

Prof. dr hab. inż. Tomasz Szmuc
Katedra Informatyki Stosowanej
Akademii Górniczo-Hutniczej
Al. Mickiewicza 30, 30-059 Kraków
tel.: (+12) 617 28 32
e-mail: tsz@agh.edu.pl

RECENZJA OSIĄGNIĘCIA NAUKOWEGO I OCENA DOROBKU NAUKOWEGO

dr inż. Teresy Mroczek

Recenzja niniejsza została przygotowana na zlecenie prof. dr hab. Idalii Kasprzyk, Prorektora ds. Kolegium Nauk Przyrodniczych Uniwersytetu Rzeszowskiego w związku z uchwałą Rady Naukowej Kolegium Nauk Przyrodniczych o powołaniu na recenzenta w postępowaniu o nadanie stopnia naukowego doktora habilitowanego dr inż. Teresie Mroczek w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja.

1. Recenzja osiągnięcia naukowego

Recenzja osiągnięcia naukowego została wykonana na podstawie dokumentacji, w tym załączonych dwóch cykli publikacji podstawowych. Na osiągnięcie naukowe składają się wyniki badań w dwóch słabo-powiązanych obszarach związanych z eksploracją danych. Pierwszy z nich zatytułowany: **Rozwój metod eksploracji danych niekompletnych** dotyczy nowych metod w przypadku niekompletnych danych. Drugi składnik osiągnięcia: **Rozwój metod dyskretyzacji danych numerycznych** opisuje badania wpływu technik dyskretyzacji (zamiana danych numerycznych na przedziały) i redukcji atrybutów na jakość i złożoność procesu uczenia.

1.1 Rozwój metod eksploracji danych niekompletnych (Osiągnięcie I)

Pierwszy składnik osiągnięcia naukowego jest reprezentowany przez 20 prac opublikowanych w latach 2016-2023. Poniżej podano (za Autoreferatem) listę uporządkowaną wg daty publikacji (od ostatnich do historycznych).

1. **T.Mroczek** (2023) *Handling the Complexity of Computing Maximal Consistent Blocks*. Electronics 12(10):2295. IF(2022):2.9. doi: 10.3390/electronics12102295
2. P.Clark, J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek**, R.Niemiec (2023) *Global and Saturated Probabilistic Approximations Based on Generalized Maximal Consistent Blocks*. Logic Journal of the IGPL 31(2):223–239. IF(2022): 0.868. doi: 10.1093/jigpal/jzac015
3. **T.Mroczek**, R.Zheng (2022) *A New Approach to Constructing Maximal Consistent Blocks for Mining Incomplete Data*. Procedia Computer Science 207:1047–1056. doi: 10.1016/j.procs.2022.09.160
4. P.Clark, J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek**, R.Niemiec (2021) *Complexity of Rule Sets Induced from Data with Many Lost Values and “Do Not Care” Conditions*. In: Abraham, A., Siarry, P., Ma, K., Kaklauskas, A. (eds) Intelligent Systems Design and Applications. ISDA 2019. Advances in Intelligent Systems and Computing 1181:376–385. Springer, Cham. doi: 10.1007/978-3-030-49342-4_36
5. P.Clark, J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek** (2021) *Mining Incomplete Data Using Global and Saturated Probabilistic Approximations Based on Characteristic Sets and Maximal Consistent Blocks*. W: Ramanna, S., Cornelis, C., Ciucci, D. (eds) Rough Sets. IJCRS 2021. Lecture Notes in Computer Science 12872:3-17. Springer, Cham. doi: 10.1007/9783-030-87334-9_1
6. P.Clark, J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek** (2021) *Complexity of Rule Sets in Mining Incomplete Data Using Characteristic Sets and Generalized Maximal Consistent Blocks*. Logic Journal of the IGPL 29(2):124–137. IF(2021): 0.868. doi: 10.1093/jigpal/jzaa041

7. P.Clark, J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek**, R.Niemiec (2020) *Complexity of Rule Sets Mined from Incomplete Data Using Probabilistic Approximations Based on Generalized Maximal Consistent Blocks*. *Procedia Computer Science* 176:1803–1812. doi: 10.1016/j.procs.2020.09.219
8. P.Clark, J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek**, R.Niemiec (2020) *Global and Saturated Probabilistic Approximations Based on Generalized Maximal Consistent Blocks*. In: de la Cal, E.A., Villar Flecha, J.R., Quintián, H., Corchado, E. (eds) *Hybrid Artificial Intelligent Systems. HAIS 2020. Lecture Notes in Computer Science* 12344:387–396. Springer, Cham. doi: 10.1007/978-3-030-61705-9_32
9. P.Clark, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2020) *Mining Data with Many Missing Attribute Values Using Global and Saturated Probabilistic Approximations Based on Characteristic Sets*. W: Lopata, A., Butkienė, R., Gudonienė, D., Sukackė, V. (eds) *Information and Software Technologies. ICIST 2020. Communications in Computer and Information Science* 1283:72–83. Springer, Cham. doi: 10.1007/978-3-03059506-7_7
10. P.Clark, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2020) *Mining Incomplete Data—A Comparison of Concept and New Global Probabilistic Approximations*. In: Czarnowski, I., Howlett, R., Jain, L. (eds) *Intelligent Decision Technologies 2019. Smart Innovation, Systems and Technologies* 142:167–178. Springer, Singapore. doi: 10.1007/978-981-13-8311-3_15
11. P.Clark, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2019) *Rule Set Complexity in Mining Incomplete Data Using Global and Saturated Probabilistic Approximations*. In: Damaševičius, R., Vasiljeviienė, G. (eds) *Information and Software Technologies. ICIST 2019. Communications in Computer and Information Science* 1078:451–462. Springer, Cham. doi: 10.1007/978-3-030-30275-7_35
12. P.Clark, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2019) *A Comparison of Global and Saturated Probabilistic Approximations Using Characteristic Sets in Mining Incomplete Data*. *INTELLI 2019: The Eighth International Conference on Intelligent Systems and Applications. IARIA*, pp. 10–15
13. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek** (2018). *A Comparison of Characteristic Sets and Generalized Maximal Consistent Blocks in Mining Incomplete Data*. In: Medina, J., et al. *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations. IPMU 2018. Communications in Computer and Information Science* 854:480–489. Springer, Cham. doi: 10.1007/978-3-319-914763_40
14. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2018). *A Comparison of Concept and Global Probabilistic Approximations Based on Mining Incomplete Data*. In: Damaševičius, R., Vasiljeviienė, G. (eds) *Information and Software Technologies. ICIST 2018. Communications in Computer and Information Science* 920:324–335. Springer, Cham. doi: 10.1007/978-3-319-99972-2_26
15. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek** (2018) *Characteristic sets and generalized maximal consistent blocks in mining incomplete data*. *Information Sciences* 453:66–79. *IF(2018)*: 5.524. doi: 10.1016/j.ins.2018.04.025
16. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2018) *Complexity of Rule Sets in Mining Incomplete Data Using Characteristic Sets and Generalized Maximal Consistent Blocks*. In: de Cos Juez, F., et al. *Hybrid Artificial Intelligent Systems. HAIS 2018. Lecture Notes in Computer Science* 10870:84–94. Springer, Cham. doi: 10.1007/978-3-31992639-1_8
17. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek**, R.Niemiec (2018) *Complexity of Rule Sets Induced by Characteristic Sets and Generalized Maximal Consistent Blocks*. In: Rutkowski, L., et al. (eds) *Artificial Intelligence and Soft Computing. ICAISC 2018. Lecture Notes in Computer Science* 10842:301-310. Springer, Cham. doi: 10.1007/978-3-319-912622_27
18. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek** (2018) *On the Number of Conditions in Mining Incomplete Data Using Characteristic Sets and Maximal Consistent Blocks*. In: Weckman,

G., Grzymała-Busse, J.W. (eds) The Fourth International Conference on Big Data, Small Data, Linked Data and Open Data. ALLDATA 2018. IARIA pp. 84–89

19. P.Clark, C.Gao, J.W.Grzymała-Busse, **T.Mroczek** (2017) *Characteristic Sets and Generalized Maximal Consistent Blocks in Mining Incomplete Data*. In: Polkowski, L., et al. Rough Sets. IJCRS 2017. Lecture Notes in Computer Science 10313:477–486. Springer, Cham. doi: 10.1007/978-3319-60837-2_39
20. J.W.Grzymała-Busse, **T.Mroczek** (2016) *Definability in Mining Incomplete Data*. Procedia Computer Science 96:179–186. doi: 10.1016/j.procs.2016.08.125

Opis Osiągnięcia I

W prowadzonych badaniach dane niekompletne dotyczą brakujących wartości atrybutów warunkowych i są dwojakiego rodzaju: dane nieistotne (*do not care condition*) i dane utracone (*lost value*) reprezentujące wartości nieznanne. O ile pierwszy rodzaj nie miał wpływu na pewne decyzje (wartość może być dowolna ze zbioru dopuszczalnych), to drugi stanowi wiedzę niekompletną i wymaga odmiennego traktowania.

W eksploracji danych niekompletnych wykorzystywane są dwa główne podejścia: bazujące na koncepcji maksymalnych bloków spójnych oraz wykorzystujące zbiory charakterystyczne. Maksymalne bloki spójne określają maksymalne zestawy obiektów nierozróżnialnych względem rozważanego zestawu atrybutów. Z kolei zbiory charakterystyczne dla danych z dowolną interpretacją brakujących wartości są kolekcjami podobnych obiektów w kontekście analizowanego zestawu atrybutów. Te dwa rodzaje zbiorów formują granule, szczególnie przy wyznaczaniu reguł wnioskowania i ocenie ich jakości.

Maksymalne bloki spójne dla danych nieistotnych zostały wprowadzone przez wcześniejszych badaczy. Badania własności tych bloków przeprowadzono m.in. w pracy [20]. Dr T. Mroczek zaproponowała [15] uogólnienie definicji maksymalnych bloków spójnych uwzględniających również dane utracone. W cytowanej pracy, bazując na maksymalnych blokach spójnych, zdefiniowano również trzy rodzaje przybliżeń probabilistycznych (B-singleton, B-podzbiór, B-koncept). Ponadto wykazano, że wspomniane przybliżenia są sobie równe i można je zredukować do jednego. W pracach [15,19] badano jakość reguł indukowanych z przybliżeń wykorzystujących wspomniane wyżej dwa rodzaje granul. Przeprowadzono liczne eksperymenty z wykorzystaniem zbiorów pobranych z repozytorium UCI (University of California Irvine). Zastosowano trzy wartości przybliżeń: dolne, środkowe (0,5) oraz górne. Dla większości zbiorów danych jakość reguł bazujących na zbiorach charakterystycznych bądź maksymalnych blokach spójnych nie różniła się istotnie. Zauważalna różnica wystąpiła przy zastosowaniu środkowego przybliżenia. W tym przypadku zastosowanie maksymalnych bloków spójnych dawało najlepsze rozwiązanie. W kolejnym kroku oceniono złożoność zbiorów reguł względem liczby warunków [18] oraz liczby reguł [17]. Wykazano, że dla zbiorów z utraconymi danymi nie ma dużych różnic w złożoności modeli uczenia. Dla zbiorów z nieistotnymi wartościami atrybutów zbiory reguł indukowane z górnych przybliżeń są prostsze (względem liczby warunków), niezależnie od stosowanych typów granul. W pracach [16, 13] przeprowadzono szereg eksperymentów z użyciem ośmiu zbiorów z repozytorium UCI zastępując 35% losowo wybranych wartości atrybutów odpowiednio wartościami utraconymi albo nieistotnymi. Uzyskano wyniki potwierdzające (na większej próbie) rezultaty wcześniejszych badań. Z kolei w [6] generowano zbiory reguł dla trzech interpretacji wartości atrybutów: utracona, nieistotna oraz atrybut-koncept. Najkrótsze zbiory reguł były indukowane przy interpretacji atrybut-koncept, natomiast najliczniejsze i najbardziej złożone były generowane dla wartości utraconych.

Istniejące metody obliczania maksymalnych bloków spójnych (*brute force*, rekurencyjna, hierarchiczna) były badane w pracy [3]. Wykazano, że niektóre z tych metod mogą generować bloki spójne, które nie są maksymalne. Wymaga to dodatkowego sprawdzania czy dodawany blok nie jest podzbiorem tworzonego, co prowadzi do złożoności NP-trudnej. W cytowanej pracy zaproponowano nowe podejście do obliczania maksymalnych bloków spójnych, efektywniejsze niż stosowana wcześniej metoda bazująca na relacji charakterystycznej. Ponadto zmodyfikowano metodę scalania przez porównywanie długości bloków i przerywanie łączenia przy pierwszym niezgodnym elemencie.

W pracy [1] oszacowano możliwą liczbę maksymalnych bloków spójnych. W tym celu zdefiniowano tzw. K-galaktykę określającą najgorszy przypadek, dla którego liczba bloków rośnie wykładniczo. Wynika stąd, że w najgorszym przypadku złożoność czasowa generowania maksymalnych bloków spójnych wynosi $O(n^n)$, gdzie n oznacza liczbę obiektów w zbiorze danych. Uzyskany wynik uzasadniał poszukiwanie efektywnych metod, w tym wykorzystujących mechanizm zrównoleglenia. W pracy [3] wykazano, że sekwencyjna aktualizacja zbioru maksymalnych bloków spójnych jest bardziej efektywna niż metoda bazująca na relacji charakterystycznej. Podstawową (ze względu na wydajność) operacją scalania bloków jest eliminacja podzbiorów w łączonych zestawach bloków spójnych. Z przechodności relacji zawierania wynika, że można realizować eliminację podzbiorów w dowolnej kolejności. Na tej podstawie w [1] zaproponowano równoległe przetwarzanie połączonych zestawów bloków spójnych. Obliczenia lokalnych bloków spójnych i procedura scalania bloków rozdzielane są na odrębne, odpowiednio synchronizowane zadania. Tym samym uzyskano istotne przyspieszenie czasu obliczeń.

W pracach [14,12] badano globalne i nasycone przybliżenia probabilistyczne określone dla zbiorów charakterystycznych oraz maksymalnych bloków spójnych. Obliczenia przybliżeń charakteryzują się wykładniczą złożonością czasową, stąd w badaniach eksperymentalnych korzystano z wersji heurystycznej bazującej na znanym algorytmie indukcji reguł MLEM2 z parametrem α . Wykonano szereg eksperymentów konfrontując wyniki proponowanych podejść dla różnych interpretacji brakujących atrybutów z rezultatami otrzymanymi dla wspomnianych przybliżeń probabilistycznych. Badania przeprowadzono z wykorzystaniem ośmiu wybranych zbiorów z repozytorium UCI. Generowano różne wersje wspomnianych zbiorów zastępując 35% losowo wybranych wartości atrybutów symbolami wartości utraconych lub nieistotnych. Celem tych badań była ocena proponowanych metod eksploracji danych. Porównywano jakość klasyfikacji indukowanych reguł i ich złożoność.

Ocena Osiągnięcia I

Główne elementy osiągnięcia naukowego scharakteryzowano poniżej.

1. Uogólnienie definicji maksymalnych bloków spójnych umożliwiające uwzględnianie utraconych wartości (*lost value*) atrybutów. Zdaniem recenzenta jest to istotne rozszerzenie względem wcześniejszej definicji dotyczącej wartości nieistotnych (*do not care condition*). Otworzyło to nowy obszar badań, istotny z punktu widzenia zastosowań praktycznych.
2. Wykonanie szczegółowej analizy własności uogólnionej definicji maksymalnych bloków spójnych. Oszacowano liczbę maksymalnych bloków spójnych oraz złożoność czasową ich obliczania. Ponadto wykazano, że niektóre metody generują bloki spójne, które nie są maksymalne.
3. Zaproponowanie nowej metody obliczania maksymalnych bloków spójnych, w tym przez odpowiednie zrównoleglenie obliczeń.
4. Opracowanie trzech rodzajów (singleton, podzbiór, koncept) przybliżeń probabilistycznych oraz ich weryfikacja względem jakości klasyfikacji i złożoności indukowanych reguł.
5. Zdefiniowanie dwóch typów przybliżeń probabilistycznych: globalne i nasycone oraz przeprowadzenie szerokich badań w kontekście jakości klasyfikacji oraz złożoności indukowanych reguł.

Wymienione elementy są wartościowe i stanowią ważny wkład w obszarze przetwarzania niekompletnych danych. Ogólnie pozytywnemu podsumowaniu towarzyszą poniższe uwagi.

1. Odczuwa się brak szerszego kontekstu porównania opracowanych metod z innymi stosowanymi poza środowiskiem działającym w szeroko rozumianym obszarze zbiorów przybliżonych. W szczególności dotyczy to odniesienia do szerokiej gamy metod klasyfikacji. Z punktu widzenia zastosowań cenne byłoby wskazanie podsumowania określającego rodzaje zbiorów, w których zastosowane metody wykazują przewagę nad istniejącymi.
2. Lista 20 publikacji tworzących osiągnięcie naukowe składa z 4 opublikowanych w czasopiśmie z listy JCR pozostałe w większości w wydawnictwach (post)konferencyjnych. Skutkuje to częstym rozdrobnieniem przyrostu nowych treści. Według recenzenta zwyczajowa liczba ok. 10 publikacji zapewnia lepsze rozłożenie treści opisujących osiągnięcie naukowe.

3. W [1] wyłącznym autorem jest dr T. Mroczek, natomiast praca [3] współautorem jest R. Zheng. W oświadczeniu współautora publikacji [3] wskazano indywidualny wkład Habilitantki w zakresie koncepcji artykułu oraz opracowania algorytmu sekwencyjnej aktualizacji maksymalnych bloków spójnych. Publikacja [1] zawiera głównie propozycję estymacji liczby maksymalnych bloków spójnych i koncepcję zrównoleglenia obliczeń maksymalnych bloków spójnych. W pozostałych 18 współautorami są (w różnych konfiguracjach) członkowie zespołu kierowanego przez prof. Jerzego W. Grzymała-Busse. W załączonych deklaracjach współautorów specyfikacje udziału Habilitantki bywają zbyt ogólne. W prawie wszystkich osiemnastu pozycjach wymienia się indywidualne prowadzenie prac eksperymentalnych oraz współautorstwo w opracowaniu: koncepcji artykułu i wybranych rozdziałów. Brak jednoznacznego potwierdzenia indywidualnego wkładu dla wszystkich wymienionych wcześniej osiągnięć Habilitantki utrudnia ocenę. Wynika to zapewne z charakteru pracy zespołowej, gdzie niekiedy trudne jest precyzyjne rozgraniczenie wkładu współautorów.

1.2 Rozwój metod dyskretyzacji danych numerycznych (Osiągnięcie II)

Drugi składnik osiągnięcia naukowego jest reprezentowany przez 5 prac opublikowanych w latach 2015-2019. Poniżej podano (za Autoreferatem) listę uporządkowaną wg daty publikacji (od ostatnich do historycznych).

1. J.W.Grzymała-Busse, Z.S.Hippe, **T.Mroczek** (2019) *Reduced Data Sets and Entropy-Based Discretization*. Entropy 21(11):1051. IF(2019): 2.494. doi: 10.3390/e21111051
2. J.W.Grzymała-Busse, **T.Mroczek** (2018) *Attribute Selection Based on Reduction of Numerical Attributes During Discretization*. In: Stańczyk, U., Zielosko, B., Jain, L. (eds) *Advances in Feature Selection for Data and Pattern Recognition*. Intelligent Systems Reference Library 138:13–24. Springer, Cham. doi: 978-3-319-67588-6_2
3. J.W.Grzymała-Busse, **T.Mroczek** (2018) *Merging of Numerical Intervals in Entropy-Based Discretization*. Entropy 20(11):880. IF(2018): 2.419. doi: 10.3390/e20110880
4. J.W.Grzymała-Busse, **T.Mroczek** (2016) *A Comparison of Four Approaches to Discretization Based on Entropy*. Entropy 18(3):69. IF(2016): 1.821. doi: 10.3390/e18030069
5. J.W.Grzymała-Busse, **T.Mroczek** (2015) *A Comparison of Two Approaches to Discretization: Multiple Scanning and C4.5*. In: Kryszkiewicz, M., Bandyopadhyay, S., Rybinski, H., Pal, S. (eds) *Pattern Recognition and Machine Intelligence. PReMI 2015. Lecture Notes in Computer Science 9124:301–310*. Springer, Cham. doi: 10.1007/978-3-319-19941-2_5

Opis i ocena Osiągnięcia II

Dyskretyzacja danych numerycznych, redukcja przedziałów i atrybutów umożliwia niekiedy istotne poprawienie efektywności i jakości procesu uczenia. Badania w tym obszarze prowadzone są od lat dwudziestych ubiegłego wieku. Opracowano wiele metod, w szczególności bazujących na entropii warunkowej. Brak jednoznacznej oceny wpływu redukcji w połączeniu z dyskretyzacją stanowił motywację do podjęcia badań tworzących Osiągnięcie II.

Teoria zbiorów przybliżonych pozwala wyznaczać redukt, czyli minimalne zbiory atrybutów reprezentujące tę samą informację co cały zbiór atrybutów. Algorytm wyznaczania wszystkich (często dużej liczby) reduktów charakteryzuje się wykładniczą złożonością, stąd w [1] zaproponowano heurystyczne procesy obliczania tzw. lewych i prawych reduktów. Przeprowadzono szereg eksperymentów zbiorów danych po ich uprzedniej dyskretyzacji znanymi metodami globalnymi. Porównywano wyniki zbiorów zredukowanych (lewy albo prawy redukt) z niezredukowanymi. Wykazano, że redukcja zbiorów zdyskretyzowanych zwiększa poziom błędów, a ponadto generowane drzewa decyzyjne dla zbiorów zredukowanych nie są prostsze.

Dyskretyzacja danych numerycznych pozwala usunąć te atrybuty, których cała dziedzina jest zawarta w jednym przedziale. W pracy [2] porównano jakość klasyfikacji numerycznych zbiorów oryginalnych i zredukowanych z zastosowaniem dyskretyzacji atrybutów dominujących. Poziom błąd dla zbiorów zredukowanych był istotnie większy, natomiast drzewa decyzyjne znacznie prostsze.

W pracach [5,4] badano wpływ technik dyskretyzacji bazujących na entropii na jakość klasyfikacji oraz złożoność procesu uczenia. W pierwszej z nich [5] porównano dwie metody dyskretyzacji – skanowanie wielokrotne połączone z systemem C4.5 oraz techniki dyskretyzacji wewnętrznej w C4.5. Z kolei w [4] porównano cztery metody dyskretyzacji: występująca w C4.5 oraz trzy globalne: skanowania wielokrotnego, równej długości przedziałów oraz częstości przedziałów. Wykazano, że technika skanowania wielokrotnego jest istotnie lepsza od pozostałych analizowanych. Ponadto drzewa decyzyjne generowane z danych dyskretyzowanych tą metodą są prostsze w porównaniu z generowanymi bezpośrednio przez C4.5. Dodatkowo stwierdzono, że różnice w wydajności analizowanych metod są nieistotne statystycznie.

Wpływ łączenia przedziałów na jakość klasyfikacji badano pracy [3]. Na podstawie oceny przeprowadzonej we wcześniejszych pracach [4,5] wybrano znaną metodę skanowania wielokrotnego do oceny wpływu łączenia przedziałów na jakość klasyfikacji. Porównywano łączenie bazujące na minimalnej wartości albo maksymalnej wartości entropii warunkowej oraz dyskretyzację bez łączenia przedziałów. Podobnie jak w poprzednich pracach korzystano ze zbiorów danych numerycznych UCI. Stwierdzono, że dla trzech wspomnianych podejść różnice uzyskanych wyników nie są statystycznie istotne. Kontynuując badania oddzielnie dla poszczególnych zbiorów danych wykazano, że istnieją różnice na poziomie istotności 1%.

W ramach badań tworzących Osiągnięcie II przeprowadzono szereg eksperymentów w celu oceny istniejących metod dyskretyzacji danych numerycznych oraz ich łączenia z redukcją atrybutów. Wyniki tych eksperymentów stanowiły podstawę o oceny. W syntetycznym ujęciu podstawowe rezultaty składające się on oceniane osiągnięcie naukowe przedstawiono poniżej.

1. Wykazano eksperymentalnie, że redukcja atrybutów (lewy albo prawy redukt) wraz z dyskretyzacją atrybutów numerycznych nie wpływa na jakość klasyfikacji. Nie stwierdzono również wpływu na uproszczenie generowanych drzew decyzyjnych.
2. Oceniono eksperymentalnie wpływ redukcji przez selekcję atrybutów podczas dyskretyzacji na jakość klasyfikacji.
3. Zbadano wpływ technik dyskretyzacji bazujących na entropii na jakość klasyfikacji i procesu uczenia.
4. Przeprowadzono badania wpływu łączenia przedziałów dyskretyzacji na jakość klasyfikacji.

Są to wyniki wartościowe, osiągnięte drogą starannie przygotowanych eksperymentów i ich statystycznej oceny. Przedstawione publikacje są w większości (za wyjątkiem [1]) dwuautorskie. W każdej na pierwszym miejscu wymieniony jest Jerzy W. Grzymała-Busse. W prezentowanych pracach istotną rolę odgrywało planowanie, kierowanie i weryfikacja eksperymentów. W tym aspekcie rola Habilitantki jest wiodąca, co potwierdziły oświadczenia współautorów. Trzy pozycje [1,3,4] opublikowane w czasopiśmie *Entropy* (MDPI, 100 punktów według wykazu MNiSzW), publikacja [2] ukazała się jako rozdział w serii *Intelligent Systems Reference Library* (Springer), natomiast [5] jest zamieszczona w materiałach międzynarodowej konferencji opublikowanych w popularnej serii *Lecture Notes in Computer Science* (Springer). Wymienione wydawnictwa są rozpoznawalne w międzynarodowym środowisku naukowców.

1.3 Podsumowanie

Oceniane osiągnięcie naukowe przedstawiono w postaci dwóch składników zatytułowanych odpowiednio: **Rozwój metod eksploracji danych niekompletnych** (Osiągnięcie I) oraz **Rozwój metod dyskretyzacji danych numerycznych** (Osiągnięcie II). Te dwa obszary badawcze są słabo powiązane, lecz są komplementarne w aspekcie metod analizy zbiorów danych. Istnieje jeszcze pewien pierwiastek wspólny – wykorzystanie elementów zbiorów przybliżonych w badaniach. Biorąc pod uwagę powyższe można było sformułować ogólny tytuł osiągnięcia, pokrywający dwa zaproponowane oraz odpowiednio ustrukturyzować opis. Jest to zabieg techniczny, który, jeśli zastosowany, poprawiłby prezentację osiągnięcia naukowego w sensie odnośnej ustawy. Z kolei powstaje pytanie, czy nie byłoby korzystniej pogłębić badania w jednym ze składników np. szersze odniesienie do metod głębokiego uczenia i tym samym przedstawić jeden, odpowiednio rozwinięty składnik.

Powyższa uwaga ma charakter dyskusyjny nie umniejsza pozytywnych ocen Osiągnięcia I i II zamieszczonych w powyższych podrozdziałach. Można zatem stwierdzić, że przedstawione dwa cykle prac składają się na dwa komplementarne, spójne wewnętrznie składniki tworzące osiągnięcie naukowe wymienione w art. 219 pkt. 2b *Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce* (Dz.U. 2018 poz. 1668, z późniejszymi zmianami).

2. Ocena dorobku naukowego

Działalność naukowa Habilitantki jest różnorodna i zawiera się w obszarze sztucznej inteligencji. W ostatnich latach badania koncentrowały się na zagadnieniach analizy zbiorów danych w tym eksploracji danych niekompletnych oraz dyskretyzacji danych numerycznych – tworzących wyżej oceniane osiągnięcie naukowe. Badania nie wymienione w osiągnięciu naukowym świadczą o szerokiej gamie zainteresowań naukowych.

W ujęciu ilościowym dorobek Habilitantki scharakteryzowano poniżej. W poszczególnych podkategoriach podano liczby po uzyskaniu stopnia uzupełniając w nawiasach liczbę przed tym stopniem. Podane dane nie uwzględniają pozycji stanowiących Osiągnięcia I i II.

1. Monografie naukowe – 1.
2. Rozdziały w monografiach – 1(16).
3. Artykuły w czasopismach – 12(1).
4. Prace w materiałach konferencyjnych – 5(8).
5. Wykłady/warsztaty zaproszone – b/d.
6. Wystąpienia na konferencjach – 8 (1).

Łączna liczba publikacji wynosi 69, co jest znaczące ilościowo. Słabiej prezentuje się ranga części wydawnictw oraz sumaryczny współczynnik *Impact Factor* (IF) dla całego dorobku, który wynosi 24,27. Liczba cytowań (w nawiasach bez autocytaowań) kształtuje się dla poszczególnych baz: Google Scholar – 383(b/d), Web of Science – 111(71), Scopus – 204(104). Indeks Hirscha wynosi odpowiednio: Google Scholar – 11 Web of Science – 6, Scopus – 7.

Wartości danych nauko-metrycznych są na akceptowalnym poziomie, choć nieco niższym od zwyczajowo przyjętego.

Dr T. Mroczek była współredaktorem 3 monografii wydawnictwa Springer, recenzentem w czasopiśmie: *Information Sciences, Applied Soft Computing, Entropy* i *Fundamenta Informaticae* oraz recenzowała prace na wielu konferencjach międzynarodowych. Była członkiem rady redakcyjnej czasopisma *Human-Intelligent Systems Integration*.

Aktywność naukowa realizowana w więcej niż jednej uczelni

Badania naukowe związane z osiągnięciem naukowym były prowadzone przy współpracy z Katedrą Elektrotechniki i Informatyki Uniwersytetu w Kansas. Należy jednak dodać, że osobą spinającą współpracę był prof. Jerzy W. Grzymała-Busse (współautor większości publikacji Osiągnięcia I i II), zatrudniony również w Wyższej Szkole Informatyki i Zarządzania – miejscu pracy Habilitantki. W tym kontekście trudno rozstrzygnąć w ramach której afiliacji Prof. J.W. Grzymała-Busse była prowadzone badania Habilitantki. Nie jest to konieczne, ze względu na współpracę z innymi instytucjami, której efekty są wystarczające do spełnienia odpowiedniego wymogu ustawy. Wymienić tu należy współpracę z:

- Uniwersytetem Rzeszowskim i Dom Opieki nad Osobami Starszymi i utworzenie systemu detekcji upadków osób
- Ośrodkami naukowymi we Francji Fastlite, Amplitude (CNRS) oraz w Niemczech (JenLab GmbH, Friedrich-Alexander Universität Erlangen-Nürnberg) przy opracowywaniu nowych źródeł światła w ramach projektu Horyzont Europa
- Konsorcjum uniwersytetów europejskich realizującym projekt (EIT): *Deep Tech in Higher Education Institutions and Ecosystems through Entrepreneurial Education+*. Habilitantka była członkiem zespołu sztucznej inteligencji

- Zespołem ekonomistów (Katedra Przedsiębiorczości i Innowacji Społecznej Uniwersytetu Pedagogicznego w Krakowie, Katedra Finansów, Bankowości i Rachunkowości Politechniki Rzeszowskiej) w obszarze zastosowań metod uczenia maszynowego w eksploracji danych wybranych problemów ekonomicznych. Wyniki tej interdyscyplinarnej współpracy zostały udokumentowane m.in. w 9 publikacjach.

3. Osiągnięcia dydaktyczne, organizacyjne i popularyzujące naukę

Działalność dydaktyczna dr T. Mroczek jest bogata i różnorodna. Habilitantka była promotorem 111 prac inżynierskich i 23 magisterskich. Praca magisterska nt. predykcji notowań akcji spółek indeksu WIG20 została wyróżniona w konkursie o nagrodę prezesa Zarządu Giełdy Papierów Wartościowych. Habilitantka prowadziła wykłady z 5 przedmiotów (Wstęp do informatyki, Bazy danych, Sztuczna inteligencja, Technologie internetowe, Eksploracja danych) oraz szereg ćwiczeń laboratoryjnych i projektowych. Opracowała dwa kursy e-learningowe, jest również współautorem skryptu z obszaru eksploracji danych.

Habilitantka pełniła funkcje organizacyjne (z-ca Kierownika Katedry Sztucznej Inteligencji, z-ca Kierownika Katedry Systemów Rozproszonych), była Pełnomocnikiem Prorektora Rektora ds. Nauki na Wydziale Informatyki Stosowanej, członkiem Wydziałowego Zespołu ds. Jakości Kształcenia, opiekunem kierunku Informatyka, koordynatorem projektu w ramach programu pomocy osobom z upośledzeniem słuchu.

W ramach działalności popularyzujących naukę prezentowała artykuły popularno-naukowe na blogu Uczelni i uczestniczyła w cyklach: *Ciekawa lekcja* oraz *Przybij piątkę nauce*.

Działalność naukowa dr T. Mroczek była wyróżniona przez władze WSiIZ (dwie nagrody indywidualne i dwie zespołowe). Wymienić również należy Best Paper Awards *on the Fourth International Conference on Big Data, Small Data, Linked Data, and Open Data* (2018) oraz wyróżnienie zespołowe w konkursie *Contest for the Most Influential Article on Rough Sets Co-authored by Polish Researches in 2020-2021*.

4. Wniosek końcowy

W przedstawionym osiągnięciu naukowym zaproponowano nowe metody eksploracji danych niekompletnych oraz przeprowadzono analizę wpływu redukcji atrybutów i/lub dyskretyzacji danych numerycznych na jakość klasyfikacji i procesu uczenia. Wyniki uzyskano i potwierdzono drogą licznych starannie przygotowanych i zweryfikowanych eksperymentów. Uzyskane rezultaty stanowią ważny i wartościowy wkład w obszarze metod eksploracji danych.

Wyniki badań nie wchodzących w skład osiągnięcia naukowego stanowią jego uzupełnienie w kontekście wcześniejszych etapów badań. Pokazują również szeroki zakres zainteresowań badawczych w różnorodnych obszarach sztucznej inteligencji. Na podkreślenie zasługuje również współpraca z instytucjami zewnętrznymi, której rezultatem były praktyczne rozwiązania.

Działalność dydaktyczna, w tym różnorodność tematów prowadzonych prac inżynierskich/magisterskich stanowi cenne uzupełnienie profilu Habilitantki.

Podsumowanie oceny osiągnięcia naukowego, dorobku po uzyskaniu stopnia doktora oraz działalności około-naukowej pozwala na stwierdzenie, że spełnione są wymagania dotyczące uzyskania stopnia naukowego doktora habilitowanego (Ustawa z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce - Dz.U. 2018 poz. 1668, z późniejszymi zmianami), co prowadzi do wniosku o dopuszczenie **dr inż. Teresy Mroczek** do kolejnego etapu postępowania habilitacyjnego w celu nadania stopnia naukowego doktora habilitowanego nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.