

dr hab. Barbara Pękala, prof. UR
Instytut Informatyki
Kolegium Nauk Przyrodniczych
Uniwersytetu Rzeszowskiego
ul. Pigoń 1, 35-310 Rzeszów

Rzeszów, 17 lipca 2024 r.

Recenzja rozprawy doktorskiej

mgr Łukasza Szymuli

*Zastosowanie teorii zbiorów rozmytych w klasyfikacji
dominującej dyscypliny naukowej autorów w
naukometrii*

*Application of Fuzzy Set Theory in Classifying Author's Dominant Scientific
Disciplines in Scientometrics*

Rozprawa napisana na Uniwersytecie Adama Mickiewicza w Poznaniu z dziedziny nauk ścisłych i przyrodniczych w dyscyplinie Informatyka pod kierunkiem:

prof. UAM dr hab. Krzysztofa Dyczkowskiego,
prof. dr hab. Marka Kwieka.

1 Ogólna charakterystyka pracy, obszar problemowy i uzyskane wyniki

Praca doktorska związana jest z dziedziną nauki zajmującą się analizą ilościową działalności naukowej, czyli naukometrią. Dziedzina ta obejmuje różne metody i techniki służące do oceny i zrozumienia procesu naukowego oraz jego wpływu na dynamikę rozwoju badań

naukowych. W naukometrii wykorzystuje się wskaźniki bibliometryczne, które pomagają ocenić jakość i wpływ badań naukowych oraz wzorców współpracy poszczególnych naukowców. Naukometria ma szerokie zastosowanie, między innymi w zarządzaniu nauką, polityce naukowej, oraz w ocenie finansowania badań i projektów naukowych. Dzięki niej możliwe jest bardziej efektywne planowanie i alokacja zasobów w nauce oraz lepsze zrozumienie dynamiki rozwoju poszczególnych dyscyplin naukowych. Praca koncentruje się na jednym z kluczowych wyzwań naukometrii jakim jest klasyfikacja dyscyplin naukowych autorów publikacji, co ma bezpośredni wpływ na ocenę ich dorobku naukowego i analizy obszarów tematycznych. Przedstawiona w pracy metoda klasyfikacji dominującej dyscypliny na podstawie wartości modalnej, nazywana w pracy podejściem bazowym, oparta jest na prostych algorytmach i ograniczonej interpretacji dla bogatych danych. Podejście to wiąże się z dużą niejednoznacznością w przypisywaniu dyscyplin do autorów. Zastosowanie teorii zbiorów rozmytych umożliwia modelowanie pojawiającej się w ww. metodzie nieprecyzyjności informacji i oferuje nowe perspektywy dla klasyfikacji dyscyplin naukowych, opierając się na rachunku i metodach dostępnych w teorii zbiorów rozmytych. Proponowana modyfikacja algorytmu bazowego z uwzględnieniem nieprecyzyjności informacji obejmuje trzy podejścia: oparte na zmiennych lingwistycznych, na sterownikach rozmytych i wykorzystaniu agregacji wartości rozmytych z wykorzystaniem operatorów OWA.

W recenzowanej pracy zbadano możliwości wykorzystania metod teorii zbiorów rozmytych do modyfikacji i ulepszenia algorytmów klasyfikacji dyscyplin naukowych autorów opartych o wartość modalną. Przeprowadzone badania zawierają praktyczne sprawdzenie skuteczności proponowanych modyfikacji w kontekście dostępnych danych bibliometrycznych. Główne hipotezy badawcze skupiają się na możliwości zwiększenia jednoznacznej klasyfikacji przez zastosowanie metod zbiorów rozmytych poprzez identyfikację kluczowych zmiennych lingwistycznych, sterowników rozmytych, metod agregacji i funkcji wagowych. Ponadto w badaniu porównano wyniki pochodzące z proponowanych podejść wykorzystujących wybrane metody modelowania nieprecyzyjności informacji z metodą bazową.

Podsumowując tę część rozprawy stwierdzam, że Autor podjął ważny, aktualny a zarazem trudny problem naukowy. Przeprowadził analizę istniejących podejść, zaproponował autorskie rozwiązania, przedstawił ich zastosowania dla pełnej bibliometrycznej bazy

danych Scopus udostępnionej w platformie ICSR Lab, Elsevier oraz zweryfikował ich skuteczność, co pozwoliło ocenić opracowane rozwiązania.

2 Zawartość pracy oraz opinia o rozprawie i oryginalność osiągnięć

Rozprawa składa się z siedmiu rozdziałów, bibliografii oraz załączników.

Rozdział pierwszy stanowi wstęp do tematyki rozprawy. Autor przedstawił motywacje do podjęcia badań dotyczących zbadania możliwości wykorzystania metod teorii zbiorów rozmytych do modyfikacji i ulepszenia algorytmów klasyfikacji dyscyplin naukowych autorów oraz strukturę rozprawy. W tej części rozprawy można zauważyć brak wyraźnego podkreślenia głównych wkładu autora w przedstawiane rozwiązania w pracy. Ich wyeksponowanie pozwoliłoby lepiej zobrazować stopień złożoności pracy badawczej niezbędnej do zrealizowania doktoratu oraz zakres rozprawy.

Rozdział drugi przedstawia podstawowe pojęcia z obszaru naukometrii wskazując występowanie w nim nieprecyzyjności. Opisano również specyfikę największych baz bibliometrycznych oraz wskaźników naukometrycznych. Opisano także algorytm klasyfikacji dominującej dyscypliny w oparciu o wartość modalną (nazywane w pracy podejściem bazowym). Przedstawiono jego najważniejsze wady wskazując skalę problemu niejednoznaczności klasyfikacji. Wskazano główne powody ich występowania wraz z przedstawieniem propozycji jego modyfikacji o aspekt nieprecyzyjności informacji, podając także metody ewaluacji jakości proponowanego podejścia rozmytego.

Rozdział trzeci omawia podstawowe pojęcia teorii zbiorów rozmytych, które są efektywnym narzędziem do modelowania niejednoznaczności informacyjnych. W tym rozdziale przedstawiono trzy kluczowe elementy, na których opiera się algorytm w podejściu rozmytym: zmienne lingwistyczne, sterowniki rozmyte oraz operatory agregacji, w tym OWA operator. Ponadto, zaprezentowano wybrane funkcje wagowe, które są także ważnym elementem analizowanej metody klasyfikacji.

W rozdziałach 4-6 przedstawione zostały propozycje algorytmu klasyfikacji dominującej dyscypliny przez wartość modalną z wykorzystaniem podejścia rozmytego.

W rozdziale czwartym zaprezentowano modyfikację algorytmu bazowego z wykorzystaniem zmiennych lingwistycznych. W badaniu zaproponowano 7 zmiennych lingwistycz-

nych: Cytowanie, FWCI 4-letnie, FWCI 5-letnie, FWCI bez ram czasowych, Percentyl czasopisma, Rok oraz Zespół. Przedstawiono proces konstrukcji zmiennych lingwistycznych, opisano szczegółowo algorytm wyznaczania dominującej dyscypliny z wykorzystaniem zmiennej lingwistycznej oraz wyszczególniono warianty funkcji wagowych użytych na etapie zliczania częstości występowania dyscyplin naukowych autora.

Rozdział piąty dotyczy zastosowania sterowników rozmytych wysokiego prestiżu publikacji. W badaniu wykorzystano cztery warianty sterownika oparte na dwóch zmiennych wejściowych: Cytowanie i Percentyl czasopisma, FWCI 4-letnie i Percentyl czasopisma, FWCI 5-letnie i Percentyl czasopisma oraz dla kombinacji FWCI bez ram czasowych i Percentyla czasopisma. Opisano proces konstruowania sterowników rozmytych, tworzenia zbioru reguł oraz zaprezentowano wersję algorytmu dla klasyfikacji dominującej dyscypliny w podejściu rozmytym opartym na systemie regułowym.

W rozdziale szóstym przedstawiono modyfikację algorytmu bazowego o wykorzystanie operatorów agregacji OWA dla wartości dwóch zmiennych lingwistycznych: Cytowania oraz Percentyla czasopisma w każdym termie. Podobnie jak w przypadku poprzednich dwóch metod opisano metodologię dla proponowanego podejścia oraz przedstawiono wyniki uzyskanych badań.

Ostatni rozdział oprócz podsumowania uzyskanych wyników badań, przedstawia perspektywy dalszego rozwoju omawianego podejścia w aspekcie eksploracji nowych rozwiązań przy wykorzystaniu teorii zbiorów rozmytych.

Przede wszystkim należy podkreślić, iż problem badawczy podejmowany w rozprawie jest bardzo ważny i złożony. Praca podejmuje kluczowe wyzwanie w obszarze naukometrii, jakim jest klasyfikacja dyscyplin naukowych autorów, proponując istotne modyfikacje w algorytmie bazowym przez uwzględnienie metod opartych na teorii zbiorów rozmytych, czyli uwzględnia nieprecyzyjność i wielowymiarowość danych naukowych. Za najważniejsze osiągnięcia przedstawione w pracy uważam metody klasyfikacji dominującej dyscypliny w podejściu rozmytym, takie jak modyfikację algorytmu bazowego z wykorzystaniem zmiennych lingwistycznych, zastosowanie sterowników rozmytych wysokiego prestiżu publikacji oraz modyfikację algorytmu bazowego o wykorzystanie operatorów agregacji OWA. Badania wykazały, że stosując podejście rozmyte możliwe jest jednoznaczne zaklasyfikowanie większej ilości obserwacji, niż w przypadku podejścia bazowego.

Podsumowując tę część recenzji uważam, iż praca jest obszerna i pokazuje dużą wiedzę

Doktoranta. Widoczne jest naukowe podejście do realizowanego zadania. Rozprawa przedstawia dużą zawartość merytoryczną, w tym naukową oraz implementacyjną związaną z przedstawionymi zastosowaniami. Uważam, że wykonana w ramach doktoratu praca stanowi istotny wkład w rozwój metod dotyczących klasyfikacji dyscyplin naukowych autorów w obszarze naukometrii. Rozprawa prezentuje wysoki poziom naukowy i w pełni zasługuje na ocenę pozytywną.

3 Uwagi i problemy do dyskusji

W pracy brakuje wyraźnie sformułowanego wkładu badawczego Autora.

W rozprawie brakuje spisu tabel, rysunków i algorytmów, co ułatwiłoby nawigację po treści pracy.

Także cytowanie prac byłoby bardziej czytelne gdyby stosowano by przyjętą numerację w spisie literatury.

Brak szerszego wyjaśnienia motywacji wyboru do anlizy i modyfikacji algorytmu bazowego spośród innych metod.

Czy Autor widzi możliwość stworzenia i wykorzystania modeli hybrydowych łączących kilka podejść dostępnych z literatury?

W metodzie opartej na sterownikach rozmytych brak precyzji w objaśnieniu elementów systemów, w szczególności brak informacji doboru parametrów fazyfikacji, informacji o technice tworzenia reguł, a także doboru metod agregacji przesłanek czy konkluzji w poszczególnych regułach (opis teoretyczny metod wyostrzania także jest nieprecyzyjny).

Ponadto, edytorsko praca ma kilka niedociągnięć, np. braki interpunkcyjne na stronach: 12, 23, 25-28, 33, 35,38, 40-43, 71.

Jednakże, na uwagę zasługuje fakt, że Autor wykazał się dużą wiedzą w zakresie tematyki rozprawy, umiejętnością pracy naukowej oraz znajomością metod badawczych. Ponadto, Autor wskazuje dalsze kierunki rozwoju rozpoczętych badań.

Ponad te wymienione w niniejszej rozprawie, warte dalszego badania, jest np. zbadanie wpływu na klasyfikacje różnych metod fazyfikacji, agregacji czy wyostrzania w tworzonych sterownikach rozmytych prestiżu publikacji.

Niemniej jednak, powyższe uwagi i wskazówki oraz kilka literówek nie wpływają na wysoką wartość merytoryczną pracy.

4 Konkluzja

Reasumując rozprawa jest cennym wkładem do zagadnienia analizy ilościowej działalności naukowej, czyli naukometrii, co za tym idzie temat pracy oceniam jako aktualny, istotny poznawczo i mający duże implikacje praktyczne. Opracowane metody/systemy zostały zastosowane i zweryfikowane dla rzeczywistych problemów. Mgr Łukasz Szymuła wykazał się dużą wiedzą w zakresie tematyki rozprawy. Osiągnięte wyniki świadczą o bardzo dobrym przygotowaniu Autora do pracy naukowej.

Rozprawa doktorska mgr Łukasza Szymuli spełnia wymagania i warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2022 r. poz. 574 z późn. zm.), zatem wnoszę o dopuszczenie mgr Łukasza Szymulę do dalszych etapów postępowania o nadanie stopnia doktora.