

RECENZJA ROZPRAWY DOKTORSKIEJ MGR INŻ. KRZYSZTOFA ROPIAKA PT. *ADAPTATION OF MERELOGICAL GRANULATION TECHNIQUES IN SELECTED DATA ANALYSIS ISSUES*

dr hab. inż. prof. KUL Krzysztof Pancerz

Katedra Podstaw Informatyki, Instytut Filozofii, Wydział Filozofii

Katolicki Uniwersytet Lubelski Jana Pawła II

Lublin, 9 maja 2025 r.

1 Wprowadzenie

Przedłożona do recenzji rozprawa doktorska mgr inż. Krzysztofa Ropiaka powstała w ramach postępowania o nadanie stopnia doktora wszczętego w dyscyplinie Informatyka Techniczna i Telekomunikacja na Wydziale Informatyki i Sztucznej Inteligencji Politechniki Częstochowskiej. Rozprawa napisana została w języku angielskim. Rozprawa liczy 172 strony. Właściwa część rozprawy składa się ze wstępu (Rozdział I), trzech głównych rozdziałów (Rozdziały II - IV) oraz zakończenia (Rozdział V). W rozprawie, oprócz spisu treści, umieszczone zostały: streszczenie w języku angielskim, streszczenie w języku polskim, wykaz rysunków, wykaz tabel, wykaz algorytmów oraz wykaz bibliografii. Szczegółowe uwagi dotyczące poszczególnych części rozprawy podane są w dalszej części recenzji. Zakres tematyczny badań Autora pozostaje cały czas aktualny, przede wszystkim w kontekście uczenia maszynowego, a w szczególności w kontekście wstępnego przetwarzania danych dla modeli klasyfikacyjnych. Zarówno w przypadku klasycznych modeli uczenia maszynowego jak i w przypadku modeli uczenia głębokiego, istotne jest przygotowanie odpowiednich danych uczących, walidacyjnych oraz testujących. W szczególności, ważnymi operacjami w procesach wstępnego przetwarzania danych są

m.in. redukcja danych, czyszczenie danych, znajdowanie danych reprezentatywnych dla poszczególnych problemów. Te operacje są istotne szczególnie w przypadku zbiorów z danymi o wartościach numerycznych ciągłych, zbiorów z danymi z brakującymi wartościami oraz zbiorów z danymi zaszumionymi lub odstającymi. Autor zwrócił w rozprawie uwagę na takie zbiory danych. Zaproponowane przez Autora w rozprawie metody mereologicznej granulacji wiedzy wspomagają te operacje. Z tego punktu widzenia, zakres zrealizowanych badań oraz ich dalszy rozwój są istotnymi elementami przyczyniającymi się do rozwoju uczenia maszynowego, a w ogólności do rozwoju sztucznej inteligencji.

2 Ocena poszczególnych części rozprawy

Rozdziały główne rozprawy podzielone zostały na trój-poziomową strukturę podrozdziałów. Dzięki takiemu podziałowi, układ pracy, z punktu widzenia rozdziałów, jest czytelny. Ogólna struktura rozprawy jest przemyślana i przejrzysta, a treści prezentowane w poszczególnych rozdziałach układają się w systematyczny i logiczny wywód naukowy, co świadczy o dobrym warsztacie metodologicznym Autora.

W streszczeniu rozprawy, Autor przedstawił motywację podjęcia badań w zakresie obliczeń granularnych wykorzystujących metody mereologiczne (w szczególności zaproponowane przez Prof. Polkowskiego), stwierdzając, że badania w tym zakresie są niszowe w obszarze analizy danych. Można zgodzić się z takim stwierdzeniem Autora, gdyż niewiele można znaleźć istotnych publikacji naukowych dotyczących mereologicznych metod granulacji wiedzy poza publikacjami autorów powiązanych ze środowiskiem Prof. Polkowskiego. W streszczeniu rozprawy, Autor sformułował cztery tezy, które w pracy zamierzał potwierdzić: (i) możliwe jest zaprojektowanie metod granulacji wiedzy nie wymagających szacowania optymalnych parametrów - promieni granulacji, (ii) granule wiedzy mogą mieć skuteczne zastosowanie we wzmacnianiu procesów klasyfikacji w zespołowych modelach klasyfikacyjnych, (iii) granule wiedzy mogą mieć szerokie zastosowanie w różnych procesach analizy danych - w tym w radzeniu sobie z wartościami brakującymi, (iv) techniki nadpróbkowania i niedopróbkowania wpływają na proces tworzenia granularnych obrazów systemów decyzyjnych. Przytoczono tutaj przeformułowane tezy w stosunku do tych, które Autor umieścił w rozprawie w streszczeniu w języku polskim. Przede wszystkim, tezy sformułowane przez Autora brzmią w języku polskim niestylistycznie: *Zastosowanie granul wiedzy może mieć ... zastosowanie ...* (tezy (ii) oraz (iii)). Część pojęć Autor pozostawił w

języku angielskim: *oversampling* i *undersampling*. Ponadto, Autor użył pojęcia, które wymaga objaśnienia, tj. *granularne refleksje systemów decyzyjnych*.

W rozdziale głównym I, stanowiącym rozdział wprowadzający, Autor przedstawił:

1. Tło badań: genezę badań dotyczących mereologii.
2. Mereologiczne metody obliczeń granularnych (granulację standardową zaproponowaną przez Prof. Polkowskiego i granulację opartą o pojęcia) oraz wyniki eksperymentów zrealizowanych przy ich użyciu na dostępnych zbiorach danych.

Uwagi merytoryczne do rozdziału I:

1. Tło badań w rozdziale 1 zostało przedstawione w formie za bardzo skrótowej (tylko trzy krótkie akapity).
2. W formule 2.4 (str. 10) Autor użył oznaczenie U_{cover} dla zbioru pokrycia granul. Pojęcie tego zbioru nie zostało wcześniej formalnie zdefiniowane.
3. Wybór zbioru danych *Iris* do przykładu w podrozdziale 2.2 nie jest najlepszym pomysłem aby przedstawić ideę granulacji opartej o pojęcia. Wynika to z faktu, że wartości atrybutów w zbiorze *Iris* należałoby potraktować jako wartości ciągłe, a nie dyskretne. W opisie zbioru *Iris* (<https://archive.ics.uci.edu/dataset/53/iris>) typ atrybutów został określony jako rzeczywisty (*real*). Konsekwencją tego jest konieczność użycia innej relacji pomiędzy obiektami w kontekście wartości atrybutów, np. relacji podobieństwa, a nie relacji nierozróżnialności (określającej czy dla danego atrybutu wartości obiektów są takie same), jak to zrobił Autor prezentując macierze nierozróżnialności (str. 17). Wybór Autora utrudnia np. znajdowanie reprezentatywnych wartości atrybutów dla granul. Ten problem został zauważony przez Autora w kolejnych rozdziałach.
4. W pkt. 4 na str. 10, Autor powinien był wspomnieć np. metody zachłanne oraz metody oparte o algorytmy genetyczne, które mogą zostać użyte do wyznaczania pokryć przez granule.
5. Nie jest jasne dlaczego Autor zastosował kodowanie *gorącej jedynki* dla atrybutów z wartościami symbolicznymi w zbiorach danych użytych w eksperymentach (podrozdział 2.4.2). W przypadku nierozróżnialności, istotna jest tylko informacja czy wartości atrybutu są takie same, czy różne.

Uwagi techniczne do rozdziału I:

1. Autor zaprezentował algorytmy obrazujące procesy klasyfikacji danych ze zbiorów oryginalnych oraz ze zbiorów poddanych granulacji standardowej (algorytm 1 na str. 55 oraz algorytm 2 na str. 66). Jednakże, te algorytmy nie różnią się niczym na zaprezentowanym poziomie abstrakcji. Stąd, prezentacja algorytmu 2 wydaje się być zbędna.

W rozdziale głównym II, stanowiącym rozdział prezentujący koncepcje i zagadnienia będące autorskimi wynikami badań, Autor przedstawił:

1. Metodę granulacji jednorodnej oraz wyniki eksperymentów zrealizowanych przy jej użyciu na dostępnych zbiorach danych.
2. Metodę granulacji jednorodnej dopuszczającą określoną maksymalną różnicę ε poszczególnych wartości atrybutów oraz wyniki eksperymentów zrealizowanych przy jej użyciu na dostępnych zbiorach danych.

Uwagi merytoryczne do rozdziału II:

1. W formule 3.1 (str. 76) Autor użył oznaczenie U_{cover} dla zbioru pokrycia granul. Pojęcie tego zbioru nie zostało wcześniej formalnie zdefiniowane.
2. Nie jest jasne co oznacza operacja eliminacji kompletnie sprzecznych obiektów realizowana w kroku 2 procesu granulacji jednorodnej przedstawionego w podrozdziale 3.1 (str. 75). Wyjaśnienie w nawiasie jest bardzo skrótowe, co skutkuje koniecznością domyślania się znaczenia takiej sprzeczności.
3. W formule 3.3 (str. 76), Autor przedstawił efekt głosowania większościowego dla określenia reprezentatywnych wartości atrybutów dla granul. Błędem jest uwzględnienie w formule również atrybutu decyzyjnego.
4. W rozdziale 4.2.2, Autor wspomniał o procedurze głosowania większościowego dla określenia reprezentatywnych wartości atrybutów dla granul. Formuła, którą autor przedstawił jest nieprecyzyjna. Jeśli i oraz j są liczbami obiektów w granulach, jak napisał Autor, to nie jest jasne co oznaczają $a(u_i)$ oraz $a(u_j)$.

W rozdziale głównym III, Autor przedstawił:

1. Nową metodę do budowy modeli zespołowych opartą o losowe obrazy granularne systemów oraz wyniki eksperymentów zrealizowanych przy jej użyciu na dostępnych zbiorach danych.
2. Sposoby radzenia sobie z brakującymi wartościami w danych w metodach granulacji jednorodnej oraz wyniki eksperymentów dla specjalnie przygotowanych zbiorów danych z brakującymi wartościami.

Uwagi merytoryczne do rozdziału III:

1. W podrozdziale 5.3.5 (str. 128), Autor użył pojęcia kompletnie nierozróżnialnych obiektów, które nie zostało zdefiniowane.
2. W podrozdziale 6.0.1, Autor opisał podstawowe strategie radzenia sobie z brakującymi wartościami w danych w metodach granulacji jednorodnej. Opis ten zawiera notację, która nie została objaśniona, np. nie zostało objaśnione co oznacza notacja $*$ = *don't care*. Te objaśnienia można dopiero wydedukować z formuł opisujących relacje nierozróżnialności.

Uwagi techniczne do rozdziału III:

1. W rozdziale 5, Autor przedstawił algorytm do budowy modeli zespołowych oparty o losowe obrazy granularne systemów w postaci opisu słownego. Bardziej czytelną formą prezentacji byłaby na przykład lista kroków, jak np. w podrozdziale 3.1.
2. W rozdziale 6, Autor umieścił bezpośrednio podrozdział 6.0.1. Brakuje podrozdziału 6.0.

W rozdziale głównym IV, Autor przedstawił wpływ technik nadpróbkiowania i niedoprobkiowania na proces granulacji wiedzy oraz wyniki eksperymentów dla specjalnie przygotowanych zbiorów danych poprzez zastosowanie technik nadpróbkiowania i niedoprobkiowania. Brak jest istotnych uwag dotyczących treści przedstawionych w tym rozdziale.

W rozdziale głównym V, stanowiącym rozdział podsumowujący rozprawę, Autor przedstawił główne osiągnięcia rozprawy w obszarze wykorzystania mereologicznych metod granulacji wiedzy w wybranych problemach analizy danych. Brak jest istotnych uwag dotyczących treści przedstawionych w tym rozdziale.

3 Ogólna ocena rozprawy

Brak jest istotnych zastrzeżeń do strony merytorycznej rozprawy. Autor wykazał się odpowiednią znajomością zagadnień poruszonych w rozprawie, w szczególności znajomością zagadnień z zakresu obliczeń granularnych, wstępnego przetwarzania danych w problemach uczenia maszynowego oraz budowy, uczenia i testowania modeli klasyfikacyjnych (pojedynczych oraz zespołowych). Drobne uwagi merytoryczne dotyczące zawartości poszczególnych rozdziałów i podrozdziałów rozprawy podane zostały w poprzedniej części recenzji. Należy bardzo pozytywnie ocenić ilość eksperymentów przeprowadzonych przez Autora w ramach części badawczej rozprawy, w tym odpowiedni dobór zbiorów danych, oraz dobrą systematykę tych eksperymentów. Realizacja eksperymentów została opisana w sposób czytelny, a wyniki eksperymentów zostały w sposób bardzo wyczerpujący przedstawione i skomentowane. Świadczy to o dobrym warsztacie badawczym Autora. Autor w sposób prawidłowy zrealizował i opisał przeprowadzone w ramach rozprawy badania. Autor rozpoczął badania od wykorzystania znanych mereologicznych metod granulacji wiedzy, zaproponowanych wcześniej w literaturze, tj. od wykorzystania metody granulacji standardowej oraz metody granulacji opartej o pojęcia. Następnie Autor zaproponował w badaniach użycie autorskiej metody granulacji jednorodnej. Metoda ta została następnie uogólniona do metody granulacji jednorodnej dopuszczającej określoną maksymalną różnicę ε poszczególnych wartości atrybutów. W kolejnych etapach badań, Autor zajął się istotnymi problemami w zakresie uczenia maszynowego: radzenia sobie z brakującymi wartościami w danych, budowę modeli zespołowych oraz radzenia sobie z niezbalansowanymi zbiorami danych. Dzięki takiemu zakresowi badań w rozprawie, powstało spójne i pełne opracowanie naukowe wybranej tematyki z zakresu obliczeń granularnych, stanowiące istotny wkład do rozwoju metod analizy danych oraz uczenia maszynowego, a w ogólności do rozwoju metod sztucznej inteligencji.

Ogólne uwagi do strony formalnej, m.in. językowej i notacyjnej pracy:

1. Sposób zapisu nazw metod nie został w rozprawie ujednolicony, np. występują `ąda-boost`"i `Ada-Boost`"(str. 116).
2. Nie wszystkie formuły w rozprawie są numerowane przez Autora.

Ogólne uwagi do wykazu literatury:

1. Wykaz literatury został uporządkowany alfabetycznie ze względu na autorów, jednak pozycja J. Łukasiewicza znalazła się na końcu wykazu.
2. Opis pozycji bibliograficznych nie został w całości ujednolicony.

Rozprawa napisana jest komunikatywnym językiem, aczkolwiek można znaleźć w tekście drobne błędy językowe i stylistyczne. Autor swobodnie i jasno formułuje myśli. Zdania układają się w logiczny ciąg, przez co nie istnieją problemy w percepcji tekstu. Patrząc na całość rozprawy, stronę formalną rozprawy należy ocenić pozytywnie. Została ona poprawnie zredagowana.

W wykazie bibliografii Autor umieścił istotny i reprezentatywny, z punktu widzenia rozważanej tematyki, zestaw źródeł literaturowych. Autor odwołuje się w tekście do publikacji, których jest współautorem. Elementy opisanych w rozprawie metod zostały przedstawione w publikacjach naukowych w czasopismach oraz w materiałach konferencji międzynarodowych.

Ogólnie, przedłożoną rozprawę doktorską należy ocenić pozytywnie. Braki i usterki techniczne i merytoryczne nie zmieniają tej oceny. Do głównych oryginalnych wyników rozprawy można zaliczyć:

- systematyczny rozwój metod mereologicznej granulacji wiedzy,
- usprawnienie metod mereologicznej granulacji wiedzy poprzez dodanie automatycznego doboru parametrów granulacji,
- przebadanie użyteczności zaproponowanych metod mereologicznej granulacji wiedzy dla różnych zbiorów danych, w tym zbiorów danych z brakującymi wartościami oraz niezbalansowanych zbiorów danych,
- przebadanie możliwości użycia zaproponowanych metod mereologicznej granulacji wiedzy w zespołowych modelach klasyfikacyjnych,
- przeprowadzenie i opisanie dużej liczby usystematyzowanych eksperymentów na odpowiednio dużej ilości zbiorów danych z wykorzystaniem zaproponowanych autorskich metod mereologicznej granulacji wiedzy.

4 Podsumowanie

Podsumowując, uważam, że przedłożona do recenzji rozprawa doktorska mgr inż. Krzysztofa Ropiaka odpowiada wymogom stawianym rozprawom doktorskim w *Ustawie Prawo o szkolnictwie wyższym i nauce* z dn. 20 lipca 2018 r. **Przedłożoną rozprawę oceniam pozytywnie.** W związku z tym wnioskuję bez zastrzeżeń o dopuszczenie mgr inż. Krzysztofa Ropiaka do dalszych etapów postępowania o nadanie stopnia doktora w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie Informatyka Techniczna i Telekomunikacja.