

dr hab. inż. Mariusz Oszust, prof. Politechniki Rzeszowskiej  
Katedra Informatyki i Automatyki  
Wydział Elektrotechniki i Informatyki  
Politechnika Rzeszowska  
al. Powstańców Warszawy 12  
35-959 Rzeszów

Rzeszów, 4.07.2022 r.

**Recenzja osiągnięcia naukowego pt. „Uczenie reprezentacji w metodach nienadzorowanych” wykazanego przez dr. Przemysława Spurka w postępowaniu o nadanie stopnia doktora habilitowanego**

Niniejsza recenzja została sporządzona w odpowiedzi na pismo IPI PAN-RN-5/2022 z dnia 09.05.2022 r., Zastępcy Dyrektora Instytutu ds. Naukowych Pani dr hab. Agnieszki Mykowieckiej informujące o powołaniu mnie na recenzenta w postępowaniu habilitacyjnym dr. Przemysława Spurka prowadzonym przez Radę Naukową Instytutu Podstaw Informatyki PAN w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.

Podstawą prawną recenzji jest Ustawa z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668).

**Sylwetka Habilitanta**

Dr. Przemysław Spurek uzyskał tytuł magistra matematyki w 2009 roku na Wydziale Matematyki i Informatyki Uniwersytetu Jagiellońskiego. Na tym samym wydziale w 2014 roku uzyskał stopień doktora nauk matematycznych w dyscyplinie informatyka za rozprawę pt. „Memory compression based on affine transformation of coordinate system”. Promotorem rozprawy był prof. dr hab. Jacek Tabor. Od 2015 roku Habilitant jest zatrudniony na stanowisku adiunkta na Wydziale Matematyki i Informatyki Uniwersytetu Jagiellońskiego.

**Ocena osiągnięcia naukowego**

Zgodnie z wnioskiem Habilitanta, podstawę do ubiegania się o stopień doktora habilitowanego w dyscyplinie informatyka techniczna i telekomunikacja stanowi cykl publikacji pt. „Uczenie reprezentacji w metodach nienadzorowanych”. W skład cyklu wchodzi 10 prac opublikowanych w latach 2014 - 2021, spośród których siedem opublikowano w czasopiśmie z listy JCR (w tym jedna przyjęta do publikacji), zaś trzy opublikowano w materiałach konferencji uwzględnionych w rankingu CORE. Sumaryczny współczynnik wpływu (Impact Factor, 2020) prac z cyklu, wyznaczony na podstawie współczynników wpływu przytoczonych przez Habilitanta, wynosi

52,35, zaś liczba punktów MNiSW/MEiN wynosi 1450 (789 uwzględniając deklarowany wkład procentowy Habilitanta w ich powstanie). Oceniane osiągnięcie [A1-10]:

[A1] Jacek Tabor, **Przemysław Spurek**, Cross-Entropy Clustering, Pattern Recognition 47, 3046–3059, 2014.

[A2] **Przemysław Spurek**, Jacek Tabor, Krzysztof Byrski, Active Function Cross Entropy Clustering, Expert Systems with Applications 72, 49–66, 2017.

[A3] **Przemysław Spurek**, Jacek Tabor, Krzysztof Byrski, Online updating of active function cross-entropy clustering, Pattern Anal. Appl. 22 (4), 1409–1425, 2019.

[A4] **Przemysław Spurek**, Wiesław Pałka, Clustering of Gaussian distributions, 2016 International Joint Conference on Neural Networks (IJCNN) 22 (4), 3346–3353, 2016.

[A5] **Przemysław Spurek**, General split Gaussian Cross-Entropy clustering, Expert Systems with Applications 68, 58–68, 2017.

[A6] **Przemysław Spurek**, Jacek Tabor, Przemysław Rola, Michał Ociepka, ICA based on asymmetry, Pattern Recognition 67, 230–244, 2017.

[A7] Szymon Knop, **Przemysław Spurek**, Jacek Tabor, Igor Podolak, Marcin Mazur, Stanisław Jastrzębski, Cramer-wold autoencoder, Journal of Machine Learning Research (JMLR) 21, 1-28 2020

[A8] **Przemysław Spurek**, Szymon Winczowski, Jacek Tabor, Maciej Zamorski, Maciej Zięba, Tomasz Trzcinski, Hypernetwork approach to generating point clouds, Proceedings of the 37th International Conference on Machine Learning ICML 2020 119, 9099–9108, 2020.

[A9] **Przemysław Spurek**, Maciej Zięba, Jacek Tabor, Tomasz Trzcinski, General hypernetwork framework for creating 3D point clouds, IEEE Transactions on Pattern Analysis and Machine Intelligence, doi:10.1109/TPAMI.2021.3131131.

[A10] Tomasz Danel, **Przemysław Spurek**, Jacek Tabor, Marek Smieja, Łukasz Struski, Agnieszka Słowik, Łukasz Maziarka, Spatial Graph Convolutional Networks, International Conference on Neural Information Processing, 2020.

Cykl składa się z prac opublikowanych w renomowanych czasopismach lub materiałach prestiżowych konferencji. W siedmiu pracach Habilitant jest pierwszym autorem z wkładem nie mniejszym niż 40%, zaś w pozostałych trzech publikacjach wkład Habilitanta wynosi pomiędzy 30 a 50%. Na uwagę zasługuje fakt, że praca [A5] jest autorskim osiągnięciem Habilitanta, zaś czasopisma, w których opublikowano prace [A8-9] otrzymały 200 punktów na liście MNiSW/MEiN. Wkład procentowy pozostałych autorów prac wchodzących w skład osiągnięcia został w większości przypadków precyzyjnie określony pisemnymi oświadczeniami, choć nie

zawsze sumuje się on do 100%, jak ma to np. miejsce dla dwuosobowej publikacji [A4] (dr Przemysław Spurek deklaruje 80% a dr Wiesław Pałka 35%). Zwykle udział Habilitanta jest dominujący lub co najmniej równy udziałowi kolejnego istotnego współautora, a prace mają kilku współautorów ze względu na prowadzenie badań w ramach prac zróżnicowanych zespołów. Prace wchodzące w skład osiągnięcia zostały zauważone przez środowisko naukowe, z odnotowanymi przez Habilitanta 34 cytowaniami (wg WoS, z pominięciem autocytowań). Nie jest to duża liczba, lecz biorąc pod uwagę, że większość cytowań otrzymały publikacje z 2014 i 2017 roku ([A1, A2, A6]) należy spodziewać się ich wzrostu w najbliższym czasie. Udostępnianie kodów źródłowych metod z pewnością może się do tego przyczynić, jak ma to miejsce w przypadku algorytmów opisanych w [A1] czy [A8].

Habilitant lokuje swoje osiągnięcie w obszarach nienadzorowanego uczenia reprezentacji w klasycznych metodach uczenia maszynowego oraz głębokich sieciach neuronowych.

W ramach badań w pierwszym obszarze, Habilitant zaproponował algorytm grupowania danych Cross-Entropy Clustering (CEC) łączący algorytm  $k$ -średnich z metodą Expectation Maximization (EM). Istotą podejścia jest modyfikacja funkcji kosztu w metodzie EM, która charakteryzuje grupy i zmniejsza prawdopodobieństwo przydziału próbek danych do części z nich, w rezultacie redukując ich liczbę. Praca [A1] zawiera opis algorytmu CEC wraz ze szczegółową analizą konsekwencji stosowania różnych rodzin rozkładów gaussowskich. Warto podkreślić, że kod źródłowy metody został udostępniony publicznie (<https://cran.r-project.org/src/contrib/Archive/CEC/>).

Praca [A2] zawiera rozszerzenie metody CEC (active function Cross-Entropy Clustering, afCEC) o możliwość grupowania danych za pomocą rozmaitości w przestrzeni  $d$ -wymiarowej dzięki wprowadzeniu nowego rozkładu  $f$ -adapted Gaussian distribution jako uogólnieniu rozkładu Gauss. Artykuł, oprócz wartościowej części analitycznej, zawiera rozbudowaną część eksperymentalną, w której porównano afCEC z innymi metodami na kilku przykładach praktycznych. Wersja algorytmu afCEC opisana w pracy [A3] zawiera aproksymacje złożonych obliczeniowo kroków algorytmu i służy do grupowania napływających danych. W pracy rozważano heurystyki Lloyda i Hartigana do minimalizacji funkcji kosztu. Opracowana metoda umożliwia aktualizację parametrów grup dla nowych próbek danych bez potrzeby ponownego obliczania parametrów metody.

W pracy [A4] opisano podejście pozwalające na znajdowanie nieliniowych struktur wykorzystując klasyczne odległości między dwoma gęstościami rozkładów normalnych. W pierwszej kolejności uruchamiany jest algorytm CEC w celu określenia rozkładu dla każdej z grup a następnie wykonywane jest grupowanie hierarchiczne. W pracy rozważano wymianę grupowania hierarchicznego na metodę  $k$ -średnich w wersji Warda, która ze względu na pominięcie kroku wyznaczania centroidów grup jest stosowana do grupowania rozkładów prawdopodobieństw.

Kolejna praca z cyklu również dotyczy grupowania [A5]. Przedstawiono w niej metodę General Split Gaussian Cross-Entropy Clustering (GSGCEC), która jest dedykowana modelowaniu danych za pomocą rozkładów skośnych lub posiadających grube ogony. W takich przypadkach efektywna optymalizacja w technice EM wymaga dużej liczby próbek danych. W metodzie GSGCEC użyto rozkładu General Split Gaussian (GSG) o stosunkowo małej liczbie parametrów i wykorzystano CEC zamiast EM uzyskując rozwiązanie o lepszej wydajności obliczeniowej.

Doświadczenia Habilitanta z rozkładami skośnymi znalazły również zastosowanie przy nowym podejściu do analizy komponentów niezależnych (ang. Independent Component Analysis – ICA). Otóż zamiast maksymalizować podatną na elementy odstające miarę niegaussowskości – kurtozę, w pracy [A6] Habilitant zaproponował algorytm ICA oparty na asymetrii danych mierzonej za pomocą skośności (modelowanie za pomocą rozkładu Split Normal). Uzyskana metoda została porównana z klasycznym FastICA oraz ProDenICA i PearsonICA.

Prace [A7-10] dotyczą uczenia nienadzorowanego w odniesieniu do metod uczenia głębokiego, a zwłaszcza modeli generatywnych opartych na architekturze autoenkodera. W pracy [A7] wprowadzony został nowy model autoenkodera generatywnego CWAE (Cramer-Wold AutoEncoder) modyfikując funkcję kosztu w klasycznym autoenkoderze WAE. Zaproponowana modyfikacja jest interesująca i polega na użyciu miary odległości Cramer-Wold pomiędzy rozkładami. Miara jest oparta na odległości MMD oraz jądrze Cramer-Wold. Odległość Cramer-Wold pomiędzy próbkami korzysta z odległości  $L_2$  pomiędzy ich estymacjami jądrowymi, zaś odległość  $L_2$  pomiędzy gęstościami jest podana jawnym wzorem. Dzięki temu autoenkoder jest pozbawiony dodatkowych kroków znanych z innych podejść. Praca [A7] zawiera porównanie CWAE z metodami z literatury na klasycznych zbiorach obrazów.

W pracach [A8-A9] rozważano generowanie chmur punktów 3D za pomocą sieci HyperNetwork uczącej się reprezentacji obiektów 3D zadanych za pomocą chmur punktów 3D. W sieci HyperNetwork jedna sieć neuronowa generuje wagi do drugiej sieci (TargetNetwork), która rozwiązuje zadanie utworzenia reprezentacji. Zastosowanie HyperNetwork jest w tym przypadku uzasadnione nieregularną strukturą chmur punktów. Wprowadzona sieć HyperCloud [A8] pozwala modelować obiekty 3D zarówno w postaci chmur punktów, jak i powierzchni (mesh) i jest pozbawiona wad klasycznego algorytmu PointFlow dzięki wykorzystaniu sieci fully-connected, zamiast Continuous Normalization Flow (CNF), jak w PointFlow. Możliwość tworzenia powierzchni w HyperCloud wynika z wykorzystania rozkładu jednostajnego na kuli jednostkowej. Ponieważ jakość uzyskiwanych powierzchni jest niewystarczająca, w pracy [A9] wykorzystano CNF zamiast TargetNetwork, a do uczenia CNF utworzono nowy rozkład prawdopodobieństwa Spherical Log-Normal o małych ogonach.

Ponieważ sieci grafowe (Graph Convolutional Networks) nie biorą pod uwagę kolejności elementów, co w przypadku modelowania grafów reprezentujących związki chemiczne uniemożliwia reprezentowanie ich geometrycznej struktury, w pracy [A10] Habilitant zbudował dedykowaną sieć grafową Spatial Graph Convolutional Network (SGCN). SGCN wykorzystuje

cechy przestrzenne węzłów do agregowania informacji od sąsiadów i pozwala na modelowanie związków chemicznych. Dzięki dostępowi do informacji geometrycznej możliwe jest wykonanie wzbogacania danych poprzez obrót danego grafu w przestrzeni i uczenie sieci na małych zbiorach danych. Praca [A10] zawiera eksperymentalne porównanie SGCN z reprezentacyjnymi podejściami z literatury, które potwierdza użyteczność zaproponowanej metody.

Artykuły przedstawione w ramach cyklu poddawane ocenie w ramach postępowania o nadanie stopnia doktora habilitowanego dr. Przemysławowi Spurka stanowią wartościowy zbiór charakteryzujący liczne i udane rozwiązania problemów występujących w obszarze nienadzorowanego uczenia reprezentacji. Rozwiązania powstałe w ramach pracy zespołowej, przy znacznym udziale Habilitanta, zostały pozytywnie zrecenzowane i opublikowane w cenionych czasopismach lub materiałach prestiżowych konferencji. Warto podkreślić, że poziom cyklu prac świadczy o głębokim zrozumieniu przez Habilitanta metod uczenia maszynowego i przyczyn ich ograniczeń. Ponadto Habilitant sprawnie posługuje się aparatem matematycznym w celu jednoznacznego opisu proponowanych zależności pomiędzy danymi i rządzącymi nimi rozkładami prawdopodobieństwa. Pozwala to na stosowanie uogólnień, tworzenie nowych rozkładów i ich wklęcie w funkcje kosztów, docelowo prowadząc do uzyskania lepszych sposobów uczenia się reprezentacji danych przez rozważane metody.

*Stwierdzam, że przedłożone przez dr. Przemysławowi Spurka dziesięć publikacji, pozytywnie zweryfikowanych w procesie recenzji i dostatecznie zauważonych przez środowisko naukowe w postaci cytowań, jest powiązanych tematycznie, wartościowym cyklem, który oceniam pozytywnie i który stanowi znaczny wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja. Jednoznacznie uważam, że przedstawione osiągnięcie jest w pełni wystarczające dla ubiegania się o stopień doktora habilitowanego w tej dyscyplinie.*

### **Ocena pozostałych osiągnięć naukowych**

Dorobek naukowy niezaliczony do osiągnięcia a wskazany przez Habilitanta, to 14 prac [B1-E14] w czterech obszarach: i) grupowanie danych w podprzestrzeniach – dwie prace, ii) ICA – dwie prace, iii) brakujące dane – trzy prace, iv) rozwój i zastosowanie sieci neuronowych - siedem prac.

Spośród prac niewliczonych do ocenianego cyklu, Habilitant opublikował 29 prac po uzyskaniu stopnia naukowego doktora. Osiem z nich to artykuły w renomowanych czasopismach, zaś pozostałe w dużej części prezentowano na prestiżowych konferencjach. Przedstawiony dorobek sprzed nadania stopnia doktora stanowi pięć prac (dwa artykuły i trzy materiały konferencyjne). W czasie pracy naukowej Habilitant otrzymał dwie zespołowe nagrody Rektora oraz nagrodę za najlepszy artykuł na konferencji CISIM 12'.

Sumaryczny Impact Factor Habilitanta wynosi 109,849, a łączna liczba punktów MNiSW/MEiN wynosi 3720. Liczba cytowań, bez autocytowań na dzień 08.11.2021 r.: Google Scholar – 241, Web of Science – 92; Indeks Hirscha, bez autocytowań: Google Scholar – 8, Web of Science – 5. Na dzień sporządzania niniejszej recenzji ogólna liczba cytowań wg Google Scholar wynosi

506 (wzrost o 118 z 388). Przedstawione informacje naukometryczne, w mojej opinii, są na odpowiednim poziomie, biorąc pod uwagę liczbę opublikowanych prac i ich współautorów. Można zatem stwierdzić, że rozpoznawalność i użyteczność prac Habilitanta jest coraz bardziej zauważana, a Habilitant znacznie powiększył swój dorobek od czasu nadania stopnia doktora. Świadczą o tym również liczne konferencje naukowe, w których brał czynny udział (10 po uzyskaniu stopnia doktora), wykłady na zaproszenie (5), udział w komitetach redakcyjnych czterech konferencji (TFML i CISIM) oraz programowych sześciu konferencji (ICLR, AAAI, NeurIPS, CVPR, ICCV i TFML). Miarą popularności i rozpoznawalności w środowisku, a także dbałości o rozwój dyscypliny, jest także liczba wykonanych recenzji. W tym przypadku dorobek Habilitanta jest zaskakująco niski, z siedmioma recenzjami artykułów dla czasopism i 25 recenzjami materiałów konferencyjnych (konferencje 2021-2022). Wygląda na to, że Habilitant, będąc autorem korespondencyjnym wielu opublikowanych prac w stosunkowo popularnej tematyce, przez kilka pierwszych lat po uzyskaniu stopnia doktora marginalizował potrzebę wykonywania recenzji. Należy jednak zauważyć, że w ostatnim roku przed złożeniem wniosku o nadanie stopnia doktora habilitowanego nastąpiło wzmożenie tej często niedocenianej aktywności do akceptowalnego poziomu.

Dr Przemysław Spurek aktywnie zdobywał finansowanie, brał udział w sześciu zrealizowanych projektach finansowanych głównie przez NCN i NCBiR (dwukrotnie był kierownikiem) oraz bierze udział w czterech projektach (w jednym jest kierownikiem). Tego typu aktywność zasługuje na podkreślenie, zwłaszcza, że Habilitant wykazał się również współpracą z sektorem gospodarczym (dwa projekty).

*Aktywność i pozostałe osiągnięcia naukowe Habilitanta oceniam pozytywnie. Potwierdzają one znaczny wkład w rozwój dyscypliny naukowej informatyka techniczna i telekomunikacja. Spełniają one wymagania stawiane kandydatom ubiegającym się o stopień doktora habilitowanego.*

### **Ocena istotnej aktywności naukowej realizowanej poza uczelnią macierzystą**

Na aktywność naukową realizowaną w więcej niż jednej jednostce badawczej składają się staże i wizyty naukowe. Należy do nich również zaliczyć granty badawcze realizowane w projektach międzyuczelnianych. Habilitant posiada obszerny dorobek w tym zakresie, również o mniej formalnym charakterze, który zaowocował wspólnymi publikacjami (12 osób, w tym pięć z zagraniczną afiliacją; 33 publikacje lub preprinty). Habilitant odbył następujące staże:

- 01.2020 – 03.2020: Postdoctoral Researcher – Faculty of Electrical Engineering and Computer Science VSB - Technical University of Ostrava, Czech Republic,
- 01.11.2018 – 31.01.2019: Stażysta w Instytucie Podstaw Informatyki, Polska Akademia Nauk, Warszawa,
- 09.2014 – 06.2015: Postdoctoral Researcher - Polska Akademia Nauk, Warszawa
- 2012 – 2013: Research Scientist, Uniwersytet Jana Kochanowskiego w Kielcach.

Ponadto brał lub bierze udział w trzech projektach międzyuczelnianych:

- 2019 – 2023: Sztuczne sieci neuronowe inspirowane biologicznie, Team-Net (FNP), nr POIR.04.04.00-00-14DE/18-00; udział bierze sześć zespołów badawczych.
- 2015 – 2015: Innowacyjna elektrownia wiatrowa z pionową osią obrotu, o mocy 1,5 MW wyposażona w układ przekształtnikowy umożliwiający uzyskanie wysokiej efektywności przetwarzania energii wiatru w energię elektryczną przy niskich prędkościach wiatru; Grant NCBiR nr WND-DEM-1-153/01, AGH, Kraków
- 2012 – 2013: Deteh – Opracowanie technologii nowej generacji czujnika wodoru i jego związków do zastosowań w warunkach ponadnormatywnych; udział w ramach konsorcjum.

*Podsumowując, uważam, że aktywność Habilitanta poza uczelnią macierzystą jest istotna i oceniam ją pozytywnie. Spełnia one ustawowe i zwyczajowe wymagania stawiane kandydatom ubiegającym się o stopień doktora habilitowanego w dyscyplinie naukowej informatyka techniczna i telekomunikacja.*

#### **Ocena osiągnięć dydaktycznych, organizacyjnych oraz popularyzujących naukę**

Poza aktywnością naukową dr Przemysław Spurek jest bardzo aktywny na polu dydaktycznym. Jest promotorem pomocniczym dwóch doktorantów, prowadził seminarium dla doktorantów, przygotował trzy autorskie wykłady dla studentów, wypromował 15 magistrantów, opiekował się dwoma studentami studiów matematyczno-przyrodniczych UJ oraz prowadził ćwiczenia i zajęcia laboratoryjne z dziesięciu przedmiotów.

W ramach osiągnięć organizacyjnych był trzykrotnie członkiem Komitetu Organizacyjnego konferencji TFML Theoretical Foundations of Machine Learning (Będlewo 2015, Kraków 2017, Kraków 2019) oraz członkiem Komitetu Organizacyjnego konferencji CISIM Computer Information Systems and Industrial Management Applications (Kraków 2013; przed uzyskaniem stopnia naukowego doktora). Do osiągnięć organizacyjnych należy również zaliczyć wykonywanie obowiązków kierownika w projektach.

W ramach popularyzacji nauki Habilitant wygłosił trzy wykłady podczas Dni Wydziału Matematyki i Informatyki Uniwersytetu Jagiellońskiego oraz cztery referaty: dwa na Studenckim Festiwalu Informatycznym oraz na Test Dive i Machine Learning Nokia Workshop.

*Uważam, że osiągnięcia dydaktyczne, organizacyjne i popularyzujących naukę dr. Przemysława Spurka są znaczące i oceniam je pozytywnie.*

## **Wniosek końcowy**

*Art. 221 pkt 8 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym (Dz. U. z 2018 r., poz. 1668) określa, że recenzja powinna oceniać czy osiągnięcia naukowe osoby ubiegającej się o stopień doktora habilitowanego odpowiadają wymaganiom określonym w art. 219 ust. 1 pkt 2. Biorąc pod uwagę powyższy wymóg, potwierdzam, że wśród wskazanych do oceny osiągnięć naukowych dr. Przemysława Spurka znajduje się cykl dziesięciu powiązanych tematycznie artykułów naukowych opublikowanych w czasopiśmie naukowych lub w recenzowanych materiałach z konferencji międzynarodowych, pozytywnie go oceniam i stwierdzam, że stanowi on znaczący wkład w rozwój dyscypliny informatyka techniczna i telekomunikacja. Ponadto moja ocena pozostałego dorobku naukowego, aktywności naukowej realizowanej poza uczelnią macierzystą oraz aktywności organizacyjnej, dydaktycznej i popularyzującej naukę jest jednoznacznie pozytywna. Dlatego zwracam się do Rady Naukowej Instytutu Podstaw Informatyki PAN o dopuszczenie Pana dr. Przemysława Spurka do dalszych etapów postępowania habilitacyjnego.*

