

Poznań, 16 kwietnia 2020 r.

Dr hab. inż. Szymon Wilk
Wydział Informatyki i Telekomunikacji
Politechnika Poznańska
ul. Piotrowo 2
60-965 Poznań

Recenzja cyklu i dorobku naukowego, dydaktycznego oraz organizacyjnego dr inż. Agnieszki Duraj

Tytuł osiągnięcia naukowego: Innowacyjne metody detekcji wyjątków w zbiorach danych różnego typu

Niniejsza recenzja została przygotowana w ramach przewodu habilitacyjnego prowadzonego przez Radę Wydziału Fizyki Technicznej, Informatyki i Matematyki Stosowanej Politechniki Łódzkiej. Wykonałem ją na podstawie pisma Dziekana Wydziału prof. dr. hab. Piotra Liczberskiego w związku z powołaniem mnie przez Centralną Komisję ds. Stopni i Tytułów Naukowych na recenzenta w tym postępowaniu. Przedmiotem recenzji jest cykl publikacji naukowych oraz dorobek naukowy, dydaktyczny i organizacyjny dr inż. Agnieszki Duraj, pracownicy Wydziału Fizyki Technicznej, Informatyki i Matematyki Stosowanej Politechniki Łódzkiej. Oceny dokonałem zgodnie z kryteriami oceny zawartymi w Rozporządzeniu Ministra Nauki i Szkolnictwa Wyższego z dnia 1.09.2011 r. w sprawie kryteriów oceny osiągnięć osoby ubiegającej się o nadanie stopnia doktora habilitowanego (Dz. U. poz. 1165).

Na przedłożoną do oceny dokumentację postępowania habilitacyjnego składają się:

- kopia dyplomu doktorskiego,
- autoreferat w języku polskim i angielskim,
- wykaz dorobku naukowego oraz informacja o osiągnięciach organizacyjnych, dydaktycznych, popularyzatorskich i współpracy międzynarodowej,
- monografia oraz kopie publikacji stanowiących osiągnięcie naukowe wraz z oświadczeniami współautorów,
- dane personalne i kontaktowe Habilitantki.

1. Sylwetka habilitantki

Dr inż. Agnieszka Duraj ukończyła w 1998 r. studia magisterskie na kierunku Informatyka (Wydział Fizyki Technicznej, Informatyki i Matematyki Stosowanej Politechniki Łódzkiej) przygotowując pracę magisterską zatytułowaną *Komputerowe wspomaganie diagnostyki medycznej*. Następnie, w 2007 r. Rada Wydziału Informatyki, Elektrotechniki i Automatyki Uniwersytetu Zielonogórskiego nadała Jej stopień doktora nauk technicznych za rozprawę *Algoritmy rozpoznawania zespołu QRS w sygnałach elektrokardiograficznych pochodzących od pacjentów z wszczepionym układem stymulującym*. Rozprawa ta otrzymała wyróżnienie Polskiego Towarzystwa Zastosowań Elektromagnetyzmu.

Po otrzymaniu stopnia magistra inżyniera Habilitantka rozpoczęła pracę na stanowisku asystenta w Instytucie Informatyki Politechniki Łódzkiej, a po obronie doktoratu na stanowisku adiunkta, na którym dotąd pracuje. Dodatkowo, w latach 2009-2016 Habilitantka prowadziła zajęcia w Zakładzie Informatyki Państwowej Wyższej Szkoły Zawodowej we Włocławku.

Zainteresowania naukowe Habilitantki koncentrują się na dwóch grupach zagadnień – wspomaganie decyzji klinicznych ze szczególnym uwzględnieniem sygnałów biomedycznych oraz detekcji wyjątków w numerycznych, tekstowych i mieszanych zbiorach danych. Tej ostatniej grupie zagadnień poświęcone jest przedłożone osiągnięcie naukowe.

2. Ocena osiągnięcia naukowego

Przetawione do oceny osiągnięcie naukowe obejmuje 10 publikacji wymienionych poniżej (zastosowano oznaczenia oraz kolejność zaproponowaną w autoreferacie):

- M.1. Duraj A. (2019): Metody analizy danych z detekcją wyjątków. Akademicka Oficyna Wydawnicza EXIT, Warszawa.
- A.1. Duraj A., Niewiadomski A., Szczepaniak P.S. (2019): Detection of outlier information by the use of linguistic summaries based on classic and interval-valued fuzzy sets. *International Journal of Intelligent Systems*, 34, 415-438.
- A.2. Duraj A., Niewiadomski A., Szczepaniak P.S. (2019): Outlier detection using linguistically quantified statements. *International Journal of Intelligent Systems*, 33, 1858-1868.
- A.3. Szczepaniak P.S., Duraj A. (2018): Case-based reasoning: the search for similar solutions and identification of outliers. *Complexity*, 2018, 9280787.
- A.4. Duraj A., Krawczyk A. (2011): Detekcja wyjątków sygnałów biomedycznych w systemach fuzji informacji. *Przegląd Elektrotechniczny*, 12b, 128-130.
- A.5. Duraj A., Krawczyk A. (2011): Dobór miar odległości w hierarchicznych aglomeracyjnych metodach wykrywania wyjątków. *Przegląd Elektrotechniczny*, 12b, 33-36.
- A.6. Duraj A., Krawczyk A. (2010): Finding outliers for large medical datasets. *Przegląd Elektrotechniczny*, 12, 188-191.
- R.1. Duraj A., Szczepaniak P.S. (2017): Information outliers and their detection. [W] Burgin M., Hofkirchner W. (red.): *Information Studies and the Quest for Transdisciplinarity*. World Scientific, 413-436
- WoS.1 Duraj A., Szczepaniak P.S., Ochelska-Mierzejewska J. (2016): Detection of outlier information using linguistic summarization. [W] Andreasen T. et al. (red.): *Flexible Query Answering Systems 2015. Advances in Intelligent Systems and Computing*, vol 400. Springer, Cham, 101-113.
- WoS.2 Duraj A., Zakrzewska D. (2015) Effective outlier detection technique with adaptive choice of input parameters. [W] Angelov P. et al. (red.). *Intelligent Systems'2014. Advances in Intelligent Systems and Computing*, vol 322. Springer, Cham, 535-546.

Sześć spośród przedstawionych publikacji zostało opublikowanych w czasopiśmie z listy JCR, a ich sumaryczny *impact factor (IF)* wynosi 9.296. Poza monografią wszystkie pozostałe pozycje z cyklu są wieloautorskie, jednak liczba autorów jest mała (2-3 osoby) i zgodnie z przedstawionym oświadczeniem w większości przypadków wkład Habilitantki jest znaczny (od 50 do 90%). Nie jest do końca jasne, jaką rolę w cyklu pełną prace opublikowane w *Przeglądzie Elektrotechnicznym* (A.4, A.5 i A.6). Habilitantka wyjaśnia, że mają one „charakter wspomagający i poszerzają audytorium czytelników”. Opisują one jednak zagadnienia, które nie zostały omówione w monografii, np. detekcję wyjątków w systemach fuzji informacji, należy więc chyba przyjąć, że mimo wszystko stanowią one istotną część osiągnięcia.

Monografia, stanowiąca najbardziej istotny element osiągnięcia, jest napisana w języku polskim i liczy 192 strony. Składa się ona z 8 rozdziałów, w tym 3 rozdziałów obejmujących wprowadzenie i podsumowanie oraz 5 zawierających wyniki, spisu tabel i rysunków oraz bibliografii liczącej 228 pozycji. Najbardziej istotne z punktu widzenia osiągnięcia są rozdziały 4, 5 i 6, które przedstawiają oryginalne propozycje Habilitantki dotyczące różnych podejść

do wykrywania wyjątków w danych. W dalszej części tego punktu omówię krótko poszczególne rozdziały z wynikami skupiając się na tych, które prezentują nowatorskie rozwiązania.

Rozdział 3 przedstawia istniejące metody stosujące techniki regresji, klasyfikacji i grupowania do wykrywania wyjątków. W pierwszej części Habilitantka szczegółowo omawia wybrane metody (m.in. regresję liniową, naiwny klasyfikator Bayesa, klasyfikator k -NN, algorytmy k -means i DBSCAN). Następnie przedstawione są przykłady ilustrujące zastosowanie części z opisanych metod, a także większe studium przypadku poświęcone wykrywaniu wyjątków w danych klinicznych pochodzących z repozytorium UCI. Należy wspomnieć, że ocenę eksperymentalną przedstawionych metod uzupełniają prace A.6 i WoS.2, poświęcone porównaniu algorytmów gęstościowych, oraz A.5 analizująca wpływ metryk odległości na działanie aglomeracyjnych metod grupowania. Ponadto w pracy A.5 zaproponowano schemat wykrywania wyjątków w systemach fuzji danych, gdzie wybrane metody są stosowane lokalnie dla poszczególnych źródeł danych, a następnie na poziomie globalnym.

Rozdział 4 prezentuje metody wykrywania wyjątków przy użyciu podsumowań lingwistycznych. Zaproponowane metody zostały opisane także w pracach A.1, A.2, R.1 oraz WoS.1 i można je uznać za najbardziej istotny element osiągnięcia. Podsumowania lingwistyczne pozwalają na wykrywanie wyjątków w danych numerycznych, tekstowych oraz mieszanych, co jest szczególnie wartościowe z praktycznego punktu widzenia (w bardzo wielu problemach i zastosowaniach pojawiają się dane heterogeniczne). W odróżnieniu od metod przedstawionych w rozdziale 3 nie identyfikują one konkretnych przypadków jako wyjątków, ale dostarczają one wartościowanych stwierdzeń na temat występowania wyjątków w analizowanym zbiorze danych. Wykorzystywane do konstrukcji podsumowań kwantyfikatory oraz zmienne lingwistyczne mogą być modelowane zarówno za pomocą klasycznych, jak i przedziałowych zbiorów rozmytych, co pozwala na modelowanie niepewności w bardziej intuicyjny sposób. Rozdział ten kończy się przykładem prezentującym zastosowanie zaproponowanych metod do wykrywania wyjątków w danych numerycznych i tekstowych.

W rozdziale 5 opisano zastosowanie algorytmów genetycznych do wykrywania wyjątków. Problem ten potraktowano jako zadanie optymalizacji wielokryterialnej, dzięki czemu zwiększa się uniwersalność podejścia i jego stabilność. Jest to kolejny istotny element osiągnięcia, który został także opisany w kilku publikacjach Habilitantki nie wchodzących w skład cyklu. W zaproponowanym podejściu jako funkcje celu wykorzystywane różne miary identyfikacji wyjątków, np. średnia odległość od innych przykładów, czy odległość od środka centroidu. W rozdziale przedstawiono wykorzystanie zaproponowanego podejścia do wykrywania wyjątków w dwóch medycznych zbiorach danych pochodzących z repozytorium UCI. W eksperymentach wykorzystano 3 dobrze znane algorytmy genetyczne – NSGA-II, PESA-II oraz SPEA-II oraz zestaw 5 kryteriów, przy czym najlepszą dokładność detekcji uzyskano dla pierwszego z algorytmów.

W rozdziale 6 opisano wykrywanie wyjątków za pomocą metod wnioskowania na przypadkach. Rozdział ten bazuje na pracy A.3 i stanowi kolejny istotny element dokonania. Zaproponowane podejście wykorzystuje mechanizm dopasowania nowych obiektów do zapamiętanych przypadków, przy czym Habilitantka zaproponowała definicje przypadków wyjątkowych dla naiwnego klasyfikatora Bayes, klasyfikatora odległościowy k -NN oraz regresji liniową. Rozdział ten kończy się przykładem, w którym zaproponowane podejście stosowane jest do danych klinicznych pochodzących z repozytorium UCI. Porównano je przy tym z tradycyjnymi podejściami opisanymi w rozdziale 3, wskazując na wyższą skuteczność nowej propozycji.

Rozdział 7 poświęcony jest wykorzystaniu wybranych metod statystycznych i odległościowych do wykrywania wyjątków w strumieniach danych. Nie prezentuje on nowych rozwiązań, stanowi jednak dobre uzupełnienie tematyki poruszonej w monografii. Podobnie jak wcześniejsze, także ta część kończy się przykładami, gdzie przedstawione metody

wykorzystane są do analizy danych ekonomicznych, przy czym w jednym z przykładów za wyjątki uznaje się szum obecny w danych.

Tematyka wykrywania wyjątków, której poświęcono oceniane osiągnięcie, jest niewątpliwie aktualna oraz ważna, zarówno z teoretycznego, jak i praktycznego punktu widzenia. Dlatego też uważam, że zaproponowane przez dr Duraj metody do wykrywania wyjątków za pomocą podsumowań lingwistycznych, algorytmów genetycznych i wnioskowania na przypadkach należy uznać za zauważalny wkład w rozwój dziedziny. Przy tej okazji chciałbym podkreślić fakt przygotowania monografii, która pozwoliła na kompleksową prezentację uzyskanych wyników rozszerzoną o przegląd metod do wykrywania wyjątków w strumieniach danych. Monografia pozwoliła też na rozwianie wątpliwości związanych z oceną wkładu Habilitantki, jakie mogły pojawić się przy publikacjach współautorskich.

Niestety, przedstawiona monografia ma też pewne niedociągnięcia, związane z ograniczonym „panowaniem nad materiałem”, do których zaliczyłbym brak spójności w stosowanych oznaczeniach i symbolach. Na przykład na str. 23 symbole x_i oraz a_i oznaczają odpowiednio obiekty i ich cechy, natomiast na str. 37 x_i oznaczają wartości cech A_i . Podobnych niespójności jest więcej i utrudniają one lekturę monografii. Szkoda, że Habilitantka nie ujednoliciła oznaczeń i nie przygotowała zbiorczego zestawiania (jak w pracy A.3) – to znacznie poprawiłoby walory edukacyjne opracowania. Poza tym w tekście znajdują się niefortunne i niejasne sformułowania (np. „miara Cosine’a (kosinusowa)” na str. 21, czy też „wyznaczone prawdopodobieństwo dla tego przypadku jest takie samo jak dla prawdopodobieństw między minimum dwoma różnymi przypadkami” na str. 141), które obniżają czytelność tekstu. Jeśli wreszcie chodzi o zawartość monografii, to muszę przyznać, że brakowało mi w niej odwołania się do metod z zakresu *one-class learning*, które są również stosowane do wykrywania wyjątków i anomalii, dyskusji na temat doboru kryteriów dla algorytmów genetycznych (obecnie jest on arbitralny i odwołuje się do wcześniejszych publikacji) oraz porównania wyników działania algorytmów genetycznych z metodami bazującymi na wnioskowaniu na przypadkach. Te niedociągnięcia nie wpływają jednak na pozytywną ocenę osiągnięcia jako całości.

Podsumowując, uważam osiągnięcie naukowe przedstawione przez dr Duraj za dostateczne w zakresie wymagań stawianych osobom ubiegającym się o stopień doktora habilitowanego nauk technicznych w dziedzinie *Informatyka*.

3. Ocena pozostałych osiągnięć naukowo-badawczych

3.1. Ocena dorobku publikacyjnego

Dorobek publikacyjny Habilitantki po uzyskaniu stopnia doktora obejmuje 42 publikacje, z czego 10 zostało opublikowanych w czasopismach z listy *Journal Citation Reports* (JCR), takich jak *International Journal of Intelligent Systems*, czy też *Complexity*. Pozostałe prace ukazały się w recenzowanych materiałach krajowych i międzynarodowych konferencji naukowych, takich jak *Information Technology in Biomedicine*, *IEEE Innovations in Intelligent Systems and Applications* czy *IEEE Beyond Databases, Architectures and Structures*.

Analiza cytowań prac Habilitantki została przedstawiona w tabeli 1. Wyniki te są na stosunkowo niskim poziomie biorąc pod uwagę zwyczajowe wymagania stawiane osobom ubiegającym się o stopień doktora habilitowanego nauk technicznych w dziedzinie *Informatyka*. Należy jednak zwrócić uwagę na fakt, że istotne prace (A.1, A2 i A3) zostały opublikowane w roku 2018 i 2019 i od momentu przygotowania wniosku liczba cytowań istotnie wzrosła, zwłaszcza biorąc pod uwagę bazę Scopus.

Tabela 1. Analiza cytowań na podstawie własnej kwerendy wykonanej w dniu 15.04.2020 r.

Baza danych	H-index	Cytowania bez autocytowań	Cytowania łącznie
Web of Science	3	15	30
Scopus	6	51	99
Google Scholar	7		158

3.2. Kierowanie i udział w projektach badawczych

Dr Duraj uczestniczyła w roli wykonawcy w dwóch projektach realizowanych w roku 2012 i 2013 w ramach Programu Operacyjnego Innowacyjna Gospodarka 2007-2013. Celem głównym projektu było stworzenie platformy informatycznej TEWI pozwalającej na transfer innowacyjnych technologii informatycznych między jednostkami naukowymi a przemysłem. Habilitantka była odpowiedzialna za rozwój aplikacji *FusionClassify* stosującej algorytmy klasyfikacji w fuzji danych oraz *GrafOutliers* służącej do graficznej reprezentacji wyjątków wykrywanych w bazach danych.

Dr Duraj brała też udział w dwóch grantach Wydziału Fizyki Technicznej, Informatyki i Matematyki Stosowanej Politechniki Łódzkiej o tematyce związanej z wykrywaniem wyjątków z użyciem kwantyfikatorów lingwistycznych oraz wielokryterialnych algorytmów genetycznych. W pierwszym z tym grantów Habilitantka pełniła rolę kierownika, a w drugim – wykonawcy.

3.3. Nagrody za działalność naukową

Dr inż. Duraj otrzymała w 2007 roku wyróżnienie Polskiego Towarzystwa Zastosowań Elektromagnetyzmu za ciekawą pracę doktorską. Ponadto Habilitantka została dwukrotnie, w roku 2008 i 2018, nagrodzona Nagrodą Rektora Politechniki Łódzkiej za osiągnięcia w pracy naukowej.

3.4. Udział w międzynarodowych lub krajowych konferencjach

Dr inż. Duraj ma doświadczenie związane z prezentowaniem swoich osiągnięć na forach krajowych oraz międzynarodowych. Habilitantka przygotowała i wygłosiła 18 wystąpień na konferencjach krajowych oraz międzynarodowych, w tym *Information Technology in Biomedicine* oraz *IEEE International Conference on Innovations in Intelligent Systems and Applications*.

Biorąc powyższe pod uwagę, uważam pozostałe osiągnięcia naukowo-badawcze przedstawione przez dr Duraj za dostateczne w zakresie wymagań stawianych osobom ubiegającym się o stopień doktora habilitowanego nauk technicznych w dziedzinie *Informatyka*. Jak już wspomniałem, wskaźniki bibliometryczne są na niskim poziomie w stosunku do zwyczajowych oczekiwań, jednak dynamika ich zmian pozwala mieć nadzieję, że w najbliższych latach ulegną poprawie.

4. Ocena dorobku dydaktycznego i popularyzatorskiego oraz współpracy międzynarodowej

4.1. Praca na rzecz czasopism i innych wydawnictw naukowych

Dr Duraj przygotowywała 5 recenzji dla czasopisma *Journal of Applied Computer Science* wydawanego przez Instytut Informatyki Politechniki Łódzkiej. Ponadto Habilitantka pełniła rolę redaktora gościnnego jednego numeru tego czasopisma poświęconego innowacyjnym metodom wykrywania wyjątków. Była także współredaktorem książki

Elektromagnetyczne techniki w ochronie zdrowia wydanej przez Centralny Instytut Ochrony Pracy – Państwowy Instytut Badawczy.

Niestety, Habilitantka nie recenzowała prac dla czasopism z listy JCR, co może świadczyć o Jej słabszej rozpoznawalności międzynarodowej.

4.2. Praca na rzecz konferencji

Dr Duraj była członkiem komitetów programowych 4 konferencji międzynarodowych, w tym *IEEE International Conference on Innovations in Intelligent Systems and Applications* oraz *International Conference on Computational Collective Intelligence*. Habilitantka brała udział w organizacji specjalnych sesji poświęconych m.in., uczeniu maszynowemu w medycynie, innowacyjnym metodom wykrywania obserwacji odstających oraz złożonym systemom decyzyjnym.

4.3. Uczestnictwo w organizacjach i towarzystwach naukowych

Dr Duraj jest członkiem Polskiego Towarzystwa Informatyki Medycznej, Polskiego Towarzystwa Zastosowań Elektromagnetyzmu oraz Polskiego Towarzystwa Cybernetycznego.

4.4. Dorobek dydaktyczny

Dr Duraj była promotorem 49 prac inżynierskich oraz 50 prac magisterskich. Ponadto Habilitantka sprawowała opiekę nad studentami indywidualnego toku studiów, a także pełniła rolę promotora pomocniczego w jednej pracy doktorskiej.

Habilitantka przygotowała treści programowe i materiały dydaktyczne oraz prowadziła zajęcia wykładowe i laboratoryjne z baz i hurtowni danych, sztucznej inteligencji, metod probabilistycznych i statystyki, inżynierii oprogramowania, projektowania systemów informatycznych oraz marketingu biznesowego. Biorąc to pod uwagę należy uznać dorobek dr Duraj w zakresie działalności dydaktycznej za bardzo dobry.

4.5. Dorobek organizacyjny i współpraca międzynarodowa

Dr Duraj pełniła lub rolę członka Wydziałowej Komisji Oceny Jakości Kształcenia oraz Ray Programowej Kolegium Logistyki Wydziału Fizyki Technicznej, Informatyki i Matematyki Stosowanej Politechniki Łódzkiej. Pełniła też funkcję członka Komisji Rekrutacyjnej w Państwowej Wyższej Szkole Zawodowej we Włocławku. Habilitantka organizowała Polsko-Białoruskie Akademickie Warsztaty Edukacyjne Współczesne Technologie Teleinformatyczne, a także współpracowała przy organizacji Targów Edukacyjnych i Festiwalu Kół Naukowych organizowanych przez Politechnikę Łódzką.

Dr Duraj nie odbyła żadnych staży międzynarodowych, co wskazuje na brak aktywności w tym zakresie. Habilitantka pełni jednak rolę eksperta Komisji Europejskiej w programie ramowym w zakresie badań naukowych i innowacji, we wniosku nie ma jednak informacji na temat recenzowanych projektów.

Biorąc powyższe pod uwagę, uważam dorobek dydaktyczny i popularyzatorski oraz współpracę międzynarodową przedstawione przez dr Duraj za dostateczne w zakresie wymagań stawianych osobom ubiegającym się o stopień doktora habilitowanego nauk technicznych w dziedzinie *Informatyka*. Chciałbym zaznaczyć, że ocena poszczególnych elementów jest zróżnicowana – dorobek dydaktyczny i organizacyjny można uznać za co najmniej dobry, natomiast współpraca międzynarodowa jest bardzo ograniczona. Wypada mieć nadzieję, że Habilitantka w najbliższym czasie będzie wykazywać większą aktywność w tym zakresie.

5. Konkluzja

Osiągnięcia naukowe dr inż. Agnieszki Duraj uzyskane po otrzymaniu stopnia doktora stanowią zauważalny wkład w rozwój dyscypliny *Informatyka*. Uwzględniając dorobek organizatorski, dydaktyczny organizacyjny oraz aktywność naukową, stwierdzam, że recenzowany wniosek spełnia wymagania *Ustawy o stopniach naukowych i tytule naukowym* z dnia 14.03.2003 r. z późniejszymi zmianami (Dz.U. poz. 1789). W związku z powyższym wnioskuje o dopuszczenie dr inż. Agnieszki Duraj do dalszych etapów powstępowania habilitacyjnego i popieram wniosek o nadanie Jej stopnia naukowego doktora habilitowanego.



(Szymon Wilk)