dr hab. Michał Baczyński
Instytut Matematyki
Wydział Matematyki, Fizyki i Chemii
Uniwersytetu Śląskiego
ul. Bankowa 14
40-007 Katowice

Katowice, 6 maja 2016 roku

Recenzja

pracy doktorskiej pana mgra Piotra Żywicy

„Miary podobieństwa przedziałowych zbiorów rozmytych w klasyfikacji danych niepewnych. Zastosowania w diagnostyce guzów jajnika."

Niniejsza recenzja została napisana na podstawie pisma prof. dra hab. Jerzego Kaczorowskiego, Dziekana Wydziału Matematyki i Informatyki UAM, z dnia 29 lutego 2016 r., w związku z prowadzonym przez Radę Wydziału Matematyki i Informatyki UAM przewodem doktorskim mgra Piotra Żywicy.

Recenzowana rozprawa doktorska liczy łącznie 112 stron. Składa się ona ze Wstępu oraz 4 podstawowych rozdziałów. Pracę uzupełnia obszerna bibliografia licząca 124 pozycje, a ponadto spis oznaczeń. Do pracy nie dołączono żadnego nośnika z oprogramowaniem, jednakże w bibliografii podano adres publicznej strony Internetowej udostępniającej repozytorium kodu wykorzystywanego w aplikacyjnej części pracy (zob. `https://github.com/bikol/phd-thesis-applications`) oraz adres strony Internetowej projektu OvaExpert (zob. `http://ovaexpert.pl/`).

Tematyka miar podobieństwa zbiorów rozmytych oraz przedziałowych zbiorów rozmytych jest bardzo ważna i aktualna nie tylko z teoretycznego punktu widzenia, lecz także w kontekście zastosowań. Badania dotyczące klasycznej sytuacji miały swoje apogeum pod koniec XX wieku, kiedy zostało opublikowanych wiele prac proponujących różną aksjomatykę podobieństwa zbiorów rozmytych. W ostatnich dwudziestoleciu badania te zostały rozszerzone na różne uogólnienia zbiorów rozmytych, i doczekały się nawet osobnej monografii poświęconej temu tematowi (poz. [75], *E. Szmidt, Distances and Similarities in Intuitionistic Fuzzy Sets, Springer, 2014*). Recenzowana praca wpisuje się w nurt tych badań i dotyczy miar podobieństwa przedziałowych zbiorów rozmytych, ale przy założeniu, że podobieństwo wyrażone jest w postaci przedziału, czyli głównym obiektem badań w tej dysertacji są funkcje postaci

$$\hat{s}\colon \hat{\mathbb{E}} \to \mathcal{I}(\mathbb{R}),$$

gdzie $\hat{\mathbb{E}} \subseteq IVFS(U) \times IVFS(U)$ (zbiór $\hat{\mathbb{E}}$ spełnia dodatkowe założenia, zob. str. 24), $IVFS(U)$ oznacza rodzinę wszystkich przedziałowych zbiorów rozmytych określonych nad danym uniwersum $U \neq \emptyset$ oraz $\mathcal{I}(\mathbb{R})$ oznacza zbiór wszystkich przedziałów domkniętych w $\mathbb{R}$.

Ah.

Możemy wyróżnić następujące cele pracy (por. Wstęp).

- Podanie i zbadanie własności, jakie powinna spełniać miara podobieństwa przedziałowych zbiorów rozmytych wyrażona w postaci przedziału, a nie tak jak to wcześniej było badane w postaci liczby rzeczywistej (rozdział 3).
- Zaproponowanie metod umożliwiających konstrukcję miar spełniających określone wcześniej własności oraz podanie algorytmów wyznaczających efektywnie uogólnioną moc względną (rozdział 3).
- Zastosowanie omawianych metod w praktycznych zagadnieniach dotyczacych klasyfikacji danych niepewnych (rozdział 4).

Jak pisze autor we Wstępie, motywacją dla powyższych zagadnień „był problem diagnostyki różnicowej guzów jajnika, gdzie niekompletność i niepewność danych jest nieunikniona". Po przeczytaniu całej rozprawy mogę stwierdzić, że powyższe cele zostały w pełni osiągnięte.

Rozdział 1 zatytułowany „Podstawowe pojęcia teorii zbiorów rozmytych" jest krótkim wprowadzeniem do zagadnień stanowiących bazę recenzowanej pracy i zawiera podstawowe definicje: normy trójkątnej, konormy trójkątnej, zbioru rozmytego, mocy skalarnej oraz przedziałowego zbioru rozmytego. Podano tutaj również pewne przykłady operacji triangularnych oraz zacytowano twierdzenie o reprezentacji mocy skalarnej (twierdzenie 1.2).

Rozdział 2 zatytułowany „Niepewność danych oraz podobieństwo zbiorów rozmytych" rozpoczyna autorska analiza pojęcia niepewności danych. Najistotniejszym wątkiem są rozważania nad interpretacją funkcji przynależności przedziałowych zbiorów rozmytych w kontekście niepewności. Autor opisuje tutaj dokładniej dwa podejścia: ontyczne i epistemiczne, oraz ilustruje je odpowiednio dobranymi przykładami (ta część pracy wpisuje się w nurt dyskusji opisanej w pracy [20]). Szczególnie te drugie podejście, czyli epistemiczne, ma istotne znaczenia dla autora rozprawy i będzie ono stanowiło punkt wyjścia do dalszych (w następnych rozdziałach) rozważań na miarami podobieństwa dla przedziałowych zbiorów rozmytych. W kolejnych częściach tego rozdziału autor przedstawił i omówił znane z literatury (zob. m.in. prace [111], [43], [32], [15] oraz niedawno wydaną monografię [75]) własności miar podobieństwa zbiorów rozmytych (podrozdział 2.5.1) oraz przedziałowych zbiorów rozmytych (podrozdział 2.5.2). Należy podkreślić, że w tym rozdziale autor omawia podobieństwa określone poprzez funkcje typu $s\colon \mathbb{E} \longrightarrow \mathbb{R}$ (miara podobieństwa zbiorów rozmytych) oraz funkcje typu $\hat{s}\colon \hat{\mathbb{E}} \longrightarrow \mathbb{R}$ (standardowa miara podobieństwa przedziałowych zbiorów rozmytych), gdzie zbiory E oraz $\hat{\mathbb{E}}$ są odpowiednio zdefiniowane na str. 24. Przegląd przykładów najważniejszych klasycznych miar podobieństwa zbiorów rozmytych zawarty jest w podrozdziale 2.6, gdzie podano wzory na: miary oparte na teorii mnogości (np. klasyczny już indeks Jaccarda), miary oparte na odległości (np. miara Minkowskiego) oraz miary oparte na logice (np. miary oparte na operatorze implikacji rozmytej). Rozdział ten nie zawiera nowych wyników matematycznych, jednak na uwagę zasługuje pierwsza część i autorskie przedstawienie omawianego zagadnienia.

ΠΜ

Rozdział 3, najdłuższy w całej rozprawie, gdyż liczący 36 stron, zatytułowany „Podobieństwo epistemicznych przedziałowych zbiorów rozmytych" stanowi najważniejszą część pracy doktorskiej mgra Żywicy. Tak naprawdę może on być podzielony na dwie części. W pierwszej (podrozdział 3.2) autor koncentruje się na własnościach podobieństwa epistemicznych przedziałowych zbiorów rozmytych. Przypomnijmy, że chodzi o takie miary podobieństwa, który zwracają przedział jako miarę podobieństwa, czyli funkcje postaci $\hat{s}\colon \hat{\mathbb{E}} \to \mathcal{I}(\mathbb{R})$. Autor nazywa je także *miarami podobieństwa uwzględniającymi niepewność*. Najpierw, w podrozdziale 3.2.1 podano te własności, które są uogólnieniem wcześniej omawianych własności dla przypadku klasycznego. Następnie, w podrozdziale 3.2.2, poddano analizie te własności, które są właściwe dla podejścia epistemicznego. W końcu w następnym podrozdziale udowodniono pewne zależności pomiędzy różnymi własnościami miar podobieństwa. Najważniejszym wynikiem matematycznym jest tutaj wniosek 3.10, podsumowujący rezultaty uzyskane we wcześniejszych lematach. Część druga (podrozdziały 3.3-3.6) dotyczy konstrukcji miar podobieństwa uwzględniających niepewność. W podrozdziale 3.3 opisano ogólną metodę (definicja 3.3) opartą na tzw. konstrukcji „wavy-slice" (inna nazwa to reprezentacja Mendel–John). Główną zaletą tej konstrukcji jest fakt, że miarę $\tilde{s}$ podobieństwa dla epistemicznych przedziałowych zbiorów rozmytych uzyskuje się z danej miary s podobieństwa zbiorów rozmytych (wzór (3.31)). Definicja ta ma też swoje ograniczenia, gdyż należy minimalizować oraz maksymalizować daną miarę s. Najważniejszym nowym wynikiem matematycznym w tym podrozdziale jest twierdzenie 3.17, które pokazuje, jak omawiane wcześniej w rozdziale 2 własności miary s przenoszą się na omawiane w rozdziale 3 własności miary $\tilde{s}$. Podrozdział 3.4 ilustruje rozszerzenia znanych miar podobieństwa (odległość Minkowskiego, miary podobieństwa oparte na operatorze implikacji, uogólniony indeks Jaccarda). W szczególności, autor zdefiniował tutaj uogólnioną przedziałową moc względną (definicja 3.5) opartą na t-normie ciągłej T i ciągłej funkcji wagowej f oraz pokazał jej związek z uogólnionym indeksem Jaccarda (wzór (3.92)). Kolejny podrozdział 3.5 dotyczy efektywnego wyliczania uogólnionej mocy względnej. Opierając się na znanych z literatury (zob. [58] oraz [35]) dwóch algorytmach dla pewnych przypadków szczególnych (pseudokody dla t-normy minimum oraz t-normy produktowej są podane w algorytmie 1) podano dwa algorytmy umożliwiające rozwiązanie problemu w ogólności (algorytm 2 dla pary funkcji (T, f) posiadających pewną dodatkową własność u oraz algorytm 3 dla par nieposiadających tej własności, gdzie T jest t-normą). Należy zaznaczyć, ze wyniki przedstawione w tym ostatnim podrozdziale zostały już opublikowane w artykule [124] opublikowanym w czasopiśmie *Fuzzy Sets and Systems*.

Rozdział 4 zatytułowany „Zastosowania w klasyfikacji" ilustruje w jaki sposób rozważania teoretyczne omówione we wcześniejszym rozdziale można zastosować do problemów praktycznych. Wpierw autor opisuje dwie metody klasyfikacji oparte na miarach podobieństwa uwzględniających niepewność: metodę k najpodobniejszych sąsiadów oraz rozmyty klasyfikator przedziałowy. W dość zwięzłym podrozdziale 4.2 zilustrowano zdefiniowane wcześniej metody klasyfikacji na jednym zbiorze danych pochodzącym z ogólnodostępnego archiwum KEEL (problem dysleksji). Jednak najwięcej miejsca autor poświecił omówieniu rzeczywistego i bardzo ważnego zagadnienia diagnostyki różnicowej

Wł

guzów jajnika. Obie metody zostały przetestowane w specjalnie napisanym pakiecie w języku R. Na uwagę zasługuje liczba przetestowanych klasyfikatorów w drugim zagadnieniu (600 wariantów metody k najpodobniejszych sąsiadów oraz 350 wariantów przedziałowgo klasyfikatora rozmytego, por. opis na str. 90). Te różne podejścia były zależne o doboru norm trójkątnych, porządków oraz innych dodatkowych parametrów. Najważniejszym wnioskiem z przeprowadzonych badań dla problemu klasyfikacji guzów jajnika jest fakt, że zaproponowane metody (w dwóch różnych scenariuszach: diagnostyka na podstawie danych pacjentki, podrozdział 4.3.2 oraz diagnostyka na podstawie modeli diagnostycznych, podrozdział 4.3.3) uzyskały znaczącą skuteczność. Praca kończy się krótkim opisem działającego w trybie rzeczywistym systemu OvaExpert wspomagającego diagnostykę guzów jajnika, gdzie jedna z zaproponowanych metod (przedziałowy klasyfikator rozmyty) została wykorzystana. Na uwagę zasługuje fakt, że wyniki omawiane w tym rozdziale były prezentowane na konferencjach dotyczących szeroko rozumianych systemów inteligentnych, jak również na konferencji zorientowanej medycznie. System OvaExpert został opisany w opublikowanym artykule [122].

Zasadniczo nie mam większych zastrzeżeń do redakcji rozprawy. Zdarzają się drobne błędy językowe (zob. np. pierwsze zdanie we własności 3.9 lub na str. 77 powinno być „nieznane" zamiast „nie znane") oraz jak w prawie każdej rozprawie naukowej, i tu występują literówki (np. w definicji 1.8 powinno być "dowolną", a nie „dowolna"), ale moim zdaniem inne takie drobne błędy nie są warte dokładnej numeracji. Mam jednakże pewne uwagi krytyczne (merytoryczne), które zamieszczam poniżej.

Uwagi ogólne i szczegółowe:

1. Brak spójności w podawaniu autorów cytowanych prac. Czasami autor pisze „et al." (np. str. 18_3), czasami „i inni' (np. str. 63_2), a czasami podaje tylko pierwszego autora (np. w obserwacji 3.12), co na pewno nie jest poprawnym rozwiązaniem.

2. Nazwa „wavy-slice representation" nie pochodzi od nazwisk autorów i powinna być pisana małymi literami.

3. str. 13: Użyty we wczesnych pracach oraz tutaj zapis μ_A jest aktualnie niestosowany, gdyż z czysto logicznego punktu widzenia nie jest on poprawny – zbiór rozmyty utożsamia się po prostu z funkcją. Tak zresztą jest to zapisane w pracy [124], której mgr Żywica jest współautorem.

4. str. 13: W definicji dopełnienia zbioru rozmytego wykorzystuje się negacje rozmyte, a nie tylko negację klasyczną $N(x) = 1 - x$.

5. Wiele definicji lub twierdzeń jest poprawnych/prawdziwych tylko przy założeniu, że uniwersum U jest zbiorem skończonym. Dotyczy to np. notacji singletonowej, definicji mocy skalarnej, twierdzenia 1.2, itp.

6. str. 17: Zdanie „Stąd też metody skuteczne w jednym zastosowaniu, okazywały się całkowicie nieprzydatne w innych" wymaga komentarza i jednak wskazania konkretnych problemów, gdzie taka sytuacja miała miejsce.

7. str. 26^7: powinno być $\mathbb{R}$ zamiast $[0,1]$.

8. str. 27, rysunek 2.2: Zamiast „$FS(\{1,2,3,4,5,6\})$" powinno być 7 punktów zapisanych następująco „$FS(\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\})$".

9. str. 33: W analizie zależności pomiędzy różnymi uogólnieniami zbiorów rozmytych brakuje istotnej informacji, że zbiory intucjonistyczne wg Atanassova są równoważne, z matematycznego punktu widzenia, przedziałowym zbiorom rozmytym (zob. *G. Deschrijver, E.E. Kerre, On the relationship between some extensions of fuzzy set theory, Fuzzy Sets and Systems 133 (2003) 227–235.*). Tym samym wszystkie definicje, przykłady i własności miar podobieństwa dla zbiorów mglistych lub intucjonistycznych wg Atanassova można w łatwy sposób przenieść na przypadek przedziałowych zbiorów rozmytych.

10. str. 36: Brak odnośnika do literatury w zdaniu „...jednakże spośród t-norm archimedesowych zachodzi on tylko dla rodziny Franka", oraz brak informacji jak ta rodzina jest określona.

11. str. 39: Jeżeli podano wyjaśnienie skrótu „S-implikacje", to autor powinien też uzasadnić skrót „R-implikacje".

12. str 43: W obserwacji 3.1 powinno być „$I([0,1])$" zamiast samego przedziału $[0,1]$.

13. str. 53: Szkoda, że nietrywialny przykład ukazany na rysunku 3.3 nie został omówiony szerzej. Bez podania dokładnych wzorów na funkcje przynależności nie można go przeanalizować i po prostu trzeba zaufać autorowi.

14. Podrozdział 3.5: Wydaje się, że niektóre przedstawione tutaj wyniki dotyczące uogólnionego indeksu Jaccarda (wzory (3.69), (3.70), twierdzenia 3.18 oraz 3.20) należało jednak umieścić w podrozdziale 2.6.1, gdyż dotyczą one klasycznej sytuacji i własności omawianych w rozdziale 2.

15. W twierdzeniu 3.18 przy badaniu aksjomatu S5 podano pewne warunki wystarczające. Czy są one również warunkami koniecznymi? Jeżeli nie, to można było podać stosowny kontrprzykład. Podobne pytanie można zadać do twierdzenia 3.20 i podanych tam warunków wystarczających.

16. Zastanawiają dziwne błędy w Bibliografii, szczególnie w przypadku książek. W pozycji [5] podany jest błędny tytuł monografii (powinno być „Fuzzy Implications"), w pozycji [36] powinno być „Kluwer" zamiast „Springer Science & Business Media", a w pozycji [75] podano państwo zamiast miasta.

Pomimo tych uwag krytycznych, **pracę doktorską Pana mgra Piotra Żywicy oceniam wysoko.** Do jej najważniejszych wyników czysto matematycznych należy zaliczyć zaproponowanie własności podobieństwa epistemicznych przedziałowych zbiorów rozmytych oraz sformułowanie i udowodnienie twierdzenia 3.17. Do najważniejszych wyników związanych z informatyką należy zaliczyć zaproponowanie i analizę dwóch algorytmów obliczających efektywnie uogólnioną przedziałową moc względną, a dokładniej algorytmy obliczające szukane wartości we wzorach (3.96) oraz (3.97). Dowody prowadzone są poprawnie. Doktorant sprawnie posługuje się aparatem i językiem matematycznym, zdaje się być osobą dociekliwą i otwartą na nowe idee. Osiąga sukces między innymi dzięki temu, że umiejętnie wykorzystuje zdobytą wiedzę z różnych dziedzin na pograniczu matematyki oraz informatyki. Niektóre wyniki zawarte w dysertacji autor przedstawiał na różnych konferencjach międzynarodowych, a część z tych wyników jest już opublikowana lub przyjęta do druku w dobrych czasopismach międzynarodowych (zob. [124], [110]).

Podsumowując uważam, że rozprawa doktorska pana mgra Piotra Żywicy stanowi wartościową analizę miar podobieństwa (epistemicznych) przedziałowych zbiorów rozmytych, wnosząc istotnie nowe wyniki w istniejące już badania. Rezultaty przedstawione w dysertacji są nie tylko interesujące same w sobie, ale mogą stanowić inspirację do dalszych prac, np. do głębszego zbadania miar podobieństwa opartych na implikacjach rozmytych. Podjęte zagadnienia oraz zastosowane metody badawcze, zarówno formalne jak i praktyczne, są właściwe dla rozpraw doktorskich w dziedzinie nauk matematycznych w zakresie informatyki. **W mojej opinii recenzowana praca doktorska spełnia wszystkie wymagane ustawą warunki, to jest art. 13 ust. 1 *Ustawy z dnia 18 marca 2011 r. o zmianie ustawy - Prawo o szkolnictwie wyższym, ustawy o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki oraz o zmianie niektórych innych ustaw*.** Stanowi ona oryginalne rozwiązanie problemu naukowego i wykazuje ogólną wiedzę teoretyczną autora w zakresie matematyki, w szczególności w teorii zbiorów rozmytych. Dlatego wnoszę o jej przyjęcie i dopuszczenie pana mgra Piotra Żywicę do dalszych etapów przewodu doktorskiego, w szczególności do publicznej obrony rozprawy doktorskiej. **Jednocześnie stawiam wniosek o wyróżnienie rozprawy doktorskiej.**

Michał Baczyński

Michał Baczyński