

Komputerowe systemy rozpoznawania

2020/2021

Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

Kacper Czernik 242371
Mateusz Grzeszczak 242398
Prowadzący: Dr inż. Marcin Kacprowicz

8 czerwca 2024

1 Cel

W ramach projektu zaimplementowano aplikację umożliwiającą lingwistyczną agregację danych liczbowych z wykorzystaniem logiki rozmytej. Poprzez stworzenie struktury klas reprezentujących różne typy zbiorów rozmytych oraz operacje na nich, aplikacja generuje podsumowania lingwistyczne w języku quasi-naturalnym [2]. Dzięki interfejsowi graficznemu użytkownicy mogą dostosować generowane podsumowania poprzez wybór predefiniowanych sumaryzatorów, kwalifikatorów i kwantyfikatorów, określanie wag dla miar jakości oraz sortowanie wyników.

2 Baza danych, zmienne lingwistyczne, kwantyfikatory lingwistyczne

2.1 Charakterystyka podsumowywanej bazy danych

W ramach projektu wybrano bazę danych [4], zawierającą pomiary parametrów zbiorników wodnych wokół Irlandii. Baza ta zawiera łącznie 281 600 rekordów, jednak po wykluczeniu wartości "Null" pozostaje ponad 210 tysięcy wyników. Zestaw danych posiada 12 kolumn możliwych do rozmycia, z których 4 zostaną połączone podczas pre-procesingu danych, w celu zachowania logicznej spójności podsumowania. Przykładowe wartości dla tego zbioru danych przedstawiono poniżej (Tabela 1). Kolumny `latitude_degrees_north` oraz `longitude_degrees_east` zostaną połączone, aby reprezentować odległość od środka Irlandii, umożliwiając identyfikację położenia pomiaru w zależności od części

kraju. Analogicznie, kolumny sea_surface_x_velocity_m_s oraz sea_surface_y_velocity_m_s zostaną połączone w celu reprezentacji bezwzględnej prędkości wody, ułatwiając podsumowywanie wyników.

time	latitude	longitude	sea_surface_temperature	sea_bottom_temperature	sea_surface_salinity	sea_bottom_salinity	sea_surface_x_velocity	sea_surface_y_velocity	mixed_layer_depth	significant_wave_height	mean_wave_direction	mean_wave_period
[UTC]	[degrees]	[degrees]	[C]	[C]	[PSU]	[PSU]	[m/s]	[m/s]	[m]	[m]	[degrees]	[s]
2016-08-15T00:00:00Z	48.0125	-17.9675	17.9150	2.1510	35.5712	34.9112	0.0142	-0.0266	24.0529	2.4823	98.1167	5.7896
2016-08-15T00:00:00Z	48.0125	-17.9625	17.9150	2.1510	35.5712	34.9112	0.0142	-0.0266	24.0529	2.4820	98.1082	5.7928

Rysunek 1: Marine Institute Monthly Model Means ϵ

2.2 Zmienne lingwistyczne (atrybuty/własności obiektów)

W ramach projektu opisano zmienne lingwistyczne, nadając im etykiety oraz prezentując wykresy funkcji przynależności oraz wzory analityczne. Poniżej przedstawiono opisy poszczególnych zmiennych wraz z ich definicjami:

2.2.1 Miejsce pomiaru

Najbardziej niestandardową zmienną lingwistyczną utworzoną na potrzeby zadania jest miejsce pobierania pomiaru, zmienna ta składać się będzie z dwóch części: szerokości geograficznej oraz długości geograficznej. Dla obu tych części utworzone zostały osobne funkcje przynależności pozwalające na dokładne opisanie miejsca, w którym dany pomiar został zebrany.

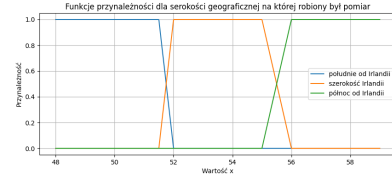
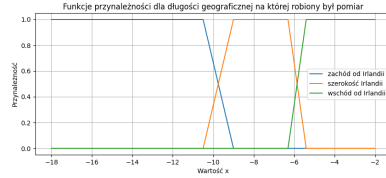
Poniżej znajduje się mapa na której zaznaczony został obszar poddany badaniom.



$$\begin{aligned}
 \text{na zachód od Irlandii}(\mathbf{X}) &= \begin{cases} 1 & \text{dla } x \leq -10.5 \\ \frac{-9-x}{1.5} & \text{dla } -10.5 < x \leq -9 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{na południe od Irlandii}(\mathbf{X}) &= \begin{cases} 1 & \text{dla } x \leq 51.5 \\ \frac{52-x}{0.5} & \text{dla } 51.5 < x \leq 52 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\
 \text{na szerokości Irlandii}(\mathbf{X}) &= \begin{cases} \frac{x-(-10.5)}{1.5} & \text{dla } -10.5 < x \leq -9 \\ 1 & \text{dla } -10.5 < x < -6.3 \\ \frac{-5.4-x}{0.9} & \text{dla } -6.3 < x \leq -5.4 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{na szerokości Irlandii}(\mathbf{X}) &= \begin{cases} \frac{x-51.5}{0.5} & \text{dla } 51.5 < x \leq 52 \\ 1 & \text{dla } 52 < x < 55 \\ 56-x & \text{dla } 55 < x \leq 56 \\ 0 & \text{w przeciwnym przypadku} \end{cases}
 \end{aligned}$$

$$\text{na wschód od Irlandii}(\mathbf{X}) = \begin{cases} 0 & \text{dla } x \leq -6.3 \\ \frac{x - (-6.3)}{0.9} & \text{dla } -6.3 < x \leq -5.4 \\ 1 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{na północ od Irlandii}(\mathbf{X}) = \begin{cases} 0 & \text{dla } x \leq 55 \\ x - 55 & \text{dla } 55 < x \leq 56 \\ 1 & \text{w przeciwnym przypadku} \end{cases}$$



2.2.2 Temperatura na powierzchni

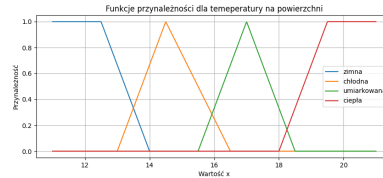
Poniżej zaprezentowany został podział temperatury na wyselekcjonowane etykiety wraz z funkcjami przynależności dla każdej z nich. Zmienna ta opisuje temperaturę uzyskaną podczas pomiaru na powierzchni tafli wody.

$$\text{zimna}(\mathbf{X}) = \begin{cases} 1 & \text{dla } x \leq 12.5 \\ \frac{14-x}{1.5} & \text{dla } 12.5 < x \leq 14 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{umiarkowana}(\mathbf{X}) = \begin{cases} \frac{x-15.5}{1.5} & \text{dla } 15.5 < x < 17 \\ 1 & \text{dla } x = 17 \\ \frac{18.5-x}{1.5} & \text{dla } 17 < x \leq 18.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{chłodna}(\mathbf{X}) = \begin{cases} \frac{x-13}{1.5} & \text{dla } 13 < x < 14.5 \\ 1 & \text{dla } x = 14.5 \\ \frac{16.5-x}{2} & \text{dla } 14.5 < x \leq 16.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{ciepła}(\mathbf{X}) = \begin{cases} 0 & \text{dla } x \leq 18 \\ \frac{x-18}{1} & \text{dla } 18 < x \leq 19.5 \\ 1 & \text{w przeciwnym przypadku} \end{cases}$$



2.2.3 Temperatura na dnie

Zmienna ta działa podobnie jak zmienna opisująca temperaturę na powierzchni z tą różnicą, że opisuje temperaturę zmierzoną na dnie zbiornika wodnego. Poniżej znajdują się wzory przynależności dla tej zmiennej lingwistycznej.

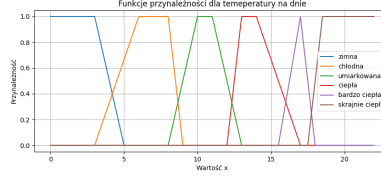
$$\text{zimna}(x) = \begin{cases} 1 & \text{dla } x \leq 3 \\ \frac{5-x}{2} & \text{dla } 3 < x \leq 5 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{umiarkowana}(x) = \begin{cases} \frac{x-8}{2} & \text{dla } 8 < x \leq 10 \\ 1 & \text{dla } 10 < x \leq 11 \\ \frac{13-x}{2} & \text{dla } 11 < x \leq 13 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{chłodna}(x) = \begin{cases} \frac{x-3}{3} & \text{dla } 3 < x \leq 6 \\ 1 & \text{dla } 6 < x \leq 8 \\ 9-x & \text{dla } 8 < x \leq 9 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{ciepła}(x) = \begin{cases} x-12 & \text{dla } 12 < x \leq 13 \\ 1 & \text{dla } 13 < x \leq 14 \\ \frac{17-x}{3} & \text{dla } 14 < x \leq 17 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{bardzo ciepła}(x) = \begin{cases} \frac{x-15.5}{1.5} & \text{dla } 15.5 < x < 17 \\ 1 & \text{dla } x = 17 \\ 18-x & \text{dla } 17 < x \leq 18 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad \text{skrajnie ciepła}(x) = \begin{cases} 0 & \text{dla } x \leq 17.5 \\ x-17.5 & \text{dla } 17.5 < x \leq 18.5 \\ 1 & \text{w przeciwnym przypadku} \end{cases}$$



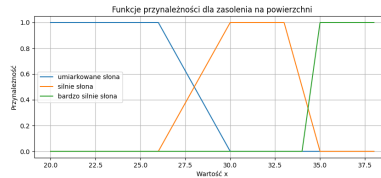
2.2.4 Zasolenie na powierzchni

Zmienna lingwistyczna opisująca zasolenie na powierzchni tafli wody, wartości przedstawione są w jednostkach zasolenia [PSU].

$$\text{umiarkowanie słona}(x) = \begin{cases} 1 & \text{dla } x \leq 26 \\ \frac{30-x}{4} & \text{dla } 26 < x \leq 30 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{silnie słona}(x) = \begin{cases} \frac{x-26}{4} & \text{dla } 26 < x \leq 30 \\ 1 & \text{dla } 30 < x < 33 \\ \frac{35-x}{2} & \text{dla } 33 < x \leq 35 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{bardzo silnie słona}(x) = \begin{cases} 0 & \text{dla } x \leq 34 \\ x-34 & \text{dla } 34 < x \leq 35 \\ 1 & \text{w przeciwnym przypadku} \end{cases}$$

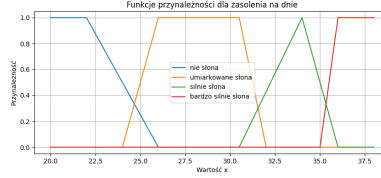


2.2.5 Zasolenie na dnie

Ta zmienna lingwistyczna opisuje zasolenie na dnie opisane w jednostkach zasolenia [PSU].

$$\text{nie słona}(x) = \begin{cases} 1 & \text{dla } x \leq 22 \\ \frac{26-x}{4} & \text{dla } 22 < x \leq 26 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad \text{umiarkowanie słona}(x) = \begin{cases} \frac{x-24}{2} & \text{dla } 24 < x \leq 26 \\ 1 & \text{dla } 26 < x \leq 30.5 \\ \frac{32-x}{1.5} & \text{dla } 30.5 < x \leq 32 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

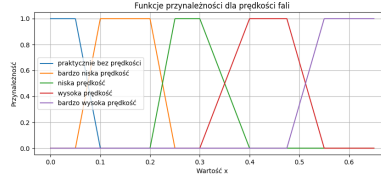
$$\text{silnie słona}(x) = \begin{cases} \frac{x-30.5}{3.5} & \text{dla } 30.5 < x < 34 \\ 1 & \text{dla } x = 34 \\ \frac{36-x}{2} & \text{dla } 34 < x \leq 36 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad \text{bardzo silnie słona}(x) = \begin{cases} 0 & \text{dla } x \leq 35 \\ x-35 & \text{dla } 35 < x \leq 36 \\ 1 & \text{w przeciwnym przypadku} \end{cases}$$



2.2.6 Prędkość wody

Wspomniana Zmienna opisuje prędkość, z jaką poruszała się woda w miejscu pomiaru w momencie jego robienia, jednostka, w której dane zostały zebrane to [m/s].

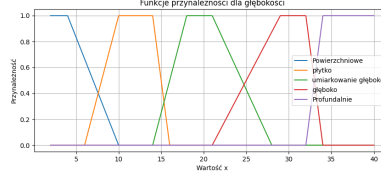
$$\begin{aligned} \text{praktycznie bez prędkości}(x) &= \begin{cases} 1 & \text{dla } x \leq 0.05 \\ \frac{0.1-x}{0.05} & \text{dla } 0.05 < x \leq 0.1 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{bardzo niska prędkość}(x) &= \begin{cases} \frac{x-0.05}{0.05} & \text{dla } 0.05 < x \leq 0.1 \\ 1 & \text{dla } 0.1 < x \leq 0.2 \\ \frac{0.25-x}{0.05} & \text{dla } 0.2 < x \leq 0.25 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{niska prędkość}(x) &= \begin{cases} \frac{x-0.2}{0.05} & \text{dla } 0.2 < x \leq 0.25 \\ 1 & \text{dla } 0.25 < x \leq 0.3 \\ \frac{0.4-x}{0.1} & \text{dla } 0.3 < x \leq 0.4 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{wysoka prędkość}(x) &= \begin{cases} \frac{x-0.3}{0.1} & \text{dla } 0.3 < x \leq 0.4 \\ 1 & \text{dla } 0.4 < x \leq 0.475 \\ \frac{0.55-x}{0.075} & \text{dla } 0.475 < x \leq 0.55 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{bardzo wysoka prędkość}(x) &= \begin{cases} 0 & \text{dla } x \leq 0.475 \\ \frac{x-0.475}{0.075} & \text{dla } 0.475 < x \leq 0.55 \\ 1 & \text{w przeciwnym przypadku} \end{cases} \end{aligned}$$



2.2.7 Głębokość dna

Zmienna opisująca, na jakiej głębokości znajduje się dno w miejscu robienia danego pomiaru, głębokość ta opisana została w metrach [m]. Etykiety wraz z funkcjami przynależności opisane zostały poniżej.

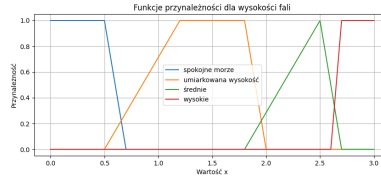
$$\begin{aligned} \text{głębokość przypowierzchniowa}(x) &= \begin{cases} 1 & \text{dla } x \leq 4 \\ \frac{10-x}{6} & \text{dla } 4 < x \leq 10 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{płytko}(x) &= \begin{cases} \frac{x-6}{4} & \text{dla } 6 < x \leq 10 \\ 1 & \text{dla } 10 < x \leq 14 \\ \frac{16-x}{2} & \text{dla } 14 < x \leq 16 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{umiarkowanie głęboko}(x) &= \begin{cases} \frac{x-14}{4} & \text{dla } 14 < x \leq 18 \\ 1 & \text{dla } 18 < x \leq 21 \\ \frac{28-x}{7} & \text{dla } 21 < x \leq 28 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{głęboko}(x) &= \begin{cases} \frac{x-21}{8} & \text{dla } 21 < x \leq 29 \\ 1 & \text{dla } 29 < x \leq 32 \\ \frac{34-x}{2} & \text{dla } 32 < x \leq 34 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{profundalnie głęboko}(x) &= \begin{cases} 0 & \text{dla } x \leq 32 \\ \frac{x-32}{2} & \text{dla } 32 < x \leq 34 \\ 1 & \text{w przeciwnym przypadku} \end{cases} \end{aligned}$$



2.2.8 wysokość fal

Wysokość fal to zmienna lingwistyczna mówiąca o tym, jak wysokie fale zostały zarejestrowane w momencie robienia pomiaru. Zmierzona wysokość ta opisana została w jednostkach si, czyli metrach [m]. Dokładny opis o tym, jak mierzy się wspomnianą wysokość opisana została w prezentacji [5].

$$\begin{aligned} \text{spokojne morze}(x) &= \begin{cases} 1 & \text{dla } x \leq 0.5 \\ \frac{0.7-x}{0.2} & \text{dla } 0.5 < x \leq 0.7 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{nie wielkie fale}(x) &= \begin{cases} \frac{x-0.5}{0.7} & \text{dla } 0.5 < x \leq 1.2 \\ 1 & \text{dla } 1.2 < x \leq 1.8 \\ \frac{2-x}{0.2} & \text{dla } 1.8 < x \leq 2 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{średnie fale}(x) &= \begin{cases} \frac{x-1.8}{0.7} & \text{dla } 1.8 < x < 2.5 \\ 1 & \text{dla } x = 2.5 \\ \frac{2.7-x}{0.2} & \text{dla } 2.5 < x \leq 2.7 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{wysokie fale}(x) &= \begin{cases} 0 & \text{dla } x \leq 2.6 \\ \frac{x-2.6}{0.1} & \text{dla } 2.6 < x \leq 2.7 \\ 1 & \text{w przeciwnym przypadku} \end{cases} \end{aligned}$$



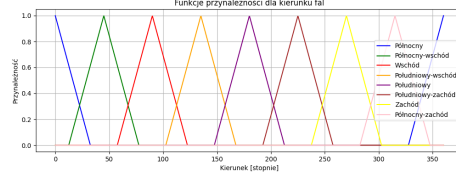
2.2.9 Kierunek fal

Kierunek fal mierzony mówi o tym, w którą stronę skierowane były fale w momencie robienia pomiaru. Wartość ta mierzona była w stopniach, jednakże dla ułatwienia analizy etykiety wybrane dla tej zmiennej zapisane zostały jako kierunki świata.

$$\begin{aligned} \text{północ}(x) &= \begin{cases} 1 & \text{dla } x \leq 0 \\ \frac{32.5-x}{32.5} & \text{dla } 0 < x \leq 32.5 \\ \frac{x-327.5}{32.5} & \text{dla } 327.5 < x \leq 360 \\ 1 & \text{dla } x \geq 360 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{północny-wschód}(x) &= \begin{cases} \frac{x-12.5}{32.5} & \text{dla } 12.5 < x \leq 45 \\ 1 & \text{dla } x = 45 \\ \frac{77.5-x}{32.5} & \text{dla } 45 < x \leq 77.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{wschód}(x) &= \begin{cases} \frac{x-57.5}{32.5} & \text{dla } 57.5 < x \leq 90 \\ 1 & \text{dla } x = 90 \\ \frac{122.5-x}{32.5} & \text{dla } 90 < x \leq 122.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{południowy-wschód}(x) &= \begin{cases} \frac{x-102.5}{32.5} & \text{dla } 102.5 < x \leq 135 \\ 1 & \text{dla } x = 135 \\ \frac{167.5-x}{32.5} & \text{dla } 135 < x \leq 167.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{południe}(x) &= \begin{cases} \frac{x-147.5}{32.5} & \text{dla } 147.5 < x < 180 \\ 1 & \text{dla } x = 180 \\ \frac{212.5-x}{32.5} & \text{dla } 180 < x \leq 212.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{południowy-zachód}(x) &= \begin{cases} \frac{x-192.5}{32.5} & \text{dla } 192.5 < x \leq 225 \\ 1 & \text{dla } x = 225 \\ \frac{257.5-x}{32.5} & \text{dla } 225 < x \leq 257.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \end{aligned}$$

$$\text{zachód}(x) = \begin{cases} \frac{x-237.5}{32.5} & \text{dla } 237.5 < x \leq 270 \\ 1 & \text{dla } x = 270 \\ \frac{302.5-x}{32.5} & \text{dla } 270 < x \leq 302.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

$$\text{północny-zachód}(x) = \begin{cases} \frac{x-282.5}{32.5} & \text{dla } 282.5 < x \leq 315 \\ 1 & \text{dla } x = 315 \\ \frac{347.5-x}{32.5} & \text{dla } 315 < x \leq 347.5 \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$



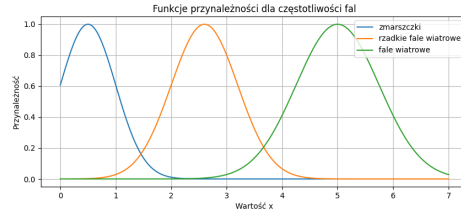
2.2.10 Częstotliwość fal

Zmienna lingwistyczna opisująca częstotliwość tworzenia się fal opisana została przy pomocy etykiet utworzonych na podstawie dokumentu [5], niewielka ilość etykiet utworzonych dla tej zmiennej spowodowana jest niewielkim przedziałem częstotliwości tworzenia się fal dla badanego obszaru.

$$\text{zmarszczki}(x) = \exp\left(-\frac{(x-0.5)^2}{2 \cdot 0.5^2}\right)$$

$$\text{rzadkie fale wiatrowe}(x) = \exp\left(-\frac{(x-2.6)^2}{2 \cdot 0.6^2}\right)$$

$$\text{fale wiatrowe}(x) = \exp\left(-\frac{(x-5)^2}{2 \cdot 0.75^2}\right)$$



2.3 Kwantyfikatory lingwistyczne

W ramach projektu opisano kwantyfikatory lingwistyczne, nadając im etykiety oraz prezentując wykresy funkcji przynależności, oraz wzory analityczne. Kwantyfikatory rozmyte względne i absolutne są narzędziami w logice rozmytej służącymi do opisu stopnia przynależności elementów do zbiorów rozmytych. Kwantyfikatory względne odnoszą się do relacji między zbiorami, podczas gdy kwantyfikatory absolutne opisują stopień przynależności wszystkich elementów do konkretnego zbioru rozmytego.

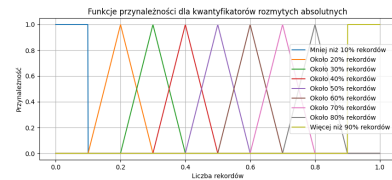
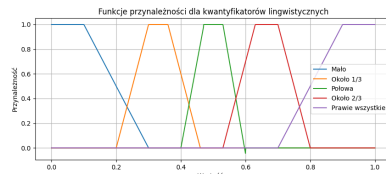
2.3.1 Kwantyfikatory rozmyte względne

$$\begin{aligned} \text{little}(\mathbf{X}) &= \begin{cases} 1 & \text{dla } x \leq 0.1 \\ \frac{0.3-x}{0.2} & \text{dla } 0.1 < x \leq 0.3 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{around_1_3}(\mathbf{X}) &= \begin{cases} 0 & \text{dla } x \leq 0.2 \\ 10(x-0.2) & \text{dla } 0.2 < x \leq 0.3 \\ 1 & \text{dla } 0.3 < x \leq 0.36 \\ 1-10(x-0.36) & \text{dla } 0.36 < x \leq 0.46 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{almost_half}(\mathbf{X}) &= \begin{cases} 0 & \text{dla } x \leq 0.4 \\ 14(x-0.4) & \text{dla } 0.4 < x \leq 0.47 \\ 1 & \text{dla } 0.47 < x \leq 0.53 \\ 1-15(x-0.53) & \text{dla } 0.53 < x \leq 0.6 \\ 0 & \text{w przeciwnym przypadku} \end{cases} & \text{around_2_3}(\mathbf{X}) &= \begin{cases} 0 & \text{dla } x \leq 0.53 \\ 10(x-0.53) & \text{dla } 0.53 < x \leq 0.63 \\ 1 & \text{dla } 0.63 < x \leq 0.7 \\ 1-10(x-0.7) & \text{dla } 0.7 < x \leq 0.8 \\ 0 & \text{w przeciwnym przypadku} \end{cases} \\ \text{almost_all}(\mathbf{X}) &= \begin{cases} 0 & \text{dla } x \leq 0.7 \\ \frac{x-0.7}{0.2} & \text{dla } 0.7 < x \leq 0.9 \\ 1 & \text{w przeciwnym przypadku} \end{cases} \end{aligned}$$

2.3.2 Kwantyfikatory rozmyte absolutne

$$\begin{aligned} \text{mniej niż 10\%}(\mathbf{X}) &= \begin{cases} 1 & \text{if } x \leq 0.1 \\ 0 & \text{otherwise} \end{cases} & \text{około 20\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.1}{0.1} & \text{if } 0.1 < x < 0.2 \\ 1 & \text{if } x = 0.2 \\ \frac{0.3-x}{0.1} & \text{if } 0.2 < x \leq 0.3 \\ 0 & \text{otherwise} \end{cases} \\ \text{około 30\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.2}{0.1} & \text{if } 0.2 < x \leq 0.3 \\ 1 & \text{if } x = 0.3 \\ \frac{0.4-x}{0.1} & \text{if } 0.3 < x \leq 0.4 \\ 0 & \text{otherwise} \end{cases} & \text{około 40\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.3}{0.1} & \text{if } 0.3 < x \leq 0.4 \\ 1 & \text{if } x = 0.4 \\ \frac{0.5-x}{0.1} & \text{if } 0.4 < x \leq 0.5 \\ 0 & \text{otherwise} \end{cases} \\ \text{około 50\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.4}{0.1} & \text{if } 0.4 < x \leq 0.5 \\ 1 & \text{if } x = 0.5 \\ \frac{0.6-x}{0.1} & \text{if } 0.5 < x \leq 0.6 \\ 0 & \text{otherwise} \end{cases} & \text{około 60\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.5}{0.1} & \text{if } 0.5 < x \leq 0.6 \\ 1 & \text{if } x = 0.6 \\ \frac{0.7-x}{0.1} & \text{if } 0.6 < x \leq 0.7 \\ 0 & \text{otherwise} \end{cases} \\ \text{około 70\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.6}{0.1} & \text{if } 0.6 < x \leq 0.7 \\ 1 & \text{if } x = 0.7 \\ \frac{0.8-x}{0.1} & \text{if } 0.7 < x \leq 0.8 \\ 0 & \text{otherwise} \end{cases} & \text{około 80\%}(\mathbf{X}) &= \begin{cases} \frac{x-0.7}{0.1} & \text{if } 0.7 < x \leq 0.8 \\ 1 & \text{if } x = 0.8 \\ \frac{0.9-x}{0.1} & \text{if } 0.8 < x \leq 0.9 \\ 0 & \text{otherwise} \end{cases} \\ \text{więcej niż 90\%}(\mathbf{X}) &= \begin{cases} 0 & \text{if } x \leq 0.9 \\ 1 & \text{otherwise} \end{cases} \end{aligned}$$

2.3.3 wykresy wybranych kwalfikatorów lingwistycznych

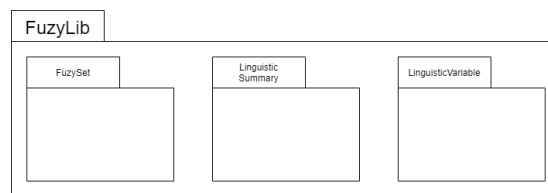


Opisy zostały przygotowane z uwzględnieniem literatury dotyczącej lingwistyki rozmytej oraz jej zastosowań.

3 Narzędzia obliczeniowe: wybór/implementacja. Diagram UML pakietu obliczeń rozmytych i generatora podsumowań. Instrukcja użytkownika

3.1 Opis pakietu obliczeń rozmytych - fuzzyLib

W celu zaspokojenia wymagań zadania utworzony został pakiet przeznaczony do obsługi logiki aplikacji. Poniżej zaprezentowany został diagram UML przedstawiający strukturę pakietu oraz jego pakietów pochodnych.



Rysunek 2: pakiet fuzzyLib wraz z jego podpakietami

Warto tutaj zaznaczyć że przedstawione w tej sekcji UML'e nie zawierają metod pomocniczych (setery i getery), które zaimplementowane są domyślnie.

3.1.1 pakiet FuzzySet

Pakiet definiujący interfejsy i implementacje funkcji przynależności, które są elementarnymi składnikami zbiorów rozmytych. Zawiera klasy i narzędzia służące do tworzenia, analizy i manipulacji tymi funkcjami. Dodatkowo klasa abstrakcyjna FuzzySet zapewnia operacje możliwe do przeprowadzenia na zbiorach rozmytych

FuzzySet: Klasa abstrakcyjna po której dziedziczą wszystkie pozostałe klasy reprezentujące zbiory rozmyte oraz ich funkcje przynależności, udostępnia operacje możliwe do wykonania na zbiorach rozmytych (dopełnienie, liczba kardynalna czy alfa cut).

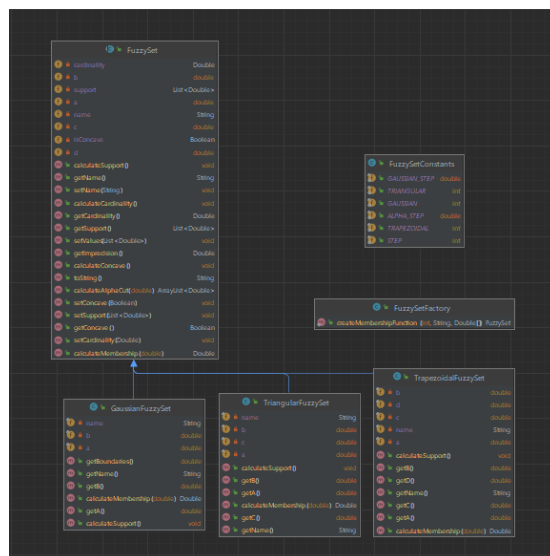
FuzzySetFactoryConsts: Zawiera stałe reprezentujące różne typy funkcji przynależności, takie jak TRIANGULAR, TRAPEZOIDAL i GAUSSIAN.

FuzzySetFactory: Fabryka do tworzenia różnych typów funkcji przynależności na podstawie podanego typu oraz parametrów. Wykorzystuje wyrażenie switch do dynamicznego tworzenia odpowiednich instancji klas na podstawie podanego typu.

TrapezoidalFuzzySet: Reprezentuje funkcję przynależności trapezoidalnej w ramach zbioru rozmytego.

TriangularFuzzySet: Reprezentuje funkcję przynależności trójkątną w ramach zbioru rozmytego.

GaussianFuzzySet: Reprezentuje funkcję przynależności Gaussa w ramach zbioru rozmytego.



Rysunek 3: Pakiet MembershipFunction

3.1.2 Pakiet LinguisticSummary

Pakiet zajmujący się generowaniem podsumowań lingwistycznych, czyli formułowania wniosków na podstawie zbiorów rozmytych. Zawiera klasy i narzędzia do tworzenia podsumowań lingwistycznych różnych stopni złożoności, w oparciu o zbiory rozmyte jako elementy wnioskowania.

Label: Klasa pomocnicza służąca do obsługi kwalifikatorów oraz kwantyfikatorów, przechowuje nazwę funkcji przynależności oraz zmienną lingwistyczną. Dzięki tem jesteśmy w stanie określić przestrzeń rozważań oraz posługiwać się odpowiednią funkcją podczas tworzenia podsumowań.

LinguisticSummary: Jest to klasa abstrakcyjna reprezentująca podsumowanie lingwistyczne. Posiada pola takie jak qualifier (kwalifikator), summarizer (sumatory), truthChecker (obiekt sprawdzający prawdziwość) i quantifier (kwantyfikator). Udostępnia metody do tworzenia podsumowań lingwistycznych. Uwaga, podmiot pobierany jest bezpośrednio z bazy danych wewnątrz "createLinguisticSummary".

TruthChecker: Jest to klasa służąca do sprawdzania stopnia prawdziwości podsumowań lingwistycznych. Jest zaimplementowana jako Singleton, aby zapewnić tylko jedną instancję tej klasy w aplikacji.

elementy *TwoSubjects : działają analogicznie co ich jedno podmiotowe odpowiedniki aczkolwiek obsługując dwa podmioty

LinguisticVariable	
private	fuzzySets List<FuzzySet>
private	name String
private	jsonConnector JsonConnector
public	toString() String
public	getFuzzySets() List<FuzzySet>
public	getName() String
public	getMembershipFunction(String) FuzzySet
public	toText(String) String
public	getJsonConnector() JsonConnector
public	deleteMembershipFunction(String) void
public	loadMembershipFunctions(String) void
public	appendMembershipFunction(FuzzySet) void
public	setFuzzySets(List<FuzzySet>) void

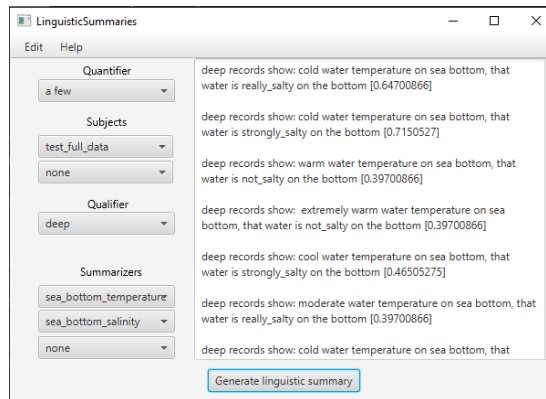
Rysunek 5: Pakiet FuzzySet

3.2 Interfejs użytkownika

Aplikacja posiada relatywnie intuicyjny interfejs użytkownika składający się z 2 paneli, pierwszy z nich służy do generowania podsumowań lingwistycznych, drugi to tak zwany interfejs użytkownika zaawansowanego pozwalający na tworzenie i dodawanie funkcji przynależności do danych zmiennych lingwistycznych. Cały front naszego programu realizowany jest przez pakiet "Gui"

3.2.1 Panel główny

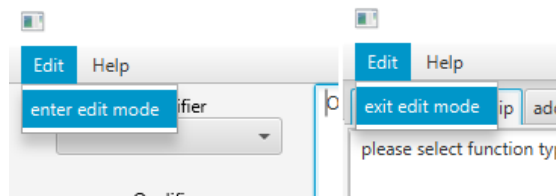
Panel główny aplikacji pozwala na wybranie z listy dostępnych obiektów kwantifikatora, kwalifikatora oraz do 3 sumatyzatorów, po zatwierdzeniu parametrów wynik wyświetlony zostanie po lewej stronie ekranu.



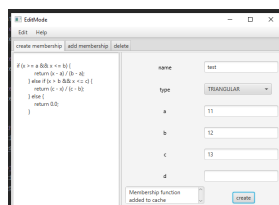
Rysunek 6: główny ekran aplikacji

3.2.2 Panel użytkownika zaawansowanego

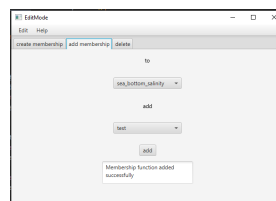
Wejście do panelu użytkownika zaawansowanego odbywa się poprzez panel Edit-
 > enter edit mode, w ten sam sposób można wspomniany panel opuścić Edit
 -> exit edit mode.



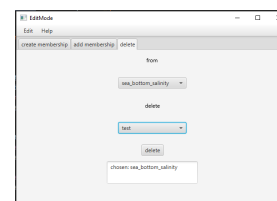
Rysunek 7: przełączanie się między panelem użytkownika zaawansowanego a panelem głównym



Rysunek 8: ekran tworzenia funkcji przynależności



Rysunek 9: ekran dodawania funkcji przynależności do zmiennej lingwistycznej



Rysunek 10: ekran usuwania funkcji przynależności ze zmiennej lingwistycznej

Utworzenie nowej funkcji przynależności odbywa się poprzez wybranie kształtu

funkcji (trójkątna, trapezoidalna oraz gausowska) oraz wprowadzenie odpowiednich parametrów. W celu dodania/usunięcia funkcji ze zmiennej lingwistycznej należy skorzystać z odpowiednich paneli.

3.3 JRE

Aby uruchomić naszą aplikację, użytkownik musi mieć zainstalowane środowisko uruchomieniowe Java (Java Runtime Environment, JRE). Aplikacja została przetestowana i działa z najnowszą wersją JRE LTS (Long-Term Support), co zapewnia stabilność i długoterminowe wsparcie.

Aplikacja do poprawnego działania wymaga dostępu do relacyjnej bazy danych. Stworzona została z myślą o użyciu PostgreSQL 13. Użytkownik musi zainstalować i skonfigurować bazę danych oraz dostarczyć odpowiednie dane dostępowe. Dodatkowo Aplikacja wykorzystuje biblioteki Java, które są zarządzane przez system budowania Maven. Użytkownik musi mieć zainstalowany Maven w wersji 3.6.3 lub nowszej.

4 Jednopodmiotowe podsumowania lingwistyczne. Miary jakości, podsumowanie optymalne

4.1 Wstęp do Eksperymentów

W ramach niniejszej sekcji sprawozdania przeprowadzono serię eksperymentów mających na celu analizę skuteczności różnych rodzajów podsumowań jednopodmiotowych. Eksperymenty te zostały zrealizowane poprzez opracowanie tabel i rankingów podsumowań dla danych atrybutów. W każdym eksperymencie szczegółowo opisano zastosowane miary jakości, a także dokonano oceny jakości podsumowania optymalnego. Należy również wspomnieć że eksperymenty wykonywane były na całej bazie danych to jest 210 tysiącach rekordów.

W trakcie wyznaczania miary jakości wygenerowanych podsumowań lingwistycznych wykorzystaliśmy wszystkie miary możliwe do wykorzystania dla danego przypadku zgodnie z materiałem źródłowym [1]. Dodatkowo w celu uzyskania wiarygodnych wyników wykorzystane zostały odpowiednie wagi nałożone na poszczególne miary jakości, ich rozkład zaprezentowany został w poniższych tabelach.

Tabela 1: Przyjęte wagi nałożone na miary jakości dla jednopodmiotowych podsumowań lingwistycznych typu pierwszego

miara jakości	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
przyjęta waga	1	1	1	1	1	1	1	1	-	-	-

Tabela 2: Przyjęte wagi nałożone na miary jakości dla jednopodmiotowych podsumowań lingwistycznych typu drugiego

miara jakości	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11
przyjęta waga	1	1	1	1	1	1	1	1	1	1	1

W celu uzyskania miary jakości z przedziału [0-1] wartość ta wyznaczona została poprzez zastosowanie odpowiedniej średniej.

4.2 Eksperyment dla posumowania jednopodmiotowego w formie pierwszej dla maksymalnej ilości sumaryzatorów

Do przeprowadzenia poniższego eksperymentu wykorzystane zostały następujące wartości sumaryzatorów, oraz kwantyfikatorów:

sumaryzatory : sea_surface_temperature, significant_wave_height, mixed_layer_depth.

kwantyfikator relatywny : about_one_third.

4.2.1 Wyniki eksperymentu

W poniższej tabeli w kolumnie zamieszczone zostało 5 najlepszych podsumowań znalezionych dla aplikacji uruchomionej z wybranymi parametrami zamieszczonymi powyżej, łączna ilość znalezionych podsumowań wynosi 80. Dodatkowo wyznaczone zostało porównanie różnicy między najlepszym znalezionym rozwiązaniem, nazwanym tutaj podsumowaniem optymalnym.

Podsumowanie	T	ΔT
about one third records show: moderate water temperature on sea surface, medium height waves, that sea is moderately shallow	0.55	0.0
about one third records show: moderate water temperature on sea surface, medium height waves, sea is almost on surface level depth	0.519	0.03
about one third records show: cold water temperature on sea surface, medium height waves, that sea is moderately shallow	0.511	0.039
about one third records show: cool water temperature on sea surface, medium height waves, that sea is moderately shallow	0.511	0.039
about one third records show: warm water temperature on sea surface, medium height waves, that sea is moderately shallow	0.511	0.039

Średnia wszystkich miar jakości dla uzyskanych podsumowań wyniosła: 0.413

4.3 Eksperyment dla posumowania jednopodmiotowego w formie drugiej przy wykorzystaniu maksymalnej dopuszczalnej ilości sumaryzatorów

Do przeprowadzenia poniższego eksperymentu wykorzystane zostały te same wartości sumaryzatorów, oraz kwantyfikatorów co w poprzednim eksperymencie, dodatkowo wykorzystany został następujący kwalifikator:

kwalifikator : east_of_Ireland z przestrzeni zmiennej lingwistycznej longitude.

4.3.1 Wyniki eksperymentu

Tak jak zostało to wykonane dla poprzedniej subsekcji poniżej zamieszczone zostało 5 najlepszych podsumowań znalezionych dla aplikacji uruchomionej z wybranymi parametrami zamieszczonymi powyżej. Dodatkowo wyznaczone zostało porównanie różnicy między najlepszym znalezionym rozwiązaniem, nazwanym tutaj podsumowaniem optymalnym.

Podsumowanie	T	ΔT
about one third records having that record was made east of Ireland show: warm water temperature on sea surface, moderately height waves, that sea is moderately shallow	0.635	0.0
about one third records having that record was made east of Ireland show: warm water temperature on sea surface, medium height waves, that sea is moderately shallow	0.611	0.025
about one third records having that record was made east of Ireland show: warm water temperature on sea surface, calm seas, that sea is moderately shallow	0.604	0.031
about one third records having that record was made east of Ireland show: warm water temperature on sea surface, height waves, that sea is moderately shallow	0.604	0.031
about one third records having that record was made east of Ireland show: warm water temperature on sea surface, moderately height waves, sea is almost on surface level depth	0.595	0.04

Średnia wszystkich miar jakości dla uzyskanych podsumowań wyniosła: 0.501

4.4 Wnioski

Podczas analizy skuteczności różnych rodzajów podsumowań jednopodmiotowych na całej bazie danych, zauważono, że istnieją pewne kombinacje sumaryzatorów i kwantyfikatorów, które skutkują podsumowaniami optymalnymi o wysokiej jakości.

Wniosek ten sugeruje, że dobór odpowiednich sumaryzatorów, kwantyfikatorów i ewentualnych kwalifikatorów może istotnie wpłynąć na skuteczność podsumowań lingwistycznych, umożliwiając generowanie optymalnych podsumowań

o wysokiej jakości.

Dodatkowo, podczas pracy nad wyliczaniem wartości poprawności podsumowania lingwistycznego, zauważono, że niektóre wartości miary T przyjmują skrajne wartości (tj. 0 i 1). W związku z tym, dobrym pomysłem byłoby nałożenie odpowiednich wag na te wartości T, co mogłoby prowadzić do bardziej zrównoważonych i realistycznych wyników. Nałożenie takich wag pozwoliłoby na uwzględnienie stopnia pewności podsumowań, a tym samym zwiększyłoby ich użyteczność i dokładność w praktycznych zastosowaniach.

5 Wielopodmiotowe podsumowania lingwistyczne i ich miary jakości

Na potrzeby tego eksperymentu zmuszeni zostaliśmy wyodrębnić jeden ze zbiorów rozmytych, po gruntownej analizie zdecydowaliśmy się wyodrębnić zmienną lingwistyczną "latitude" ze względu na jej mniej więcej równy podział pomiędzy wartościami. Dzięki tej operacji uzyskaliśmy 3 podmioty:

- records_from_north
- records_on_Ireland
- records_from_south

5.1 Wstęp do Eksperymentów

W tej sekcji sprawozdania przeprowadzono serię eksperymentów mających na celu analizę skuteczności różnych rodzajów podsumowań wielopodmiotowych. Nasza aplikacja automatycznie generuje wszystkie cztery rodzaje podsumowań, bez możliwości wyboru konkretnego typu. Eksperymenty te zostały zrealizowane poprzez opracowanie tabel i rankingów podsumowań dla danych atrybutów. W każdym eksperymencie szczegółowo opisano zastosowane miary jakości. Eksperymenty były wykonywane na całej bazie danych, obejmującej 210 tysięcy rekordów.

W celu zapewnienia rzetelności wyników, dane dla dwóch podmiotów były przetwarzane i analizowane w kontekście różnych lingwistycznych zmiennych i ich wartości. Proces generowania podsumowań obejmował:

Generowanie Podsumowań Typu 1 : Analiza porównawcza pomiędzy dwoma podmiotami bez dodatkowych kwalifikatorów, uwzględniająca wyłącznie lingwistyczne zmienne podsumowujące.

Generowanie Podsumowań Typu 2 : Porównanie pomiędzy podmiotami z uwzględnieniem kwalifikatora, który filtrował dane na podstawie określonego kryterium.

Generowanie Podsumowań Typu 3 : Porównanie z uwzględnieniem kwalifikatora w odniesieniu do jednego z podmiotów.

Generowanie Podsumowań Typu 4 : Ocena inkluzji zbiorów rozmytych typu 2, mająca na celu określenie relacji porządkującej pomiędzy danymi podmiotami.

5.1.1 Przyjęte wartości niezbędne do wykonania eksperymentu

podmioty : records_from_north, records_from_south

sumaryzatory : sea_surface_temperature, significant_wave_height, mixed_layer_depth.

kwantyfikator relatywny : about_one_third.

kwalifikator : east_of_Ireland z przestrzeni zmiennej lingwistycznej longitude.

5.1.2 Wyniki eksperymentu

Poniżej zamieszczone zostały najlepsze podsumowania dla każdego typu podsumowań lingwistycznych wygenerowanych dla dwóch podmiotów.

Podsumowanie	T
about one third test_south_data comparing to test_north_data show: moderate water temperature on sea surface, medium height waves, that sea is deep	0.5

Średnia wszystkich miar jakości dla uzyskanych podsumowań wyniosła: 0.458

Podsumowanie	T
about one third test_south_data comparing to test_north_data having that record was made east of Ireland show: moderate water temperature on sea surface, moderately height waves, that sea is moderately shallow	0.212

Średnia wszystkich miar jakości dla uzyskanych podsumowań wyniosła: 0.177

Podsumowanie	T
about one third test_south_data having that record was made east of Ireland comparing to test_north_data show: cool water temperature on sea surface, medium height waves, that sea is deep	0.788

Średnia wszystkich miar jakości dla uzyskanych podsumowań wyniosła: 0.722

Podsumowanie	T
more test_south_data than test_north_data show: warm water temperature on sea surface, medium height waves, that sea is moderately shallow	0.266

Średnia wszystkich miar jakości dla uzyskanych podsumowań wyniosła: 0.109

5.2 Wnioski

Ponadto, zaobserwowano, że metoda wielopodmiotowych podsumowań lingwistycznych jest szczególnie skuteczna w kontekście analizy porównawczej różnych grup danych. Ta metodologia dostarcza cennych informacji o relacjach między podmiotami, wskazując na potencjalne dominacje cech w poszczególnych grupach.

Ogólnie rzecz biorąc, wielopodmiotowe podsumowania lingwistyczne wykazały swoją skuteczność w generowaniu intuicyjnych i relatywnie wartościowych wniosków, co może mieć zastosowanie w różnych dziedzinach.

6 Dyskusja, wnioski

W ramach przeprowadzonych eksperymentów dokonano analizy skuteczności różnych rodzajów podsumowań jednopodmiotowych oraz wielopodmiotowych, z uwzględnieniem różnych miar jakości. Eksperymenty przeprowadzone na bazie danych zawierającej 210 tysięcy rekordów wykazały, że dobór odpowiednich sumaryzatorów, kwantyfikatorów oraz kwalifikatorów jest kluczowy dla uzyskania optymalnych podsumowań lingwistycznych. Wyniki eksperymentów sugerują, że istnieją konkretne kombinacje tych elementów, które prowadzą do generowania podsumowań o wysokiej jakości.

Jednopodmiotowe podsumowania lingwistyczne, szczególnie te generowane przy użyciu formy pierwszej, wykazały różnorodność w jakości podsumowań w zależności od zastosowanych sumaryzatorów. W eksperymencie z maksymalną ilością sumaryzatorów, najlepsze podsumowania miały miarę jakości T na poziomie 0.55, co wskazuje na umiarkowaną skuteczność w ujęciu ogólnym. Jednakże, uwzględnienie kwalifikatorów w formie drugiej podniosło wartość miar jakości, co sugeruje, że dodatkowe warunki mogą znacząco wpływać na precyzję podsumowań.

W kontekście wielopodmiotowych podsumowań lingwistycznych, zaobserwowano, że metodologia ta jest szczególnie przydatna w analizie porównawczej różnych grup danych. Przeprowadzone eksperymenty pokazały, że porównania między różnymi podmiotami (np. danych z różnych regionów geograficznych) mogą dostarczać cennych informacji na temat dominujących cech i ich wzajemnych relacji. Generowanie wielopodmiotowych podsumowań lingwistycznych wykazało swoją skuteczność w uzyskiwaniu intuicyjnych i relatywnie wartościowych wniosków, co może być korzystne w różnych zastosowaniach praktycznych.

Podsumowując, przeprowadzone badania wykazały, że:

- Dobór odpowiednich sumaryzatorów, kwantyfikatorów i kwalifikatorów ma kluczowe znaczenie dla jakości podsumowań lingwistycznych.
- Jednopodmiotowe podsumowania lingwistyczne mogą dostarczać wartościowych wniosków, jednak ich skuteczność zależy od użytych parametrów i warunków.
- Wielopodmiotowe podsumowania lingwistyczne są szczególnie przydatne w analizie porównawczej i mogą dostarczać cennych informacji o relacjach między różnymi grupami danych.
- Użycie odpowiednich wag dla miar jakości może prowadzić do bardziej zrównoważonych i realistycznych wyników, co zwiększa użyteczność podsumowań lingwistycznych w praktyce.
- W przyszłości warto rozważyć dalsze badania nad optymalizacją doboru parametrów dla podsumowań lingwistycznych, a także nad ich zastosowaniami w różnych dziedzinach, takich jak analiza danych, zarządzanie informacją czy systemy wspomagania decyzji.

Literatura

- [1] A. Niewiadomski, Zbiory rozmyte typu 2. Zastosowania w reprezentowaniu informacji. Seria „Problemy współczesnej informatyki” pod redakcją L. Rutkowskiego. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2019.
- [2] S. Zadrozny, Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych, EXIT, 2006, Warszawa
- [3] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2008.
- [4] Marine Institute Monthly Model Means:
<https://data.world/marineinstitute/b2348da6-873f-4ebb-a03c-5280ee2cf025>
- [5] Mechanizm falowania wód morskich Kryteria podziału i typy falowania morza