13.03.2014r. na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego. Rozprawa stanowiąca podstawę doktoratu nosiła tytuł „Algorithms for Similarity Relation Learning from High Dimensional Data”.

Dr Andrzej Janusz jest cały czas związany z macierzystą Uczelnią, Uniwersytetem Warszawskim (UW). Na długo przed obroną doktoratu, pierwszego listopada 2010r. został asystentem naukowym na Wydziale Matematyki, Informatyki i Mechaniki UW. Oznacza to, że do końca pracy na tym stanowisku, tzn. do końca października 2014 roku, nie musiał prowadzić zajęć dydaktycznych. Od 1 listopada 2014 roku pracuje na stanowisku adiunkta (domyślam się, że badawczo-dydaktycznego) na tym samym Wydziale i tej samej Uczelni – Wydziale Matematyki, Informatyki i Mechaniki UW. Habilitant nie przedstawił wiele informacji na temat Jego aktywności zawodowej, ograniczam się do informacji dostarczonej w dokumentacji.

Tematyka badań dra A. Janusza realizowanych po doktoracie nadal pozostaje w obszarze maszynowego uczenia. Czy procedura habilitacji powinna być wszczęta w dziedzinie nauk technicznych w dyscyplinie informatyka techniczna i telekomunikacja? Spójrzmy na tematykę publikacji naukowych Habilitanta, notowanych w bazie naukowej Web of Science (Rys. 1).

Field: Web of Science Categories	Record Count	% of 51
Computer Science Artificial Intelligence	31	60.784%
Computer Science Information Systems	29	56.863%
Computer Science Theory Methods	26	50.980%
Engineering Electrical Electronic	7	13.725%
Computer Science Software Engineering	6	11.765%
Mathematics Applied	4	7.843%
Computer Science Interdisciplinary Applications	2	3.922%
Robotics	2	3.922%
Automation Control Systems	1	1.961%
Computer Science Cybernetics	1	1.961%
Engineering Geological	1	1.961%
Engineering Multidisciplinary	1	1.961%

Rys. 1. Obszar tematyczny prac pana Andrzeja Janusza, notowanych w bazie Web of Science. Dane pobrane w dniu 14.01.2024 roku.

Ponad 60% tych prac zostało zakwalifikowanych do obszaru *Computer Science Artificial Intelligence*. Patrząc na tematykę wszystkich publikacji naukowych Habilitanta można stwierdzić, że mieści się ona w ważnym i rozwijanym w środowisku naukowym obszarze badań, jak również mieści się w dyscyplinie *Informatyka techniczna i telekomunikacja*.

2. Ocena osiągnięcia naukowego stanowiącego podstawę ubiegania się o stopień doktora habilitowanego

Dr. Andrzej Janusz przedstawił osiągnięcie naukowe zatytułowane „Wybrane metody reprezentacji obiektów i pojęć w danych dla algorytmów uczenia maszynowego”. Tematyka ta nie jest nowa, reprezentacja danych jest jednym z kluczowych aspektów uczenia maszynowego¹. Sposób reprezentacji danych silnie wpływa na wydajność i skuteczność algorytmów maszynowego uczenia – obiekty świata rzeczywistego muszą być przekształcone w reprezentację numeryczną na potrzeby opracowywania modeli uczenia maszynowego. Tematyka osiągnięcia naukowego Habilitanta jest aktualnym i ważnym obszarem badań. Swoje osiągnięcie naukowe Habilitant zdefiniował w postaci cyklu jedenastu publikacji:

¹ Guoqiang Zhong, Li-Na Wang, Xiao Ling, Junyu Dong. An overview on data representation learning: From traditional feature learning to recent deep learning. *The Journal of Finance and Data Science*, Vol. 2, Issue 4, 2016, Pages 265-278, <https://doi.org/10.1016/j.jfds.2017.05.001>.

1. Andrzej Janusz, Sebastian Stawicki, Hung Son Nguyen: **Adaptive Learning for Improving Semantic Tagging of Scientific Articles**. In FedCSIS 2014: 27-34 (70 pkt.)
2. Andrzej Janusz, Dominik Slezak: **Rough Set Methods for Attribute Clustering and Selection**. *Appl. Artif. Intell.* 28(3): 220-242 (2014) (40 pkt., IF=0,527 w 2014 r.)
3. Andrzej Janusz, Dominik Slezak: **Computation of Approximate Reducts with Dynamically Adjusted Approximation Threshold**. ISMIS 2015: 19-28 (20 pkt.)
4. Andrzej Janusz, Marek Grzegorowski, Marcin Michalak, Łukasz Wróbel, Marek Sikora, Dominik Slezak: **Predicting seismic events in coal mines based on underground sensor measurements**. *Eng. Appl. Artif. Intell.* 64: 83-94 (2017) (140 pkt., IF=2,819, w 2024 r. IF= 7,802)
5. Sebastian Stawicki, Dominik Slezak, Andrzej Janusz, Sebastian Widz: **Decision bireducts and decision reducts – a comparison**. *Int. J. Approx. Reason.* 84: 75-109 (2017) (140 pkt., IF= 1,766, w 2023 r. IF=4,452)
6. Andrzej Janusz, Dominik Slezak: **Investigating Similarity between Hearthstone Cards: Text Embeddings and Interchangeability Approaches**. *SMC*, 2018: 3421-3426 (70 pkt., IF= 7,351, w 2024 r. IF=11,471)
7. Mateusz Przyborowski, Tomasz Tajmajer, Łukasz Grad, Andrzej Janusz, Piotr Biczysk, Dominik Slezak: **Toward Machine Learning on Granulated Data – a Case of Compact Autoencoder-based Representations of Satellite Images**. *IEEE BigData* 2018: 2657-2662 (70 pkt., IF= 0, w 2024 r. IF= 4,271)
8. Dominik Slezak, Marek Grzegorowski, Andrzej Janusz, Michał Kozielski, 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) (200 pkt., IF= 5,524, w 2023-2024r. IF= 8,233)
9. Andrzej Janusz, Dominik Slezak: **Analytics over Multi-sensor Time Series Data – A Case-Study on Prediction of Mining Hazards**. *Information Technology in Geo-Engineering*. ICITG 2019: 815-830 (20 pkt.)
10. Andrzej Janusz, Daniel Kałuza, Maciej Matraszek, Łukasz Grad, Maciej Swiechowski, Dominik Slezak: **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) (200 pkt., w 2023-2024 r. IF= 8,233)
11. Andrzej Janusz, Dominik Slezak, Sebastian Stawicki, Krzysztof Stencel: **A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings using Decision Bireducts**. *Inf. Sci.*, In print (2023) (200 pkt., w 2023-2024 r. IF= 8,233)

Łączna suma punktów (tzw. ministerialnych) dla publikacji wymienionych jako osiągnięcie naukowe wynosi 1060, licząc obecną punktację prac konferencyjnych, a łączna suma IF (z dokładnością, z którą udało mi się ustalić wartości IF na rok publikacji pracy) to: 34,453.

W swoim osiągnięciu habilitacyjnym, pan Andrzej Janusz wydzielił trzy główne podobszary badań, są to:

1. Metody selekcji cech (ang. *feature selection*),
2. Metody ekstrakcji cech (ang. *feature extraction*), zwane również inżynierią cech (ang. *feature engineering*),
3. Metody uczenia się zanurzeń danych (ang. *embedding learning*).

Poniżej omawiam krótko osiągnięcia w powyższych podobszarach. W omówieniu skupię się na wybranych elementach pracy Habilitanta, które **uznamę za ważniejsze z punktu widzenia oceny nowatorskości i wagi osiągnięć.**

Ad. 1. Metody selekcji cech

Selekcja cech, istotnych z punktu widzenia zadania maszynowego uczenia, to klasyczne, popularne podejście do uczenia się zwężonej reprezentacji obiektów ze świata rzeczywistego. Wyróżnia się trzy typy selekcji cech: metody opakowujące (ang. *Wrapper methods*), np. selekcja do przodu (ang. *forward*), do tyłu (ang. *backward*) i krokowa (ang. *stepwise*); metody filtrowania (ang. *Filter methods*), np. ANOVA, korelacja Pearsona, progowanie wariancji; metody osadzone (ang. *Embedded methods*), np. Lasso, Ridge, drzewo decyzyjne. Prace Habilitanta, mieszczące się w tej tematyce, dotyczą aproksymacyjnych reduktów decyzyjnych wywodzących się z teorii zbiorów przybliżonych. Podejście to było rozwijane stopniowo; w pracy A. Janusz, D. Slezak: *Computation of Approximate Reducts with Dynamically Adjusted Approximation Threshold*. ISMIS 2015, zawarte jest podsumowanie badań dotyczących obliczania reduktów aproksymacyjnych z dynamicznie określanym współczynnikiem aproksymacji (DAAR, *dynamically adjusted approximate reducts*). Konstruowanie aproksymacyjnych reduktów jest zatrzymywane na podstawie testu statystycznego. Efektem tego otrzymuje się lepiej dopasowany do danych współczynnik aproksymacji oraz cechy słabo powiązane z atrybutem decyzyjnym nie występują w wynikowych podzbiorach (ściślej mówiąc, prawdopodobieństwo znalezienia cech niepowiązanych w wynikowych zbiorach jest bardzo małe). Dodatkową zaletą jest wysoka efektywność obliczeniowa. Warto podkreślić, że w ramach współpracy z zespołem badawczym z University of Granada, algorytm został zaimplementowany w R i udostępniony (<https://CRAN.R-project.org/package=RoughSets>), omówienie pakietu zostało opublikowane w renomowanym czasopiśmie *Information Sciences*.

Metoda DAAR była wykorzystana do rzeczywistych problemów a wyniki zostały opublikowane w liczących się czasopismach naukowych, np. D. Slezak, M. Grzegorowski, A. Janusz, M. Kozielski, S. H. Nguyen, M. Sikora, S. Stawicki, Ł. 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*. *Information Sciences*, 451-452: 112-133 (2018); A. Janusz, M. Grzegorowski, M. Michalak, Ł. Wróbel, M. Sikora, D. Slezak: *Predicting seismic events in coal mines based on underground sensor measurements*. *Eng. Appl. Artif. Intell.* 64: 83-94 (2017) oraz na międzynarodowej konferencji *ICITG 2019*: 815-830. Zastosowania te powstały w ramach projektu DISESOR – Zintegrowany, szkieletowy system wspomagan

decyzji dla systemów monitorowania procesów, urządzeń i zagrożeń, realizowanego w ramach konsorcjum.

Praca A. Janusz, D. Slezak: Rough Set Methods for Attribute Clustering and Selection. *Appl. Artif. Intell.* 28(3): 220-242 (2014) dotyczy problemu wyznaczania skupisk przestrzeni atrybutów warunkowych w kontekście konstruowania modeli predykcyjnych, czyli synergii algorytmów grupowania atrybutów w danych z algorytmami selekcji cech. Autorzy zajmują się danymi wielowymiarowymi (mikromacierzami), dla których tradycyjne podejścia konstrukcji reduktów są mało wydajne (wysoka złożoność obliczeniowa i niezadawalający rozmiar generowanych zbiorów). Autorzy proponują heurystyczne podejście i grupują atrybuty które są wymienne, co skraca czas obliczeń i polepsza jakość wyników. Autorzy zmodyfikowali proces generowania permutacji tak, aby nie uwzględniał atrybutów z klastra większościowego – jest to ciekawe spostrzeżenie. Wyniki grupowania atrybutów zastosowano do obliczenia randomizowanych reduktów zachłannych. Dało to poprawę efektywności i zmniejszyło średnią i minimalną wielkość otrzymanych reduktów. Niektóre z metod z tej pracy były dalej rozwijane i publikowane, np. prace pod numerem 6 i 9 z listy osiągnięć Habilitanta.

Za ciekawy i wartościowy wynik uważam sposób mierzenia stopnia wymienialności (zastępowalności) jednego atrybutu innym, zaprezentowany w pracy A. Janusz, D. Slezak: Analytics over Multi-sensor Time Series Data – A Case-Study on Prediction of Mining Hazards. *Information Technology in Geo-Engineering. ICITG 2019*. Przesłankę stanowiło tu spostrzeżenie, że jeden atrybut można zastąpić drugim, jeśli rzadko współwystępuje on w tych samych reduktach decyzyjnych, jednocześnie rozkład jego współwystąpień w reduktach z innymi atrybutami jest podobny.

Ad. 2. Metody ekstrakcji cech

Jednym z ważnych zagadnień dotyczących przygotowania zbioru danych uczących jest zmniejszenie wymiarowości danych (problem znany jako *klątwa wymiarowości*). Do tego celu, poza selekcją atrybutów, może też służyć transformacja zbioru danych na przestrzeń mniej wymiarową (konstrukcja nowych cech obiektów na podstawie cech wyjściowych). Wadą transformacji jest często utrata semantyki cech, niemniej jest to dość powszechnie stosowane podejście. Problem konstrukcji cech wystąpił przy pracy Habilitanta z danymi rzeczywistymi, w ramach projektu DISESOR. Dr Andrzej Janusz zajmował się opracowaniem modułu wykorzystującego model obliczeń Map-Reduce do skalowalnej ekstrakcji cech opisujących fragmenty szeregów czasowych oraz zastosowań tego modułu do monitorowania zagrożeń występujących w kopalniach węgla kamiennego. Zadanie modułu to automatyzacja procesu tworzenia reprezentacji wielowymiarowych szeregów czasowych, złożonych z odczytów

sensorów monitorujących bezpieczeństwo w kopalniach. Opracowany moduł uważam za ważne osiągnięcie Habilitanta. Konstruowane cechy na różnych poziomach złożoności opisują szerszy kontekst niż wyznaczone dla pojedynczych okien czasowych.

W pracy A. Janusz, D. Slezak, S. Stawicki, K. Stencel: A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings using Decision Bireducts, *Information Sciences*, Vol. 645, 2023, będącej rozszerzeniem wcześniejszych prac współautorstwa Habilitanta, rozwiązywany jest problem konstrukcji reprezentacji dla danych wielomodalnych. Praca powstała w ramach projektu Toolbox4HR, którego celem było opracowanie prototypowego systemu wspierającego rekruterów w ocenie kandydatów dla zadanych ofert pracy w branży IT. Analizowane dane to opisy ofert pracy i dane o potencjalnych kandydatach, pozyskane od rekruterów (CV) oraz z profili z serwisów społecznościowych. Cel systemu to przewidywanie chęci kandydatów do zmiany pracy. Autorzy zaproponowali procedurę oceny i porównania rankingów ważności atrybutów pod kątem ich wpływu na decyzję. Procedura oceny rankingów atrybutów uwzględnia dwa kryteria: (i) k najważniejszych atrybutów powinno być faktycznie istotnych dla rozważanego problemu decyzyjnego, oraz (ii) istotne atrybuty nie powinny być nisko w rankingu. Nie mając danych o 'ważności' atrybutów, autorzy zaproponowali modyfikację zbiorów danych poprzez rozszerzenie ich o sztuczne atrybuty, zwane atrybutami tasowanymi (ang. *shuffled attributes*).

Ad.3. Metody uczenia się zanurzeń danych

Zanurzanie (osadzanie, ang. *Embedding*) to metoda uczenia się reprezentacji danych, gdzie informacje o obiekcie są kondensowane w wektory o mniejszym wymiarze. Ważne jest, by zachowana była relacja podobieństwa pomiędzy obiektami i pomiędzy ich zanurzeniami.

Pierwszy efekt badań dotyczących zanurzeń to ulepszenie techniki analizy wyraźnej semantyki ESA (ang. *explicit semantic analysis*). Ulepszenie nazwane adaptacyjną analizą semantyki (ang. *adaptive semantic analysis*, ASA) jest zaprezentowane w pracy A. Janusz, S. Stawicki, H. S. Nguyen: Adaptive Learning for Improving Semantic Tagging of Scientific Articles. In FedCSIS 2014. Praca ta przyniosła Habilitantowi nagrodę IFSA Awards for Young Scientist przyznaną przez International Fuzzy Set Association.

Do uczenia się zanurzeń obrazów, Habilitant z współautorami zaproponował architekturę sieci neuronowej wykorzystującą wielowarstwowe autoenkodery splotowe. Praca M. Przyborowski, T. Tajmayer, Ł. Grad, A. Janusz, P. Biczuk, D. Slezak: Toward Machine Learning on Granulated Data – a Case of Compact Autoencoder-based Representations of Satellite Images. IEEE BigData 2018 pokazuje zastosowanie tych sieci do trenowania

reprezentacji o wysokim współczynniku kompresji dla wielokanałowych zdjęć satelitarnych. Habilitant wykorzystywał również techniki analogiczne do word2vec do uczenia się reprezentacji innych typów danych niż dokumenty tekstowe, m.in. do uczenia się reprezentacji ofert pracy dotyczących poszczególnych procesów rekrutacyjnych, np. publikacja A. Janusz, D. Slezak, S. Stawicki, K. Stencel: A Practical Study of Methods for Deriving Insightful Attribute Importance Rankings using Decision Bireducts. *Information Sciences*, Vol. 645, 2023. Opracowane efektywne obliczeniowo algorytmy konstruowania bireduktów decyzyjnych były porównywane z XGBoost dla zadania, w którym XGBoost dostarcza dokładne modele, ale nie identyfikuje wielu atrybutów ważnych dla użytkownika (tu: ekspertów merytorycznych). Biredukty decyzyjne są łatwiejsze do interpretacji, bardziej niż XGBoost skupiają się na różnorodności atrybutów, oraz zespoły (ang. *ensembles*) bireduktów decyzyjnych zapewniają porównywalną dokładność do XGBoost.

Metody opracowane dla przetwarzania języka naturalnego są dość popularne, Habilitant czerpał z nich również w badaniach dotyczących uczenia się reprezentacji pojęć występujących w grach wideo, projekt SENSEI – System wspomagający rozwój umiejętności graczy. Pan Janusz zaproponował trzy metody uczenia się zanurzeń kart i talii z gry *Hearthstone: Heroes of Warcraft*. Nie znam tych gier, ale przeglądając publikację A. Janusz, D. Slezak: Investigating Similarity between Hearthstone Cards: Text Embeddings and Interchangeability Approaches. *SMC*, 2018: 3421-3426 oceniam, że nawet zastosowanie pierwszej techniki, czyli word2vec, nie było prostym zadaniem. Reprezentacja kart była uczona na podstawie ich opisów w języku naturalnym. Wagi z pierwszej warstwy wytrenowanej sieci odpowiadały zanurzeniom poszczególnych słów, które następnie agregowano do reprezentacji dla kart i talii. W drugim podejściu, jako źródło danych wykorzystano repozytorium talii wykorzystywanych przez graczy w prawdziwych rozgrywkach. Zadaniem uczonej sieci neuronowej było przewidywanie dla wskazanej karty prawdopodobieństw wystąpienia innych kart w talii.

Publikacja A. Janusz, D. Kałuża, M. Matraszek, Ł. Grad, M. Swiechowski, D. Slezak: 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) przedstawia najbardziej zaawansowane badania w tym obszarze. Przedmiotem badań jest wpływ łączenia wielu reprezentacji pochodzących z heterogenicznych źródeł danych na wydajność modeli uczenia maszynowego. Porównywane są dwa podejścia: (i) niezależne modele są uczone na reprezentacjach danych z różnych źródeł, (ii) pojedynczy model jest uczone na połączonych reprezentacjach danych. Metody uczenia się reprezentacji są stosowane w systemie doradczym SENSEI dla graczy gier wideo. Zaproponowano również nową

architekturę głębokiego uczenia się end2end z mechanizmem uwagi mającym na celu odzwierciedlenie znaczących interakcji między podmiotami.

Podsumowanie – ocena osiągnięcia naukowego

Habilitant wyróżnił trzy podobszary, które są ze sobą ściśle powiązane: metody selekcji cech, metody ekstrakcji cech oraz metody uczenia się zanurzeń danych. Przedstawiony cykl jedenastu publikacji prezentuje nowe rozwiązania w wymienionych wyżej obszarach. Za ważny wynik badań w obrębie pierwszego obszaru uważam **wykorzystanie statystycznego testu permutacyjnego jako warunku zatrzymania** w metodach selekcji cech. Sposób mierzenia **stopnia wymienialności jednego atrybutu innym** (bez pogorszenia jakości analizy danych) może okazać się wartościowym rozwiązaniem w problemach rzeczywistych, gdzie koszty pomiaru poszczególnych atrybutów są istotnie różne. **Metody generowania reduktów decyzyjnych** wykorzystujące grupowanie atrybutów w tablicach decyzyjnych to kolejny ważny wynik badań pana Janusza. Zadanie Habilitanta przy opracowywaniu systemu monitorowania zagrożeń w kopalniach węgla kamiennego to **automatyzacja procesu tworzenia reprezentacji wielowymiarowych szeregów czasowych**, złożonych z odczytów sensorów monitorujących bezpieczeństwo w kopalniach. Wymagało to opracowania **metody skalowalnej ekstrakcji cech opisujących fragmenty szeregów czasowych**. Wysoko należy ocenić efekt tych prac – konstruowane cechy opisują szerszy kontekst niż wyznaczone dla pojedynczych okien czasowych. Ważnym wynikiem badań są **efektywne obliczeniowo algorytmy konstruowania bireduktów decyzyjnych**.

Konkludując tę część recenzji stwierdzam, że dr Andrzej Janusz zaproponował ważne i nowatorskie elementy naukowe – algorytmy w dziedzinie maszynowego uczenia. Jest rozpoznawalny w światowym środowisku naukowym.

3. Ocena aktywności naukowej Habilitanta

Dr Andrzej Janusz współpracuje z różnymi zespołami, którą to umiejętność uważam za ważną w badaniach naukowych. Jest to szczególnie ważne w obszarze maszynowego uczenia, czy szerzej – sztucznej inteligencji, gdzie badania znajdują szerokie zastosowania w różnych obszarach społeczno-gospodarczych. Wg podanych w dokumentacji informacji, Habilitant współpracował z ponad czterdziestoma naukowcami z polskich i zagranicznych ośrodków badawczych (ośrodki te są wymienione w dalszej części recenzji).

Habilitant publikuje wyniki swoich badań w liczących się czasopismach naukowych, prezentuje wyniki badań na międzynarodowych konferencjach. Opracowywane algorytmy były weryfikowane podczas realizacji projektów badawczo-rozwojowych, które dotyczyły

rzeczywistych problemów w różnych dziedzinach. Opracowane pakiety oprogramowania są udostępnione publicznie, chętnie pobierane przez użytkowników z różnych krajów. Pan Andrzej Janusz otrzymał nagrodę IFSA Awards for Young Scientist przyznaną przez International Fuzzy Set Association.

Poza cyklem prac złożonych jako ‘monografia’, dr Andrzej Janusz przedstawił listę pozostałych publikacji, lista zawiera 34 prace. Są wśród nich publikacje w renomowanych czasopismach, jak *Rough Sets and Intelligent Systems Paradigms*, *Fundam. Informaticae*, *Inf. Sci.*, *Rough Sets and Knowledge Technology*, *Big Data Research*, *Appl. Soft Comput*, oraz w materiałach konferencyjnych. Nie będę tu przytaczać pełnych danych wszystkich wymienionych prac, omówię je tylko ogólnie.

Prace te dotyczą w większości zastosowań metod maszynowego uczenia do rzeczywistych problemów, w ramach realizowanych projektów. Bardzo często, jeśli nie zawsze, rozwiązywanie rzeczywistych problemów wymaga analizy metod znanych jako najskuteczniejsze (tzw. *state of the art*) oraz opracowania metody referencyjnej dla rozważanego problemu. Publikacje Habilitanta zawierają wyniki takich prac.

Habilitant wielokrotnie organizował międzynarodowe konkursy eksploracji danych, dane były powiązane z realizowanymi projektami. Aż 14 prac spoza listy osiągnięcia naukowego prezentuje podsumowania konkursów – najważniejsze zgłoszone rozwiązania oraz rozwiązania referencyjne. Są to publikacje konferencyjne, co wydaje się dość naturalne, jako że konkursy są organizowane przy konferencjach, tak też robił Habilitant.

Konkurs przeprowadzony w 2019 r. w ramach IEEE BigData, we współpracy z amerykańską firmą Security on Demand (SOD) dotyczył cyberbezpieczeństwa, wymagał anonimizacji danych, które pochodziły od klientów firmy. To zagadnienie jest nieco spoza głównego nurtu badań Habilitanta, który rozwiązał ten problem przy współpracy z ekspertami firmy SOD. Ważne było, aby anonimizacja nie spowodowała utraty potencjalnie istotnych informacji.

Nie sposób nie wspomnieć tu o bibliotece *RoughSets* języka *R*, powstałej we współpracy z zespołem badawczym z University of Granada. Biblioteka ta jest opisana w publikacji w liczącym się czasopiśmie: Lala Septem Riza, Andrzej Janusz, Christoph Bergmeir, Chris Cornelis, Francisco Herrera, Dominik Slezak, José Manuel Benítez: Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "RoughSets". *Inf. Sci.* 287: 68-89 (2014). Biblioteka zawiera implementacje algorytmów wywodzących się z teorii zbiorów przybliżonych oraz rozmytych zbiorów przybliżonych. Wkład Habilitanta w tworzenie tej biblioteki jest duży, odpowiadał za implementację wielu metod, m.in. metody konstrukcji dolnych i górnych aproksymacji pojęć, redukcji wymiarowości tablic decyzyjnych i in. Praca

opublikowana w czasopiśmie *Rough Sets and Knowledge Technology* z 2015. roku zawiera przegląd narzędzi eksploracji danych wywodzących się z teorii zbiorów przybliżonych, dostępnych w bibliotece *RoughSets*, a także rozszerzenie dla platformy RapidMiner.

Pozostałe publikacje często zawierają wcześniejsze wersje opracowywanych nowatorskich algorytmów, dotyczą różnych zastosowań. W mojej ocenie jest to typowe podejście w pracy naukowej. Pan Janusz był współredaktorem książki będącej zbiorem prac z międzynarodowej konferencji International Joint Conference on Rough Sets, IJCRS 2016, na której pełnił rolę współprzewodniczącego.

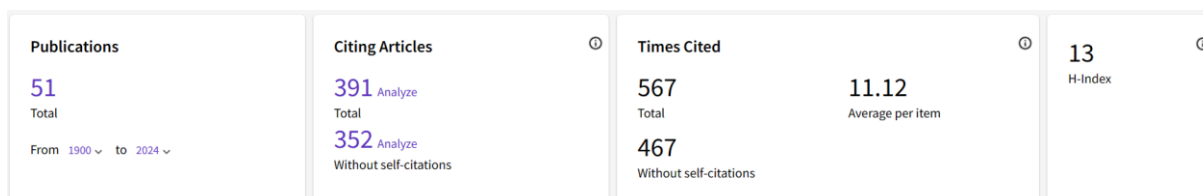
Podsumowując tę część stwierdzam, że **Habilitant ma interesujący dorobek naukowy poza głównym osiągnięciem naukowym.**

Podsumowanie ilościowe dorobku naukowego Habilitanta

Habilitant dostarczył tabelę podstawowych wskaźników bibliometrycznych, pobranych 12 czerwca 2023 roku. W bazie Web of Science jest 50 prac Habilitanta, które są cytowane 530 razy, a tzw. h-indeks wynosi 12; w bazie Scopus jest to odpowiednio 68, 891 i 17, a w bazie Google Scholar – 72, 1197 i 18. Ponieważ minęło pół roku od pobrania tych danych, to sprawdziłam stan obecny. W dniu 14.01.2024 roku zadałam pytania o dorobek Habilitanta do tych trzech baz danych dorobku naukowego, otrzymałam następujące odpowiedzi:

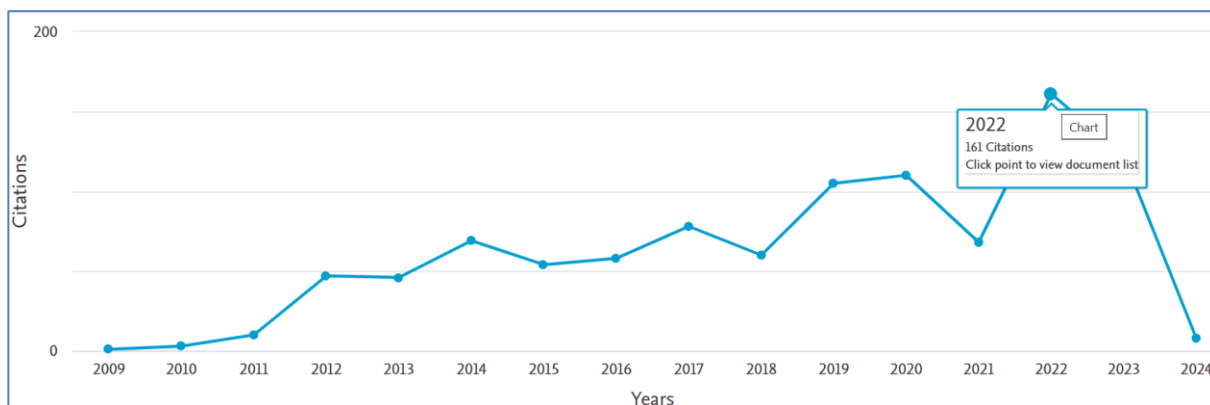
Baza Web of Science, Results for JANUSZ ANDRZEJ (Author), bez ograniczania czasowego:

Najwięcej cytowań, 108, ma praca: Riza, LS; Janusz, A; (...); Benítez, JM: Implementing algorithms of rough set theory and fuzzy rough set theory in the R package "RoughSets", Dec 10 2014, *Information Sciences* 287, pp.68-89 (są to wyniki z autocytoowaniami, nie sprawdzam stanu bez autocytoowań) (Rys. 2). Wyniki bibliometryczne Habilitanta, zawarte w tej bazie uległy poprawie w ciągu ostatniego półrocza.



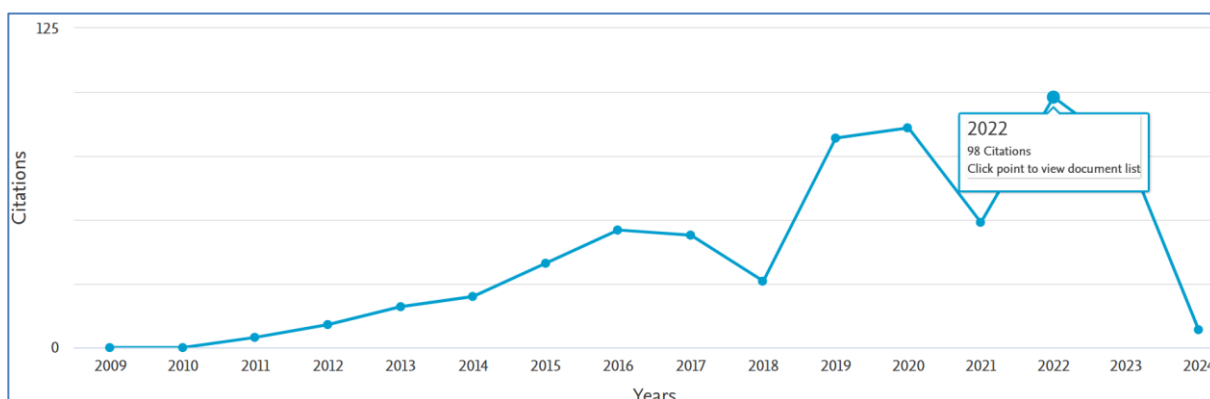
Rys. 2. Analiza cytowani dorobku Habilitanta wg bazy Web of Science, z autocytoowaniami, bez ograniczenia czasowego, dane pobrane w dniu 14.01.2024 roku.

Baza Scopus, zapytanie o prace autora 'Janusz AND A', z afiliacją ograniczoną do *University of Warsaw*. Baza generuje dane o cytowaniach za 15 lat, od 2009 roku (Habilitant obronił doktorat w 2014 roku, okres obejmuje cały dorobek naukowy po doktoracie):



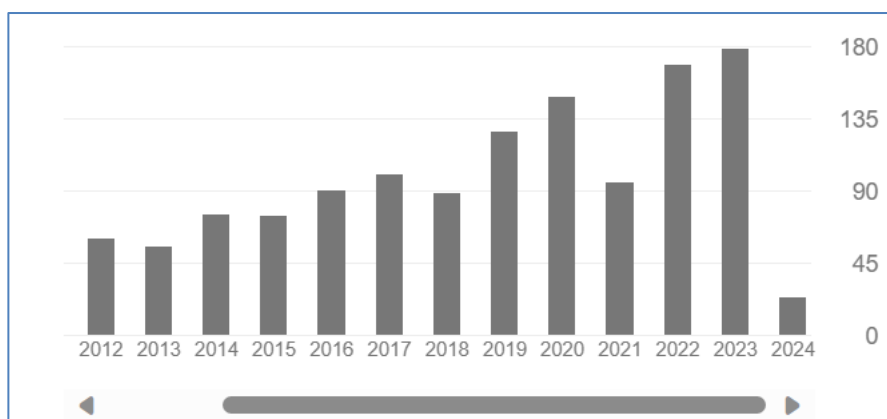
Rys. 4. Analiza cytowań dorobku Habilitanta wg bazy Scopus, z autocytowaniami, dane pobrane w dniu 14.01.2024 roku.

Otrzymałam następujące odpowiedzi: 76 documents found. Cytowania: Total: 1005, h-index: 17 (Rys. 4); bez autocytowań h-index : 13 (Rys. 3). Wyniki różnią się od wskazanych przez habilitanta, zwłaszcza liczbą znalezionych publikacji (nie 68 a 76. Niemniej patrząc na wykresy widać ciągle zainteresowanie środowiska naukowego pracami dra Andrzeja Janusza.



Rys. 3. Analiza cytowań dorobku Habilitanta wg bazy Scopus, bez autocytowań, dane pobrane w dniu 14.01.2024 roku.

Baza Scholar Google, autor Andrzej Janusz, afiliacja Uniwersytet Warszawski:



Wszystkie prace:

- cytowania: 1328;
- h-indeks: 20
- i10-indeks: 36

Od 2019 roku:

- cytowania: 745;
- h-indeks: 13
- i10-indeks: 20

Rys. 5. Cytowania Habilitanta rocznie wg bazy Scholar Google, dane pobrane 14.01.2024 roku.

W tej bazie również widać przyrost zainteresowania się pracami Habilitanta w ciągu ostatniego półrocza.

Podsumowując ten punkt recenzji stwierdzam, że dorobek publikacyjny Habilitanta wskazuje, że dr Andrzej Janusz jest rozpoznawalny w międzynarodowym środowisku naukowym. Potrafi przenosić autorskie rozwiązania w świat rzeczywistych problemów. Jego prace są licznie cytowane. Analiza załącznika 4. pozwala mi stwierdzić, że Andrzej Janusz ma istotny wkład merytoryczny w publikacjach z Jego współautorstwem.

5. Inna aktywność (naukowa i dydaktyczna)

Uczestnictwo w projektach badawczo-rozwojowych

Dr Andrzej Janusz uczestniczył w siedmiu projektach realizowanych w różnych zespołach:

1. ICRA: *Nowoczesne narzędzia inżynierskie do wspomaganie decyzji przeznaczone dla dowódców podczas działań ratowniczo-gaśniczych PSP w obiektach budowlanych*. NCBR: O ROB/0010/03/001, 2013 – 2015, konsorcjum, rola: wykonawca badań.
2. DISESOR – *Zintegrowany, szkieletowy system wspomaganie decyzji dla systemów monitorowania procesów, urządzeń i zagrożeń*. NCBR: PBS2/B9/20/2013, 2015 – 2016, konsorcjum, rola: wykonawca badań.
3. Toolbox4HR: *Opracowanie innowacyjnego systemu automatyzującego wczesne procesy rekrutacji opartego o samouczące się algorytmy i Big Data*. PARP: POIR.02.03.02-14-0009/15, 2016 – 2017, konsorcjum, rola: wykonawca badań.
4. Grail: *Opracowanie modułowego narzędzia do implementacji i optymalizacji zaawansowanej sztucznej inteligencji szerokiego zastosowania w grach wideo*. NCBR: POIR.01.02.00-00-0150/16, 2017 – 2018, realizowany w: firma Silver Bullet Labs Sp. z o.o., rola: członek kadry kluczowej projektu.
5. SENSEI: *System wspomagający rozwój umiejętności graczy*. NCBR: POIR.01.02.00-00-0184/17-00, 2018 – 2020, realizowany w: firma ESENSEI Sp. z o.o., rola: kierownik naukowy części badawczej i członek kadry kluczowej projektu.
6. LitL: *Labelling in the Loop*. *Opracowanie opartego na uczeniu maszynowym systemu automatycznego i półautomatycznego oznaczania danych w dużych zbiorach danych*. NCBR: POIR.01.01.01-00-0213/19, 2020 – 2022, rola: wykonawca badań i członek kadry kluczowej projektu.
7. BrightBox: *Narzędzia klasy explainable AI służącego do poprawy interpretowalności i przewidywalności działania metod uczących oraz diagnostyki poprawności działania wyuczonych modeli AI/ML*. NCBR: MAZOWSZE/0198/19, realizowany w: firma QED Software, rola: kierownik badań i członek kadry kluczowej projektu.

Współpraca z innymi instytucjami naukowymi

Szerzej współpraca jest omówiona w dokumentacji Habilitanta, tutaj wymienię jednostki, z którymi Habilitant współpracował i w jakim obszarze.

Współpraca w ramach projektów (uczelnie i firmy):

- ICRA: Szkoła Głównej Służby Pożarniczej w Warszawie, Centrum Naukowo-Badawcze Ochrony Przeciwpożarowej – Państwowy Instytut Badawczy, firma Dituel.

- DISESOR: Politechnika Śląska, Instytut Technik Innowacyjnych EMAG, firma Sevitel Sp. z o.o.
- Toolbox4HR: firma Toolbox for HR Sp. z o.o.
- Grail: Silver Bullet Labs Sp.
- SENSEI: ESENSEI Sp. z o.o.
- LitL: QED Software.
- BrightBox: QED Software.

Współpraca z międzynarodowymi uczelniami i organizacjami

- University of Granada w Hiszpanii, zespół kierowany przez prof. Francisco Herrera – efektywna implementacja algorytmów z teorii zbiorów przybliżonych oraz rozmytych zbiorów przybliżonych w języku R.
- Dalhousie University w Kanadzie, zespół prof. Stanisława Matwina i Politechnika Warszawska – uczenie się zanurzeń indeksów spółek z giełdy w Nowym Jorku na podstawie rekomendacji inwestycyjnych ekspertów.
- COST – European Cooperation in Science and Technology – Habilitant prowadził m.in. zajęcia w ramach letniej szkoły Advanced Big Data Training School for Life Sciences w Barcelonie. Zajęcia te dotyczyły praktycznych metod selekcji cech w wielowymiarowych zbiorach danych.
- European Alliance for Innovation (EAI) – zaproszenie w roku 2021 do poprowadzenia wykładu zaproszonego (ang. keynote speech) na konferencji EAI GOODTECHS 2021.

Wizyty naukowe

- Dalhousie University – lipiec - sierpień 2017 wizyta naukowa

Organizacja konferencji naukowych

- Przewodniczący komitetu programowego (Program Chair)
 - lata 2015 — 2017: International Symposium on Advances in Artificial Intelligence and Applications w ramach serii Conference on Computer Science and Intelligence Systems (FedCSIS)
 - Rok 2016: International Joint Conference on Rough Sets (IJCRS'16),
- Członek komitetów programowych: IEEE ICME, IEEE BigData, FedCSIS, IJCRS, WIC, RIVF, CS&P
- Przewodniczący komitetów organizacyjnych sesji specjalnych w ramach konferencji poświęconych organizowanym przez Habilitanta konkursom eksploracji danych

Osiągnięcia dydaktyczne

Dr Andrzej Janusz prowadzi dydaktykę na Wydziale Matematyki, Informatyki i Mechaniki Uniwersytetu Warszawskiego.

- Opieka nad pracami dyplomowymi:
 - Magisterskich: 9
 - Licencjackich: 2

- Promotor pomocniczy rozpraw doktorskich:
 - Marek Grzegorowski: Selected aspects of interactive feature extraction (Wybrane aspekty interaktywnej ekstrakcji cech). Rozprawa obroniona: 2 grudnia 2021.
 - Mateusz Przyborowski: Modele mikstur gaussowskich i metryka Wassersteina w zagadnieniu porównywania dużych zbiorów danych. Planowany termin obrony: rok akademicki 2023/2024.
 - Daniel Kałuża: Aktywne Uczenie Maszynowe. Planowany termin obrony: rok akademicki 2024/2025.
- Organizacja międzynarodowych konkursów eksploracji danych

Wymienię konkursy zorganizowane przez Habilitanta, nie jest to częste by jedna osoba była zaangażowana w organizację tak wielu różnorodnych, międzynarodowych konkursów.

1. PAKDD'15 Data Mining Competition: Gender Prediction Based on E-commerce Data. 330 drużyn. <https://knowledgepit.ai/pakdd15-dazainteresowaniem>.
2. AAIA'14 Data Mining Competition: Key risk factors for Polish State Fire Service. 116 drużyn. <https://knowledgepit.ai/aaia14-dm-competition/>
3. AAIA'15 Data Mining Competition: Tagging Firefighter Activities at a Fire Scene. 149 drużyn. <https://knowledgepit.ai/aaia15-data-mining-competition/>
4. IJCRS'15 Data Challenge: Mining Data from Coal Mines. 131 drużyn. <https://knowledgepit.ai/ijcrs15-data-challenge/>
5. AAIA'16 Data Mining Challenge: Predicting Dangerous Seismic Events in Active Coal Mines. 203 drużyn. <https://knowledgepit.ai/aaia16-data-mining-challenge/>
6. ISMIS'17 Data Mining Competition: Trading Based on Recommendations. 159 drużyn. <https://knowledgepit.ai/ismis17-data-mining-competition/>
7. AAIA'17 Data Mining Challenge: Helping AI to Play Hearthstone. 296 drużyn. <https://knowledgepit.ai/aaia17-data-mining-challenge/>
8. AAIA'18 Data Mining Challenge: Predicting Win-rates of Hearthstone Decks. 208 drużyn. <https://knowledgepit.ai/predicting-winrates-of-hearthstone-decks/>
9. IEEE BigData 2019 Cup: Suspicious Network Event Recognition. 249 drużyn. <https://knowledgepit.ai/suspicious-network-event-recognition/>
10. Clash Royale Challenge: How to Select Training Decks for Win-rate Prediction. 115 drużyn. <https://knowledgepit.ai/clash-royale-challenge/>
11. IEEE BigData 2020 Cup: Predicting Escalations in Customer Support. 254 drużyn. <https://knowledgepit.ai/predicting-escalations-in-customer-support/>
12. FedCSIS 2020 Challenge: Network Device Workload Prediction. 150 drużyn. <https://knowledgepit.ai/fedcsis20-challenge/>
13. IEEE BigData 2021 Cup: Predicting Victories in Video Games. 125 drużyn. <https://knowledgepit.ai/predicting-victories-in-video-games/>
14. FedCSIS 2022 Challenge: Predicting the Costs of Forwarding Contracts. 135 drużyn. <https://knowledgepit.ai/fedcsis-2022-challenge/>
15. IEEE BigData 2022 Cup: Privacy-preserving Matching of Encrypted Images. 43 drużyny. <https://knowledgepit.ai/privacy-preserving-matching-of-images/>
16. ICME 2023 Grand Challenge: Predicting Frags in Tactic Games. 55 drużyn. <https://knowledgepit.ai/icme-2023-grand-challenge/>
17. FedCSIS 2023 Challenge: Cybersecurity Threat Detection in the Behavior of IoT Devices. 85 drużyn. <https://knowledgepit.ai/fedcsis-2023-challenge/>

Habilitant regularnie organizuje mniejsze konkursy w ramach prowadzonych zajęć dydaktycznych dla studentów. Na stronie https://mimuw.fandom.com/pl/wiki/Janusz,_Andrzej są opinie studentów o zajęciach z dr. A. Januszem, bardzo pozytywne, choć nie są najnowsze.

6. Wniosek końcowy

Podsumowując niniejszą recenzję stwierdzam, że dr Andrzej Janusz konsekwentnie rozwijał metody związane z maszynowym uczeniem, skupiał się na algorytmach selekcji i konstrukcji atrybutów, oraz uczenia się zanurzeń danych. Zaproponowane algorytmy były stosowane do rozwiązywania rzeczywistych problemów, głównie podczas realizacji różnych projektów. Prace były publikowane, w tym w liczących się czasopismach naukowych. Są też licznie cytowane przez naukowców z różnych krajów. Udostępniona biblioteka *RoughSets* cieszy się popularnością, liczba pobrań jest duża i stale rośnie. Kandydat współpracuje z innymi uczelniami, organizacjami europejskimi i firmami. Równie pozytywnie oceniam pozostałą aktywność naukową, dydaktyczną i organizacyjną Habilitanta. **Biorąc pod uwagę powyższe uważam, że osiągnięcia naukowe dra Andrzeja Janusza spełniają kryteria określone w art. 219 ust. 1 pkt 2) Ustawy z dnia 20 lipca 2018r. - Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668), oraz uzasadniają nadanie stopnia doktora habilitowanego nauk inżynierijno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja. Wysoko oceniam dorobek naukowy oraz inne aktywności naukowe dra Andrzeja Janusza i wnoszę o dopuszczenie dra Andrzeja Janusza do dalszych etapów przewodu habilitacyjnego.**



Halina Kwaśnicka