UNIWERSYTET IM. ADAMA MICKIEWICZA
W POZNANIU

Wydział Matematyki i Informatyki

Miary podobieństwa i zawierania zbiorów rozmytych

Similarity and subsethood measures of fuzzy sets

Patryk Żywica
Numer albumu: 329575
Kierunek: matematyka

PRACA MAGISTERSKA
wykonana pod kierunkiem
prof. UAM dr hab. Macieja Wygralaka

Poznań 2014

Poznań, dnia 11 stycznia 2015

OŚWIADCZENIE

Ja, niżej podpisany Patryk Żywica student Wydziału Matematyki i Informatyki Uniwersytetu im. Adama Mickiewicza w Poznaniu oświadczam, że przedkładaną pracę dyplomową pt:

„*Miary podobieństwa i zawierania zbiorów rozmytych*"

napisałem samodzielnie. Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałem z pomocy innych osób, a w szczególności nie zlecałem opracowania rozprawy lub jej części innym osobom, ani nie odpisywałem tej rozprawy lub jej części od innych osób.

Oświadczam również, że egzemplarz pracy dyplomowej w wersji drukowanej jest całkowicie zgodny z egzemplarzem pracy dyplomowej w wersji elektronicznej.

Jednocześnie przyjmuję do wiadomości, że przypisanie sobie, w pracy dyplomowej, autorstwa istotnego fragmentu lub innych elementów cudzego utworu lub ustalenia naukowego stanowi podstawę stwierdzenia nieważności postępowania w sprawie nadania tytułu zawodowego.

TAK – wyrażam zgodę na udostępnianie mojej pracy w czytelni Archiwum UAM

TAK – wyrażam zgodę na udostępnianie mojej pracy w zakresie koniecznym do ochrony mojego prawa do autorstwa lub praw osób trzecich

Streszczenie

Praca poświęcona jest zagadnieniu zawierania i podobieństwa zbiorów rozmytych. Dokonano przeglądu metod i modeli, zarówno w przypadku klasycznej teroii mnogości, jak i teorii zbiorów rozmytych. Pokazano niektóre relacje pomiędzy miarami uzyskanymi różnymi metodami. Zaprezentowano autorską metodę wizualizacji wraz z jej rezultatami dla wybranych miar zawierania zbiorów rozmytych.

Słowa kluczowe zbiory rozmyte, zawieranie, podobieństwo, miara podobieństwa, miara zawierania

Summary

This work is devoted to the issue of inclusion and similarity of fuzzy sets. An overview of methods and models were presented, both in the case of classical set theory and fuzzy set theory. The relationships between some measures obtained by various methods were shown. An original method for the visualization and its results for selected subsethood measures of fuzzy sets were presented.

Keywords fuzzy sets, inclusion, similarity, subsethood measure, inclusion grade, inclusion measure, similarity measure

Spis treści

Streszczenie **iv**

Wstęp **vii**

1. Podejście klasyczne **9**

1.1. Aspekty psychologiczne podobieństwa obiektów 9

1.2. Przykłady 10

1.2.1. Taksonomia 11

1.2.2. Psychologia 12

2. Teoria zbiorów rozmytych **16**

2.1. Elementy teorii zbiorów rozmytych 16

2.1.1. Definicja zbioru rozmytego 16

2.1.2. Operacje na zbiorach rozmytych 19

2.1.3. Moc zbiorów rozmytych 22

3. Rozmyte miary zawierania i podobieństwa **23**

3.1. Podejście oparte na teorii mnogości 23

3.1.1. Miary zawierania 24

3.1.2. Miary podobieństwa 26

3.1.2.1. Własności 29

3.2. Podejście oparte na bliskości 30

3.2.1. Miary podobieństwa 31

3.2.1.1. Uogólnione metryki 31

3.2.1.2. Odległość kątowa 32

3.2.1.3. Odległość przedziałowa 33
3.2.1.4. Wektory parametrów 35

3.3. Podejście oparte na logice 36

3.3.1. Miary zawierania 36
3.3.2. Miary podobieństwa 37

4. Wizualizacja wybranych miar zawierania **39**

4.1. Konstrukcja wykresów 39

4.2. Wizualizacje 40

5. Podsumowanie **51**

Literatura **53**

Wstęp

We wszystkich dziedzinach nauki od dawna istniała potrzeba porównania pewnych obiektów. Podczas, gdy niektóre gałęzie nauki starały się odpowiedzieć na pytanie o naturę podobieństwa, inne potrzebowały precyzyjnej, formalnej jego definicji. Porównanie dwóch obiektów lub zdarzeń może być postrzegane jako próba określenia relacji pomiędzy nimi. Najważniejszymi i najczęściej wykorzystywanymi relacjami pomiędzy obiektami są podobieństwo, różność i zawieranie. W literaturze najwięcej uwagi poświęcone jest zagadnieniu podobieństwa obiektów.

W ostatnich dziesięcioleciach teoria zbiorów rozmytych znalazła zastosowanie w wielu obszarach nauki jak i życia codziennego. Potrzeba porównywania zbiorów rozmytych pojawiła się nautralnie już na samym początku istnienia teorii. Opracowano wiele metod często bazujących na tych wykorzystywanych w przypadku klasycznych zbiorów. Intensywny rozwój logiki rozmytej oraz jej zastosowań często wymaga określenia nowych sposobów porównywania obiektów. Zagadnienie to jest szczególnie istotne w komputerowym wspomaganiu decyzji, klasyfikacji czy przetwarzaniu języka naturalnego.

Pomimo, że zagadnienie porównywania obiektów jest kluczowe dla wielu zastosowań tej teorii, wciąż nie udało się jednoznacznie sformalizować podstawowych pojeć takich jak podobieństwo czy zawieranie. Podczas, gdy część badaczy zajmujących się logiką rozmytą dąży do precyzyjnego zdefiniowania pojęć, inni kwestionują to podejście tłumacząc, że narzucanie sztywnych ram ogranicza możliwość zastosowań praktycznych.

Klasyfikacja miar podobieństwa, jak również terminologia i część oznaczeń została zaczerpnięta z książki autorstwa Valerie Cross oraz Thomasa Sudkampa [10]. W miarę możliwości klasyfikacje wzbogacono o wyniki zawarte w nowych pracach.

Pierwszy rozdział pracy zawiera przegląd metod służących do mierzenia podobieństwa korzystających z klasycznej teorii mnogości. Przedstawiono również dyskusję nad psychologicznymi aspektami podobieństwa. Rozdział drugi wprowadza podstawowe pojęcia teorii zbiorów rozmytych. Szczególnie istotne jest wprowadzenie ogólnych definicji operacji przekroju i sumy zależnych od t-normy i t-konormy.

Trzeci rozdział zawiera przegląd miar zawierania i podobieństwa. Został on podzielony na trzy podrozdziały, każdy poświęcony jest miarom skonstruowanym z

wykorzystaniem innych metod. Rozdział czwarty zawiera opis autorskiej metody wizualizacji miar zawierania zbiorów rozmytych wraz z wykresami wybranych miar.

Rozdział 1

Podejście klasyczne

1.1. Aspekty psychologiczne podobieństwa obiektów

Podobieństwo jest najprawdopodobniej najczęściej używaną i najtrudniejszą do precyzyjnego określenia miarą zgodności. Ten rozdział przedstawia matematyczne techniki określania podobieństwa bazujące na klasycznej teorii mnogości. Analiza podobieństwa dwóch obiektów jest podstawowym narzędziem w biologii, taksonomii, psychologi i stanowi podstawę do rozumowania przez analogię. Problem przypisania unikalnego znaczenia pojęciu podobieństwa, został podjęty przez psychologa Roberta Gregsona [20]. Pojęcie podobieństwa nie jest tak proste jak to się wydaje, gdyż składa się na nie kilka różnych procesów i operacji [44].

Zazwyczaj przyjmuje się, że określenie różnic lub odmienności pomiędzy obiektami jest równoważne określeniu ich podobieństwa [34, 39]. Jeśli podobieństwo dwóch obiektów A i B oznaczmy przez $Sim(A, B)$, to ich odmienność równa jest $1 - Sim(A, B)$. Niektóre badania wykazały jednak, że człowiek inaczej postrzega każde z tych pojęć. Podczas określania podobieństwa, większy nacisk kładziony jest na cechy wspólne obiektów, podczas gdy do określania odmienności wykorzystywane są głównie różnice pomiędzy obiektami [56]. W teorii zbiorów rozmytych najczęściej przyjmuje się założenie o odwrotności podobieństwa i odmienności [1, 41, 62]. Czasem dla zaznaczenia rozróżnialności tych pojęć wartość $1 - Sim(A, B)$ nazywana jest niepodobieństwem [12].

Powszechnie uważa się, że podobieństwo jest symetryczne. W języku miar podobieństwa oznacza to, że

$$Sim(A, B) = Sim(B, A)\,.$$

Formalne badania wykazały, że podobieństwo nie zawsze może być traktowane jako relacja symetryczna [55]. Ma to miejsce w sytuacji, gdy jeden z obiektów staje się obiektem odniesienia. Jawne określenie odniesienia może nastąpić przykładowo poprzez sformułowanie zadania ("Proszę porównać obiekt A z obiektem B"). Odniesie-

nie może zostać również określone niejawnie. W takiej sytuacji najczęściej obiektem odniesienia staje się obiekt bardziej poprawny, lepszy albo o bardziej złożonej strukturze. Przykładowo, osoba ma za zadanie porównać jakość dwóch obrazów przedstawiających tę samą scenę. Obraz o lepszej jakości staje się implicite obiektem odniesienia. Z tego powodu, założenie symetrii zostało osłabione w miarach wykorzystywanych w niektórych dziedzinach takich jak psychologia bodźców.

W badaniach nad podobieństwem konieczne jest również rozróżnienie pomiędzy prostym podobieństwem opartym na wielowymiarowej pojedynczej cesze, od podobieństwa będącego pochodną relacji poznawczych pomiędzy wieloma różnymi cechami [53]. Przykładem pojedynczej wielowymiarowej cechy obiektu jest kolor, którego składowymi są jasność, czy saturacja. Z drugiej strony, kształt geometryczny ma wiele różnych cech takich jak typ figury, rozmiar czy orientacja. W pierwszym przypadku podobieństwo zachowuje się podobnie jak miara odległości w przestrzeniach Euklidesowych. Jednak w przypadku wielu różnych cech, podobieństwo jest bardziej skomplikowane i zależy nie tylko od samych obiektów, ale również od kontekstu podejmowania decyzji.

Jak wykazały badania psychologiczne podobieństwo może być zależne od kontekstu w jakim jest ono oceniane [20, 53]. W określeniu $Sim(A, B)$ biorą udział nie tylko cechy obiektów A i B, ale również cała charakterystyka otoczenia. Choć wydaje się to niepoprawne, na oszacowanie podobieństwa dwóch obiektów wpływ mają również cechy innych badanych obiektów. Podczas oceniania podobieństwa wielu obiektów przez człowieka na istotność poszczególnych cech obiektu wpływa również zmienność ich wartości. Charakter tego wpływu może być jednak różny w zależności od tego jaki jest cel porównania. W problemach klasyfikacji obiektów wykazano, że rzadkie wartości cechy są bardziej istotne [51], podczas gdy na potrzeby taksonomii korzystniej jest przyjąć przeciwne założenie [18].

1.2. Przykłady

Oczywiście powyżej wspomniane problemy nie muszą występować we wszystkich obszarach badawczych, w których zgodność pomiędzy obiektami wykorzystywana jest do analizy informacji i wspomagania podejmowania decyzji. W tej sekcji przedstawiony zostanie problem porównywania obiektów w dziedzinie taksonomii oraz psychologii. Wszystkie omówione metody oparte są na klasycznej teorii mnogości. Jak się później okaże, będą one często stanowić punkt wyjściowy do wielu uogólnień pozwalających na mierzenie podobieństwa zbiorów rozmytych.

1.2.1. Taksonomia

Taksonomia, poddyscyplina systematyki biologicznej, to nauka o zasadach i metodach klasyfikowania, katalogowania oraz opisywania organizmów żywych w oparciu o badania ich różnorodności, pochodzenia i pokrewieństwa. Organizmy żywe (obiekty w taksonomii) opisane są poprzez posiadanie lub brak wybranych cech. Uniwersum U stanowi zbiór wszystkich atrybutów wykorzystywanych w klasyfikacji taksonomicznej. Obiekt jest charakteryzowany jest przez zbiór atrybutów który posiada. Dla uproszczenia będziemy zakładać, że obiekt jest zdefiniowany przez wartości swoich atrybutów. Dzięki temu porównanie dwóch obiektów może być oparte na posiadaniu wspólnych atrybutów.

Niech X i Y będą dwoma obiektami. Dla każdego atrybutu może zachodzić jeden z czterech przypadków: występuje w obu obiektach, występuje w X ale nie w Y, występuje w Y ale nie w X oraz nie występuje w obu. Liczba atrybutów dla każdego przypadku to odpowiednio: $|X \cap Y|$, $|X \cap \overline{Y}|$, $|\overline{X} \cap Y|$ oraz $|\overline{X} \cap \overline{Y}|$. Występowanie wspólnych atrybutów stanowiło podstawę do zaproponowania przez Jaccarda [25] następującej metody określania podobieństwa dwóch obiektów

$$Sim_{XY} = \frac{|X \cap Y|}{|X \cap Y| + |X \cap \overline{Y}| + |\overline{X} \cap Y|} = \frac{|X \cap Y|}{|X \cup Y|}\,.$$

Indeks Jaccarda jest również często nazywany współczynnikiem podobieństwa lub indeksem współwystępowania [49].

Indeks Jaccarda nie uwzględnia przypadku braku atrybutu w obu obiektach. Jednakże czasem badacze systematyki biologicznej, mogą uważać dwa organizmy jako podobne ponieważ nie posiadają one pewnego atrybutu. Prosty współczynnik dopasowania rozwiązuje ten problem [50]

$$\hat{Sim}_{XY} = \frac{|X \cap Y| + |\overline{X} \cap \overline{Y}|}{|X \cup Y| + |\overline{X} \cap \overline{Y}|} = \frac{|X \cap Y| + |\overline{X} \cap \overline{Y}|}{|U|}\,. \tag{1.2.1}$$

Prosty współczynnik dopasowania został rozszerzony tak, aby mógł uwzględniać nie tylko atrybuty binarne lecz również te o dowolnej liczbie stanów pośrednich [52].

W obu przypadkach podobieństwo wyrażone jest przy pomocy wartości liczbowej z przedziału $[0, 1]$, gdzie 1 oznacza najwyższy stopień dopasowania. Zasadnicza różnica pomiędzy Sim_{XY} oraz $\hat{Sim}_{XY}$ jest taka, że Sim_{XY} bazuje tylko na liczbie występujących atrybutów. Pominięcie atrybutów niewystępujących w obu obiektach pozwala zredukować rozmiar uniwersum do $|U| - |\overline{X} \cap \overline{Y}|$. Podobieństwo mierzone za pomocą $\hat{Sim}_{XY}$ oparte jest zarówno na "pozytywnych" jak i "negatywnych" dopasowaniach wykorzystując w ten sposób wszystkie dostępne atrybuty, których dobór w tym przypadku jest krytyczny. Zbyt wielki zbiór atrybutów może prowadzić do sytuacji, w której dwa obiekty zostaną uznane za podobne tylko dlatego, że oba nie posiadają dużej ilości cech. Obie miary podobieństwa są używane w

taksonomii. Przykładowo, negatywne dopasowania nie są brane pod uwagę w bakteriologii [49].

Zupełnie inne podejście wykorzystywane jest w numerycznej taksonomii, gdzie obiekt reprezentowany jest przez wektor wartości atrybutów wyrażanych jako liczby rzeczywiste [7]. Obiekt X definiowany przez n atrybutów jest reprezentowany jako wektor $[x_1, \cdots, x_n]$, gdzie x_i jest wartością i-tego atrybutu. Miara podobieństwa pomiędzy dwoma obiektami $X = [x_1, \cdots, x_n]$ i $Y = [y_1, \cdots, y_n]$ wyrażona jest jako cosinus kąta θ pomiędzy dwoma wektorami

$$\cos\theta = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \cdot \sqrt{\sum_{i=1}^{n} y_i^2}} \,.$$

Taka miara podobieństwa wykorzystuje tylko kierunek wektorów, ignorując ich długość. Przemnażanie wektorów X i Y przez dowolne stałe nie zmienia ich podobieństwa.

Odległość Hamminga

$$d_{MCD}(X, Y) = \frac{1}{n} \sum |x_i - y_i| \,,$$

została wykorzystana jako miara podobieństwa taksonomicznego dla obiektów reprezentowanych przez wektory liczb rzeczywistych [8]. Ciekawą modyfikacją tej metryki jest metryka Canberra [35]

$$d_{CAN}(X, Y) = \sum \frac{|x_i - y_i|}{x_i + y_i} \,,$$

która dokonuje przeskalowania bezwzględnej różnicy przez czynnik zależny od porównywanych wartości. Oczywiście w przypadku miar podobieństwa opartych na metrykach, wartość 0 oznacza maksymalne podobieństwo, które maleje wraz ze wzrostem odległości pomiędzy obiektami.

Przeprowadzono liczne badania porównawcze na temat skuteczności różnych podejść do określenia miar podobieństwa na potrzeby klasyfikacji w taksonomii. Pewne zalecenia dotyczące wyboru odpowiedniej miary podobieństwa w taksonomii zostały podane na podstawie wcześniejszych doświadczeń, ale nie ma ogólnie przyjętych norm [49].

1.2.2. Psychologia

Pojęcie podobieństwa jest bardzo istotne w psychologii. Większość teorii psychologicznych odwołuje się do niego bezpośrednio lub pośrednio, tym samym zależy na jego definicji [2]. Porównywanie obiektów zawsze następuje wobec pewnych ustalonych kryteriów. Zasadniczo można je podzielić na dwa typy, obiekty mogą być porównywane pod kątem posiadania/niepodsiadania pewnych wspólnych cech

(atrybutów) lub pod kątem swojej odległości w pewnej ustalonej wielowymiarowej przestrzeni. W literaturze psychologicznej modele oparte na pierwszym podejściu określa się mianem opartych na treści, natomiast drugim – na odległości [20].

Spośród modeli opartych na treści, szeroką grupę stanowią teoriomnogościowe miary, które w większości mogą zostać uogólnione do następującej postaci [55]

$$Sim_{\alpha,\beta}(X,Y) = \frac{f(X \cap Y)}{f(X \cap Y) + \alpha \cdot f(X - Y) + \beta \cdot f(Y - X)} \,. \tag{1.2.2}$$

Wzór ten stanowi parametryzację indeksu Jaccarda, gdzie $X - Y = X \cap \overline{Y}$ oraz $Y - X = \overline{X} \cap Y$. f może być dowolną funkcją spełniającą warunek addytywności na rozłącznych zbiorach

$$f(X \cup Y) = f(X) + f(Y)\,, \qquad \text{jeżeli } X \cup Y = \emptyset\,.$$

Jeśli $\alpha, \beta \geq 0$ to miara podobieństwa jest znormalizowana tak, że

$$0 \leq Sim_{\alpha,\beta}(X,Y) \leq 1\,.$$

Dobór odpowiednich wartości parametrów α i β zależy od celu badań. Poniższe trzy miary są najczęściej stosowane [14, 15, 25]:

- dla $\alpha = \beta = 1$
$$Sim_{1,1}(X,Y) = \frac{f(X \cap Y)}{f(X \cup Y)}\,,$$
- dla $\alpha = \beta = 0.5$
$$Sim_{0.5,0.5}(X,Y) = \frac{2 \cdot f(X \cap Y)}{f(X) + f(Y)}\,,$$
- dla $\alpha = 1$, $\beta = 0$
$$Sim_{1,0}(X,Y) = \frac{f(X \cap Y)}{f(X)}\,.$$

Modele wektorowe były również badane w kontekście badania podobieństwa obiektów w eksperymentach psychologicznych. Początkowo rozważano jednowymiarowy model, w którym obiekt X reprezentowany był przez wartość m_X. Najprostszą miarę podobieństwa można określić w następujący sposób

$$Sim(X,Y) = \frac{m_X}{m_Y}\,, \qquad \text{gdzie } m_X \leq m_Y\,.$$

Model ten jednak nie wykazywał liniowej zależności względem danych eksperymentalnych [15]. Zaproponowano nowy model, który znacznie lepiej odzwierciedlał dane eksperymentalne

$$Sim(X,Y) = \frac{2 \cdot m_X}{m_X + m_Y}\,.$$

Dalsze prace doprowadziły do wprowadzenia dwuwymiarowego modelu podobieństwa, w którym długość wektora modeluje ilość, a kierunek jakość obserwowanego zjawiska [16]. Stopień zawierania jednego pojęcia w drugim może zostać wyliczony jako projekcja jednego wektora na drugi

$$Inc(X,Y) = \frac{m_X \cdot \cos\theta}{m_Y} ,$$

gdzie θ jest kątem pomiędzy tymi wektorami. Jeśli $m_X = m_Y$ miara zawierania jest równa $\cos\theta$, natomiast, gdy $\cos\theta = 1$ wektory mają ten sam kierunek i miara zawierania jest równa prostemu ilorazowi wielkości. Miarę podobieństwa można skonstruować z miary zawierania poprzez połączenie stopnia zawierania X w Y oraz Y w X

$$Sim(X,Y) = \frac{m_X \cos\theta + m_Y \cos\theta}{m_X + m_Y} .$$

Powstało wiele modyfikacji modelu wektorowego mających na celu lepsze odzwierciedlenie danych eksperymentalnych. Wiele z nich oparte było na zastosowaniu uogólnionych operacji na wektorach.

Modele oparte na odległości definiują podobieństwo jako negację odległości obiektów w pewnej przestrzeni. Tradycyjnie odległość definiuje się w przestrzeni Euklidesowej, w której odległość dwóch punktów $X = (x_1, \cdots, x_n)$ i $Y = (y_1, \cdots, y_n)$ mierzona jest za pomocą metryki Minkowskiego

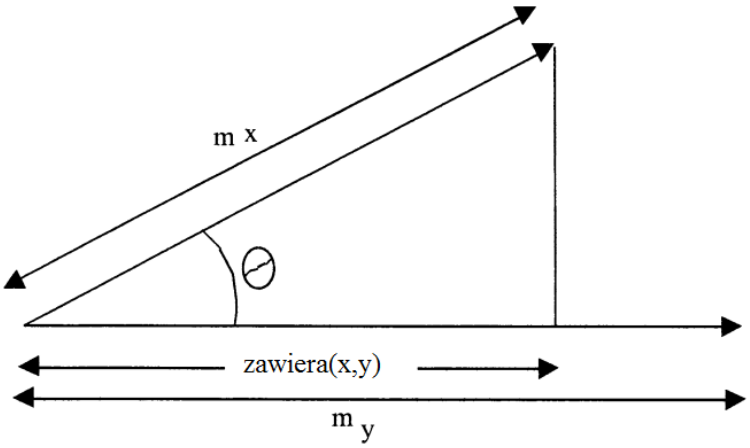
$$d_r(X,Y) = (\sum_{i=1}^{n} |x_i - y_i|^r)^{\frac{1}{r}} .$$

Dla $r \geq 1$ metryki Minkowskiego są nieujemne, symetryczne, niezmienne ze względu na przesunięcia i spełniają nierówność trójkąta. Tylko gdy $r = 2$ metryka Minkowskiego jest niezmienna ze względu na obroty, dlatego odległość Euklidesowa jest najczęściej wykorzystywana w psychologii [42].

Niektórzy badacze kwestionują jednak założenie, że przestrzeń psychologiczna jest Euklidesowa, właśnie głównie z powodu niezmienności ze względu na obroty [2, 23]. Własność ta pozwala stwierdzić, że dany układ odniesienia jest równie dobry jak każdy inny. Jako alternatywę zaproponowano metrykę miejską (metryka Minkowskiego dla $r = 1$), gdzie odległość dwóch punktów jest sumą odległości po wszystkich współrzędnych.

Wspomniane wcześniej rozróżnienie na bodźce wielowymiarowe i wielocechowe znajduje swoje odzwierciedlenie w doborze odpowiedniej metryki. Okazuje się, że dla zjawisk o prostej i łatwej do zrozumienia strukturze najbardziej dopasowana jest odległość Euklidesowa, podczas gdy wraz ze wzrostem skomplikowania struktury coraz bardziej adekwatne jest modelowanie przy użyciu metryki miejskiej [53].

Metryka nieskończoności (metryka Minkowskiego dla $r = \infty$) również jest wykorzystywana jako miara niepodobieństwa obiektów [9, 24]. Ogólnie parametr r w



Rys. 1.2.1. Ilustracja modelu wektorowego. Źródło [10].

metrykach Minkowskiego może być interpretowany jako wskaźnik znaczenia różnic w pojedynczym wymiarze w stosunku do całkowitej odległości punktów. Inna interpretacja wartości r mówi, że im większa jest interakcja pomiędzy poszczególnymi wymiarami tym wartość r powinna być większa [24].

Rozdział 2

Teoria zbiorów rozmytych

Teoria zbiorów rozmytych, ciesząca się obecnie ogromną popularnością, została zapoczątkowana w 1965 roku przez Loftiego Zadeha w artykule pod tytułem „Fuzzy sets”[61]. Dynamiczny rozwój, początkowo niezauważanej teorii, nastąpił dopiero w drugiej połowie lat 70-tych, kiedy udało się rozwiązać problem sterowania piecem do wytwarzania cementu – z wykorzystaniem logiki rozmytej (Mamdani). Kluczowe okazały się również spektakularne sukcesy w Japonii. Najgłośniejszym z nich było opracowanie układu sterowania metrem w mieście Sendai. Dzięki zastosowaniu wnioskowania rozmytego, wykorzystano wiedzę doświadczonych motorniczych, która w połączeniu z możliwościami obliczeniowymi komputerów, pozwoliła między innymi zmniejszyć czasy opóźnień oraz koszty utrzymania metra.

Dalszy rozwój spowodował zastosowanie logiki rozmytej nie tylko w urządzeniach przemysłowych, lecz również codziennego użytku. Jednym z ciekawszych przypadków jest stabilizacja obrazu w kamerach. Problem ten jest bardzo trudny, chociażby ze względu na konieczność odróżnienia zamierzonego ruchu od drgań ręki operatora. Sama teoria również została znacznie rozwinięta, umożliwiając w ten sposób zastosowanie w kolejnych obszarach przemysłu i nauki.

2.1. Elementy teorii zbiorów rozmytych

2.1.1. Definicja zbioru rozmytego

Podstawowym pojęciem opisywanej tu teorii jest zbiór rozmyty. W zastosowaniach matematyki, najczęściej celem definiowania zbioru jest określenie lub sprecyzowanie pewnego pojęcia, takiego jak na przykład liczby parzyste czy wartości dopuszczalne pewnego procesu. Jako ilustracja opisywanych tu pojęć posłuży zbiór określający pojęcie „niewielka względna zmiana”, który mógłby zostać przedstawiony w następującej postaci: $\{x \in \mathbb{R} : -0.2 \leq x \leq 0.2\}$. Jest to podzbiór zbioru

liczb rzeczywistych, który w tym przypadku stanowi tak zwane *uniwersum* (obszar rozważań).

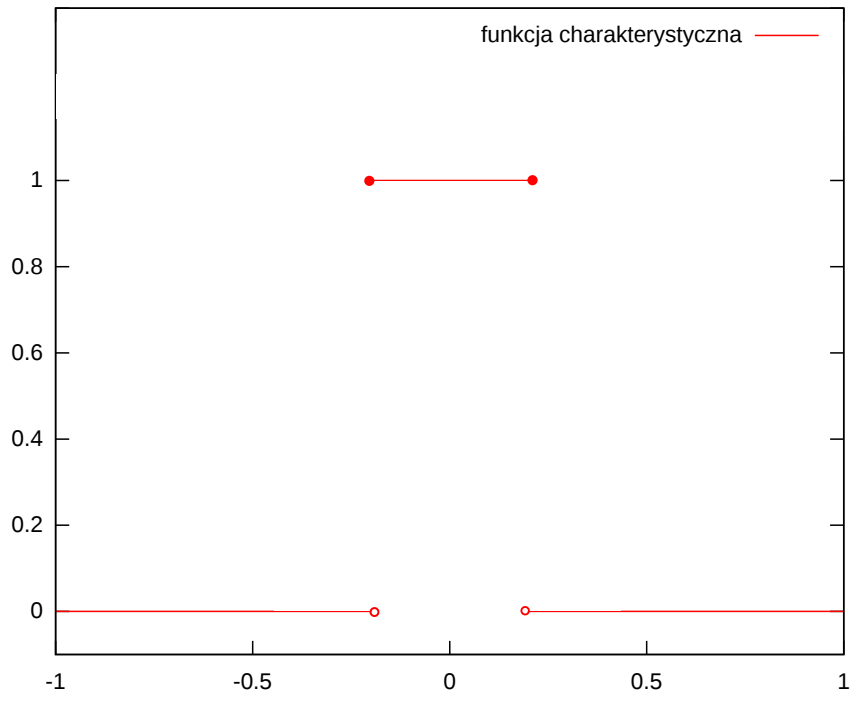
W klasycznej teorii mnogości, z każdym zbiorem A powiązana jest *funkcja charakterystyczna*

$$\chi_A : X \to \{0, 1\},$$

która dla dowolnego elementu x w uniwersum X przyjmuje wartość 1, jeśli x należy do zbioru A, oraz 0 w przeciwnym przypadku. Wykres funkcji charakterystycznej, która odpowiada zbiorowi z przykładu przedstawiony jest na rysunku 2.1.1.

Zastosowanie zbiorów do reprezentacji pojęć (wartości) precyzyjnych, z jakimi mamy do czynienia w matematyce, jest naturalne i nie stanowi problemu. Okazuje się jednak, że człowiek rzadko operuje informacją precyzyjną. Jeśli zapytaćby pasażerów tramwaju, ile czasu trwała podróż, uzyskamy odpowiedzi: około 10 minut, mniej niż kwadrans, dłużej niż zwykle, a nawet nie wiem. Wszystkie odpowiedzi opisują tę samą wartość rzeczywistego czasu przejazdu.

Pomimo tego, że człowiek doskonale radzi sobie z informacją nieprecyzyjną, w wielu przypadkach próba jej wyrażenia za pomocą precyzyjnych zbiorów skazana jest na porażkę. Podstawowym ograniczeniem zbiorów, wykorzystywanych do modelowania pojęć nieprecyzyjnych, jest to, że dla każdego elementu możliwe są tylko dwa stany – może on należeć do zbioru lub nie. Obserwacja ta była podstawą wprowadzenia przez Zadaha pojęcia zbioru nieostrego (rozmytego)[61]. Funkcję charak-



Rys. 2.1.1. Funkcja charakterystyczna zbioru $\{x \in \mathbb{R} : -0.2 \leq x \leq 0.2\}$.

terystyczną zbioru χ_A rozszerzono do *funkcji przynależności* μ_A:

$$\mu_A : X \to [0, 1],$$

która przyjmuje dowolną wartość pomiędzy 0 (całkowite nieprzynależenie) oraz 1 (pełne przynależenie).

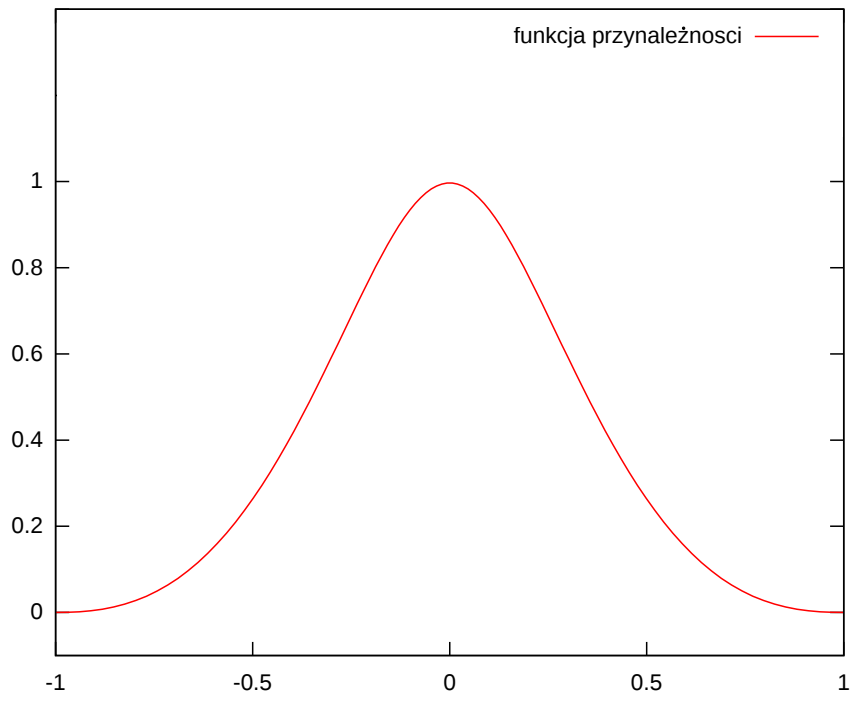
W przypadku przykładowego pojęcia „małych zmian względnych" odpowiedni zbiór rozmyty może zostać przedstawiony jak na rysunku 2.1.2.

Choć funkcje przynależności mogą być (i z reguły są), analogicznie jak w przypadku klasycznym, utożsamiane z odpowiadającymi im zbiorami rozmytymi, najczęściej przyjmuje się jednak inną formalną definicję zbioru rozmytego[26]. Zbiór rozmyty A w uniwersum X określa się jako zbiór (klasyczny) par uporządkowanych

$$A = \{(\mu_A(x), x) : x \in X\},$$

gdzie μ_A jest funkcją przynależności zbioru rozmytego A, a $\mu_A(x) \in [0, 1]$ nazywany jest stopniem przynależności elementu x do zbioru rozmytego A. Tak zdefiniowany zbiór rozmyty jest tożsamy z funkcją przynależności, gdyż ta w świetle teorii mnogości określona jest jako zbiór par uporządkowanych.

W praktyce najczęściej wykorzystywane są dwie reprezentacje zbioru rozmytego: za pomocą funkcji przynależności lub w *postaci singletonowej*. Pierwsza z nich została już omówiona. Druga jest szczególnie przydatna do reprezentacji zbiorów rozmytych, dla których zbiór $supp(A) = \{x \in X : \mu_A(x) > 0\}$, nazywany *nośnikiem*,



Rys. 2.1.2. Funkcja przynależności zbioru przykładowego.

jest skończony lub przeliczalny. Poniższy zapis odpowiada zbiorowi rozmytemu z rysunku 2.1.3

$$A = {}^{0.1}/_{-0.2} + {}^{0.5}/_{-0.1} + {}^{1.0}/_{0.0} + {}^{0.5}/_{0.1} + {}^{0.1}/_{0.2}\,.$$

Korzystając z tej notacji można przedstawić dowolny zbiór klasyczny B jako zbiór rozmyty $\tilde{B}$

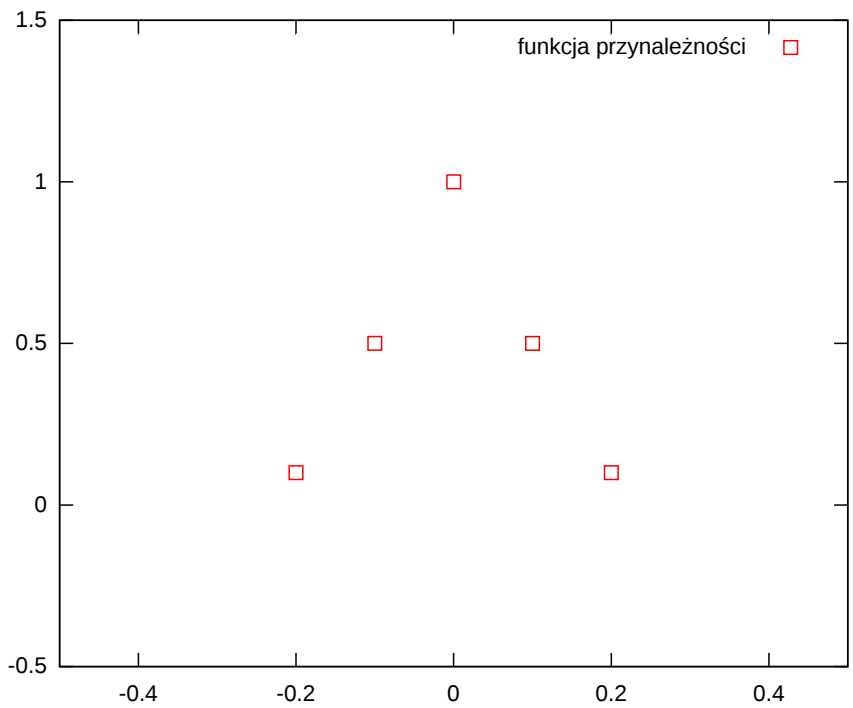
$$\tilde{B} = \sum_{b \in B} {}^{1.0}/_{b}\,.$$

2.1.2. Operacje na zbiorach rozmytych

W klasycznej teorii mnogości podstawowe operacje na zbiorach to: dopełnienie, suma oraz iloczyn. W przypadku zbiorów rozmytych, jak zostanie to dalej opisane, nie są to operacją określone w jeden możliwy sposób. Na początku jednak zostanie omówiony przypadek najprostszy, który jest zarazem najczęściej wykorzystywany w praktyce.

Dopełnienie zbioru rozmytego A określa się jako zbiór rozmyty, którego funkcja przynależności dana jest w następujący sposób:

$$\mu_{\neg A}(x) = 1 - \mu_A(x)\,.$$



Rys. 2.1.3. Zbiór rozmyty $A = {}^{0.1}/_{-0.2} + {}^{0.5}/_{-0.1} + {}^{1.0}/_{0.0} + {}^{0.5}/_{0.1} + {}^{0.1}/_{0.2}$.

Suma dwóch zbiorów rozmytych A i B stanowi zbiór rozmyty, którego funkcja przynależności ma postać

$$\mu_{A\cup B}(x) = \max(\mu_A(x), \mu_B(x))\,,$$

natomiast przekrój (iloczyn) zbiorów określa następujący wzór:

$$\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x))\,.$$

Powyższe definicje sumy i przekroju, mimo że powszechnie stosowane i zgodne z intuicją, nie są jedynymi możliwymi. Poniżej zostaną opisane dwie rodziny funkcji, które dzięki spełnieniu pewnych warunków, mogą zastąpić operację maksimum i minimum w definicjach sumy i przekroju zbioru.

Funkcję $t:\ [0,1]\times[0,1]\to[0,1]$ taką, że dla każdego $a,b,c\in[0,1]$:

$$\begin{aligned}
&t(a,1)=1 && \text{(element neutralny 1)},\\
&a\le b \Rightarrow t(a,c)\le t(b,c) && \text{(monotoniczność)},\\
&t(a,b)=t(b,a) && \text{(przemienność)},\\
&t(a,t(b,c))=t(t(a,b),c) && \text{(łączność)}
\end{aligned}$$

nazywamy *t-normą*.

Funkcję $s:\ [0,1]\times[0,1]\to[0,1]$ taką, że dla każdego $a,b,c\in[0,1]$:

$$\begin{aligned}
&s(a,0)=a && \text{(element neutralny 0)},\\
&a\le b \Rightarrow s(a,c)\le s(b,c) && \text{(monotoniczność)},\\
&s(a,b)=s(b,a) && \text{(przemienność)},\\
&s(a,s(b,c))=s(s(a,b),c) && \text{(łączność)}
\end{aligned}$$

nazywamy *t-konormą*.

Podstawowymi przykładami t-norm są:

- minimum
$$t_{\min}(a,b)=\min(a,b)\,,$$
- iloczyn algebraiczny
$$t_{\text{prod}}(a,b)=a\cdot b\,,$$
- t-norma Łukasiewicza
$$t_{\text{Łuk}}(a,b)=\max(0,a+b-1)\,.$$

Podstawowymi przykładami t-konorm są:

- maksimum
$$s_{\max}(a,b)=\max(a,b)\,,$$

- suma probabilistyczna
$$s_{\text{prod}}(a,b) = a + b - ab\,,$$
- t-konorma Łukasiewicza
$$s_{\text{Łuk}}(a,b) = \min(a+b,1)\,.$$

Ogólne definicje sumy i przekroju zbiorów mają postać:

$$\mu_{A\cup B}(x) = s(\mu_A(x), \mu_B(x))\,,$$

$$\mu_{A\cap B}(x) = t(\mu_A(x), \mu_B(x))\,.$$

W praktyce ważne jest, aby t-norma t i t-konorma s były odpowiednio dopasowane, tj. by spełniony był następujący warunek de Morgana:

$$\forall_{a,b\in\mathbb{R}} \quad s(a,b) = 1 - t(1-a, 1-b)\,.$$

Zapewnia to poprawne zachowanie operacji przekroju i sumy, gdy są one wykonywane łącznie.

Rodzina t-norm i t-konorm Franka jest bardzo ważna ze względu na to, że obejmuje ona wszystkie ciągłe t-normy i t-konormy spełniające następujący warunek

$$s(x,y) + t(x,y) = x + y\,.$$

Jego spełnienie jest szczególnie istotne przy definiowaniu operacji na zbiorach rozmytych. Rodzina ta jest określona dla dowolnego $\lambda \geq 0$ w następujący sposób

$$t_\lambda^F(x,y) = \begin{cases} t_{\min}(x,y)\,, & \text{jeśli } \lambda = 0\,, \\ t_{\text{prod}}(x,y)\,, & \text{jeśli } \lambda = 1\,, \\ t_{\text{Łuk}}(x,y)\,, & \text{jeśli } \lambda = \infty\,, \\ \log_\lambda(1 + \frac{(\lambda^x-1)(\lambda^y-1)}{\lambda-1}) & \text{w przeciwnym przypadku,} \end{cases}$$

oraz

$$s_\lambda^F(x,y) = \begin{cases} s_{\max}(x,y)\,, & \text{jeśli } \lambda = 0\,, \\ s_{\text{prod}}(x,y)\,, & \text{jeśli } \lambda = 1\,, \\ s_{\text{Łuk}}(x,y)\,, & \text{jeśli } \lambda = \infty\,, \\ \log_\lambda(1 + \frac{(\lambda^{1-x}-1)(\lambda^{1-y}-1)}{\lambda-1}) & \text{w przeciwnym przypadku.} \end{cases}$$

Jak łatwo można zuważyć wszyskie trzy podstawowe t-normy i t-konormy należą do rodziny operacji Franka.

2.1.3. Moc zbiorów rozmytych

Moc zbioru to jedna z podstawowych jego charakterystyk. Podczas, gdy w klasycznej teorii mnogości elementy należą do zbioru lub nie, w przypadku zbiorów rozmytych sytuacja znacznie się komplikuje. W literaturze można znaleźć wiele podejść do tego zagadnienia. Dominującą rolę zdają się odgrywać moce skalarne oraz wektorowe zbiorów rozmytych [59]. W tej pracy wykorzystane zostaną tylko moce skalarne.

Funkcję $\sigma : FS \to [0, \infty)$ nazywamy mocą skalarną, gdy σ spełnia następujące warunki dla każdego $a, b \in [0, 1]$, $A, B \in FS$ oraz $x, y \in U$:

1. $\sigma(1/x) = 1$,
2. $a \leq b \implies \sigma(a/x) \leq \sigma(b/y)$,
3. $A \cap B = \emptyset \implies \sigma(A \cup B) = \sigma(A) + \sigma(B)$.

Istnieje prosta charakterystyka wszystkich mocy skalarnych. σ jest mocą skalarną wtedy i tylko wtedy, gdy istnieje niemalejąca funkcja $f : [0, 1] \to [0, 1]$, dla której $f(0) = 0$ i $f(1) = 1$, taka że

$$\sigma(A) = \sum_{x \in \mathrm{supp}(A)} f(A(x))$$

dla dowolnego zbioru rozmytego A. Taką funkcję f nazywamy funkcją wzorocową (alternatywnie wagową). Najpopularniejszą funkcją wagową jest funkcja identycznościowa id.

Moc względna zbioru rozmytego reprezentuje proporcje elementów zbioru rozmytego, które jednocześnie należą do innego zbioru rozmytego, stąd też mówi się o mocy względnej zbioru rozmytego A względem zbioru rozmytego B i oznacza przez $\sigma(A|B)$

$$\sigma(A|B) = \frac{\sigma(A \cap B)}{\sigma(B)}\,.$$

Rozdział 3

Rozmyte miary zawierania i podobieństwa

3.1. Podejście oparte na teorii mnogości

Podejście to ma swoje korzenie w modelu opartym na treści w psychologii oraz w występowaniu wspólnych cech w taksonomii. Podstawowy indeks Jaccarda został uogólniony do postaci podanej w (1.2.2) i posłużył jako podstawa do konstrukcji wielu miar dla zbiorów rozmytych. Zastąpienie klasycznych operacji na zbiorach ich rozmytymi odpowiednikami, pozwala na przeniesienie znanych metod mierzenia podobieństwa klasycznych zbiorów, na zbiory rozmyte [11].

Okazuje się jednak, że korzystając z (1.2.2) nie jest możliwe otrzymanie prostego współczynnika dopasowania 1.2.1. Stąd potrzeba dalszego uogólnienia do następującej postaci

$$Sim_{\alpha,\beta,\gamma}(X,Y) = \frac{g(X \cap Y) + \gamma \cdot g(\overline{X} \cap \overline{Y})}{g(X \cap Y) + \alpha \cdot g(X \cap \overline{Y}) + \beta \cdot g(\overline{X} \cap Y)} \,. \tag{3.1.1}$$

Dla wartości $\alpha = \beta = \gamma = 1$ otrzymujemy prosty współczynnik dopasowania.

Zastępując w powyższym wzorze klasyczne zbiory X i Y przez zbiory rozmyte A i B, otrzymamy model rozmyty. Jako g można przyjąć dowolną funkcję spełniającą następujące warunki:

1. $g(\emptyset) = 0$,
2. $g(U) = 1$,
3. jeśli $A \subseteq B$ to $g(A) \leq g(B)$.

Taka funkcja, nazywana dalej skalarnym ewaluatorem zbioru rozmytego, pozwala na zastąpienie zbioru rozmytego pojedynczą wartością skalarną. Ewaluator jest egzystencjalny jeśli $g(A) = 0$ wtedy i tylko wtedy, gdy $A = \emptyset$ oraz uniwersalny jeśli

$g(A) = 1$ wtedy i tylko wtedy, gdy $A = U$. Przykładem ewaluatora egzystencjalnego jest funkcja

$$g_{\sup}(A) = \sup_{u \in U} \mu_A(u)\,,$$

natomiast

$$g_{\inf}(A) = \inf_{u \in U} \mu_A(u)$$

jest uniwersalnym ewaluatorem. Uogólnione skalarne moce względne [45, 59, 60] tworzą bardzo szeroką rodzinę skalarnych ewaluatorów

$$g_f(A) = \sigma_{f,t}(A|U) = \frac{\sigma_f(A \cap_t U)}{\sigma_f(U)} = \frac{\sigma_f(A)}{\sigma_f(U)}\,,$$

gdzie t jest dowolną t-normą, a f funkcją wagową skalarnej mocy względnej. Uniwersalność i egzystencjalność tak skonstruowanego ewaluatora zależy od doboru funkcji wagowej. Przykładowo g_{id} jest jednocześnie egzystencjalny jak i uniwersalny.

Do konstrukcji operacji sumy i przekroju można wykorzystać dowolną t-normę wraz ze sprzężoną z nią t-konormą. Należy jednak pamiętać, że znany z klasycznej teorii mnogości fakt

$$|X \cup Y| = |X| + |Y| - |X \cap Y|$$

nie zawsze jest spełniony w przypadku rozmytym. Popularna t-norma minimum spełnia ten warunek, jednakże spośród t-norm archimedesowych zachodzi on tylko dla rodziny t-norm Franka.

3.1.1. Miary zawierania

Najprościej stopień zawierania klasycznego zbioru X w zbiorze Y, gdzie $X \nsubseteq Y$ można określić jako stosunek liczby wspólnych elementów X i Y do liczby elementów występujących w jednym lub drugim zbiorze. Przy takim określeniu stopnia zawierania, jest on równy 1 wtedy i tylko wtedy gdy $X \subset Y$. Takie podejście jest zgodne z definicją zawierania zbiorów rozmytych podaną przez Zadeha

$$A \subset B\,, \qquad \text{jeśli } \forall_{u \in U} \mu_A(u) \leq \mu_B(u)\,. \tag{3.1.2}$$

Jednakże, wartość współczynnika jest równa 0 tylko wtedy gdy zbiory X i Y są rozłączne, co nie jest zgodne z (3.1.2). Takie zachowanie jest jednak oczekiwane, gdyż miara zawierania powinna móc przyjmować dowolne wartości pośrednie z przedziału $[0, 1]$.

Zastosowanie innych niż minimum t-norm i dowolnych ewaluatorów, może spowodować nawet większe odstępstwa od definicji zawierania podanej przez Zadeha. Wartości brzegowe 0 i 1 mogą być osiągane dla innych zbiorów rozmytych jak to ma miejsce w przypadku klasycznych zbiorów. Przykładowo dla wielu t-norm $A \cap B = \emptyset$ nie oznacza, że zbiory rozmyte A i B są rozłączne.

Dubois oraz Prade zaproponowali następujący układ aksjomatów jaki powinien być spełniony przez miarę zawierania [12]:

1. $Sub(A, B) = 1$ wtedy i tylko wtedy gdy $\overline{A} \cup B = U$,
2. jeśli zbiory A i B są rozłączne to $Sub(A, B) = 0$,
3. $Sub(A, B)$ zależy od wartości $g(\overline{A} \cap B)$.

Założenia te doprowadziły do sformułowania ogólnego wzoru dla miary zawierania zbiorów rozmytych

$$Sub_{g,s}(A, B) = \frac{g(\overline{A} \cup_s B) - g(\overline{A})}{1 - g(\overline{A})}\,. \tag{3.1.3}$$

Dwa pierwsze warunki wykluczają się wzajemnie w przypadku gdy $A = \emptyset$. Z jednej strony $Sub(\emptyset, B) = 1$, gdyż $\overline{\emptyset} = U$. Z kolei $\emptyset$ oraz B są rozłączne, a zatem z drugiego warunku $Sub(\emptyset, B) = 0$. Dla zachowania zgodności z klasyczną teorią mnogości przyjmujemy $Sub(\emptyset, B) = 1$ oraz $Sub(A, \emptyset) = 0$ za wyjątkiem przypadku, gdy $A = \emptyset$. Nie wszystkie ewaluatory skalarne zbiorów rozmytych oraz operacje sumy generowane przez t-konormy podstawione do (3.1.3) gwarentują spełnienie tych aksjomatów.

Jeśli operacja sumy zbiorów będzie generowana przez t-konormę Łukasiewicza $s_{\text{Łuk}}$, a jako ewaluator wykorzystamy g_{id}, to z (3.1.3) otrzymamy miarę zawierania zaproponowaną przez Kosko [33]

$$\begin{aligned}
Sub_{g_{\text{id}},s_{\text{Łuk}}}(A, B) &= \frac{g(\overline{A} \cup B) - g(\overline{A})}{1 - g(\overline{A})} = \frac{\frac{\sum_{u\in U} s(1-\mu_A(u),\mu_B(u)) - \sum_{u\in U}(1-\mu_A(u))}{\sum_{u\in U} f(1)}}{1 - \frac{\sum_{u\in U} 1-\mu_A(u)}{\sum_{u\in U} f(1)}} = \\
&= \frac{\sum_{u\in U} \min\left(1 - \mu_A(u) + \mu_B(u), 1\right) - \left(1 - \mu_A\left(u\right)\right)}{\sum_{u\in U} f(\mu_A(u))} = \\
&= \frac{\sum_{u\in U} \min\left(\mu_A(u), \mu_B(u)\right)}{\sum_{u\in U} f(\mu_A(u))} = \frac{\sigma_f(A \cap_{t_{\min}} B)}{\sigma_f(A)} = Sub_{g_{\text{id}},\cap_{\min}}(A, B)\,.
\end{aligned}$$

Miara zawierania Kosko może zostać uogólniona na inne operacje przekroju generowane przez dowolne t-normy, otrzymując w ten sposób bardziej restrykcyjne miary zawierania takie jak $Sub_{g_{\text{id}},\cap_{\text{prod}}}(A, B)$ czy $Sub_{g_{\text{id}},\cap_{\text{Łuk}}}(A, B)$.

Ewaluatory g_{inf} oraz g_{sup} również mogą posłużyć do konstrukcji miar zawierania. Przykładowo $Sub_{g_{\text{sup}},s_{\text{Łuk}}}(A, B)$ spełnia wszystkie zaproponowane aksjomaty, podobnie jak $Sub_{g_{\text{inf}},s_{F,\lambda}}(A, B)$.

Aksjomat numer 2 wymaga, aby $Sub(A, B) = 0$, jeśli A i B są rozłączne. Zaostrzenie tego warunku do postaci $Sub(A, B) = 0$ wtedy i tylko wtedy, gdy A i B są rozłączne wykluczyłoby pewne miary zawierania. $Sub_{g_{\text{id}},s_{min}}(A, B)$ może być równe 0 nawet jeśli oba zbiory nie są rozłączne. Taka własność jednak może być pożądana w pewnych zastosowaniach, gdzie konieczne jest odrzucenie elementów, dla których

$\mu_A(u) + \mu_B(u) \leq 1$. Zastąpienie warunku rozłączności, równoważnym w klasycznej teorii mnogości warunkiem $A \cap B = \emptyset$ również nie rozwiązuje problemu. Przekrój zbiorów generowanych przez t-normę Łukasiewicza, dopuszcza sytuację w której dla dwóch nierozłącznych zbiorów zachodzi $A \cap B = \emptyset$. Takie podejście wykluczyłoby też wszystkie miary oparte na ewaluatorze g_{inf}.

Niektóre zależności pomiędzy t-konormami przenoszą się na miary zawierania. Tak jest w przypadku ich uporządkowania

$$Sub_{g_{\text{id}},s_{\min}}(A,B) \leq Sub_{g_{\text{id}},s_{\text{prod}}}(A,B) \leq Sub_{g_{\text{id}},s_{\text{Łuk}}}(A,B)$$

oraz

$$Sub_{g_{\text{inf}},s_{\min}}(A,B) \leq Sub_{g_{\text{inf}},s_{\text{prod}}}(A,B) \leq Sub_{g_{\text{inf}},s_{\text{Łuk}}}(A,B)\,.$$

Jednak zależność pomiędzy $Sub_{g_{\text{id}},s}(A,B)$ i $Sub_{g_{\text{inf}},s}(A,B)$ nie jest jednoznacznie określona. Pomimo, że wartości funkcji g są uporządkowane, to z powodu normalizacji miary zawierania już nie są. To, która z tych wartości jest większa zależy od doboru zbiorów rozmytych.

Relacja zawierania jest zwrotna. Niestety, własność ta nie jest spełniona dla większości miar zawierania określonych za pomocą (3.1.3). Wynika to z faktu, że większość popularnych rodzin t-norm jest Archimedesowa co oznacza, że $t(a,a) < a$, a co za tym idzie $Sub(A,A) < 1$. Przykładem zwrotnej miary zawierania jest $Sub_{g_{\text{id}},s_{\text{Łuk}}}$.

Kolejną własnością miar zawierania jest przechodniość. Jeśli spełnione jest $A_1 \subseteq A_2 \subseteq A_3$ to można by było oczekiwać, że $Sub(A_1,A_3) \geq Sub(A_2,A_3)$. W przypadku $Sub_{g_{\text{inf}},s_{F,\lambda}}(A,B)$ rzeczywiście ma to miejsce, jednak już dla $Sub_{g_{\text{id}},s}(A,B)$ nie.

3.1.2. Miary podobieństwa

W klasycznej teorii mnogości do określenia stopnia podobieństwa wykorzystuje się moc dopełnienia różnicy symetrycznej obu zbiorów. Przy takim podejściu zbiory uznaje się za całkowicie podobne jeśli ich różnica symetryczna jest pusta, a gdy różnica symetryczna równa jest sumie tych zbiorów, uznaje się je za całkowicie różne.

Oryginalna definicja równości zbiorów podana przez Zadeha mówi, że zbiory A i B są równe wtedy i tylko wtedy gdy

$$\forall_{u \in U} \mu_A(u) = \mu_B(u)\,.$$

Oczywiście miara podobieństwa musi być mniej restrykcyjna, aby podobieństwo mogło przyjmować dowolne wartości z przedziału $[0,1]$.

Pojęcie różnicy symetrycznej pełni bardzo ważną rolę w określaniu stopnia podobieństwa. W klasycznej teorii mnogości może ono zostać zdefiniowane na następujące

dwa sposoby

$$X \oplus Y = (X \cup Y) \cap (\overline{X} \cup \overline{Y}) = (\overline{X} \cup Y) \cap (X \cup \overline{Y}) \,.$$

Przeniesienie tego pojęcia do teorii zbiorów rozmytych okazuje się jednak trudne, właśnie ze względu na te dwie równoważne definicje. Okazuje się bowiem, że w przypadku rozmytym ich równoważność nie zawsze zachodzi. Dlatego też wprowadzono dwie różne definicje różnicy symetrycznej

$$A \oplus^{+} B = (A \cup B) \cap (\overline{A} \cup \overline{B})$$

oraz

$$A \oplus^{-} B = (\overline{A} \cup B) \cap (A \cup \overline{B}) \,.$$

Oczywiście dla klasycznych zbiorów $X \oplus^{+} Y = X \oplus^{-} Y$. Ponadto, gdy sumy i przekroje zbiorów generowane są przez t-nomę minimum oraz odpowiadającą jej t-konormę maksimum własność ta również jest spełniona.

Dubois oraz Prade [12] zaproponowali cztery aksjomaty, które powinny być spełnione przez miarę podobieństwa skonstruowaną w oparciu o różnicę symetryczną zbiorów

1. $Sim(A, B) = 1$ wtedy i tylko wtedy, gdy $A \oplus B = \emptyset$,
2. jeśli A i B są rozłącznymi zbiorami rozmytymi to $Sim(A, B) = 0$,
3. $Sim(A, B) = Sim(B, A)$,
4. $Sim(A, B)$ jest zależne od $g(A \oplus B)$ lub od wartości symetrycznej funkcji o argumentach $g(A \cap \overline{B})$ i $g(\overline{A} \cap B)$.

Warunki te uogólniają własności znane z klasycznej teorii mnogości. Pierwszy warunek nie jest równoważny wymogowi, aby $Sim(A, B) = 1$ wtedy i tylko wtedy, gdy $A = B$, gdyż nie wszystkie rozmyte różnice symetryczne dają zbiór pusty, gdy wejściowe zbiory są sobie równe. Przykładowo zastosowanie $\oplus^{+}$ z przekrojem i sumą generowanymi przez t-normę i t-konormę Łukasiewicza, daje w rezultacie zbiór pusty tylko gdy $A = B$ oraz zbiory te redukują się do klasycznych zbiorów. Warunek drugi wymaga aby $Sim(A, B) = 0$, gdy zbiory A i B są rozłączne.

W literaturze występują trzy typy miar podobieństwa oparte o te aksjomaty

1. miary uzyskane poprzez zastosowanie skalarnego ewaluatora g na zbiorze $A \oplus B$,
2. miary uzyskane poprzez zastosowanie symetrycznej dwuargumentowej funkcji r na wartościach $g(A \cup \overline{B})$ i $g(\overline{A} \cup B)$,

3. miary uzyskane poprzez zastosowanie symetrycznej dwuargumentowej funkcji h na wartościach $Sub(A, B)$ i $Sub(B, A)$.

W dalszej części pracy stosowane będzie oznaczenie Sim^1, Sim^2 oraz Sim^3 na miary pierwszego, drugiego i trzeciego typu.

Pierwszy typ miar można ogólnie zapisać w następujący sposób

$$Sim^1_{\oplus,\cup,g}(A, B) = \frac{g(\overline{A \oplus B}) - g(\overline{A \cup B})}{1 - g(\overline{A \cup B})} \tag{3.1.4}$$

co po przekształceniach równoważne jest

$$Sim^1_{\oplus,\cup,g}(A, B) = 1 - \frac{g(A \oplus B)}{g(A \cup B)}\,. \tag{3.1.5}$$

W niektórych pracach stwierdzano, że suma zbiorów rozmytych w powyższym wzorze oraz ta wykorzystana w konstrukcji różnicy symetrycznej mogą być wybrana niezależnie. Takie postępowanie może jednak prowadzić do otrzymania ujemnej wartości podobieństwa.

Jedną z najczęściej stosowanych miar podobieństwa jest indeks Jaccarda zdefiniowany w (1.2.2). Miara $Sim^1_{\oplus^-_{\text{Łuk}},\cup_{\max},g_{\text{id}}}$ rozszerza go do zbiorów rozmytych [37]. W tym przypadku (3.1.4) można przekształcić do postaci

$$Sim^1_{\oplus,\cup,g} = 1 - \frac{g(A) + g(B) - 2 \cdot g(A \cap B)}{g(A \cup B)}\,,$$

a następnie korzystając z własności t-normy i t-konormy Łukasiewicza $g_{\text{id}}(A) + g_{\text{id}}(B) = g_{\text{id}}(A \cap B) + g_{\text{id}}(A \cup B)$ otrzymujemy

$$Sim^1_{\oplus,\cup,g} = \frac{g(A \cap B)}{g(A \cup B)}\,.$$

W tej postaci od razu widać podobieństwo do indeksu Jaccarda. Pozwala to na określenie nowej rodziny miar podobieństwa $Sim^1_{\cap,\cup,g}$.

Jeśli jako ewaluator skalarny wykorzystamy g_{inf} to dla dowolnej różnicy symetrycznej otrzymamy następującą miarę podobieństwa

$$Sim^1_{\oplus,\cup,g_{\text{inf}}}(A, B) = \inf_{u \in U}(1 - \mu_{A \oplus B}(u)) = 1 - sup_{u \in U}\mu_{A \oplus B}(u)\,.$$

Drugi typ miar można ogólnie zapisać jako

$$Sim^2_{\cup,r,g} = \frac{r\left(g(A \cup \overline{B}), g(\overline{A} \cup B)\right) - r\left(g(\overline{A}), g(\overline{B})\right)}{1 - r\left(g(\overline{A}), g(\overline{B})\right)}\,, \tag{3.1.6}$$

gdzie r jest dowolną symetryczną funkcją spełniającą $r(0, 0) = 0$ oraz taką, że $r(a, b) = 1$ wtedy i tylko wtedy, gdy $a = b = 1$. Własności te spełnone są przez dowolną t-normę oraz dowolny operator sumy symetrycznej[48] w tym przez średnią arytmetyczną $r_M(a, b) = \frac{a+b}{2}$.

Biorąc $g_{\inf}$ jako ewaluator skalarny otrzymujemy następującą miarę podobieństwa

$$Sim^2_{\cup,r,g_{\inf}} = r(g_{\inf}(A \cup \overline{B}), g_{\inf}(\overline{A} \cup B)) \,,$$

który może być interpretowany jako symetryczna agregacja wartości reprezentujących konieczność należenia do B wiedząc, że element należy do A, oraz vice versa.

Miary trzeciego typu powstają przez symetryczną agregację miary zawierania

$$Sim^3_{Inc,r}(A, B) = r(Inc(A, B), Inc(B, A)) \,, \tag{3.1.7}$$

gdzie r podobnie jak poprzednio jest dowolną symetryczną funkcją spełniającą $r(0, 0) = 0$ oraz taką, że $r(a, b) = 1$ wtedy i tylko wtedy, gdy $a = b = 1$. Miary tego typu agregują stopień zawierania A w B oraz B w A. Takie podejście zostało zaproponowane w [37] i doprowadziło to powstania wprowadzonego wcześniej indeksu Jaccarda.

3.1.2.1. Własności

Nie wszyskie miary podobieństwa są zwrotne. Powodem tego, podobnie jak w przypadku miar zawierania, jest Archimedesowość t-norm. Przykładem zwrotnej miary podobieństwa jest $Sim^1_{\cap_{\min},\cup_{\max},g_{\mathrm{id}}}$. Jednak dla innych popularnych t-norm i t-konorm zachodzi

$$Sim^1_{\cap,\cup,g}(A, A) = \frac{A \cap A}{A \cup A} < 1 \,.$$

Podobnie $Sim^1_{\oplus,\cup,g_{\inf}}$ nie jest zwrotna za wyjątkiem przypadku, gdy $\oplus = \oplus^-_{\text{Łuk}}$.

Podobnie jak miary zawierania, niektóre rodziny miar podobieństwa również są uporządkowane. W przypadku miar typu pierwszego, uporządkowanie wynika bezpośrednio z relacji pomiędzy t-normami i t-konormami. Analogicznie do miar zawierania nie jest możliwe porówanie wartości miar uzyskanych z użyciem różnych ewaluatorów skalarnych. Dla miar typu drugiego zachodzi

$$Sim^2_{\cup,r,g_{\inf}}(A, B) \leq Sim^2_{\cup,r,g_{\mathrm{id}}} \,.$$

Dla miar typu trzeciego przy ustalonej funkcji r, uporządkowanie wynika z relacji pomiędzy miarami zawierania. Przy ustalonych miarach zawierania relacje pomiędzy miarami wynikają natomiast z relacji pomiędzy użytymi funkcjami r. Dla najpopularniejszych funkcji r uporządkowanie jest następujące

$$Sim^3_{Inc,t_{\text{Łuk}}} \leq Sim^3_{Inc,t_{\text{prod}}} \leq Sim^3_{Inc,t_{\min}} Sim^3_{Inc,r_M}$$

Zachodzą również pewne zależności pomiędzy reprezentantami poszczególnych typów miar podobieństwa. Przykładowo dla miar $Sim^1_{\cap,\cup,g_{\mathrm{id}}}$ oraz $Sim^3_{Inc_{\cup},g_{\mathrm{id}}}$ zachodzi

$$Sim^1_{\cap,\cup,g_{\mathrm{id}}}(A, B) \leq Sim^3_{Inc_{\cup},g_{\mathrm{id}}}(A, B) \,.$$

Proste przekształcenia pozwalają zauważyć, że rzeczywiście tak jest

$$\begin{aligned} Sim^1_{\cap,\cup,g_{\mathrm{id}}}(A,B) &= \\ &= \frac{g(A\cap B)}{g(A\cup B)} \leq \frac{g(A\cap B)}{\max\left(g(A),g(B)\right)} = \\ &= \min\left(Inc_{\cup,g_{\mathrm{id}}}(A,B), Inc_{\cup,g_{\mathrm{id}}}(B,A)\right) = Sim^3_{Inc_{\cup,g_{\mathrm{id}}}}(A,B)\,. \end{aligned}$$

Niektóre miary różnych typów są tożsame. Tak jest w przypadku $Sim^2_{\cup,t_{\min},g_{\mathrm{id}}}$ i $Sim^3_{Inc_{\cup,g_{\mathrm{id}}},t_{\min}}$ jak również w przypadku $Sim^2_{\cup,t_{\min},g_{\inf}}$ i $Sim^3_{Inc_{\cup,g_{\inf},t_{\min}}}$, dla których

$$Sim^2_{\cup,t_{\min},g_{\inf}}(A,B) = Sim^3_{Inc_{\cup,g_{\inf},t_{\min}}}(A,B) = \min\left(g_{\inf}(A\cup\overline{B}), g_{\inf}(\overline{A}\cup B)\right)\,.$$

3.2. Podejście oparte na bliskości

Tversky zauważył, że w teoretycznych jak i doświadczalnych badaniach podobieństwa bardzo często przyjmuje się, że obiekty są dobrze reprezentowane przez punkty w pewnym układzie współrzędnych [55]. Przy takim założeniu relacja niepodobieństwa zachowuje się dokładnie jak funkcja odległości. Pomimo, że oryginalne badania dotyczyły klasycznych zbiorów pojęcie bliskości obiektów może zostać również wykorzystane do określenia podobieństwa zbiorów rozmytych. W przypadku zbiorów rozmytych odległość zamiast pomiędzy punktami w przestrzeni, musi zostać zdefiniowane pomiędzy funkcjami przynależności.

Pierwsze podejście uogólnia pojęcie metryki na zbiory rozmyte. Dzięki zasadzie rozszerzania Zadeha określono rozmytą odległość pomiędzy zbiorami rozmytymi. Podstawowymi reprezentami tej rodziny miar są ogólnione metryki Minkowskiego, Hausdorffa oraz współczynnik różności. Inne podejście zakłada powiązanie zbioru rozmytego z punktem w n-wymiarowej przestrzeni. Każdy wymiar tej przestrzeni odpowiada pewnej określonej własności zbioru. Przykladowo, moc oraz środek ciężkości zbioru rozmytego są bardzo często wykorzystywane jako jedne z wymiarów.

Nie wszystkie formuły wykorzystywane w do konstrukcji miar podobieństwa spełniają formalne warunki bycia odległością (metryką), stąd aby uniknąć nieporozumień używany jest termin bliskość obiektów. Jednakże, większość miar pochodzi od pewnych metryk. Jako metrykę lub funkcję odległości w uniwersum U rozumiemy dwuargumentową funkcję $d : U \times U \to [0,\infty)$ spełniającą:

1. $d(u,v) = d(y,u)$ (symetryczność),
2. $d(u,w) \leq d(u,v) + d(v,w)$ (nierównośc trójkąta),
3. jeśli $u \neq v$ to $d(u,v) > 0$,
4. jeśli $u = v$ to $d(u,v) = 1$,

dla dowolnych punktów $u, v, w \in U$. Funkcja nazywana jest półmetryką lub pseudometryką jeśli trzeci warunek nie jest spełniony.

Odległość obiektów jest z definicji symetryczna, w przeciwieństwie do miar zawierania. Stąd metody oparte na bliskości obiektów pozwalają na konstrukcje tylko miar podobieństwa. Im większa odleglość tym mniejsze jest podobieństwo badanych obiektów, skąd wynika potrzeba normalizacji. W tym celu wykorzystuje się funkcję normalizującą wartość odległości do przedzialu $[0, 1]$.

3.2.1. Miary podobieństwa

3.2.1.1. Uogólnione metryki

Interpretując zbiór rozmyty A nad skończonym uniwersum $U = \{u_1, u_2, \cdots, u_n\}$ jako punkt $(\mu_A(u_1), \mu_A(u_2), \cdots, \mu_A(u_n))$ w n-wymiarowej przestrzeni, do wyznaczania odległości pomiędzy zbiorami można wykorzystać dowolną r-metrykę Minkowskiego

$$d_r(A, B) = \left(\sum_{i=1}^{n} |\mu_A(u_i) - \mu_B(u_i)|^r \right)^{\frac{1}{r}},$$

gdzie $r \geq 1$. Dla $r = 1$, d_r staje się metryką miejską, gdy $r = 2$ otrzymujemy metrykę Euklidesową. Dla $r = \infty$ d_r przyjmuje następującą postać

$$d_\infty(A, B) = \sup_{1 \leq i \leq n} |\mu_A(u_i) - \mu_B(u_i)|$$

zwaną metryką nieskończoności.

Jak wcześniej zauważono odległość obiektów przyjmuje wartości odwrotne do miar podobieństwa. Stąd też aby zdefiniować miarę podobieństwa należy wcześniej znormalizować ją do przedzialu $[0, 1]$, a nastepnie odjąć od 1

$$Sim_{d_r,|U|} = 1 - \frac{d_r(A, B)}{|U|}.$$

Jak łatwo można zauważyć niektóre miary podobieństwa otrzymane w ten sposób są równoważne miarom opartym na podejściu teoriomnogościowym. Przykładowo

$$\begin{aligned} Sim_{d_\infty,-}(A, B) &= 1 - d_\infty(A, B) = 1 - \sup_{1 \leq i \leq n} |\mu_A(u_i) - \mu_B(u_i)| \\ &= \inf_{1 \leq i \leq n} (1 - |\mu_A(u_i) - \mu_B(u_i)|) = Sim^1_{\oplus^-_{\text{Łuk}}, s_{\text{Łuk}}, g_{\inf}}(A, B). \end{aligned}$$

Ta miara podobieństwa nie wymaga stosowania normalizacji odległości, co zostało zaznaczone przez $-$ w indeksie. Kolejnym przykładem jest $Sim_{d_1,|A \cup_{\text{Łuk}} B|}$

$$\begin{aligned} Sim_{d_1,|A \cup_{\text{Łuk}} B|} &= 1 - \frac{d_1(A, B)}{|A \cup_{\text{Łuk}} B|} = 1 - \frac{\left(\sum_{u \in U} |\mu_A(u_i) - \mu_B(u_i)|\right)}{|A \cup_{\text{Łuk}} B|} \\ &= Sim^1_{\oplus_{\text{Łuk}}, s_{\text{Łuk}}, g_{f_{\text{id}}}}(A, B). \end{aligned}$$

Do normalizacji odległości można wykorzystać różne wartości. Najczęściej używana jest moc całego uniwersum $|U|$ oraz moc sumy obu zbiorów $|A \cup B|$. Podstawowa różnica polega na uwzględnianiu dopasowań na brakujących cechach. W przypadku mocy uniwersum na ocenę podobieństwa wpływają zarówno elementy należace do obu zbiorów jak i te nienależące. Natomiast, gdy do normalizacji wykorzystana zostanie moc sumy zbiorów elementy niewystępujące w obu zbiorach nie mają wpływu na wartość podobieństwa.

Inną metodą na otrzymanie miary podobieństwa z odległości pomiędzy dwoma zbiorami rozmytymi jest zastosowanie nastepującej transformacji [28, 43]

$$Sim_{d_r,\text{trans}}(A,B) = \left(1 + \left(\frac{d_r(A,B)}{s}\right)^t\right)^{-1}.$$

Najczęsciej przyjmuje się $s = t = 1$.

Metryki Minkowskiego nie są jedynymi metrykami, które można rozszerzyć na zbiory rozmyte. Funkcja zdefiniowana w następujący sposób jest metryką na przestrzeni zbiorów rozmytych

$$d_{\oplus^-}(A,B) = |A \oplus^- B|.$$

Metryki te można wykorzystać w analogiczny sposób jak metryki Minkowskiego do konstrukcji miar podobieństwa zbiorów rozmytych.

Zachodzą natępujące zależności pomiędzy powyżej zdefiniowanymi miarami podobieństwa

$$Sim_{d_1,\text{trans}}(A,B) \leq \cdots \leq Sim_{d_r,\text{trans}}(A,B) \leq \cdots \leq Sim_{d_\infty,\text{trans}}(A,B)$$

oraz

$$Sim_{d_1,|U|}(A,B) \leq \cdots \leq Sim_{d_r,|U|}(A,B) \leq \cdots \leq Sim_{d_\infty,|U|}(A,B)$$

Ze względu na normalizację nie jest możliwe natomiast uporządkowanie miar $Sim_{d_r,|A\cup B|}$.

3.2.1.2. Odległość kątowa

Uniwersum U może reprezentować zbiór cech posiadanych przez pewien obiekt. Przy takiej interpretacji możliwe jest zastosowanie podejścia znanego z taksonomi – miarę podobieństwa można określić na podstawie kąta pomiędzy dwoma wektorami

$$Sim_{\cos\theta}(A,B) = \frac{\sum_{u\in U} \mu_A(u) \cdot \mu_B(u)}{\sqrt{\sum_{u\in U} \mu_A(u)^2} \cdot \sqrt{\sum_{u\in U} \mu_B(u)^2}}.$$

$Sim_{\cos\theta}(A,B)$ przyjmuje największą wartość, gdy kąt pomiędzy wektorami reprezentującymi zbiory rozmyte wynosi 90°. Ta miara podobieństwa wykorzystywana jest w przypadku, gdy szczególnie istotne jest porównanie kształtu zbiorów. $Sim_{\cos\theta}(A,B)$

jest niezmienne ze względu na przemnażanie przez stałą [7]. Przykładowo dla zbiorów rozmtych A i B postaci $\mu_B(u) = \alpha \cdot \mu_A(u)$ miara podobieństwa jest równa 1. Podobnie tak zdefiniowane podobieństwo nie jest ważliwe na przesunięcia wektorów. Zbiory A i B postaci $\mu_B(u) = \mu_A(u) + \beta$ są całkowicie podobne zgodnie z tą miarą.

Cosinus kąta nie jest jedyną miarą kątową. W literaturze można znaleźć wiele modyfikacji tego wzroru takich jak [27]

$$Sim_B(A,B) = \sqrt{1 - \frac{\sum_{u \in U} \sqrt{\mu_A(u) \cdot \mu_B(u)}}{\sqrt{\sum_{u \in U} \mu_A(u)^2} \cdot \sqrt{\sum_{u \in U} \mu_B(u)^2}}}\,.$$

3.2.1.3. Odległość przedziałowa

Wiele obiektów może być reprezentowanych przez zbiory rozmyte będące liczbami lub przedziałami rozmytymi. W takim przypadku możliwe jest wykorzystanie metod stworzonych do porównywania przedziałów rzeczywistych. Spośród najpopularniejszych sposobów mierzenia bliskości przedziałów rzeczywistych należy wymienić metrykę Hausdorffa oraz współczynnik różności.

Metryka Hausdorffa mierzy odległość pomiędzy dwoma zwartymi niepustymi podzbiorami przestrzeni rzeczywistej X i Y

$$q(X,Y) = \max\left(\sigma(X,Y), \sigma(Y,X)\right)$$

gdzie

$$\sigma(X,Y) = \sup_{x \in X} \inf_{y \in Y} d_2(x,y)\,.$$

Wartość $\sigma(X,Y)$ określana jest poprzez punkt z X, który maksymalizuje odległość od zbioru Y. W przypadku gdy zbiory X i Y są przedziałami metryka Hausdorffa przyjmuje następującą postać

$$q([x_1,y_1],[x_2,y_2]) = \max(|x_1 - x_2|, |y_1 - y_2|)\,.$$

Oczywiście metrykę Euklidesową można zastąpić dowolną metryką. Zaproponowano wiele uogólnień metryki Hausdorffa do przedziałów oraz liczb rozmytych. Niektóre z nich wykorzystują tak zwane α-przekroje [47, 62], inne korzystają z zasady rozszerzania Zadeha [46].

Zastosowanie zasady rozszerzania do σ pozwala zdefiniować następujący zbiór rozmyty

$$\mu_{\sigma(A,B)}(r) = \sup_{r=\sigma(A,B)} \alpha$$

Rozmyta metryka Hausdorffa obliczna jest w następujący sposób

$$H(A,B) = \max(\sigma(A,B), \sigma(B,A))\,.$$

Wartość ta jest nadal liczbą rozmytą. Zaproponowana metoda defuzyfikacji [46] wykorzystuje pojęcie rozmytego progu odległości L. Podobieństwo zbiorów A i B określane jest na podstawie stopnia spełnienia

$$H(A, B) \subset L\,.$$

Kolejne podejście wykorzystuje pojęcie α-przekroju [47, 62]. Przekroje liczb oraz przedziałów rozmytych są przedziałami $A_\alpha = [\underline{A}_\alpha, \overline{A}_\alpha]$. Dzięki temu możliwe jest zastosowanie metryki Hausdorffa do każdego przekroju oddzielnie. Tak otrzymany ciąg liczb należy następnie zagregować. Zaproponowano kilka metod agregacji, jedną z najpopularniejszych jest całkowanie względem wartości α

$$q_{H,\int}(A, B) = \int_0^1 q(A_\alpha, B_\alpha)\mathrm{d}\alpha\,.$$

Inna metoda wybiera tylko największą wartość odległości Hausdorffa

$$q_{H,\infty}(A, B) = \sup_{\alpha \geq 0} q(A_\alpha, B_\alpha)\,.$$

W [62] wykorzystano metrykę Hausdorffa do porównywania samych rdzeni zbiorów rozmytych

$$q_{H^*}(A, B) = q(A_{\alpha=1}, B_{\alpha=1})\,.$$

Wszystkie te odległości nie są znormalizowane stąd potrzeba zastosowania transformacji analogicznych jak w przypadku miar podobieństwa opartych o metryki Minkowskiego. Przykładowo

$$Sim_{H,\int}(A, B) = \left[1 + \left(\frac{q_{H,\int}(A, B)}{s}\right)^t\right]^{-1}$$

Współczynnik różności określa odległość pomiędzy dwoma przedziałami $[a, b]$ i $[c, d]$ w następujący sposób

$$D([a, b], [c, d]) = \frac{|a - c| + |b - d|}{2 \cdot (\beta_2 - \beta_1)}\,,$$

gdzie oba przedziały zawarte są w przedziale $[\beta_1, \beta_2]$. Dzięki podzieleniu przez dwukrotność długości tego przediału otrzymany współczynnik jest znormalizowany do przedzialu $[0, 1]$. Dobór odpowiedniego przedziału normalizującego zależy od charakterystyki problemu oraz porządanych własności podobieństwa. Sam przedział nie powinien jednak zależeć od porównywanych przedziałów, gdyż prowadzi to do pewnych anomalii. Przykładowo wybierając najmniejszy przedział zawierający oba porównywane zbiory, współczynnik różności nie rozróżnia przedziałów jednopunktowych. Niech $a < b$ wtedy $D([a, a], [b, b]) = \frac{|a-b|+|a-b|}{2\cdot(b-a)} = 1$. Problemu tego można uniknąć wykorzystując całe uniwersum jako przedział normalizujący.

Rozszerzenie współczynnika różności na liczby rozmyte zostało oparte na α-przekroju [29]

$$D_{\int}(A,B) = \int_0^1 D(A_\alpha, B_\alpha)\mathrm{d}\alpha\,,$$

gdzie przedział normalizujący $[\beta_1, \beta_2]$ musi spełniać $supp(A) \cup supp(B) \subset [\beta_1, \beta_2]$. W odróżnieniu od metryki Hausdorffa, współczynnik różności przyjmuje wartości w przedziale $[0, 1]$, jednak ich znacznie jest odwrócone. Miarę podobieństwa otrzymuje się ze współczynnika różności w następujacy sposób

$$Sim_D(A,B) = 1 - D(A,B)\,.$$

Analogicznie jak w przypadku Sim_H możliwe jest określenie miar Sim_{D^*} oraz Sim_{D_∞}.

3.2.1.4. Wektory parametrów

Podejście to zostało opracowane na potrzeby zagadnienia aproksymacji lingwistycznej. Mając dany zbiór etykiet reprezentowanych przez zmienne lingwistyczne należy znaleźć tą, która najlepiej oddaje znaczenie nieznanego, nowego obiektu. Ze względów wydajnościowych proces ten często rozdzielany jest na dwie fazy. W pierwszej odrzucane są wszystkie etykiety, których reprezentacja jest znacząco różna od badanego obiektu. W drugiej najbliższe etykiety porównywane są przy pomocy wprowadzonych wcześniej metryk opertych na odległości.

Porównywanie zbiorów w pierwszej fazie odbywa się w oparciu o dowolną miarę odległości. Od poprzednio wprowadzonych metod odróżnia ją natomiast konstrukcja punktów, pomiędzy którymi odległość jest liczona. We wszystkich poprzednich podejściach odległość była liczona pomiędzy wektorami składającymi się ze stopni przynależności. Przy takim podejściu wektory, a co za tym idzie wymiar przestrzenie może stać się bardzo duży. Podejście oparte o wektory parametów, zakłada, że reprezentacja zbioru rozmytego w przestrzeni jest znacznie prostsza – co najwyżej kilkuwymiarowa. Składowe wektora stanowią obliczane na podstawie funkcji przynależności parametry. Najczęściej wykorzystuje się moc zbioru rozmytego, entropię, środek ciężkości

$$COG(A) = \frac{\sum_{u\in U} u \cdot \mu_A(u)}{|A|}$$

oraz skośność

$$skew(A) = \frac{\sum_{u\in U} \mu_A(u) \cdot (u - COG(A))^3}{|A|}\,.$$

Przykładowo [57] proponuje aby zbiory rozmyte reprezentować w przestrzeni dwuwymiarowej za pomocą wektorów postaci $[|A|, COG(A)]$, a do oceny ich podobieństwa wykorzystać standardową metrykę Euklidesową. W ten sposób miara podobieństwa przyjmuje następującą postać

$$Sim_{||,COG}(A,B) = 1 - \frac{\sqrt{(|A| - |B|)^2 + (COG(A) - COG(B))^2}}{|U|}\,.$$

3.3. Podejście oparte na logice

Miary oparte na logice [21, 22] wykorzystują interpretację funkcji przynależności zbioru rozmytego jako stopnień prawdziwości wniosku reprezentowanego przez ten zbiór. Podstawowa metoda zakłada użycie operatora implikacji co umożliwia konstrukcję zarówno miar zawierania jak i podobieństwa.

3.3.1. Miary zawierania

Operatory implikacji mogą zostać wykorzystane do konstrukcji miar zawierania. W tym celu przyjmuje się, że stopień zawierania $A \subset B$ jest równy stopniowi w jakim A implikuje B. W klasycznej logice operator implikacji może zostać zdefiniowany na kilka równoważnych sposobów. Uogólnienie do logiki rozmytej, gdzie dopuszczane jest nieskończenie wiele stopni prawdziwości, spowodowało powstanie wielu nierównoważnych definicji tego pojęcia. Najczęściej wykorzystywnymi operatorami implikacji są S-implikacje i R-implikacje [3, 54]. Pierwszą rodzinę implikacji otrzymuje się poprzez bezpośrednie uogólnienie implikacji znanej z logiki

$$a \Rightarrow b \equiv \overline{a} \vee b\,.$$

S-implikacja otrzymana jest poprzez zastąpienie operatora alternatywy poprzez t-konormę

$$a \Rightarrow b = S(1-a, b)\,.$$

Nazwa S-implikacji pochodzi od użycia t-konormy s w jej definicji. Często S-implikacje nazywa się również silnymi implikacjami.

Wiele znanych implikacji rozmytych okazało się być S-implikacjami. Operator implikacji Łukasiewicza powstaje poprzez zastosowanie t-konormy Łukasiewicza

$$a \Rightarrow_{\text{Łuk}} b = \min(1, 1-a+b)\,.$$

Implikacja Kleene–Dienesa powstaje poprzez zastosowanie t-konormy maksimum

$$a \Rightarrow_{\max} b = \max(1-a, b)\,.$$

Rodzina R-implikacji powstaje poprzez uogólnienie implikacji znanej z logiki zdań

$$a \Rightarrow b \equiv \sup\{x \in [0,1] : a \wedge x \leq b\}\,.$$

W nazwie R-implikacji, R pochodzi od nazwy półgrup (ang. *residuated*), które powstają poprzez zastąpienie koniunkcji t-normą

$$a \Rightarrow b = \sup\{x \in [0,1] : t(a,x) \leq b\}\,.$$

Wiele implikacji znanych z logiki nieskończenie wielowartościowej to R-implikacje. Przykładowo zastosowanie t-normy produktowej pozwala uzyskać implikację Goguen'a [17, 19]

$$a \Rightarrow_G b = \begin{cases} \min(1, \frac{b}{a}), & \text{jeśli } a \neq 0 \\ 1 & \text{w przeciwym przypadku} \end{cases}$$

Implikacja Łukasiewicza może również zostać uzyskana jako R-implikacja przy użyciu t-normy Łukasiewicza.

Dzięki zastosowaniu t-norm i t-konorm możliwe jest określenie nieskończenie wielu operatorów implikacji. Ponad 40 różnych zostało zaproponowanych i zbadanych w literaturze [31, 36]. Istnieją również badania nad kryteriami doboru i stosowalnością poszczególnych implikacji w zastosowaniach [4–6, 13, 30, 32, 38, 40, 58].

W celu porównania dwóch zbiorów rozmytych zdefiniujemy współczynnik zawierania dwóch stopni przynależności $\equiv_{\subset}$. W najprostszym przypadku za współczynnik zawierania można bezpośrednio przyjąć operator implikacji

$$a \equiv_{\subset} b = a \Rightarrow b\,.$$

Takie podejście odpowiada badaniu stopnia w jakim A pociąga B. Gdy to nie jest wystarczające możliwe jest rozszerzenie współczynnika zawierania do następującej postaci

$$a \equiv_{\subset} b = \frac{1}{2} \cdot \big(a \Rightarrow b + \overline{b} \Rightarrow \overline{a}\big)\ .$$

Współczynnik zawierania pozwala na porównywanie poszczególnych stopni przynależności. Porównując dwa zbory rozmyte A i B otrzymamy zatem, wektor postaci

$$v_{A,B} = [\mu_A(u_1) \equiv_{\subset} \mu_B(u_1), \cdots, \mu_A(u_n) \equiv_{\subset} \mu_B(u_n)]\,.$$

Aby wyliczyć miarę zawierania, należy go zagregować do pojedynczej liczby rzeczywistej z przedziału $[0, 1]$. Wektor ten można w pewnym sensie traktować jako zbiór rozmyty. Dlatego też do jego agregacji można wykorzystać przykładowo dowolny ewaluator skalarny.

3.3.2. Miary podobieństwa

Operatory implikacji mogą również posłużyć do konstrukcji miar podobieństwa. W tym celu zamiast współczynnika zawierania, należy wykorzystać współczynnik podobieństwa $\equiv_{=}$. Najprostszy z nich to tak zwany opeartor ko-implikacji

$$a \equiv_{=} b = a \Rightarrow b \wedge b \Rightarrow a\,.$$

Został on ogólniony do następującej postaci [22]

$$a \equiv_{=} b = \frac{1}{2} \cdot \big((a \Rightarrow b \wedge b \Rightarrow a) + (\overline{a} \Rightarrow \overline{b} \wedge \overline{b} \Rightarrow \overline{a})\big)\ . \tag{3.3.1}$$

Dalsze ogólnienie tego wzoru jest możliwe poprzez użycie dowolnej t-normy zamiast $\wedge$. Jeśli w (3.3.1) użyjemy S-implikacji to otrzymany wskaźnik podobieństwa będzie identyczny z ko-implikacją

$$\begin{aligned} a \equiv_{=} b &= \frac{1}{2} \cdot \big((a \Rightarrow b \wedge b \Rightarrow a) + (\overline{a} \Rightarrow \overline{b} \wedge \overline{b} \Rightarrow \overline{a})\big) \\ &= \frac{1}{2} \cdot (((1 - a \vee b) \wedge (1 - b \vee a)) + ((a \vee 1 - b) \wedge (b \vee 1 - a))) \\ &= ((1 - a \vee b) \wedge (1 - b \vee a)) \\ &= (a \Rightarrow b) \wedge (b \Rightarrow a)\,. \end{aligned}$$

Równość ta nie zachodzi w przypadku R-implikacji.

Rozdział 4

Wizualizacja wybranych miar zawierania

W pracy przedstawiono wiele podejść do problemu zawierania zbiorów rozmytych. Jak zostało to pokazane czasem miary zdefiniowane z wykorzystaniem całkowicie różnych metod są tożsame. Porównywanie poszczególnych miar jest bardzo trudne, szczególnie jeśli nie są one tożsame. Rozdział ten zawiera opis metody służacej do wizualizacji poszczególnych miar zawierania przy pomocy stosunkowo prostych wykresów konturowych jak również wykresy wybranych miar.

4.1. Konstrukcja wykresów

Dla dowolnego zbioru rozmytego A w uniwersum $U \subset \mathbb{R}$ zdefiniujmy zbiór rozmyty $A_{x,y}$ w następujacy sposób

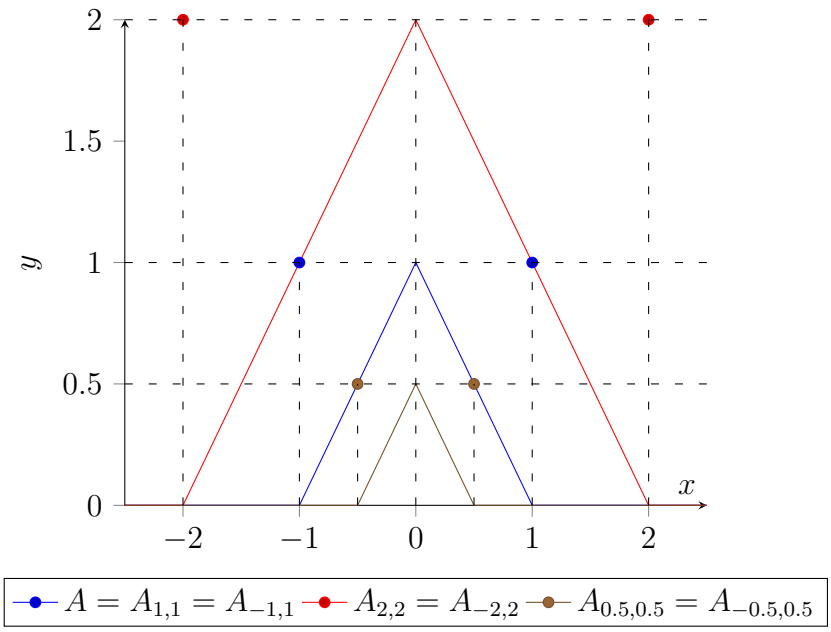
$$\mu_{A_{x,y}}(u) = y * \mu_A(\frac{1}{x} \cdot u)\,.$$

Zbiór rozmyty $A_{x,y}$ jest zatem przeskalowaną w pionie i poziomie wersją zbioru A. Oczywiście zachodzi następująca równość $A_{1,1} = A$. Rysunek 4.1.1 przedstawia zbiory $A_{x,y}$ dla różnych wartości x i y.

Do wizualizacji miar zawierania posłużą wykresy konturowe następującej dwuargumentowej funkcji $f_A : \mathbb{R} \times [0, \frac{1}{\alpha}] \to [0, 1]$

$$f_A(x, y) = Sub(A_{|x|,y}, A)\,,$$

gdzie $\alpha = \max_{u \in U} \mu_A(u)$. Wartość $f_A(x, y)$ jest równa stopniowi zawierania przeskalowanego zbioru w zbiorze A. Jeśli jako miarę zawierania przyjąć klasyczną binarną definicję zawierania zbiorów zaproponowaną przez Zadeha, to funkcja f_A przyjmuje



Rys. 4.1.1. test

następującą postać niezależnie od wyboru zbioru A

$$f_A(x,y) = \begin{cases} 1\,, & \text{jeśli } |x|, y \leq 1 \\ 0 & \text{w przeciwnym przypadku} \end{cases}.$$

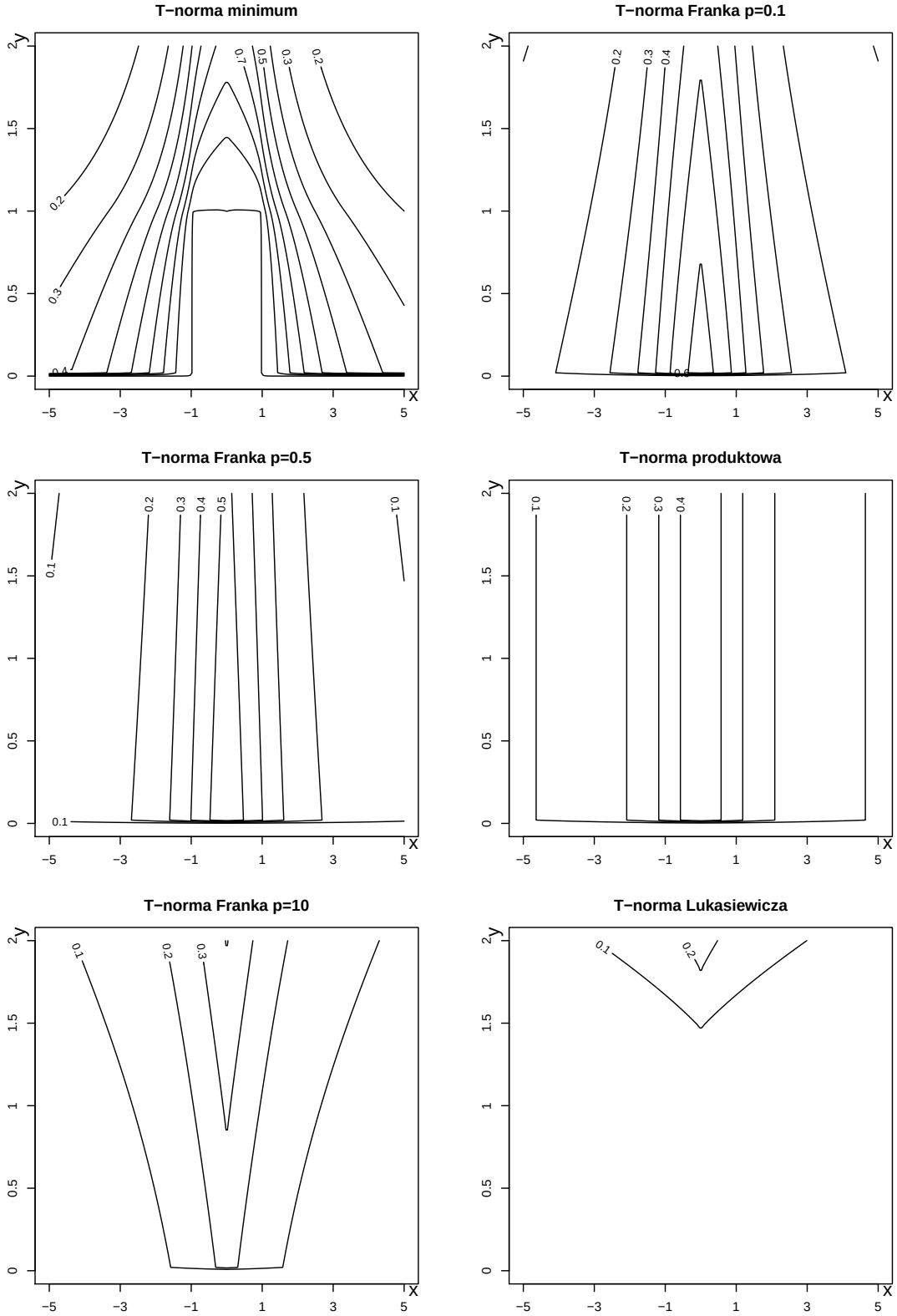
Jako zbiór A przyjąć można dowolny zbiór rozmyty. Należy jednak zadbać aby był on wystarczająco reprezentatywny, a zarazem na tyle prosty aby nie wprowadzał nadmiernej złożoności do wykresów. Z tego powodu w tej pracy jako A wybrano triangularny zbiór rozmyty, którego funkcja przynależności dana jest następującym wzorem

$$\mu_A(x) = \begin{cases} \max(0, \frac{1+x}{2})\,, & \text{jeśli } x \leq 0 \\ \max(0, \frac{1-x}{2})\,, & \text{jeśli } x > 0 \end{cases}.$$

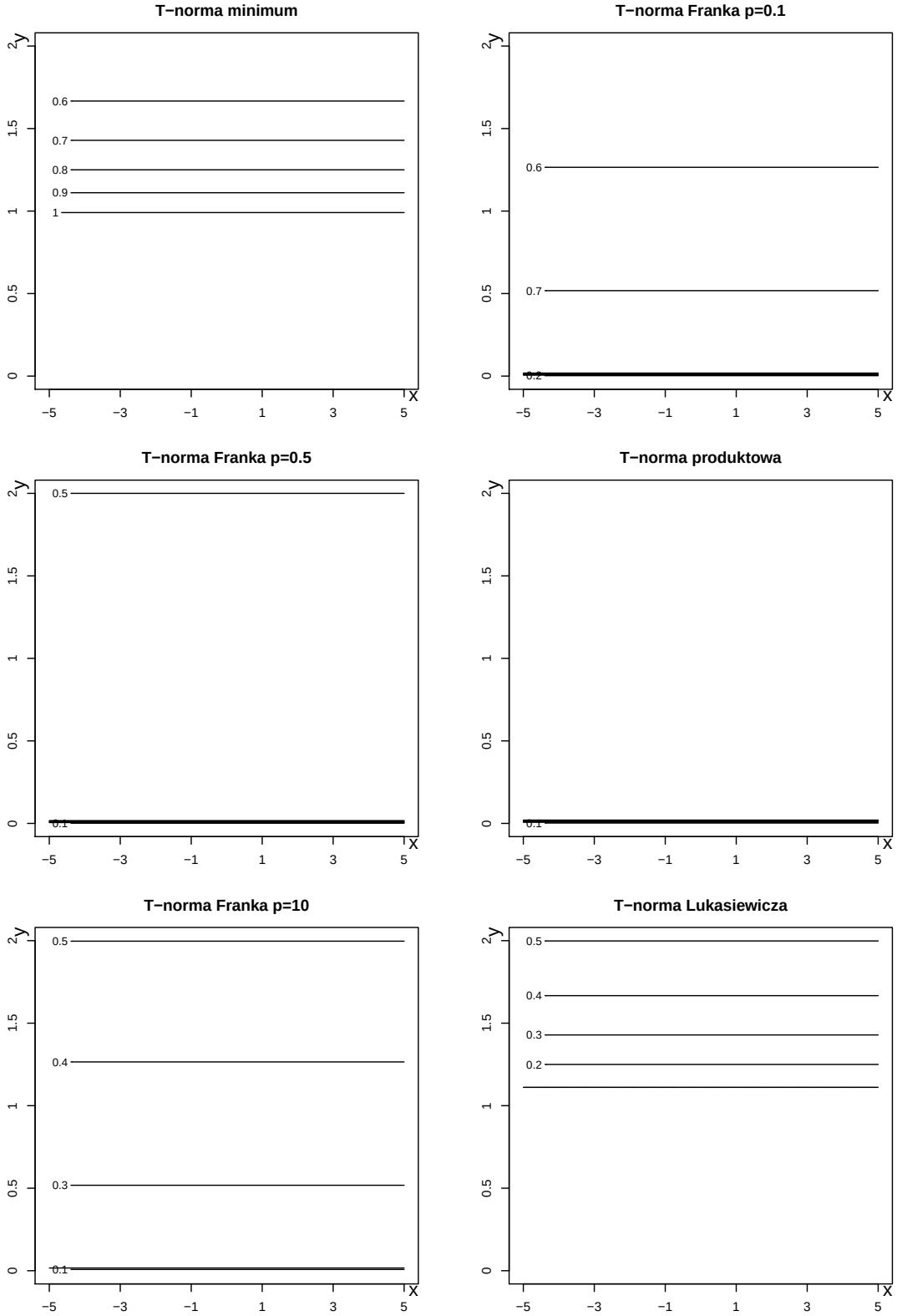
4.2. Wizualizacje

Sekcja ta zawiera wykresy konturowe funkcji f_A różnych miar zawierania zdefiniowanych w rozdziale trzecim pracy. Uwzględniono t-normy i t-konormy z rodziny Franka przy różnych wartościach parametrów. Podstawowe trzy t-normy i t-konormy (minimum, produktowa oraz Łukasiewicza) również należą do tej rodziny (odpowiednio $p = 0$, $p = 1$ oraz $p = \infty$).

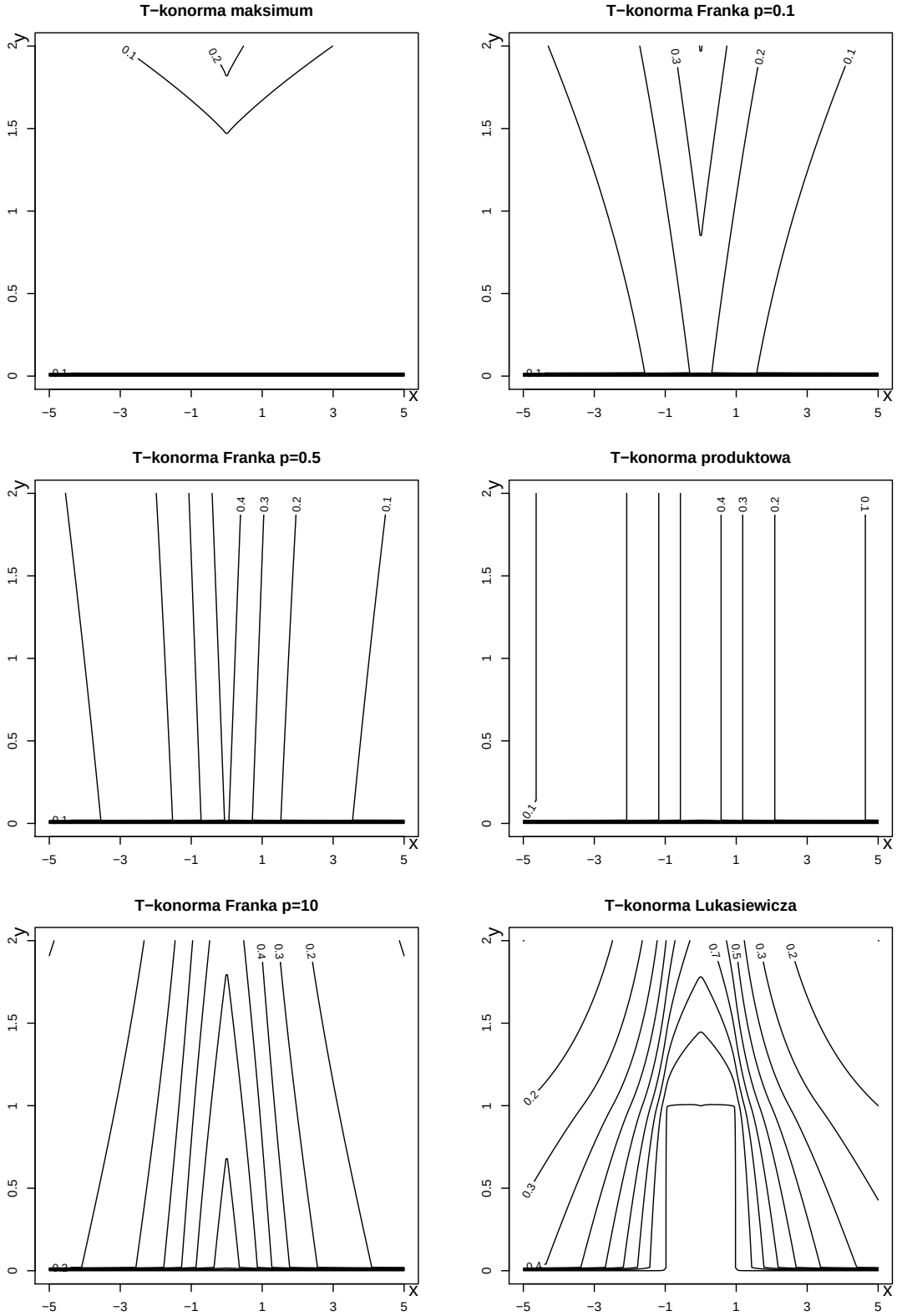
Pominięto wykresy miar zawierania, w przypadku których funkcja f_A jest stała. Ma to miejsce dla niektórych miar korzystających z ewaluatorów skalarnych g_{inf} (stale równa 0) oraz g_{sup} (stale równa 1).



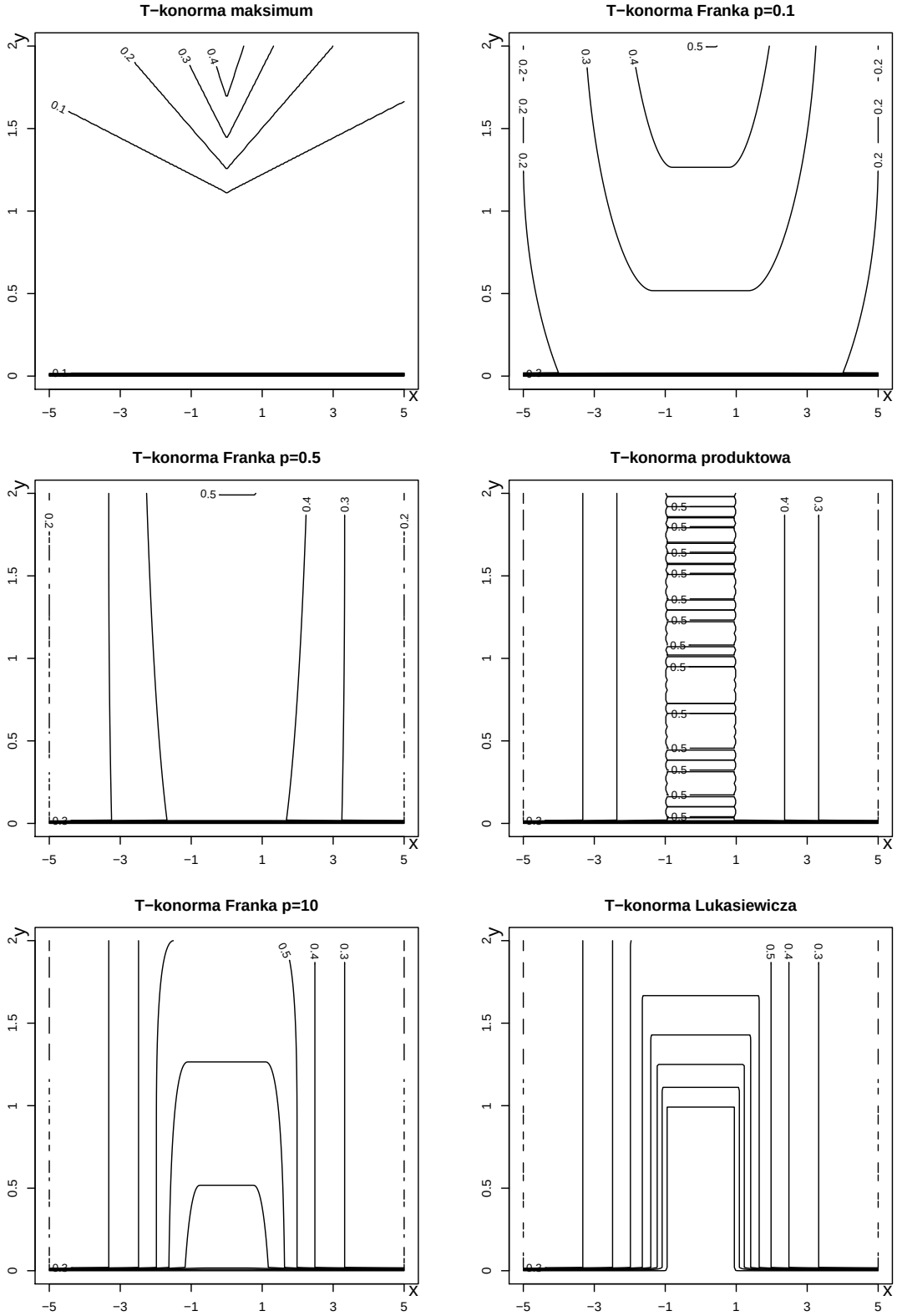
Rys. 4.2.1. Wykres konturowy miar zawierania $Sub_{g_{\mathrm{id}},\cap_t}$ otrzymanych za pomocą operacji przekroju generowanych przez różne t-normy.



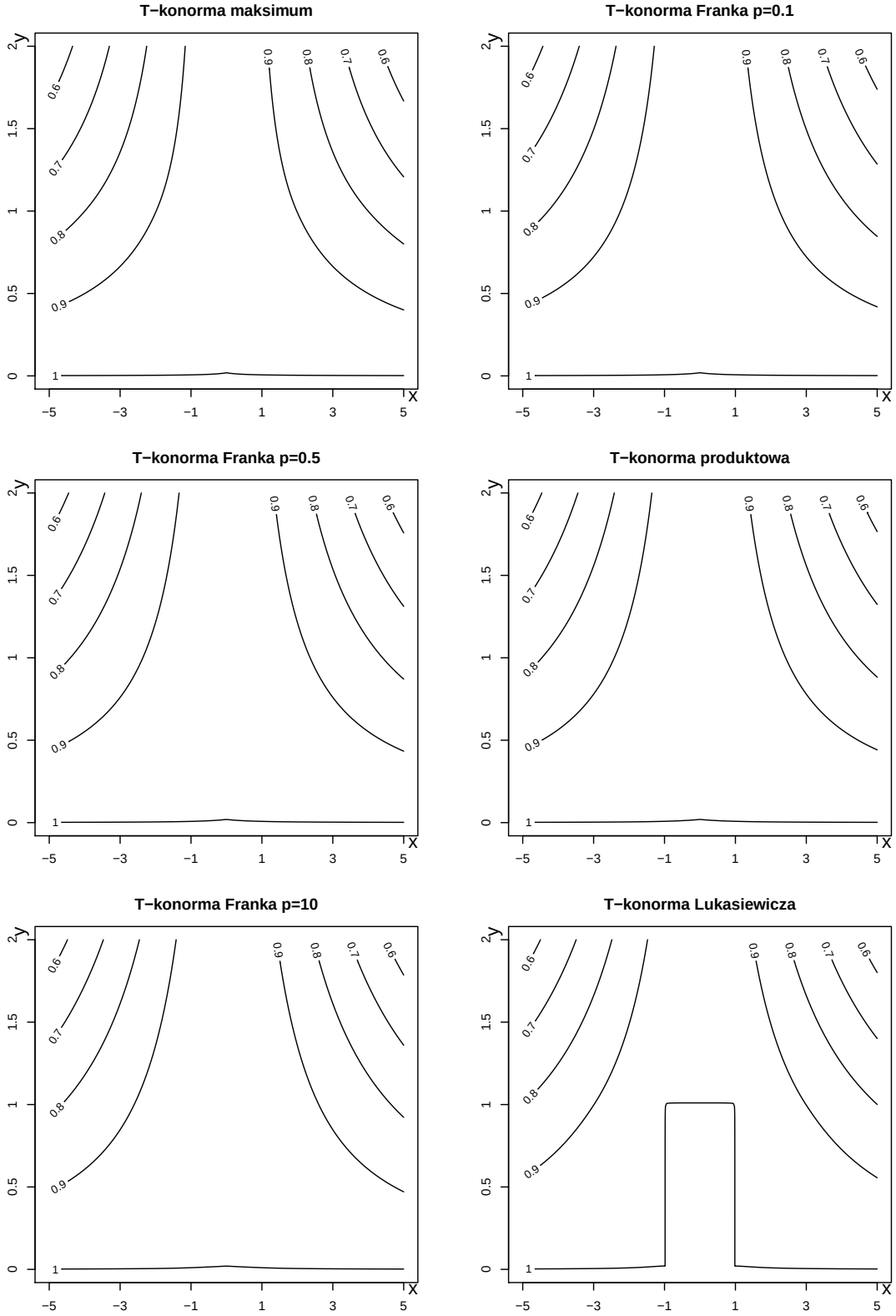
Rys. 4.2.2. Wykres konturowy miar zawierania $Sub_{g_{\sup},\cap_t}$ otrzymanych za pomocą operacji przekroju generowanych przez różne t-normy.



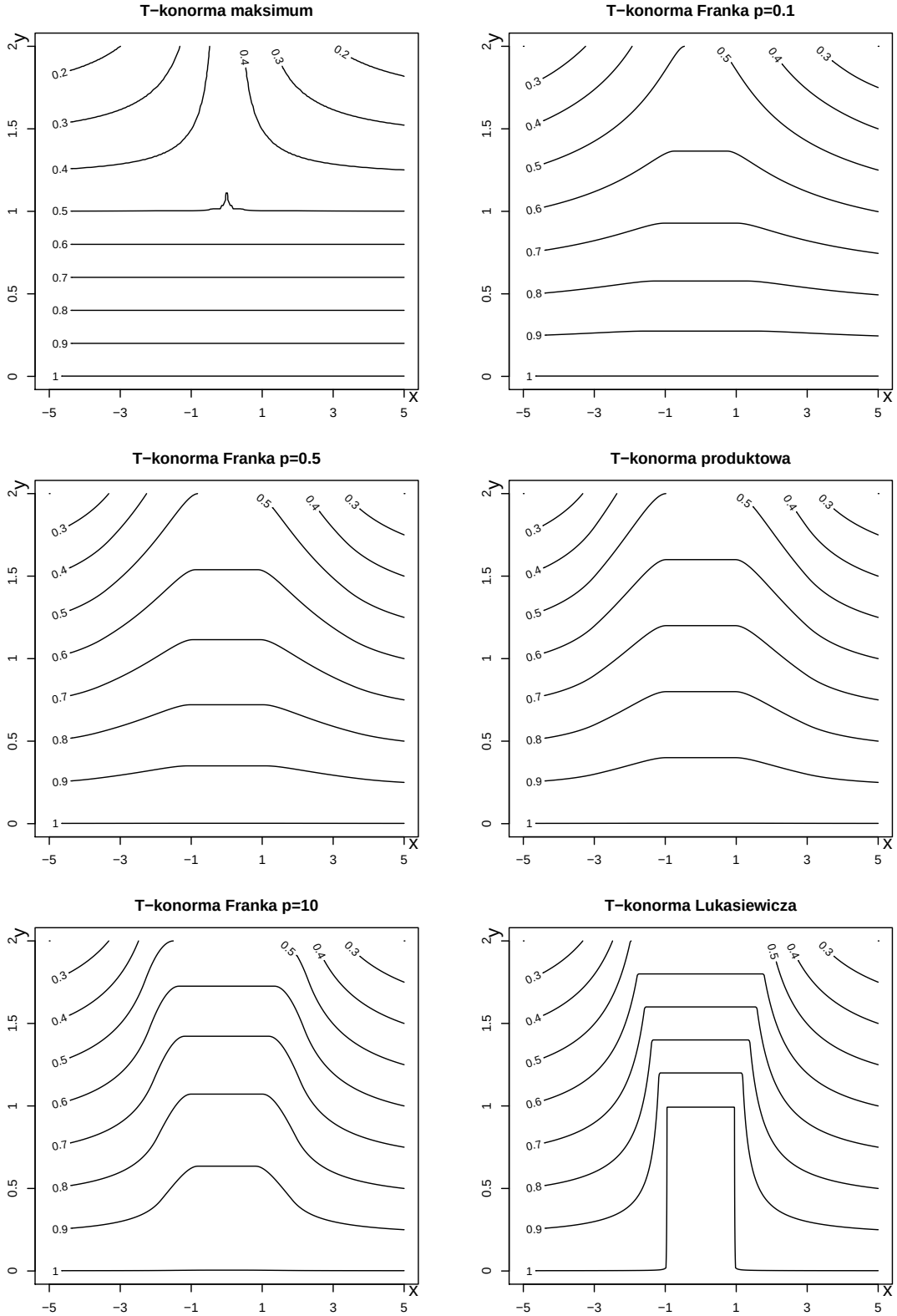
Rys. 4.2.3. Wykres konturowy miar zawierania $Sub_{g_{\mathrm{id}},s}$ otrzymanych za pomocą różnych t-konorm.



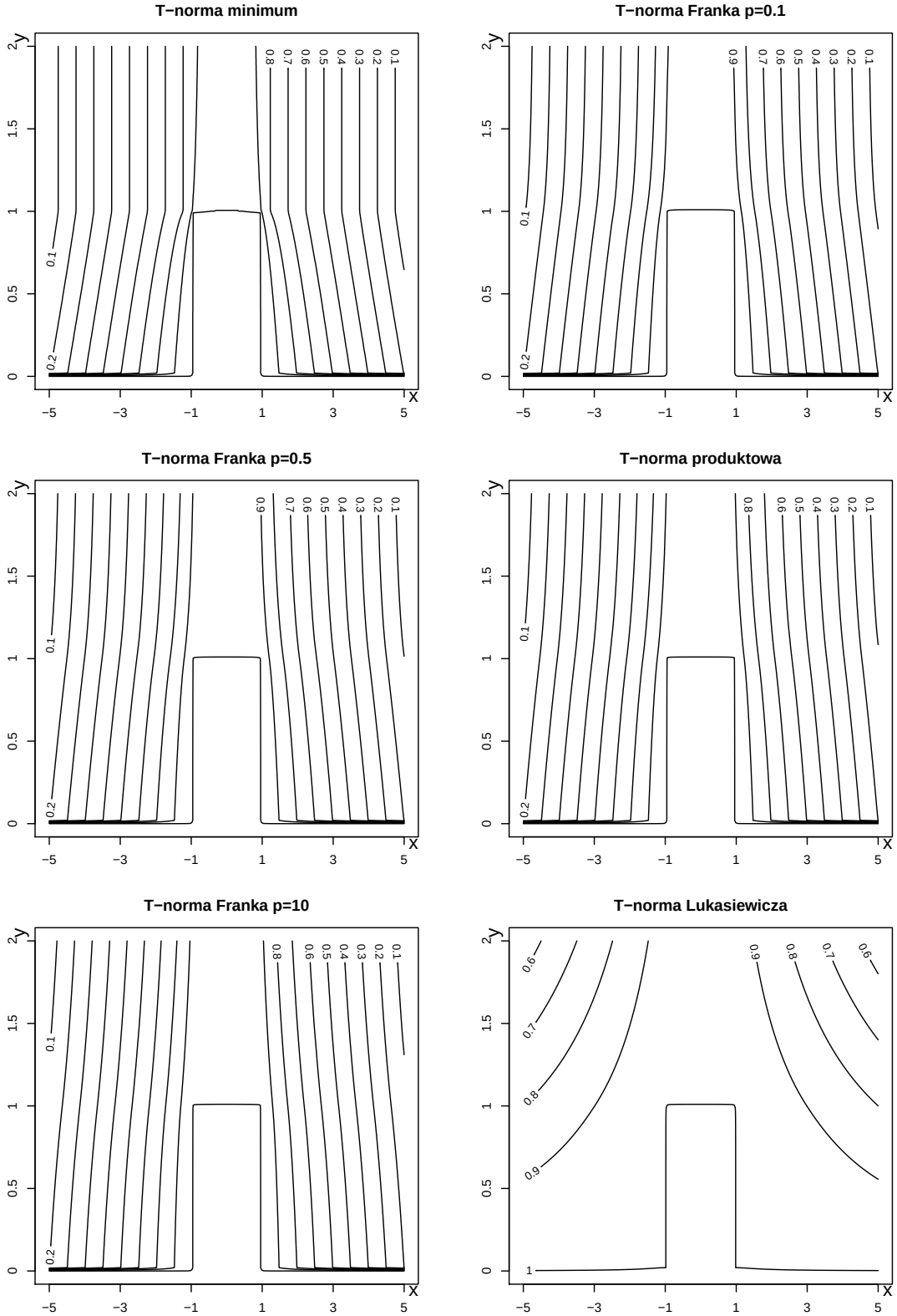
Rys. 4.2.4. Wykres konturowy miar zawierania $Sub_{g_{\inf},s}$ otrzymanych za pomocą różnych t-konorm.



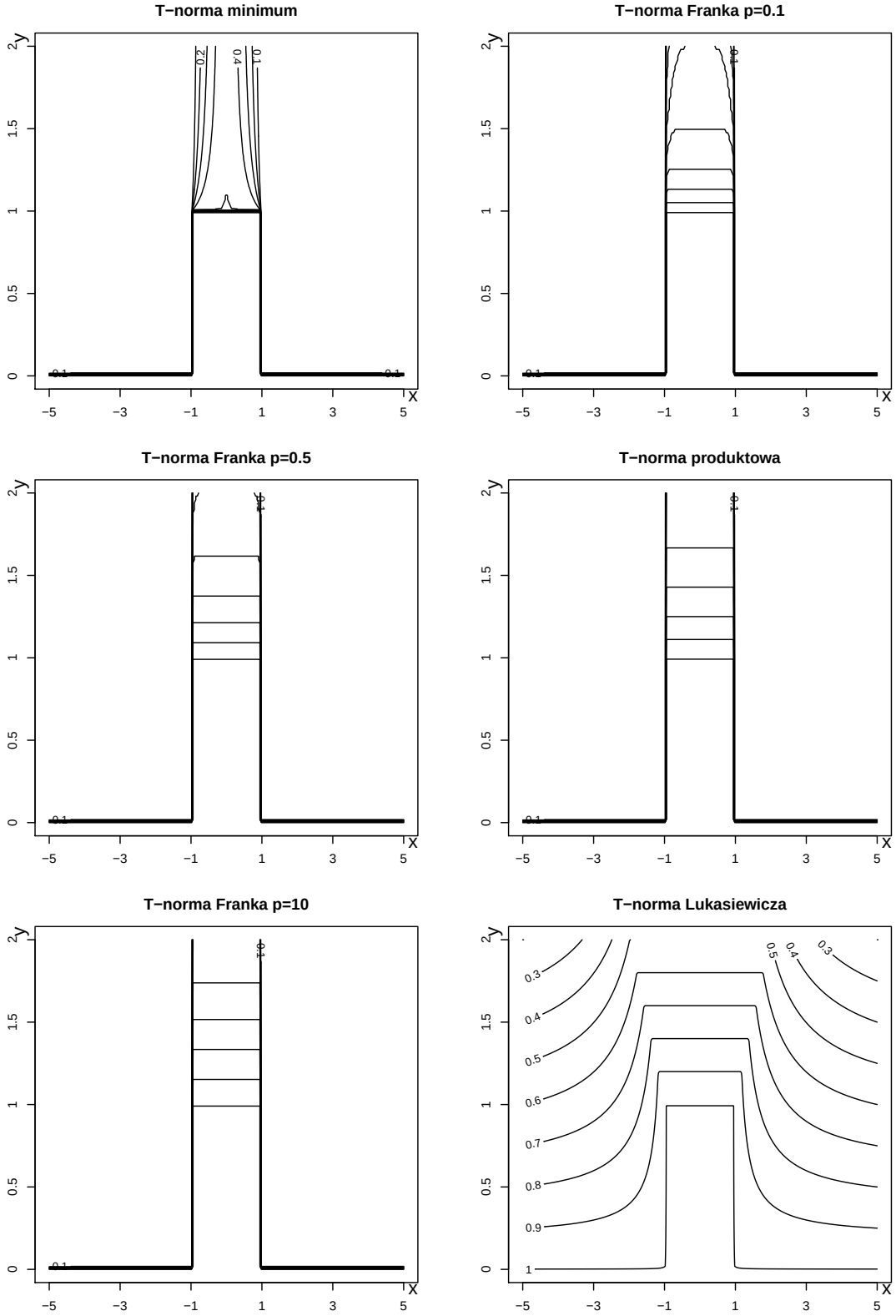
Rys. 4.2.5. Wykres konturowy miar zawierania skonstruowanych na podstawie prostego współczynnika zawierania i S–implikacji generowanych przez różne t-konormy. Do agregacji wykorzystano ewaluator skalarny g_{id}.



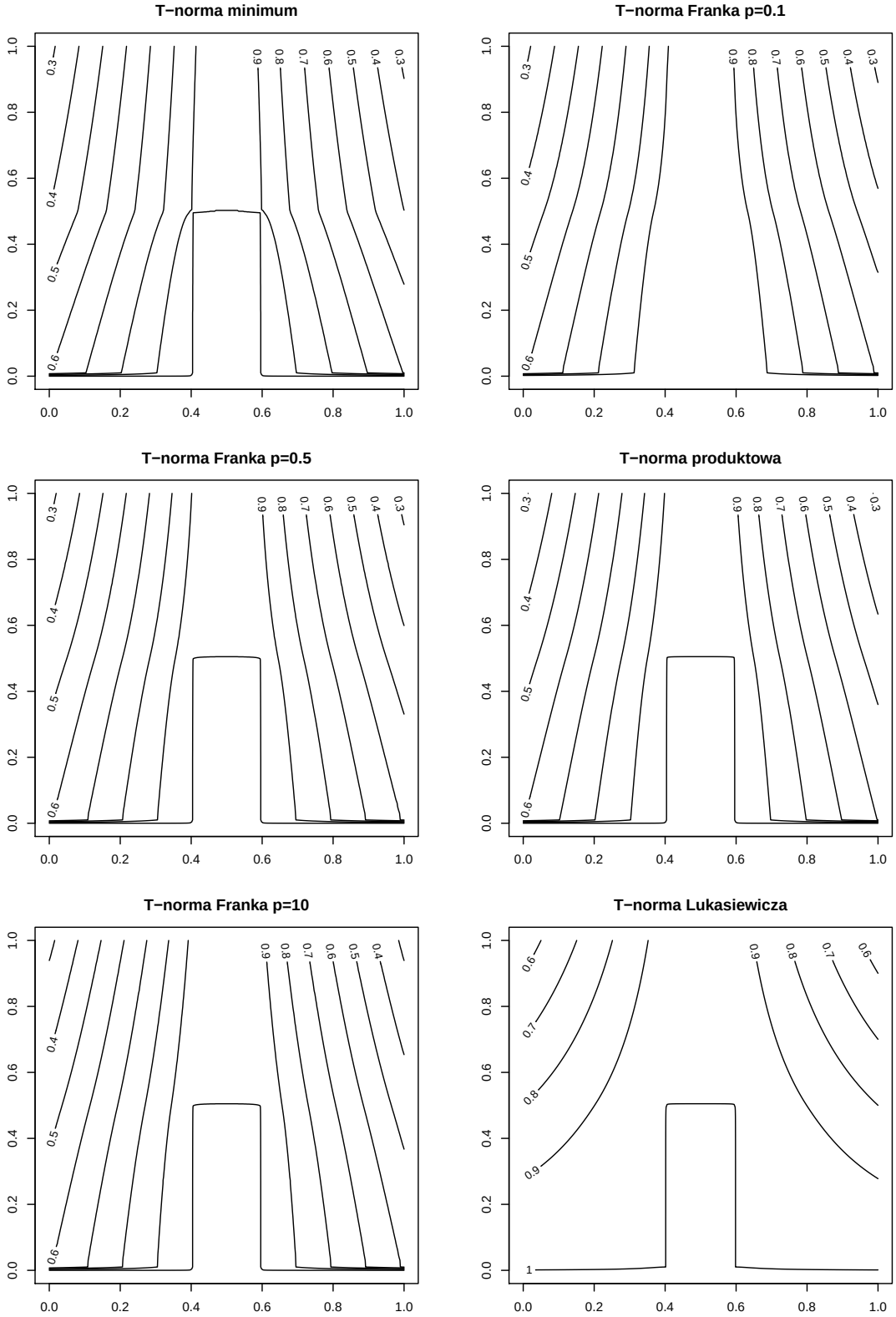
Rys. 4.2.6. Wykres konturowy miar zawierania skonstruowanych na podstawie prostego współczynnika zawierania i S–implikacji generowanych przez różne t-konormy. Do agregacji wykorzystano ewaluator skalarny $g_{\inf}$.



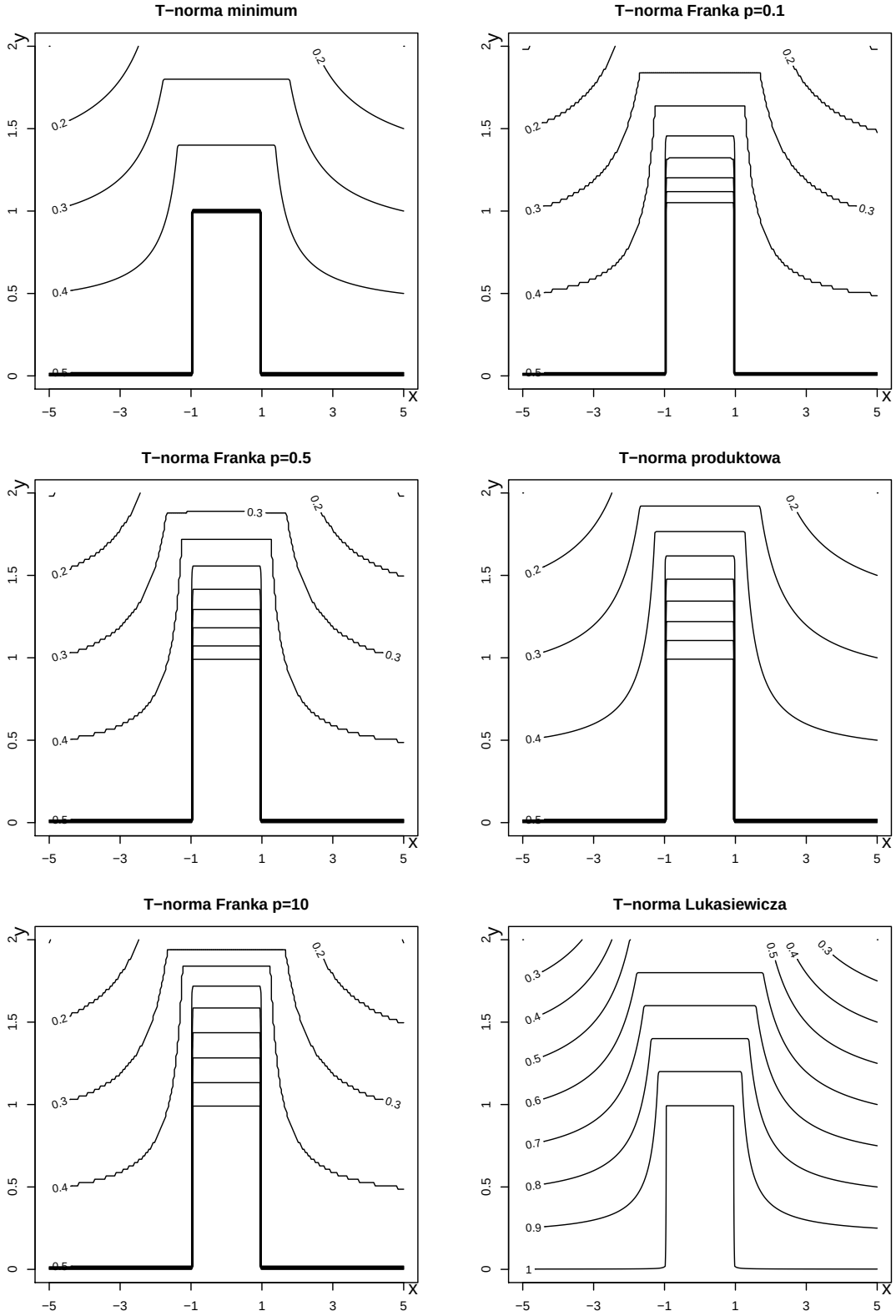
Rys. 4.2.7. Wykres konturowy miar zawierania skonstruowanych na podstawie prostego współczynnika zawierania i R–implikacji generowanych przez różne t-normy. Do agregacji wykorzystano ewaluator skalarny g_{id}.



Rys. 4.2.8. Wykres konturowy miar zawierania skonstruowanych na podstawie prostego współczynnika zawierania i R–implikacji generowanych przez różne t-normy. Do agregacji wykorzystano ewaluator skalarny $g_{\inf}$.



Rys. 4.2.9. Wykres konturowy miar zawierania skonstruowanych na podstawie rozszerzonego współczynnika zawierania i R–implikacji generowanych przez różne t-normy. Do agregacji wykorzystano ewaluator skalarny g_{id}.



Rys. 4.2.10. Wykres konturowy miar zawierania skonstruowanych na podstawie rozszerzonego współczynnika zawierania i R–implikacji generowanych przez różne t-normy. Do agregacji wykorzystano ewaluator skalarny $g_{\inf}$.

Rozdział 5

Podsumowanie

Niniejsza praca poświęcona jest problemowi mierzenia stopinia zawierania i podobieństwa zbiorów rozmytych. Przedstawiono tło historyczne tego problemu oraz podstawowe miary podobieństwa oparte na klasycznej teorii mnogości. Wiele modeli dla zbiorów rozmytych stanowi uogólnienie tych miar.

Ze względu na metody wykorzystane do ich konstrukcji wyróżnione zostały trzy główne grupy miar. Miary oparte na teorii mnogości z reguły stanowią uogólnienie metod dostępnych w klasycznej teorii mnogości. Wiele z nich jak chociażby $Sub_{g_{\mathrm{id}},\cap_{\mathrm{min}}}$ jest powszechnie wykorzystywane w praktycznych zastosowaniach logiki rozmytej. Do grupy tej należą zarówno miary zawierania jak i podobieństwa, ponadto jedna z omówionych metod umożliwia generowanie nowych miar podobieństwa z dowolnej miary zawierania.

Drugą omówioną grupą były miary oparte na bliskości. Ze względu na ich symetryczność, do grupy tej mogą należeć tylko miary podobieństwa. Omówione zostały znane metody oparte o odległość Minkowskiego oraz Hausdorffa. Należyta uwaga została poświęcona również podobieństwu bazującemu na wektorach parametrów, które dzięki zastosowaniu w intensywnie rozwijanej dziedzinie aproksymacji lingwistycznej, zyskują coraz bardziej na znaczeniu.

Ostatnią grupę stanowią miary oparte na logice. Zastosowanie rozmytych operatorów implikacji pozwala na stworzenie nieskończenie wielu miar zawierania i podobieństwa. W pracy omówiono dwie podstawowe rodziny implikacji rozmytych – S i R-implikacje. Przedstawiono dwie metody generowania miar podobieństwa i zawierania z operatorów implikacji oraz pokazano podstawowe ich własności.

W pracy przedstawiono wybrane znane wyniki dotyczące równości wybranych miar zawierania i podobieństwa. Porównywanie poszczególnych miar jest bardzo trudne. O ile formalne wykazanie równości dwóch miar jest możliwe, o tyle sprawdzenie, czy dwie miary zachowują się podobnie dla zbliżonych zbiorów rozmytych jest praktycznie niemożliwe. W rozdziale czwartym podjęto problem wizualizacji miar

zawierania. Zaproponowana w tej pracy metoda pozwala na porównanie różnych miar zawierania. Dalsze badania nad własnościami i porównaniem różnych miar są jednak niezbędne. Z przyczyn technicznych w niniejszej pracy przedstawiono tylko wykresy wybranych miar zawierania.

Bibliografia

[1] Morteza Anvari and Gene F Rose. Fuzzy relational databases. *analysis of Fuzzy Information*, 2:203–212, 1984.

[2] Fred Attneave. Dimensions of similarity. *The American journal of psychology*, pages 516–556, 1950.

[3] Michał Baczyński and Balasubramaniam Jayaram. *An Introduction to Fuzzy Implications*. Springer, 2008.

[4] Jim F Baldwin and BW Pilsworth. Axiomatic approach to implication for approximate reasoning with fuzzy logic. *Fuzzy sets and systems*, 3(2):193–219, 1980.

[5] Wyllis Bandler and Ladislav J Kohout. The use of checklist paradigm in inference systems. *Fuzzy Logic in Knowledge Engineering*, pages 95–111, 1986.

[6] B Bouchon. Inferences with imprecisions and uncertainties in expert systems. *Fuzzy Expert Systems. Addison-Wesley, New York*, 1987.

[7] AJ Boyce. Mapping diversity: a comparative study of some numerical methods. *Numerical taxonomy*, pages 1–31, 1969.

[8] AJ Cain and GA Harrison. An analysis of the taxonomist's judgment of affinity. In *Proceedings of the Zoological Society of London*, volume 131, pages 85–98. Wiley Online Library, 1958.

[9] David V Cross. Metric properties of multidimensional stimulus generalization. *Stimulus generalization*, pages 72–93, 1965.

[10] Valerie V Cross and Thomas A Sudkamp. *Similarity and compatibility in fuzzy set theory: assessment and applications*, volume 93. Springer, 2002.

[11] Didier Dubois and Henri Prade. New results about properties and semantics of fuzzy set-theoretic operators. In *Fuzzy Sets*, pages 59–75. Springer, 1980.

[12] Didier Dubois and Henri Prade. A unifying view of comparison indices in a fuzzy set-theoretic framework. *Fuzzy Sets and Possibility Theory: Recent Developments. Pergamon, New York*, 1982.

[13] Amitava Dutta. Reasoning with imprecise knowledge in expert systems. *Information Sciences*, 37(1):3–24, 1985.

[14] Harald Dyckhoff and Witold Pedrycz. Generalized means as model of compensative connectives. *Fuzzy sets and Systems*, 14(2):143–154, 1984.

[15] Hannes Eisler and Gösta Ekman. A mechanism of subjective similarity. *Nordisk Psykologi*, 11(1):1–10, 1959.

[16] Gösta Ekman and Ralf Lindman. *Multidimensional ratio scaling and multidimensional similarity.* University of Stockholm, the Psychological Laboratory, 1961.

[17] Brian R Gaines. Foundations of fuzzy reasoning. *International Journal of Man-Machine Studies*, 8(6):623–668, 1976.

[18] Paval Gambaryan. *Taxonomic analysis of the genus pinus l.* British Library, Lending Division, 1966.

[19] Joseph A Goguen. The logic of inexact concepts. *Synthese*, 19(3):325–373, 1969.

[20] Robert Anthony Mills Gregson. *Psychometrics of similarity.* Academic Press New York, 1975.

[21] K Hirota and W Pedrycz. Handling fuzziness and randomness in process of matching fuzzy data. In *Proceedings of the Third IFSA Congress*, pages 97–100, 1989.

[22] Kaoru Hirota and Witold Pedrycz. Matching fuzzy quantities. *Systems, Man and Cybernetics, IEEE Transactions on*, 21(6):1580–1586, 1991.

[23] Alston S Householder. Mathematical biophysics and the central nervous system. *Acta Biotheoretica*, 8(1):67–76, 1946.

[24] R. Hyman and A. Well. Judgments of similarity and spatial models. 1967.

[25] Paul Jaccard. *Nouvelles recherches sur la distribution florale.* 1908.

[26] Janusz Kacprzyk. *Wieloetapowe sterowanie rozmyte.* Wydawnictwa Naukowo-Techniczne, 2001.

[27] Thomas Kailath. The divergence and bhattacharyya distance measures in signal selection. *Communication Technology, IEEE Transactions on*, 15(1):52–60, 1967.

[28] Abraham Kandel. *Fuzzy techniques in pattern recognition.* Cambridge Univ Press, 1982.

[29] A Kauffman and Madan M Gupta. Introduction to fuzzy arithmetic: Theory and application. *New York: VanNostrand Reinhold*, 1991.

[30] JB Kiszka, ME Kochanska, and DS Sliwinska. The influence of some parameters on the accuracy of a fuzzy model. *Industrial Applications of Fuzzy Control*, pages 187–230, 1985.

[31] George Klir and Bo Yuan. *Fuzzy sets and fuzzy logic*, volume 4. Prentice Hall New Jersey, 1995.

[32] Ladislav J Kohout and Wyllis Bandler. How the checklist paradigm elucidates the semantics of fuzzy inference. In *Fuzzy Systems, 1992., IEEE International Conference on*, pages 571–578. IEEE, 1992.

[33] Bart Kosko. Fuzziness vs. probability. *International Journal of General System*, 17(2-3):211–240, 1990.

[34] Oswald Kulpe and Edward Bradford Titchener. Outlines of psychology. 1895.

[35] GN Lance and WT Williams. Mixed-data classificatory programs i - agglomerative systems. *Australian Computer Journal*, 1(1):15–20, 1967.

[36] Chuen-Chien Lee. Fuzzy logic in control systems: fuzzy logic controller. ii. *Systems, Man and Cybernetics, IEEE Transactions on*, 20(2):419–435, 1990.

[37] CT Lin and CS George Lee. Real-time supervised structure/parameter learning for fuzzy neural network. In *Fuzzy Systems, 1992., IEEE International Conference on*, pages 1283–1291. IEEE, 1992.

[38] P Magrez and Philippe Smets. Fuzzy modus ponens: A new model suitable for applications in knowledge-based systems. *International Journal of Intelligent Systems*, 4(2):181–200, 1989.

[39] James Mill. *Analysis of the Phenomena of the Human Mind*, volume 1. Longmans, Green, Reader and Dyer, 1869.

[40] Masaharu Mizumoto and Hans-Jürgen Zimmermann. Comparison of fuzzy reasoning methods. *Fuzzy sets and systems*, 8(3):253–283, 1982.

[41] Kiyohiko Nakamura and Sosuke Iwai. A representation of analogical inference by fuzzy sets and its application to information retrieval systems. *Fuzzy information and decision processes*, pages 373–386, 1982.

[42] Donald A Norman and Harley A Bernbach. *Models of human memory.* Academic Press New York, 1970.

[43] Sankar K Pal and D Dutta Majumder. Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Trans. Syst., Man, Cybern*, 7(8):625–629, 1977.

[44] Stephen C Pepper. Concept and quality. 1967.

[45] Daniel Pilarski. Generalized relative cardinalities of fuzzy sets. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 13(01):1–10, 2005.

[46] Henri Prade and Claudette Testemale. Fuzzy relational databases: representational issues and reduction using similarity measures. *Journal of the American Society for Information Science*, 38(2):118–126, 1987.

[47] Anca L Ralescu and Dan A Ralescu. Probability and fuzziness. *Information Sciences*, 34(2):85–92, 1984.

[48] W Silvert. Symmetric summation-class of operations on fuzzy-sets. *IEEE Transactions on Systems Man and Cybernetics*, 9(10):657–659, 1979.

[49] Peter HA Sneath, Robert R Sokal, et al. *Numerical taxonomy. The principles and practice of numerical classification.* 1973.

[50] Peter Henry Andrews Sneath and Robert R Sokal. *Principles of numerical taxonomy.* WH Freeman San Francisco, 1963.

[51] PHA Sneath. The application of numerical taxonomy to medical problems. *Mathematics and Computer Science in Biology and Medicine*, pages 81–91, 1965.

[52] RR Sokal and CD Michener. A statistical method for evaluating systematic relationships. i'niv. kansas sci. bull., 38: 1409-1438. *Primary productivity and ecological factors in Lake Maggiore*, 127, 1958.

[53] Warren S Torgerson. Multidimensional scaling of similarity. *Psychometrika*, 30(4):379–393, 1965.

[54] Enric Trillas and Llorenç Valverde. On mode and implication in approximate reasoning. *Approximate reasoning in expert systems*, pages 157–166, 1985.

[55] Amos Tversky and Itamar Gati. Studies of similarity. *Cognition and categorization*, 1(1978):79–98, 1978.

[56] Laurent Ughetto, Didier Dubois, and Henri Prade. Fuzzy interpolation by convex completion of sparse rule bases. In *Proceedings of 9th IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2000), San Antonio, Tx, USA*, pages 465–470, 2000.

[57] Fred Wenstøp. Quantitative analysis with linguistic values. *Fuzzy Sets and Systems*, 4(2):99–115, 1980.

[58] Thomas Whalen and Brian Schott. Alternative logics for approximate reasoning in expert systems: A comparative study. *International journal of man-machine studies*, 22(3):327–346, 1985.

[59] Maciej Wygralak. *Cardinalities of fuzzy sets*, volume 118. Springer, 2003.

[60] Maciej Wygralak. Cardinalities of fuzzy sets with triangular norms. In *EUSFLAT Conf.*, pages 691–696, 2003.

[61] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[62] Rami Zwick, Edward Carlstein, and David V Budescu. Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, 1987.