

dr hab. inż. Julian Szymański, prof. PG
Politechnika Gdańska
Wydział Elektroniki Telekomunikacji i Informatyki
Katedra Architektury Systemów Komputerowych

RECENZJA

Rozprawy doktorskiej mgr inż. Krzysztofa Ropiaka
pt.: Adaptation of mereological granulation techniques
in selected data analysis issues
promotor: dr hab. inż. Piotr Artiemjew, prof. UWM

1. Formalna podstawa wykonania recenzji

Recenzja przygotowana została na zlecenie przewodniczącego Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Częstochowskiej, prof. dr hab. inż. Roberta Nowickiego (Uchwała nr 10/2023/2024 z dnia 25.06.2024) .

Ponowna recenzja została przygotowana na prośbę prodziekana ds. Nauki Informatyka Techniczna i Telekomunikacja Wydziału Informatyki i Sztucznej Inteligencji Politechniki Częstochowskiej dr hab. Adama Kulawika, prof. PCz. (pismo R-WIMII-BOD-510-7/21 z dnia 24.02.2025r.)

2. Uwagi ogólne

Poniższa recenzja jest rozwinięciem mojej opinii przedstawionej do poprzedniej wersji doktoratu mgr inż. Krzysztofa Ropiaka, która skierowana została do poprawy decyzją komisji doktorskiej z dnia 18 grudnia 2024r. (Uchwała 3/2024/2025).

Rozprawa doktorska mgr inż. Krzysztofa Ropiaka poświęcona jest adaptacji technik granulacji mereologicznej do wybranych zagadnień analizy danych. Praca ta przedstawia szczegółowe omówienie nowych technik aproksymacji dla systemów decyzyjnych, zwłaszcza w kontekście klasyfikacji danych. Autor koncentruje się na rozwinięciu metod wywodzących się z teorii zbiorów przybliżonych Pawlaka, a w szczególności metod Polkowskiego.

Celem pracy jest zaprezentowanie nowatorskich wyników uzyskanych za pomocą znanych metod granulacji oraz nowej techniki, która nie wymaga estymacji parametrów – granulacji jednorodnej. Autor bada również zastosowanie tej techniki w radzeniu sobie z brakującymi wartościami oraz jej efektywność w modelach ensemble, takich jak boosting i bagging. Praca ma także na celu sprawdzenie wydajności tych metod w połączeniu z technikami oversamplingu.

Głównym wkładem pracy doktorskiej mgr inż. Krzysztofa Ropiaka jest rozwinięcie i zastosowanie technik granulacji mereologicznej w analizie danych, które obejmują:

1. Opracowanie metody jednorodnej granulacji i jej wariantu epsilon. Podejścia te nie wymagają estymacji parametrów, co stanowi innowację w dziedzinie granulanych technik analizy danych.
2. Zastosowanie metod granulacji w tworzeniu modelu typu ensemble poprzez wprowadzenie modelu Ensemble of Random Granules, który wykorzystuje granulację do poprawy procesów klasyfikacji. Model ten okazał się być konkurencyjny w stosunku do innych modeli ensemble, takich jak boosting i bagging.
3. Absorpcja brakujących wartości proponuje techniki granulacji wiedzy, które zostały przetestowane pod kątem efektywności w radzeniu sobie z brakującymi wartościami w danych, co jest istotnym zadaniem w analizie danych.
4. Praca bada wpływ oversamplingu i undersamplingu na proces granulacji danych, wykazując, że kombinacja metod undersamplingu i oversamplingu może poprawić jakość klasyfikacji.

Wyniki, wraz z teoretycznymi rozważaniami przeprowadzonymi w pracy, potwierdzają postawione tezy i otwierają nowe możliwości zastosowania technik granulacji wiedzy w różnych obszarach analizy danych.

Część wyników zaprezentowanych w pracy zostały docenione w konkursie PP-RAI Contest for the Most Influential Article on Rough Sets, co również pozytywnie świadczy o ich wartości naukowej.

Reasumując uwagi ogólne uważam, że wybór tematu rozprawy jest trafny i uzasadniony aktualnym stanem wiedzy i pozytywnie oceniam wielkość wkładu Autora w dziedzinę.

Przedstawiona do ponownej recenzji praca zawiera szereg poprawek, które uwzględniają uwagi, które pojawiły się na wcześniejszych etapach postępowania o nadanie stopnia doktora.

3. Struktura rozprawy, ocena merytoryczna i metodyki badawczej

Obecnie rozprawa obejmuje 172 strony co jest pomniejszeniem w stosunku do 177 stron w wersji poprzedniej. Praca zorganizowana w sposób przejrzysty i logiczny, umożliwiając czytelnikowi płynne zapoznanie się z badanymi zagadnieniami, podzielona jest logicznie na 5 rozdziałów.

Do poprawionej wersji pracy załączono listę wprowadzonych poprawek, jednakże nie zawierają one numerów stron, na których zostały naniesione zmiany co znacząco utrudnia identyfikację zmienionych miejsc, zwłaszcza w wersji papierowej doktoratu przedłożonego do ponownej recenzji.

Rozprawa rozpoczyna się od wprowadzenia w tematykę mereologii i obliczeń granularnych oraz przeglądu literatury opisującego obecny stan wiedzy. Następnie Autor poświęca rozdział opisowi rozwoju technik granulacji wiedzy z wykorzystaniem adaptacyjnego mechanizmu do wyznaczania promienia granulacji. Wprowadzona zostaje tu koncepcja jednorodnej granulacji, która bazuje na granulkach tworzonych wokół centralnych obiektów. Proces ten kontynuuje rozszerzanie promienia granulacji aż do momentu, gdy granulki zawierają wyłącznie obiekty z tej samej klasy decyzyjnej.

Rozdział trzeci koncentruje się na wybranych zastosowaniach technik granulacji wiedzy w analizie danych. Obejmuje on szczegółowe omówienie nowego modelu ensemble nazwanego "Random Granular Reflections" oraz eksperymenty związane z jego efektywnością w porównaniu do innych popularnych metod ensemble, takich jak bagging i boosting.

Rozdział czwarty analizuje wpływ technik oversamplingu i undersamplingu na jakość procesu granulacji danych. W szczególności, rozdział ten skupia się na tym, jak te techniki wpływają na równowagę klas w zbiorach danych oraz na efektywność klasyfikacji po przeprowadzeniu procesu granulacji.

Praca kończy się rozdziałem podsumowującym osiągnięcia w zakresie opracowania nowych technik granulacji wiedzy oraz ich zastosowaniu w analizie danych. Wskazana tu została skuteczność jednorodnej granulacji i jej wariantu epsilon w klasyfikacji danych i radzeniu sobie z brakującymi wartościami, a także pozytywny wpływ technik oversamplingu i undersamplingu na jakość procesu granulacji. Rozdział ten podkreśla również, że nowe metody mogą być efektywnie wykorzystane w modelach ensemble oraz wskazuje na przyszłe kierunki badań, w tym rozwój nowych technik klasyfikacji i zastosowanie granulacji do różnych typów danych.

Zastosowane piśmiennictwo jest adekwatne i obejmuje kluczowe prace z dziedziny teorii zbiorów przybliżonych, obliczeń granularnych oraz technik klasyfikacyjnych. Przegląd literatury mógłby jednak być bardziej wszechstronny i uwzględniać więcej nowszych badań.

Rozprawa mgr inż. Krzysztofa Ropiaka cechuje się dobrym poziomem merytorycznym. Praca przedstawia nowatorskie podejście do adaptacji technik granulacji mereologicznej w analizie danych, co jest zgodne z aktualnym stanem wiedzy w dziedzinie. Autor szczegółowo omawia teoretyczne podstawy swoich metod, a następnie przechodzi do ich praktycznego zastosowania i oceny efektywności. Wyniki badań są przedstawione w sposób klarowny, a wnioski uzasadnione rezultatami eksperymentów.

Celem pracy było rozwinięcie i zastosowanie technik granulacji mereologicznej w analizie danych, w szczególności w kontekście klasyfikacji, radzenia sobie z brakującymi wartościami oraz poprawy jakości klasyfikacji przy użyciu technik oversamplingu i undersamplingu. Cel został jasno sformułowany i osiągnięty poprzez przedstawienie nowatorskich metod oraz przeprowadzenie licznych eksperymentów.

Metodyka badawcza zastosowana w rozprawie jest poprawna i odpowiednio dobrana do badanych zagadnień. Autor wykorzystuje różnorodne zbiory danych z repozytorium UCI oraz kaggle.com, co pozwala na wszechstronną ocenę efektywności proponowanych metod. Badania eksperymentalne są przeprowadzone zgodnie ze standardami naukowymi,

a uzyskane wyniki są poddane szczegółowej analizie i porównaniu z innymi modelami ensemble, takimi jak boosting i bagging.

Ekspertyzacje obejmują zarówno standardowe techniki granulacji, jak i podejścia opracowane przez Autora, takie jak jednorodna granulacja i jej wariant epsilon. Analiza wpływu technik oversamplingu i undersamplingu na jakość granulacji dodatkowo wzbogaca wartość naukową pracy.

Omówienie wyników badań jest szczegółowe i dobrze uzasadnione. Autor prezentuje wyniki w sposób klarowny, porównując efektywność proponowanych metod z innymi popularnymi technikami klasyfikacyjnymi. Wyniki eksperymentów potwierdzają skuteczność proponowanych rozwiązań.

Wyniki badań mają potencjalne zastosowanie w różnych dziedzinach analizy danych, w tym w klasyfikacji, radzeniu sobie z brakującymi wartościami oraz poprawie jakości klasyfikacji danych niezbalansowanych. Metody te mogą znaleźć zastosowanie w systemach decyzyjnych, analizie medycznej, finansach oraz innych obszarach wymagających efektywnego przetwarzania dużych zbiorów danych.

Pomimo powyżej wymienionych pozytywnych aspektów, praca doktorska mgr inż. Krzysztofa Ropiaka nie jest wolna od pewnych niedociągnięć, które warto podkreślić w celu dalszego doskonalenia warsztatu badawczego Autora.

W poprzedniej recenzji zgłosiłem uwagi, że przegląd literatury mógłby być bardziej szczegółowy i obszerny. Choć autor odnosi się do kluczowych prac z dziedziny teorii zbiorów przybliżonych i obliczeń granularnych, brakuje bardziej wszechstronnego omówienia najnowszych badań i ich wpływu na rozwój technik granulacji. W bibliografii było jedynie 55 pozycji literatury z czego bardzo niewielka część jest pozycjami wydanymi po 2020 roku. W chwili obecnej liczba ta została zredukowana do 51, i nie zawiera znaczących nowych pozycji co nie istotną poprawą przeglądu literaturowego dziedziny.

Przypisy bibliograficzne zawierają strony, na których są cytowania – jest to mało spotykany sposób tworzenia literatury, i sam w sobie jest nadmiarowy ponieważ na podstawie samych indeksów można je odnaleźć w tekście pracy. Część przypisów nie zawiera istotnych informacji np.: przy pozycji [36] brakuje wskazania gdzie artykuł został opublikowany.

W poprzedniej wersji pracy Autor pisał „results that were accepted or recognised in the competition - PP-RAI Contest for the Most Influential Article on Rough Sets co-authored by Polish Researchers in 2020-2021 papers [1] and [2].” jednakże prace [1] i [2] w bibliografii nie odnosiły się do artykułów promotora opublikowanych w innych materiałach niż PP-RAI 20/21 co dla mnie nie było jednoznaczne.

Obecnie indeksy przypisów zostały skorygowane, jednak cały czas określenie „wyniki zostały zaakceptowane lub wyróżnione” (i dalej) nie zawiera jednoznacznej informacji gdzie i która część tego osiągnięcia została opublikowana.

Pierwotne polskie streszczenie pracy sprawiało wrażenie niedopracowanego tłumaczenia z języka angielskiego. Szczególnie rzucało się w oczy częste używanie liczby mnogiej w odniesieniu do autora, co jest akceptowalne w języku angielskim, ale w języku polskim

brzmi nienaturalnie. Innym przykładem jest określenie „Jej zastosowanie do wchłaniania brakujących wartości”, które miało nawiązywać do angielskiego „Its use for missing values handling”. Jednak w języku polskim brzmi to niezręcznie.

W nowej wersji te niezgrabności zostały poprawione.

W pierwotnej wersji rozprawy moje wątpliwości wzbudził również sposób postawienia hipotez badawczych: (i) „Is it possible to design knowledge granulation methods that do not require estimation of optimal parameters - granulation radii?” co jest raczej pytaniem badawczym. Teza badawcza powinna być jednoznacznym stwierdzeniem, które można udowodnić lub obalić w trakcie badań. Teza (iv) „Oversampling and undersampling techniques affect the process of creating granular reflections of decision-making systems powinna być bardziej precyzyjna i określać, jak techniki oversamplingu i undersamplingu wpływają na proces tworzenia granularnych systemów decyzyjnych.

Nowa wersja pracy zawiera poprawione hipotezy, które są postawione w sposób akceptowalny.

W pracy zabrakło mi przykładów zastosowania opracowanych technik w rzeczywistych scenariuszach biznesowych lub przemysłowych. Choć autor przeprowadza eksperymenty na danych z repozytorium UCI oraz kaggle.com, to włączenie studiów przypadków z rzeczywistego świata zdecydowanie zwiększyłoby praktyczną wartość badań.

Prezentacja wyników mogłaby być bardziej przystępna dla czytelnika. Miejscami praca jest bardzo techniczna a liczba przedstawionych wyników eksperymentów przyćmiewa istotę badań. Warto tu było rozważyć dodanie bardziej przystępnych wyjaśnień i sumarycznej wizualizacji wyników.

Warto byłoby również przeprowadzić bardziej rozbudowaną analizę porównawczą z innymi nowoczesnymi metodami klasyfikacji i analizy danych. Chociaż autor porównuje swoje metody z popularnymi technikami *ensemble*, to szersze porównanie z innymi nowatorskimi podejściami mogłoby lepiej wykazać unikalność i przewagi proponowanych rozwiązań.

W niektórych miejscach opisy metodologii eksperymentów powinny być bardziej szczegółowe. Dokładniejsze wyjaśnienia dotyczące doboru parametrów oraz procesu walidacji wyników byłyby pomocne dla lepszego zrozumienia przeprowadzonych badań.

W rozprawie opisano nowe techniki granulacji wiedzy, które nie były wcześniej stosowane w takim kontekście, oraz przedstawiono nowatorskie podejście do analizy danych przy użyciu adaptacyjnych mechanizmów granulacji. W mojej opinii rozprawa stanowi oryginalne rozwiązanie problemu naukowego i spełnia wymagania stawiane rozprawom doktorskim w dziedzinie Informatyki Technicznej i Telekomunikacji.

Rozprawa potwierdza dobrą wiedzę teoretyczną kandydata w dziedzinie analizy danych, w szczególności w obszarze teorii zbiorów przybliżonych. Kandydat wykazał się umiejętnością samodzielnego prowadzenia pracy naukowej, tworząc nowe metody oraz przeprowadzając szerokie eksperymenty potwierdzające skuteczność proponowanych rozwiązań.

Podsumowując, pomimo przedstawionych powyżej uwag krytycznych w mojej opinii, praca mgr inż. Krzysztofa Ropiaka wnosi istotny wkład do dziedziny analizy danych i technik granulacji wiedzy.

4. Uwagi końcowe

Dorobek Autora na podstawie baz Scopus i Web of Science uważam za dobre osiągnięcie na tym etapie rozwoju naukowego.

Pozytywnie oceniam zawartość merytoryczną rozprawy. Wykorzystana metodologia badawcza jest poprawna, a uzyskane wyniki popierają postawione hipotezy.

Podsumowując stwierdzam, że przedstawiona rozprawa doktorska mgr. inż. Krzysztofa Ropiaka zatytułowana „Adaptation of mereological granulation techniques in selected data analysis issues” stanowi oryginalne rozwiązanie problemu naukowego, potwierdzając wiedzę oraz umiejętność samodzielnego prowadzenia badań naukowych przez Doktoranta i w mojej opinii spełnia wymagania stawiane stosownej ustawie. Na tej podstawie wnioskuję do Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Politechniki Częstochowskiej o dopuszczenie jej do dalszych etapów postępowania o nadanie stopnia doktora w dziedzinie nauk inżynieryjno-technicznych w dyscyplinie Informatyka Techniczna i Telekomunikacja.

dr hab. inż. Julian Szymański, prof. PG