

RECENZJA ROZPRAWY DOKTORSKIEJ MGR INŻ.
DAWIDA KOSIORA PT. *NOWE METODY
UWZGLĘDNIENIA NIEPEWNOŚCI I
NIEPRECYZYJNOŚCI DANYCH W KLASYFIKACJI:
ZASTOSOWANIA W MEDYCYNIE*

dr hab. inż. prof. KUL Krzysztof Pancerz
Katedra Podstaw Informatyki, Instytut Filozofii, Wydział Filozofii
Katolicki Uniwersytet Lubelski Jana Pawła II

Lublin, 21 stycznia 2026 r.

1 Wprowadzenie

Przedłożona do recenzji rozprawa doktorska mgr inż. Dawida Kosiora powstała w ramach postępowania o nadanie stopnia doktora wszczętego w dyscyplinie Informatyka Techniczna i Telekomunikacja w Instytucie Badań Systemowych Polskiej Akademii Nauk w Warszawie. Rozprawa napisana została w języku polskim. Rozprawa liczy 137 stron. Właściwa część rozprawy składa się ze wstępu (Rozdział 1), pięciu głównych rozdziałów (Rozdziały 2 - 6) oraz podsumowania (Rozdział 7). W rozprawie, oprócz spisu treści, umieszczone zostały: streszczenie w języku polskim, streszczenie w języku angielskim, wykaz podstawowych oznaczeń, skróty, wykaz bibliografii, wykaz rysunków oraz wykaz tabel. Szczegółowe uwagi dotyczące poszczególnych części rozprawy podane są w dalszej części recenzji. W ogólności, tematyka badań Autora dotyczyła formalnych metod radzenia sobie z wiedzą niedoskonałą w narzędziach komputerowych. Ten obszar badań pozostaje cały czas aktualny, przede wszystkim w kontekście sztucznej inteligencji, a w szczególności w kontekście uczenia maszynowego. Zaproponowane przez Autora w rozprawie metody uwzględniania niepewności i nieprecyzyjności danych wspomagają procesy wstępnego przetwarzania danych oraz budowę modeli klasyfikacyjnych. 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-poziomą 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.

We wstępie do rozprawy (rozdział 1), Autor przedstawił zakres tematyczny rozprawy, motywację podjęcia badań w zakresie wykorzystania przedziałowych zbiorów rozmytych w problemach wstępnego przetwarzania danych oraz problemach klasyfikacji danych, cel rozprawy, główne wyniki rozprawy oraz jej strukturę. Autor odniósł się do wcześniejszych badań w zakresie metod klasyfikacji w warunkach niepewnej i nieprecyzyjnej informacji. Autor we wstępie użył trzech pojęć dotyczących wiedzy niedoskonałej: dane niepewne, dane nieprecyzyjne, dane niepełne. Autor użył również w tym kontekście pojęć informacja i wiedza. Generalnie zaś, najczęściej użyte zostały określenia niepewność i nieprecyzyjność. W takim przypadku, wskazane byłoby dokładniejsze wyjaśnienie tych różnych określeń, przede wszystkim: dane, informacja, wiedza, ale również: niepewność, nieprecyzyjność, niepełność. Autor wskazał algorytm klasyfikacyjny k najbliższych sąsiadów (k-NN) jako ten, dla którego głównie zostały opracowane w pracy metody. Uzasadnienie, że algorytm k-NN daje zwykle bardzo dobre wyniki wydaje się być trochę na wyrost. Dla wielu zbiorów danych można wskazać inne podejścia do budowy klasyfikatorów dające lepsze rezultaty niż k-NN. We wstępie, Autor wskazał możliwości wykorzystania opracowanych metod do klasyfikacji danych w trzech obszarach medycznych: diagnostyka raka piersi, diagnostyka problemów sercowo-naczyniowych, diagnostyka depresji. Stwierdzenie Autora, że opracowane podejścia spowodują, że pacjenci będą mogli być objęci odpowiednią opieką medyczną, co przełoży się na zwiększenie szans na skuteczne leczenie i poprawę rokowań, wydaje się przedwczesne. Droga od eksperymentów na danych do rzeczywistego wdrożenia systemu wspomagającego decyzje diagnostyczne w medycynie jest znacznie dłuższa.

W rozdziale 2, Autor przedstawił:

1. Podstawowe pojęcia teorii klasycznych zbiorów rozmytych.
2. Podstawowe pojęcia teorii przedziałowych zbiorów rozmytych.

Uwagi merytoryczne do rozdziału 2:

1. Autor wymienił wśród operacji na zbiorach rozmytych dopełnienie, następnie wskazał na negację rozmytą. Zabrakło informacji, czy według Autora dopełnienie i negacja oznaczają tę samą operację.
2. W podrozdziałach 2.2.1 oraz 2.2.2, Autor nie wyjaśnił czym są agregacja rozmyta i rozmyta funkcja agregacji.
3. W podrozdziale 2.2.1, Autor nie wyjaśnił czym jest własność Ferrersa.
4. W podrozdziale 2.2.4, Autor zastosował zapis $\max\{\bullet\}$ (analogicznie $\min\{\bullet\}$), wcześniej minimum i maksimum traktowane były jako funkcje, tj. $\max(\bullet)$ (analogicznie $\min(\bullet)$).

Uwagi techniczne do rozdziału 2:

1. Autor użył tych samych krojów czcionek do oznaczenia rodzin obiektów i samych obiektów, np. w podrozdziale 2.2.3, Autor oznaczył rodziny negacji możliwości i negacji konieczności tak samo jak negacje możliwości i konieczności, tj. N_{pos} i N_{nec} .

W rozdziale 3, Autor przedstawił:

1. Definicje miar zawierania się przedziałów.
2. Definicje miar zawierania się przedziałowych zbiorów rozmytych.
3. Definicje miar podobieństwa dla przedziałowych zbiorów rozmytych.
4. Definicje entropii dla przedziałowych zbiorów rozmytych.

Uwagi merytoryczne do rozdziału 3:

1. Nie jest jasne czy Twierdzenie 3.2 zostało zaproponowane przez Autora. Brak jest odnośnika do źródła bibliograficznego. Jeśli jest to twierdzenie Autora, brak jest dowodu lub jego szkicu.
2. W twierdzeniu 3.3, nie jest zrozumiały zapis $\mathcal{A} \leq \vee$.
3. Nie jest jasne czy Twierdzenie 3.5 zostało zaproponowane przez Autora. Brak jest odnośnika do źródła bibliograficznego. Jeśli jest to twierdzenie Autora, brak jest dowodu lub jego szkicu.
4. Nie jest jasne czy Twierdzenie 3.6 zostało zaproponowane przez Autora. Brak jest odnośnika do źródła bibliograficznego. Jeśli jest to twierdzenie Autora, brak jest dowodu lub jego szkicu.

Brak jest uwag technicznych do rozdziału 3.

W rozdziale 4, Autor przedstawił:

1. Wybrane metody klasyfikacji danych (algorytm k najbliższych sąsiadów, drzewa decyzyjne, naiwne algorytmy Bayesa).
2. Miary oceny jakości klasyfikacji danych.

Uwagi merytoryczne do rozdziału 4:

1. Przy opisie klasyfikatorów regułowych, Autor przyjął, że w poprzedniku reguł występuje koniunkcja warunków. Teoretycznie reguły mogą zawierać różne operacje logiczne w poprzedniku (często jest to również alternatywa).
2. Autor przytoczył definicję pokrycia reguły. Użyte przez Autora w tym kontekście pojęcie pokrycia jest dyskusyjne. Generalnie, pokrycie reguły odnosi się do stopnia pokrycia wartości atrybutu decyzyjnego, natomiast to co Autor nazwał pokryciem powinno zostać nazwane wsparciem reguły, a dokładnie siłą reguły (względna miara wsparcia reguły).
3. Przy opisie drzew decyzyjnych, Autor stwierdził, że do liścia drzewa trafiają wyłącznie obiekty z jednej klasy decyzyjnej. Generalnie, tak być nie musi. W szczególności, dla sprzecznych tablic decyzyjnych, znajdują się w drzewie liście zawierające obiekty z różnych klas decyzyjnych.

Brak jest uwag technicznych do rozdziału 4.

W rozdziale 5, Autor przedstawił:

1. Modyfikację algorytmu k-NN polegającą na wykorzystaniu wcześniej zdefiniowanych miar zawierania, podobieństwa i entropii dla przedziałowych zbiorów rozmytych oraz wyniki eksperymentów klasyfikacji danych ze zbioru Wisconsin Breast Cancer za pomocą zmodyfikowanego algorytmu.
2. Problem uzupełniania brakujących wartości atrybutów z wykorzystaniem wcześniej zdefiniowanych miar entropii dla przedziałowych zbiorów rozmytych oraz wyniki eksperymentów klasyfikacji danych ze zbioru Wisconsin Breast Cancer walidujących zaproponowane podejścia.
3. Modyfikację algorytmu Relief selekcji cech polegającą na wykorzystaniu wcześniej zdefiniowanych miar podobieństwa dla przedziałowych zbiorów rozmytych oraz wyniki eksperymentów klasyfikacji danych ze zbioru Wisconsin Breast Cancer walidujących zaproponowane podejścia.
4. Wykorzystanie wcześniej zdefiniowanych miar podobieństwa dla przedziałowych zbiorów rozmytych w zaproponowanym Systemie Prewencji Depresji (SPD) opartym o test składający się z 22 pytań.

Uwagi merytoryczne do rozdziału 5:

1. Autor wymienił algorytm k-NN jako jeden z dziesięciu algorytmów uznawanych za najlepsze w Data Mining, ale powołał się na ranking z 2008 roku.
2. W podrozdziale 5.4.1, Autor odwołał się do Międzynarodowej Klasyfikacji Chorób ICD-10, podczas gdy od roku 2022 obowiązuje klasyfikacja ICD-11.

Brak jest uwag technicznych do rozdziału 5.

W rozdziale 6, Autor przedstawił:

1. Wykorzystanie przedziałowego wnioskowania rozmytego opartego o reguły w diagnostyce chorób sercowo-naczyniowych.
2. Bibliotekę dla języka Python (IFIS) wspomagającą implementację systemów przedziałowego wnioskowania rozmytego.

Brak jest uwag merytorycznych do rozdziału 6.

Uwagi techniczne do rozdziału 6:

1. Autor umieścił błędny podpis dla rysunku 6.4. Rysunek przedstawia diagram klas zaimplementowanej biblioteki IFIS.

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 teorii zbiorów rozmytych oraz uczenia maszynowego. Drobne uwagi merytoryczne dotyczące zawartości poszczególnych rozdziałów i podrozdziałów rozprawy podane zostały w poprzedniej części recenzji. Istotnym elementem pracy, stanowiącym naukowy wkład Autora,

jest zaproponowanie nowych miar zawierania, podobieństwa i entropii dla przedziałowych zbiorów rozmytych. Autor nie ograniczył się tylko do aspektów teoretycznych. Zaproponowane miary zostały wykorzystane przez Autora do modyfikacji algorytmów uczenia maszynowego: algorytmu k-NN oraz algorytmu Relief, a następnie zostały wykorzystane do rozwiązania praktycznych problemów wstępnego przetwarzania danych oraz klasyfikacji danych. Należy pozytywnie ocenić eksperymentalną walidację zaproponowanych podejść na danych medycznych. Autor w sposób prawidłowy zrealizował i opisał przeprowadzone w ramach rozprawy badania. Dzięki takiemu zakresowi badań, w rozprawie powstało spójne i pełne opracowanie naukowe wybranej tematyki z zakresu teorii zbiorów rozmytych, stanowiące istotny wkład do rozwoju metod sztucznej inteligencji. Ważnym elementem badań przeprowadzonych przez Autora jest implementacja części zaproponowanych metod jako pakietu w języku Python. Przede wszystkim, umożliwia to upowszechnienie wyników rozprawy w środowisku naukowym.

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

1. Autor nie zachował w tekście rozprawy jednolitej liczby gramatycznej. Wykorzystana została zarówno pierwsza osoba w liczbie pojedynczej jak i pierwsza osoba w liczbie mnogiej.
2. Autor nie zachował w tekście rozprawy jednolitej numeracji jej elementów: w numeracji twierdzeń uwzględnione zostały numery głównych rozdziałów, pozostałe elementy (definicje, przykłady) ponumerowane zostały kolejnymi liczbami.
3. Nie wszystkie formuły matematyczne zostały ponumerowane przez Autora.

Zakres wykorzystanych w rozprawie źródeł literaturowych należy uznać za poprawny. W wykazie bibliografii Autor umieścił istotny i reprezentatywny, z punktu widzenia rozważanej tematyki, zestaw źródeł literaturowych. Autor w tekście rozprawy odwoływał się wielokrotnie do różnych źródeł, w tym do publikacji, których był współautorem. Elementy opisanych w rozprawie metod zostały przedstawione w publikacjach naukowych w czasopiśmie oraz w materiałach konferencji międzynarodowych.

Rozprawa napisana jest komunikatywnym językiem, aczkolwiek można znaleźć w tekście drobne błędy literowe, 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.

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

- opracowanie nowych miar zawierania, podobieństwa i entropii dla przedziałowych zbiorów rozmytych oraz propozycje ich wykorzystania do rozwiązania praktycznych problemów uczenia maszynowego takich jak wstępne przetwarzanie danych (uzupełnianie brakujących wartości, selekcja cech) oraz klasyfikacja danych,
- przebadanie użyteczności zaproponowanych miar w klasyfikacji danych medycznych,
- opracowanie Systemu Prewencji Depresji wykorzystującego zaproponowane miary dla przedziałowych zbiorów rozmytych,

- opracowanie mechanizmu wnioskowania rozmytego wykorzystującego nowe metody agregacji przedziałowych.

4 Podsumowanie

Podsumowując, uważam, że przedłożona do recenzji rozprawa doktorska mgr inż. Dawida Kosiora 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. Dawida Kosiora do dalszych etapów postępowania o nadanie stopnia doktora w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie Informatyka Techniczna i Telekomunikacja.

Krzysztof Pawan