

Data oddania: _____

Ocena: _____

Hubert Gawłowski 224298
Kamil Kiszko-Zgierski 224328

Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

1. Cel

Celem zadania jest zaimplementowanie lingwistycznej agregacji, tj. przedstawienie danych liczbowych za pomocą sformułowań w języku pozornie naturalnym, czyli takim, w którym sformułowania mają z góry określony schemat typu: " Q obiektów są S_1 [T]" lub " Q obiektów będących S_2 są S_1 [T]" [3] (np. *Większość zawodników jest wysoka* [0.63] lub *Okolo 9000 zawodników mających idealny wpływ na drużynę, zdobywa bardzo dużo punktów* [0.82]; liczba w nawiasie oznacza stopień prawdziwości podanych zdań). Do tego celu posłużą nam zbiory rozmyte. Zbiory rozmyte to takie zbiory, w których funkcją charakterystyczną jest funkcja przynależności, która określa w jakim stopniu dany element przynależy do zbioru. Owa implementacja zostanie wykonana w technologii Java z graficznym interfejsem użytkownika oraz z wykorzystaniem systemu zarządzania bazą danych o nazwie MySQL, a badania zostaną przeprowadzone w oparciu o bazę danych zawierającą statystyki zawodników ligi koszykówki NBA od sezonu 1996/1997 do 2019/2020 [4].

2. Charakterystyka podsumowywanej bazy danych

Jako bazę danych w zadaniu wykorzystaliśmy bazę danych dotyczącą informacji o zawodnikach NBA (najbardziej prestiżowej lidze koszykarskiej na świecie) z lat 1996 - 2019. Cała baza danych składa się z 22 kolumn zawierających informacje o danych zawodnika oraz jego statystykach różnego rodzaju.

Baza danych ma charakter pliku csv i zawiera 11144 rekordy. Spośród kolumn wybraliśmy te, które zawierają wartości, naszym zdaniem, najlepiej nadające się do rozmycia. Atrybuty do rozmycia dobraliśmy tak, aby zgadzały się z definicją zmiennej lingwistycznej [1].

Ostatecznie spośród 22 kolumn wybraliśmy następujące:

- Kolumny zawierające informacje pozwalające zidentyfikować zawodnika (nie wykorzystywane w obliczeniach):
 1. nr identyfikacyjny zawodnika (index)
 2. imię i nazwisko zawodnika (player_name)
 3. skrót nazwy drużyny (team_abbreviation)
 4. kraj urodzenia zawodnika (country)
- Kolumny wykorzystywane w obliczeniach:
 1. wiek zawodnika (age) - atrybut przyjmuje wartości liczbowe całkowite od 18 do 44. Przekładając ten atrybut na język naturalny możemy użyć wartości: *junior*, *młody*, *w średnim wieku*, *doświadczony*, *stary*.
 2. wzrost zawodnika (player_height) - atrybut przyjmuje wartości liczbowe zmiennoprzecinkowe od 160.02 do 231.14. Przekładając ten atrybut na język naturalny możemy użyć wartości: "bardzo niski", "niski", "średniego wzrostu", "wysoki", "bardzo wysoki".
 3. waga zawodnika (player_weight) - atrybut przyjmuje wartości liczbowe zmiennoprzecinkowe od 60.33 do 163.29. Przekładając ten atrybut na język naturalny możemy użyć wartości: *bardzo lekki*, *lekki*, *o przeciętnej wadze*, *ciężki*, *"bardzo ciężki"*.
 4. kolejność wyboru w drafcie (draft_number) - oznacza, który w kolejności został wybrany zawodnik w organizowanym co roku tzw. drafcie, który polega na wybieraniu przez drużyny młodych zawodników na następny sezon. Atrybut przyjmuje wartości liczbowe całkowite od 1 do 60 oraz wartość *undrafted* - nie wybrany w drafcie. Wartość *undrafted* będzie traktowana jako wartość maksymalna. Przekładając ten atrybut na język naturalny możemy użyć wartości: *błyskawicznie*, *szybko*, *średnio*, *późno*, *na koniec*.
 5. rozegrane gry (gp) - liczba gier zawodnika w danym sezonie. Atrybut ten przyjmuje wartości liczbowe całkowite od 1 do 85. Przekładając ten atrybut na język naturalny możemy użyć wartości: *znikoma liczba*, *mało*, *średnio*, *dużo*, *maksymalnie*.
 6. zdobyte punkty (pts) - średnia liczba punktów zdobytych przez zawodnika w meczach w danym sezonie. Atrybut ten przyjmuje wartości liczbowe zmiennoprzecinkowe od 0.0 do 36.1. Przekładając ten atrybut na język naturalny możemy użyć wartości: *bardzo mało*, *mało*, *dostatecznie*, *dużo*, *bardzo dużo*.
 7. liczba zbiórek (reb) - średnia liczba zbiórek zawodnika na mecz - zbiórka jest to złapanie piłki przez zawodnika po nieudanym rzucie do kosza. Atrybut ten przyjmuje wartości liczbowe zmiennoprzecinkowe od 0.0 do 16.3. Przekładając ten atrybut na język naturalny możemy użyć wartości: *bardzo mało*, *mało*, *dostatecznie*, *dużo*, *bardzo dużo*.
 8. liczba asyst (ast) - średnia liczba asyst zawodnika na mecz - asysta jest to ostatnie podanie między zawodnikami tej samej drużyny, po któ-

rym zdobyty zostaje punkt. Atrybut ten przyjmuje wartości liczbowe zmiennoprzecinkowe od 0.0 do 11.7. Podobnie jak w atrybucie wyżej, przekładając ten atrybut na język naturalny możemy użyć wartości: *bardzo mało, mało, dostatecznie, dużo, bardzo dużo*.

9. wpływ na drużynę (*net_rating*) - jaki wpływ miał zawodnik na punkty drużyny, gdy znajdował się na parkiecie. Atrybut ten przyjmuje wartości liczbowe zmiennoprzecinkowe od -100.0 do 100.0 ze zdecydowaną większością wartości mieszczących się w przedziale $[-25.0, 25.0]$. Przekładając ten atrybut na język naturalny możemy użyć wartości: *fatalny, negatywny, neutralny, pozytywny, idealny*
10. skuteczność rzutów (*ts_pct*) - jak efektywnie zawodnik rzucał do kosza, czyli ile spośród jego rzutów kończyło się punktami. Atrybut ten przyjmuje wartości liczbowe zmiennoprzecinkowe od 0.0 do 1.0, gdzie 0.0 oznacza, że żaden rzut nie kończył się punktem, a 1.0 - każdy rzut zawodnika kończył się punktem. Najwięcej wartości znajduje się w przedziale: $[0.30, 0.67]$. Przekładając ten atrybut na język naturalny możemy użyć wartości: *fatalny, nieskuteczny, przeciętny, skuteczny, idealny*.
11. procent asyst (*ast_pct*) - procent punktów przy jakich asystował zawodnik, gdy znajdował się na boisku - atrybut ten mówi dużo o wpływie zawodnika na drużynę. Przyjmuje on wartości liczbowe zmiennoprzecinkowe z przedziału $[0.0, 1.0]$ ze zdecydowaną większością wartości w przedziale $[0.0, 0.40]$. Przekładając ten atrybut na język naturalny możemy użyć wartości: *fatalny, mały, przeciętny, duży, idealny*.

Oczywiście wartości takie jak: wzrost, waga, czy wiek będą odnosiły się jedynie do koszykarzy, ponieważ np. słowo *stary* w kontekście gracza koszykówki będzie związane z zupełnie innym wiekiem, niż słowo *stary* używane na co dzień w stosunku do opisu wieku ludzi.

W rankingu najbardziej dochodowych lig sportowych liga NBA zajmuje trzecie miejsce na świecie ¹ (wyżej w rankingu są tylko ligi NFL - hokej i MLB - baseball). W związku z dużym zainteresowaniem wokół niej zasadnym jest, aby dane liczbowe koszykarzy przedstawiać również za pomocą zmiennych lingwistycznych.

Po pierwsze, jako iż owa dyscyplina jest bardzo popularna, wielu kibiców jest również zainteresowanych statystykami koszykarzy. Dla przeciętnego widza dane liczbowe mogą być jednak mało zrozumiałe. Rozwiązaniem tego problemu wydaje się wprowadzenie zmiennych lingwistycznych w celu ułatwienia interpretacji danych, co wiąże się z lepszym przyswojeniem informacji przez odbiorcę, jak również zaoszczędzeniem czasu na próbie ich zrozumienia.

Po drugie, przedstawienie danych liczbowych w postaci zmiennych lingwistycznych zapewnia grupowanie rekordów. Dzięki temu, dużo łatwiejsze jest dokonanie filtracji zawodników, co skutkuje szybszym wyszukaniem graczy o podobnych parametrach fizycznych lub osiągniętych statystykach.

W końcu, popularność oraz statystyki przychodów klubów NBA powo-

¹ <https://globalsportmatters.com/business/2019/03/07/tv-is-biggest-driver-in-global-sport-league-revenue/>

dują zainteresowanie wśród potencjalnych sponsorów, którzy niekoniecznie muszą być związani z koszykówką. Dane przedstawione w formie zmiennych lingwistycznych znacznie pomogłyby w analizie statystyk oraz ocenie ryzyka związanego z inwestycją w dany klub.

#	Field	Schema	Table	Type
1	index	nba_players	nba_players	INT
2	player_name	nba_players	nba_players	VARCHAR
3	team_abbreviation	nba_players	nba_players	VARCHAR
4	country	nba_players	nba_players	VARCHAR
5	age	nba_players	nba_players	INT
6	player_height	nba_players	nba_players	DOUBLE
7	player_weight	nba_players	nba_players	DOUBLE
8	draft_number	nba_players	nba_players	VARCHAR
9	gp	nba_players	nba_players	INT
10	pts	nba_players	nba_players	DOUBLE
11	reb	nba_players	nba_players	DOUBLE
12	ast	nba_players	nba_players	DOUBLE
13	net_rating	nba_players	nba_players	DOUBLE
14	ts_pct	nba_players	nba_players	DOUBLE
15	ast_pct	nba_players	nba_players	DOUBLE

Rysunek 1. Nazwy kolumn (atrybutów) w systemie zarządzania bazami danych MySQL

3. Atrybuty i liczności obiektów wyrażone zmiennymi lingwistycznymi

3.1. Zmienne lingwistyczne

Definicja zmiennej lingwistycznej została już wprowadzona w sekcji 1. W bieżącej sekcji skupimy się dokładniej na tej definicji podając dla każdego z 11 atrybutów z sekcji 1 nazwę zmiennej lingwistycznej, zbiór terminów lingwistycznych oraz przestrzeń rozważań. Uważamy, że podając konkretny zbiór terminów lingwistycznych, jaki wykorzystamy, reguły gramatyczna oraz semantyczna będą widoczne i nie ma potrzeby podawania ich wprost. Wszak np. dla wzrostu nie podamy jako termin lingwistyczny słowa *zielony* lub *szybki* oraz wszystkie terminy we wszystkich zmiennych będą uporządkowane w sposób semantyczny (np. słowo *ciężki* nie wystąpi przed słowem *lekki*, a słowo *idealny* przed *fatalny*). Poniżej przedstawiamy wykorzystane zmienne lingwistyczne wraz ze wzorami oraz wykresami:

1. wiek zawodnika - przestrzeń rozważań: liczby naturalne z przedziału $[18, 44]$
— junior

$$\mu_{junior}(x) = \begin{cases} 1 & \text{dla } x \in [18, 20) \wedge x \in \mathbf{N} \\ -0.25x + 6 & \text{dla } x \in [20, 24] \wedge x \in \mathbf{N} \\ 0 & \text{dla } x \in (24, 44] \wedge x \in \mathbf{N} \end{cases} \quad (1)$$

— młody

$$\mu_{młody}(x) = \begin{cases} 0 & \text{dla } x \in ([18, 20) \cup (28; 44]) \wedge x \in \mathbf{N} \\ 0.25x - 5 & \text{dla } x \in [20, 24) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [24, 26) \wedge x \in \mathbf{N} \\ -0.5x + 14 & \text{dla } x \in [26, 28] \wedge x \in \mathbf{N} \end{cases} \quad (2)$$

— w średnim wieku

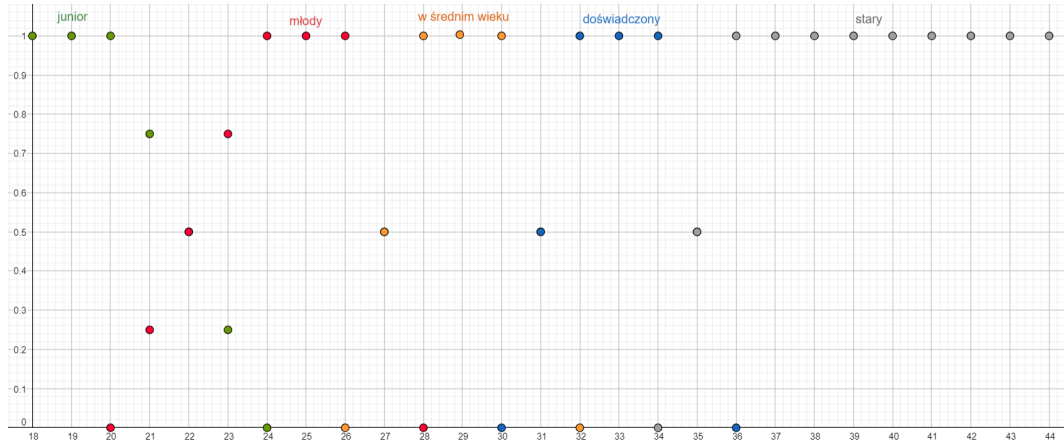
$$\mu_{wsrednimwieku}(x) = \begin{cases} 0 & \text{dla } x \in ([18, 26) \cup (32, 44]) \wedge x \in \mathbf{N} \\ 0.5x - 13 & \text{dla } x \in [26, 28) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [28, 30) \wedge x \in \mathbf{N} \\ -0.5x + 16 & \text{dla } x \in [30, 32] \wedge x \in \mathbf{N} \end{cases} \quad (3)$$

— doświadczony

$$\mu_{doswiadczony}(x) = \begin{cases} 0 & \text{dla } x \in ([18, 30) \cup (34, 44]) \wedge x \in \mathbf{N} \\ 0.5x - 15 & \text{dla } x \in [30, 32) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [32, 34) \wedge x \in \mathbf{N} \\ -0.5x + 18 & \text{dla } x \in [34, 36] \wedge x \in \mathbf{N} \end{cases} \quad (4)$$

— stary

$$\mu_{stary}(x) = \begin{cases} 0 & \text{dla } x \in [18, 34) \wedge x \in \mathbf{N} \\ 0.5x - 17 & \text{dla } x \in [34, 36) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [36, 44] \wedge x \in \mathbf{N} \end{cases} \quad (5)$$



Rysunek 2. Wykres funkcji przynależności dla zmiennej lingwistycznej *wiek zawodnika*.

2. wzrost zawodnika - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[160.03, 231.14]$

— niski

$$\mu_{niski}(x) = \begin{cases} 1 & \text{dla } x \in [160.03, 185.0) \\ -0.1x + 19.5 & \text{dla } x \in [185.0, 195.0] \\ 0 & \text{dla } x \in (195.0, 231.14] \end{cases} \quad (6)$$

— średniego wzrostu

$$\mu_{\text{średniogwzrostu}}(x) = \begin{cases} 0 & \text{dla } x \in (160.03, 185.0] \cup (205.0, 231.14] \\ 0.1x - 18.5 & \text{dla } x \in [185.0, 195.0] \\ -0.1x + 20.5 & \text{dla } x \in [195.0, 205.0] \end{cases} \quad (7)$$

— wysoki

$$\mu_{\text{wysoki}}(x) = \begin{cases} 0 & \text{dla } x \in (160.03, 200.0] \\ 0.1x - 20 & \text{dla } x \in [200.0, 210.0] \\ 1 & \text{dla } x \in [210.0, 231.14] \end{cases} \quad (8)$$

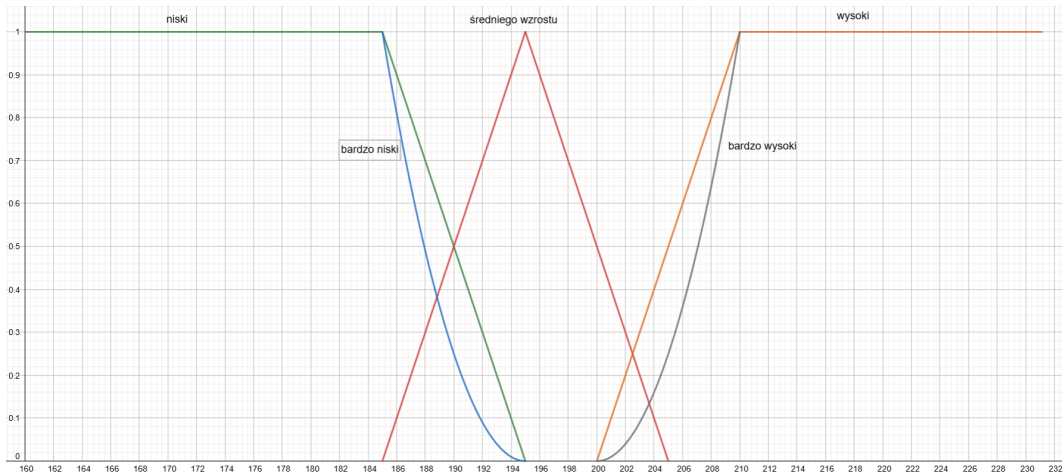
— bardzo niski

$$\mu_{\text{bardzoniski}}(x) = \mu_{\text{niski}}(x)^2 \quad (9)$$

— bardzo wysoki

$$\mu_{\text{bardzowysoki}}(x) = \mu_{\text{wysoki}}(x)^2 \quad (10)$$

Oczywiście, zgodnie z regułą semantyczną termin *bardzo niski* występuje przed terminem *niski* w zmiennej lingwistycznej. Powyżej został on wymieniony później, ponieważ definiując funkcję przynależności dla terminu *bardzo niski* korzystamy z funkcji przynależności dla terminu *niski* (konkretnie podnosimy ją do kwadratu).



Rysunek 3. Wykres funkcji przynależności dla zmiennej lingwistycznej *wzrost zawodnika*.

3. waga zawodnika - przestrzeń rozważań: Liczby rzeczywiste z przedziału: $[60.33, 163.29]$

— lekki

$$\mu_{\text{lekki}}(x) = \begin{cases} 1 & \text{dla } x \in [60.33, 85.0] \\ -0.1x + 9.5 & \text{dla } x \in [85.0, 95.0] \\ 0 & \text{dla } x \in (95.0, 163.29] \end{cases} \quad (11)$$

— o przeciętnej wadze

$$\mu_{\text{oprzecietnejwadze}}(x) = \begin{cases} 0 & \text{dla } x \in [60.33, 90.0] \cup (110.0, 163.29] \\ 0.1x - 9 & \text{dla } x \in [90.0, 100.0] \\ -0.1x + 11 & \text{dla } x \in [100.0, 110.0] \end{cases} \quad (12)$$

— ciężki

$$\mu_{ciężki}(x) = \begin{cases} 0 & \text{dla } x \in [60.33, 105.0) \\ 0.1x - 10.5 & \text{dla } x \in [105.0, 115.0) \\ 1 & \text{dla } x \in [115.0, 163.29] \end{cases} \quad (13)$$

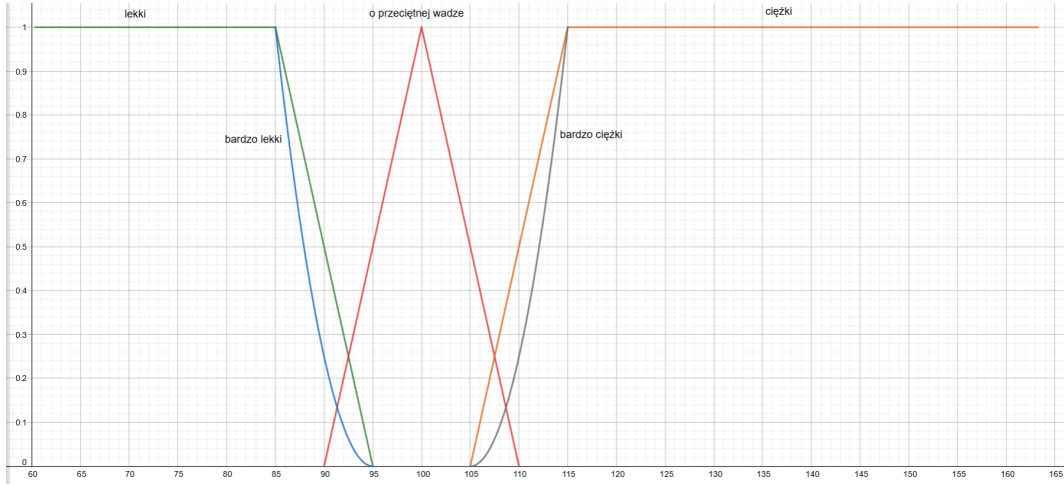
— bardzo lekki

$$\mu_{bardzolekki}(x) = \mu_{lekki}(x)^2 \quad (14)$$

— bardzo ciężki

$$\mu_{bardzociężki}(x) = \mu_{ciężki}(x)^2 \quad (15)$$

Oczywiście, zgodnie z regułą semantyczną termin *bardzo lekki* występuje przed terminem *lekki* w zmiennej lingwistycznej. Powyżej został on wymieniony później, ponieważ definiując funkcję przynależności dla terminu *bardzo lekki* korzystamy z funkcji przynależności dla terminu *lekki* (konkretnie podnosimy ją do kwadratu).



Rysunek 4. Wykres funkcji przynależności dla zmiennej lingwistycznej *waga zawodnika*.

4. kolejność wyboru w drafcie - przestrzeń rozważań: liczby naturalne z przedziału: $[0, 60]$

— błyskawicznie

$$\mu_{błyskawicznie}(x) = \begin{cases} 1 & \text{dla } x \in [0, 6) \wedge x \in \mathbf{N} \\ -0.125x + 1.75 & \text{dla } x \in [6, 14] \wedge x \in \mathbf{N} \\ 0 & \text{dla } x \in (14, 60] \wedge x \in \mathbf{N} \end{cases} \quad (16)$$

— szybko

$$\mu_{szybko}(x) = \begin{cases} 0 & \text{dla } x \in ([0, 10) \cup (30; 60]) \wedge x \in \mathbf{N} \\ 0.25x - 2.5 & \text{dla } x \in [10, 14] \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [14, 22] \wedge x \in \mathbf{N} \\ -0.125x + 3.75 & \text{dla } x \in [22, 30] \wedge x \in \mathbf{N} \end{cases} \quad (17)$$

— średnio

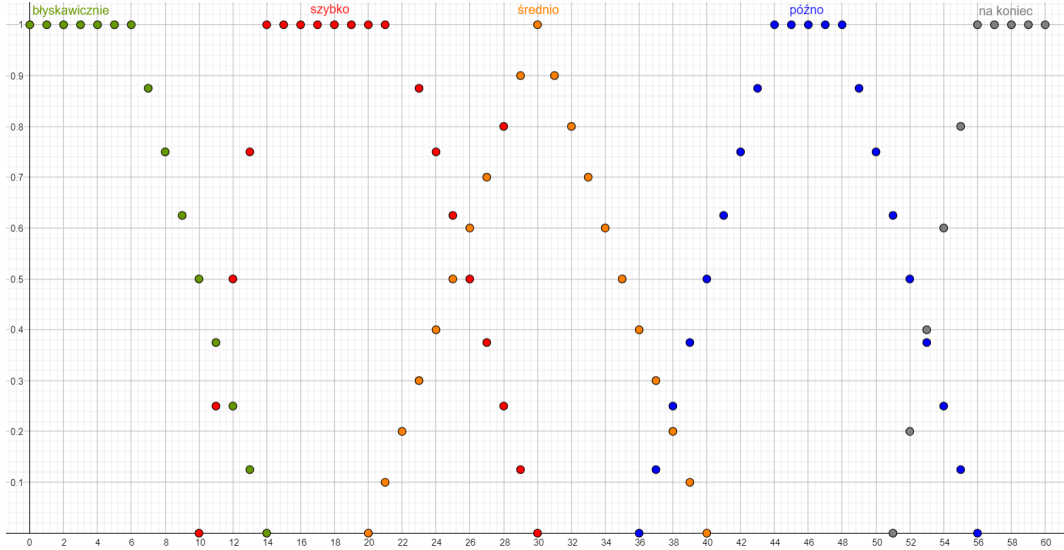
$$\mu_{\text{średnio-draft}}(x) = \begin{cases} 0 & \text{dla } x \in ([0, 20) \cup (40, 60]) \wedge x \in \mathbf{N} \\ 0.1x - 2 & \text{dla } x \in [20, 30) \wedge x \in \mathbf{N} \\ -0.1x + 4 & \text{dla } x \in [30, 40] \wedge x \in \mathbf{N} \end{cases} \quad (18)$$

— późno

$$\mu_{\text{późno}}(x) = \begin{cases} 0 & \text{dla } x \in ([0, 36) \cup (56, 60]) \wedge x \in \mathbf{N} \\ 0.125x - 4.5 & \text{dla } x \in [36, 44) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [44, 48) \wedge x \in \mathbf{N} \\ -0.125x + 7 & \text{dla } x \in [48, 56] \wedge x \in \mathbf{N} \end{cases} \quad (19)$$

— na koniec

$$\mu_{\text{na koniec}}(x) = \begin{cases} 0 & \text{dla } x \in [0, 51) \wedge x \in \mathbf{N} \\ 0.2x - 10.2 & \text{dla } x \in [51, 56) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [56, 60] \wedge x \in \mathbf{N} \end{cases} \quad (20)$$



Rysunek 5. Wykres funkcji przynależności dla zmiennej lingwistycznej *kolejność wyboru w drafcie*.

5. rozegrane gry w sezonie - przestrzeń rozważań: liczby naturalne z przedziału: $[1, 85]$

— znikoma liczba

$$\mu_{\text{znikomaliczba}}(x) = \begin{cases} 1 & \text{dla } x \in [1, 10) \wedge x \in \mathbf{N} \\ -0.0625x + 1.625 & \text{dla } x \in [10, 26) \wedge x \in \mathbf{N} \\ 0 & \text{dla } x \in [26, 85] \wedge x \in \mathbf{N} \end{cases} \quad (21)$$

— mało

$$\mu_{\text{mało-gier}}(x) = \begin{cases} 0 & \text{dla } x \in ([1, 20) \cup (40; 85]) \wedge x \in \mathbf{N} \\ 0.1x - 2 & \text{dla } x \in [20, 30) \wedge x \in \mathbf{N} \\ -0.1x + 4 & \text{dla } x \in [30, 40] \wedge x \in \mathbf{N} \end{cases} \quad (22)$$

— średnio

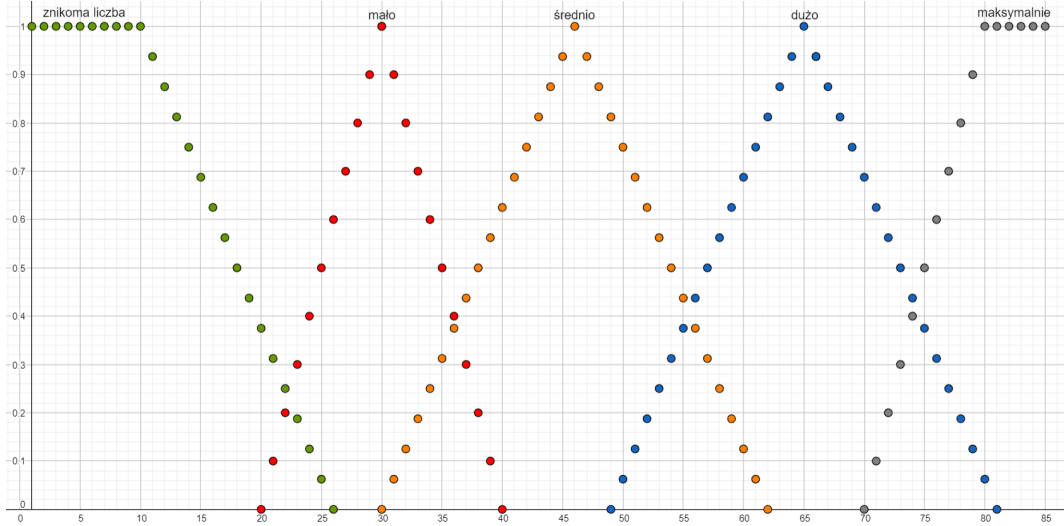
$$\mu_{\text{średnio-gier}}(x) = \begin{cases} 0 & \text{dla } x \in ([1, 30) \cup (62, 85]) \wedge x \in \mathbf{N} \\ 0.0625x - 1.875 & \text{dla } x \in [30, 46) \wedge x \in \mathbf{N} \\ -0.0625x + 3.875 & \text{dla } x \in [46, 62] \wedge x \in \mathbf{N} \end{cases} \quad (23)$$

— dużo

$$\mu_{\text{duzo-gier}}(x) = \begin{cases} 0 & \text{dla } x \in ([1, 49) \cup (81, 85]) \wedge x \in \mathbf{N} \\ 0.0625x - 3.0625 & \text{dla } x \in [49, 65) \wedge x \in \mathbf{N} \\ -0.0625x + 5.0625 & \text{dla } x \in [65, 81] \wedge x \in \mathbf{N} \end{cases} \quad (24)$$

— maksymalnie

$$\mu_{\text{maksymalnie}}(x) = \begin{cases} 0 & \text{dla } x \in [1, 70) \wedge x \in \mathbf{N} \\ 0.1x - 7 & \text{dla } x \in [70, 80) \wedge x \in \mathbf{N} \\ 1 & \text{dla } x \in [80, 85] \wedge x \in \mathbf{N} \end{cases} \quad (25)$$



Rysunek 6. Wykres funkcji przynależności dla zmiennej lingwistycznej *rozegrane gry*.

6. średnia liczba zdobytych punktów na mecz - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[0.0, 36.1]$

— mało

$$\mu_{\text{malo-punktow}}(x) = \begin{cases} 1 & \text{dla } x \in [0.0, 1.0) \\ -0.1x + 1.1 & \text{dla } x \in [1.0, 11.0] \\ 0 & \text{dla } x \in (11.0, 36.1] \end{cases} \quad (26)$$

— dostatecznie

$$\mu_{\text{dostatecznie-punktow}}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 1.0) \cup (16.0, 36.0] \\ 0.2x - 0.2 & \text{dla } x \in [1.0, 6.0) \\ 1 & \text{dla } x \in [6.0, 10.0) \\ -0.125x + 2.25 & \text{dla } x \in [10.0, 16.0] \end{cases} \quad (27)$$

— dużo

$$\mu_{duzo_punktow}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 11.0) \\ 0.125x - 1.375 & \text{dla } x \in [11.0, 19.0) \\ 1 & \text{dla } x \in [19.0, 36.1] \end{cases} \quad (28)$$

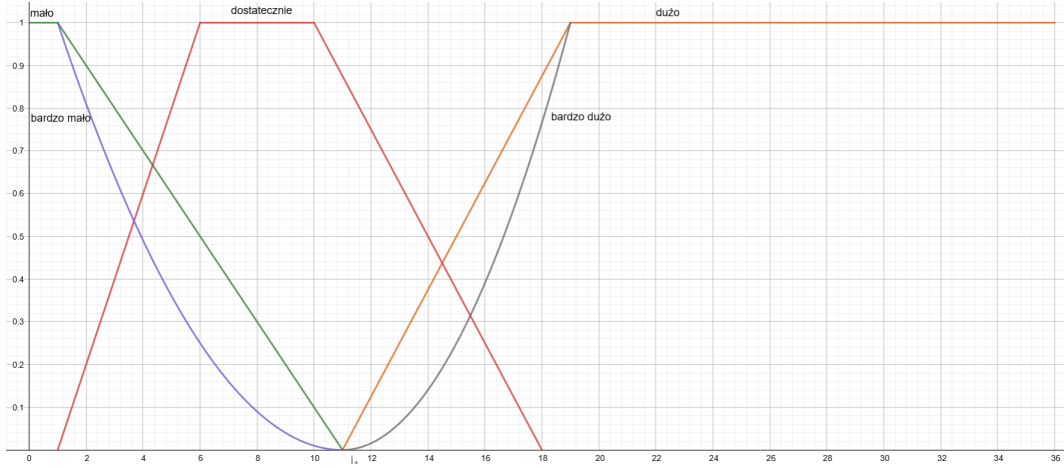
— bardzo mało

$$\mu_{bardzomalo_punktow}(x) = \mu_{malo}(x)^2 \quad (29)$$

— bardzo dużo

$$\mu_{bardzoduzo_punktow}(x) = \mu_{duzo}(x)^2 \quad (30)$$

Oczywiście, zgodnie z regułą semantyczną termin *bardzo mało* występuje przed terminem *mało* w zmiennej lingwistycznej. Powyżej został on wymieniony później, ponieważ definiując funkcję przynależności dla terminu *bardzo mało* korzystamy z funkcji przynależności dla terminu *mało* (konkretnie podnosimy ją do kwadratu).



Rysunek 7. Wykres funkcji przynależności dla zmiennej lingwistycznej *zdołyte punkty*.

7. średnia liczba zbiórek na mecz - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[0.0, 16.3]$

— mało

$$\mu_{malo_zbiorek}(x) = \begin{cases} 1 & \text{dla } x \in [0.0, 1.0) \\ -0.25x + 1.25 & \text{dla } x \in [1.0, 5.0] \\ 0 & \text{dla } x \in (5.0, 16.3] \end{cases} \quad (31)$$

— dostatecznie

$$\mu_{dostatecznie_zbiorek}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 2.0) \cup (6.0, 16.3] \\ 0.5x - 1 & \text{dla } x \in [2.0, 4.0) \\ 1 & \text{dla } x \in [4.0, 5.0) \\ -x + 6 & \text{dla } x \in [5.0, 6.0] \end{cases} \quad (32)$$

— dużo

$$\mu_{duzo_zbiorek}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 4.0) \\ 0.4x - 1.6 & \text{dla } x \in [4.0, 6.5) \\ 1 & \text{dla } x \in [6.5, 16.3] \end{cases} \quad (33)$$

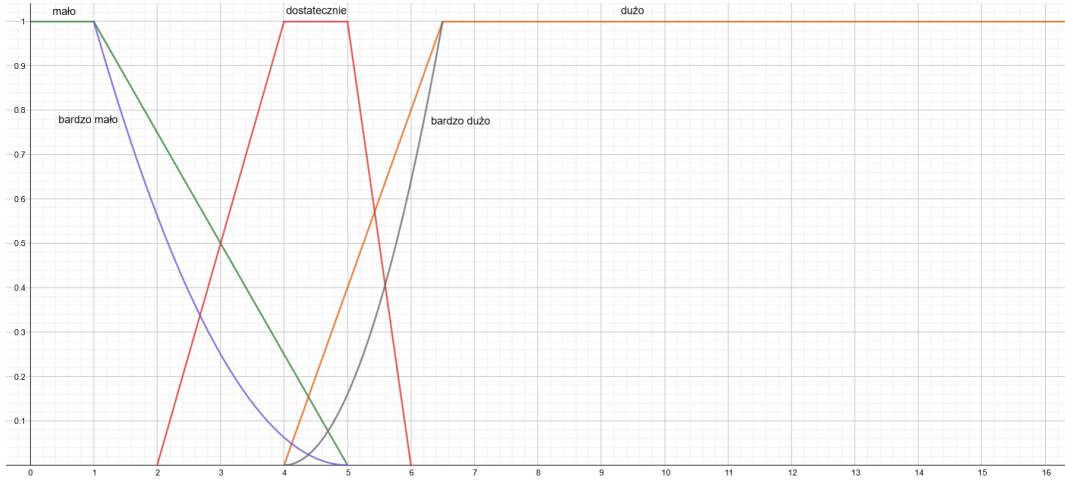
— bardzo mało

$$\mu_{bardzomalo_zbiorek}(x) = \mu_{malo}(x)^2 \quad (34)$$

— bardzo dużo

$$\mu_{bardzoduzo_zbiorek}(x) = \mu_{duzo}(x)^2 \quad (35)$$

Oczywiście, zgodnie z regułą semantyczną termin *bardzo mało* występuje przed terminem *mało* w zmiennej lingwistycznej. Powyżej został on wymieniony później, ponieważ definiując funkcję przynależności dla terminu *bardzo mało* korzystamy z funkcji przynależności dla terminu *mało* (konkretnie podnosimy ją do kwadratu).



Rysunek 8. Wykres funkcji przynależności dla zmiennej lingwistycznej *liczba zbiórek*.

8. średnia liczba asyst na mecz - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[0.0, 11.7]$

— mało

$$\mu_{malo_asyst}(x) = \begin{cases} 1 & \text{dla } x \in [0.0, 0.5) \\ -x + 1.5 & \text{dla } x \in [0.5, 1.5] \\ 0 & \text{dla } x \in (1.5, 11.7] \end{cases} \quad (36)$$

— dostatecznie

$$\mu_{dostatecznie_asyst}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 0.5) \cup (3.0, 11.7] \\ x - 0.5 & \text{dla } x \in [0.5, 1.5] \\ 1 & \text{dla } x \in [1.5, 2.0] \\ -x + 3 & \text{dla } x \in [2.0, 3.0] \end{cases} \quad (37)$$

— dużo

$$\mu_{duzo_asyst}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 2.0) \\ 0.625x - 1.25 & \text{dla } x \in [2.0, 3.6] \\ 1 & \text{dla } x \in [3.6, 11.7] \end{cases} \quad (38)$$

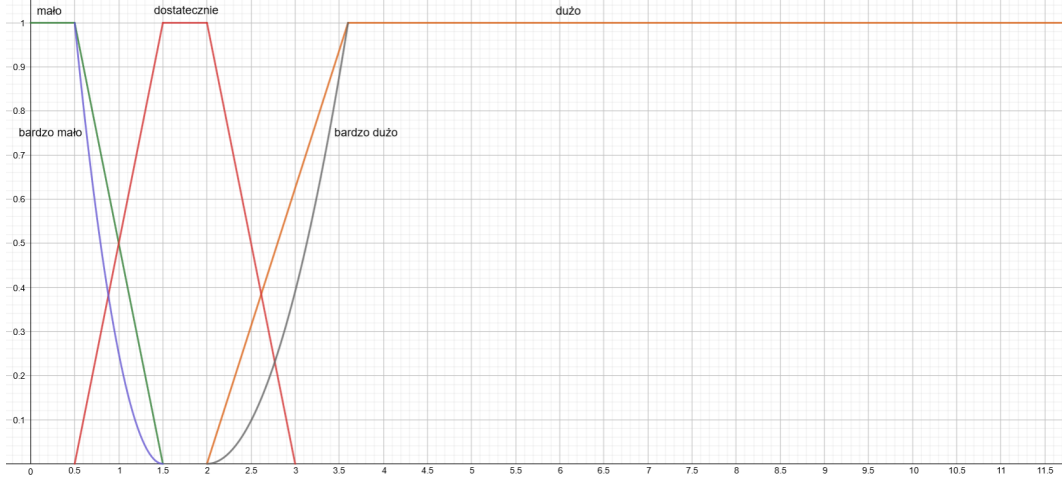
— bardzo mało

$$\mu_{bardzomalo_asyst}(x) = \mu_{malo_asyst}(x)^2 \quad (39)$$

— bardzo dużo

$$\mu_{bardzoduzo_asyst}(x) = \mu_{duzo_asyst}(x)^2 \quad (40)$$

Oczywiście, zgodnie z regułą semantyczną termin *bardzo mało* występuje przed terminem *mało* w zmiennej lingwistycznej. Powyżej został on wymieniony później, ponieważ definiując funkcję przynależności dla terminu *bardzo mało* korzystamy z funkcji przynależności dla terminu *mało* (konkretnie podnosimy ją do kwadratu).



Rysunek 9. Wykres funkcji przynależności dla zmiennej lingwistycznej *liczba asyst.*

9. wpływ na drużynę - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[-100.0, 100.0]$

— fatalny

$$\mu_{fatalny_wpływ}(x) = \begin{cases} 1 & \text{dla } x \in [-100.0, -15.0) \\ -0.1x - 0.5 & \text{dla } x \in [-15.0, -5.0] \\ 0 & \text{dla } x \in (-5.0, 100.0] \end{cases} \quad (41)$$

— negatywny

$$\mu_{negatywny_wpływ}(x) = \begin{cases} 0 & \text{dla } x \in [-100.0, -15.0) \cup (0.0, 100.0] \\ 0.125x + 1.875 & \text{dla } x \in [-15.0, -7.0] \\ 1 & \text{dla } x \in [-7.0, -2.5] \\ -0.4x & \text{dla } x \in [-2.5, 0.0] \end{cases} \quad (42)$$

— neutralny

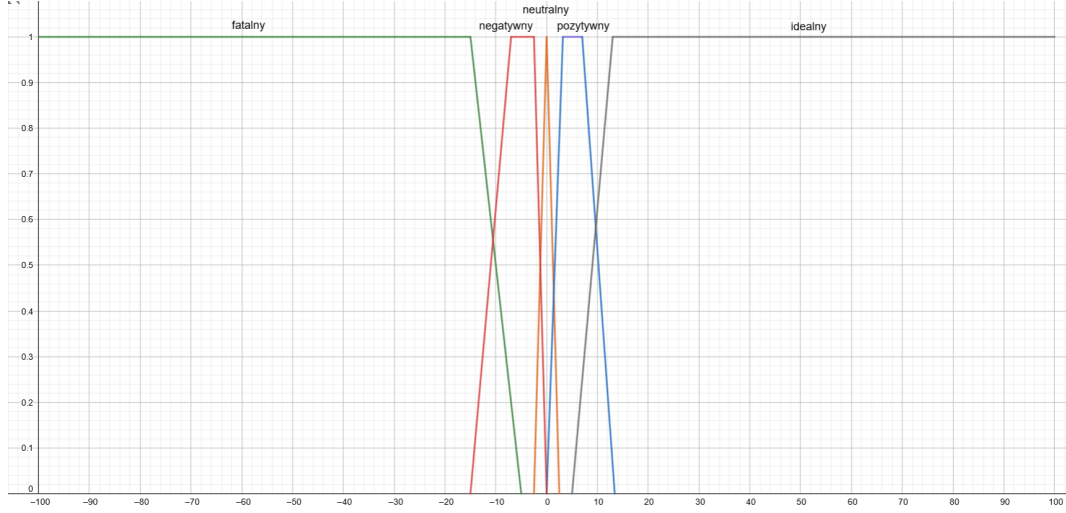
$$\mu_{neutralny_wpływ}(x) = \begin{cases} 0 & \text{dla } x \in [-100.0, -2.5) \cup (2.5, 100.0] \\ 0.4x + 1 & \text{dla } x \in [-2.5, 0.0] \\ -0.4x + 1 & \text{dla } x \in [0.0, 2.5] \end{cases} \quad (43)$$

— pozytywny

$$\mu_{pozytywny_wpływ}(x) = \begin{cases} 0 & \text{dla } x \in [-100.0, 0.0) \cup (13.4, 100.0] \\ 0.3125x & \text{dla } x \in [0.0, 3.2] \\ 1 & \text{dla } x \in [3.2, 7.0] \\ -0.15625x + 2.09375 & \text{dla } x \in [7.0, 13.4] \end{cases} \quad (44)$$

— idealny

$$\mu_{idealny_wplyw}(x) = \begin{cases} 0 & \text{dla } x \in [-100.0, 5.0) \\ 0.125x - 0.625 & \text{dla } x \in [5.0, 13.0) \\ 1 & \text{dla } x \in [13.0, 100.0] \end{cases} \quad (45)$$



Rysunek 10. Wykres funkcji przynależności dla zmiennej lingwistycznej *wpływ na drużynę*.

10. skuteczność rzutów - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[0.0, 1.0]$

— fatalny

$$\mu_{fatalny_skuteczność}(x) = \begin{cases} 1 & \text{dla } x \in [0.0, 0.35) \\ -10x + 4.5 & \text{dla } x \in [0.35, 0.45] \\ 0 & \text{dla } x \in (0.45, 1.0] \end{cases} \quad (46)$$

— nieskuteczny

$$\mu_{nieskuteczny}(x) = e^{-\frac{(x-0.4)^2}{0.003}} \text{ dla } x \in [0.0, 1.0] \quad (47)$$

— przeciętny

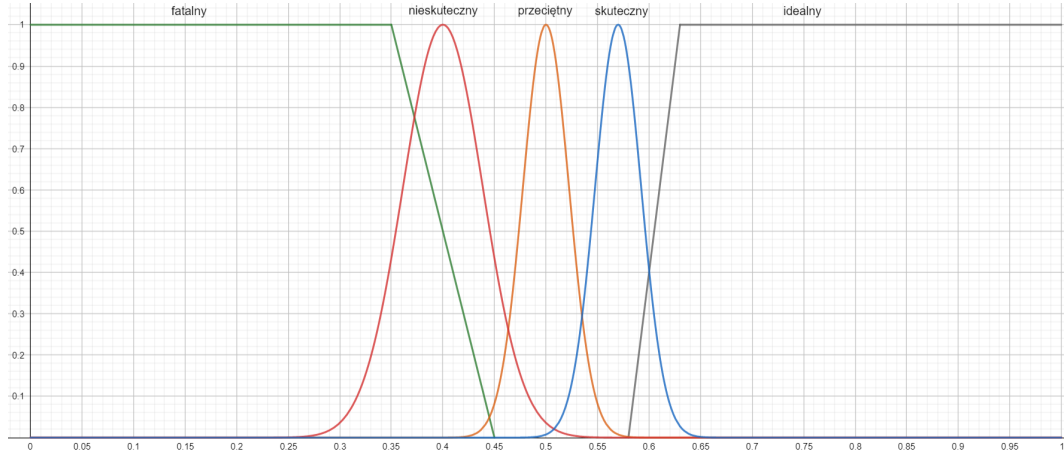
$$\mu_{przecietny_skuteczność}(x) = e^{-\frac{(x-0.5)^2}{0.001}} \text{ dla } x \in [0.0, 1.0] \quad (48)$$

— skuteczny

$$\mu_{skuteczny}(x) = e^{-\frac{(x-0.57)^2}{0.001}} \text{ dla } x \in [0.0, 1.0] \quad (49)$$

— idealny

$$\mu_{idealny_skuteczność}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 0.58) \\ 20x - 11.6 & \text{dla } x \in [0.58, 0.63) \\ 1 & \text{dla } x \in [0.63, 1.0] \end{cases} \quad (50)$$



Rysunek 11. Wykres funkcji przynależności dla zmiennej lingwistycznej *skuteczność rzutów*.

11. procent asyst - przestrzeń rozważań: liczby rzeczywiste z przedziału: $[0.0, 1.0]$
— fatalny

$$\mu_{fatalny_asyst}(x) = \begin{cases} 1 & \text{dla } x \in [0.0, 0.01) \\ -25x + 1.25 & \text{dla } x \in [0.01, 0.05] \\ 0 & \text{dla } x \in (0.05, 1.0] \end{cases} \quad (51)$$

— mały

$$\mu_{mały_asyst}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 0.01) \cup (0.06; 1] \\ 25x - 0.25 & \text{dla } x \in [0.01, 0.05] \\ 1 & \text{dla } x \in [0.05, 0.06] \\ -25x + 2.5 & \text{dla } x \in [0.06, 0.10] \end{cases} \quad (52)$$

— przeciętny

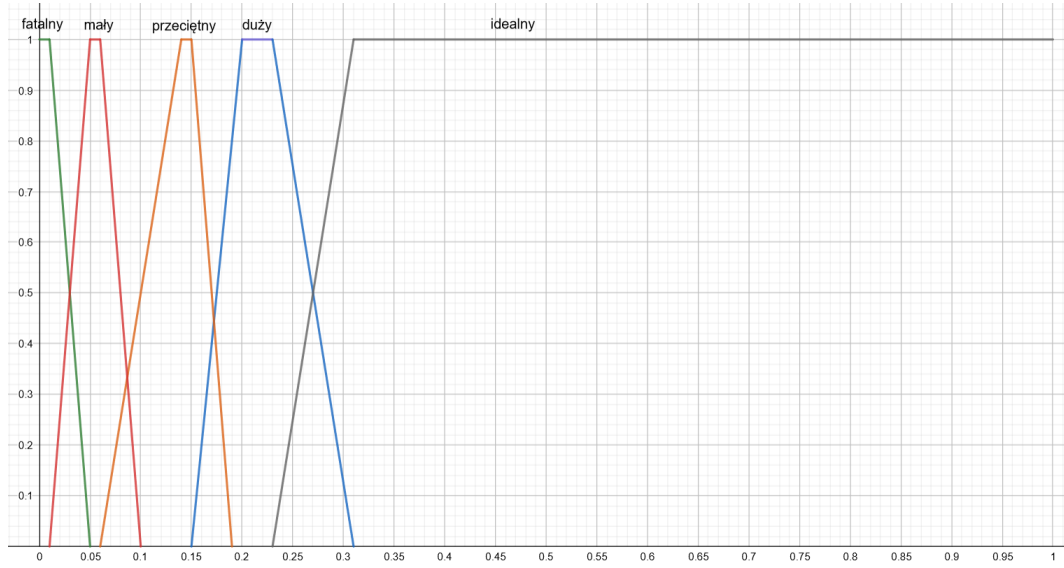
$$\mu_{przecietny_asyst}(x) = \begin{cases} 0 & \text{dla } x \in [0; 0.06) \cup (0.19; 1] \\ 12.5x - 0.75 & \text{dla } x \in [0.06; 0.14] \\ 1 & \text{dla } x \in [0.14; 0.15] \\ -25x + 4.75 & \text{dla } x \in [0.15; 0.19] \end{cases} \quad (53)$$

— duży

$$\mu_{duży_asyst}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 0.15) \cup (0.31, 1.0] \\ 20x - 3 & \text{dla } x \in [0.15, 0.20] \\ 1 & \text{dla } x \in [0.20, 0.23] \\ -12.5x + 3.875 & \text{dla } x \in [0.23, 0.31] \end{cases} \quad (54)$$

— idealny

$$\mu_{idealny_asyst}(x) = \begin{cases} 0 & \text{dla } x \in [0.0, 0.23) \\ 12.5x - 2.875 & \text{dla } x \in [0.23, 0.31] \\ 1 & \text{dla } x \in [0.31, 1.0] \end{cases} \quad (55)$$



Rysunek 12. Wykres funkcji przynależności dla zmiennej lingwistycznej *procent asyst.*

3.2. Kwantyfikatory lingwistyczne

W programie zostały wykorzystane 2 typy kwantyfikatorów: absolutne oraz względne.

Kwantyfikatory absolutne wraz ze wzorami oraz wykresem prezentują się następująco:

— Więcej niż 0

$$\mu_{wiecej_niz_0}(x) = \begin{cases} -0.000667x + 1 & \text{dla } x \in [0, 1500] \\ 0 & \text{dla } x \in (1500, 11144] \end{cases} \quad (56)$$

— Około 2000

$$\mu_{okolo_2000}(x) = \begin{cases} 0 & \text{dla } x \in [0; 500) \\ 0.000667x - 0.333 & \text{dla } x \in [500; 2000) \\ -0.000667x + 2.333 & \text{dla } x \in [2000; 3500) \\ 0 & \text{dla } x \in [3500, 11144] \end{cases} \quad (57)$$

— Około 4000

$$\mu_{okolo_4000}(x) = \begin{cases} 0 & \text{dla } x \in [0, 2500) \\ 0.000667x - 1.667 & \text{dla } x \in [2500, 4000) \\ -0.000667x + 3.667 & \text{dla } x \in [4000, 5500) \\ 0 & \text{dla } x \in [5500, 11144] \end{cases} \quad (58)$$

— Około 6000

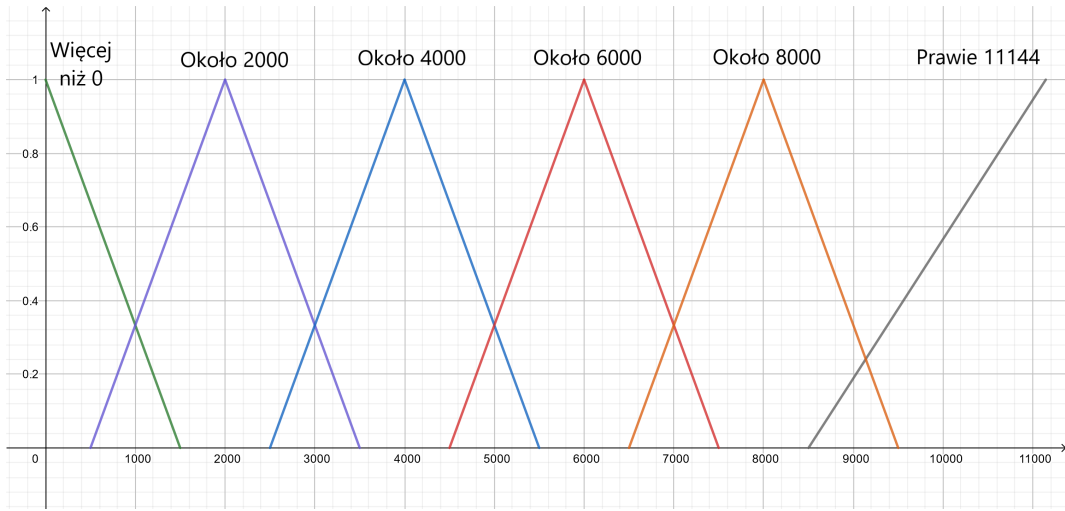
$$\mu_{okolo_6000}(x) = \begin{cases} 0 & \text{dla } x \in [0, 4500) \\ 0.000667x - 3 & \text{dla } x \in [4500, 6000) \\ 0.000667x + 5 & \text{dla } x \in [6000, 7500) \\ 0 & \text{dla } x \in [7500, 11144] \end{cases} \quad (59)$$

— Około 8000

$$\mu_{okolo_8000}(x) = \begin{cases} 0 & \text{dla } x \in [0, 6500) \\ 0.000667x - 4.333 & \text{dla } x \in [6500, 8000) \\ -0.000667x + 6.333 & \text{dla } x \in [8000, 9500) \\ 0 & \text{dla } x \in [9000, 11144] \end{cases} \quad (60)$$

— Prawie 11144

$$\mu_{prawie_11144}(x) = \begin{cases} 0 & \text{dla } x \in [0, 8500) \\ 0.000378x - 3.215 & \text{dla } x \in [8500, 11144] \end{cases} \quad (61)$$



Rysunek 13. Wykres funkcji przynależności dla kwantyfikatora absolutnego.

Kwentyfikatory względne wraz z wzorami oraz wykresem wyglądają następująco:

— Prawie żaden

$$\mu_{prawie_zaden}(x) = e^{-\frac{x^2}{0.01}} \quad \text{dla } x \in [0, 1]. \quad (62)$$

— Około 1/4

$$\mu_{okolo_1/4}(x) = e^{-\frac{(x-0.25)^2}{0.01}} \quad \text{dla } x \in [0, 1]. \quad (63)$$

— Około 1/2

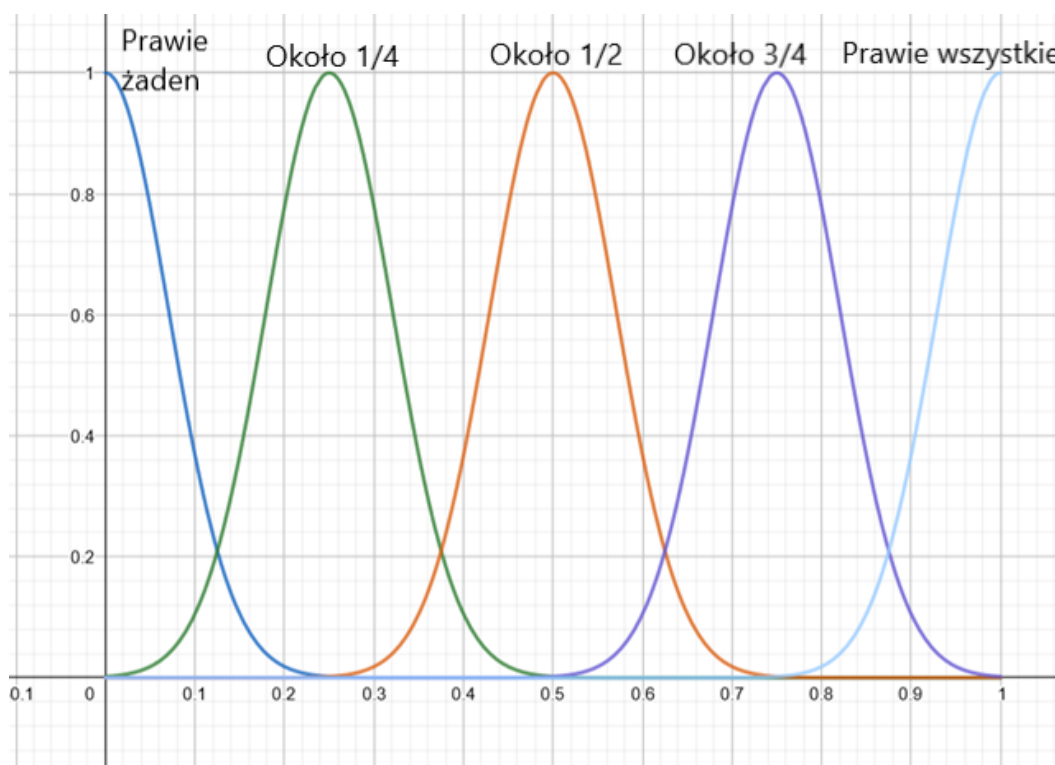
$$\mu_{okolo_1/2}(x) = e^{-\frac{(x-0.5)^2}{0.01}} \quad \text{dla } x \in [0, 1]. \quad (64)$$

— Około 3/4

$$\mu_{okolo_3/4}(x) = e^{-\frac{(x-0.75)^2}{0.01}} \quad \text{dla } x \in [0, 1]. \quad (65)$$

— Prawie wszystkie

$$\mu_{prawie_wszystkie}(x) = e^{-\frac{(x-1)^2}{0.01}} \quad \text{dla } x \in [0, 1]. \quad (66)$$



Rysunek 14. Wykres funkcji przynależności dla kwantyfikatora relatywnego.

4. Narzędzia obliczeniowe: projekt (wybór, implementacja) i diagram UML pakietu obliczeń rozmytych. Diagram UML generatora podsumowań

4.1. Diagram pakietu obliczeń rozmytych

Jako iż podsumowania lingwistyczne opierają się na teorii zbiorów rozmytych, na potrzeby tego zadania postanowiliśmy zaimplementować własny pakiet przeznaczony do obliczeń rozmytych. Pakiet nazywa się *FuzzySetCalculations* i składa się z następujących elementów:

FuzzySet - klasa, która reprezentuje zbiór rozmyty. Metody tej klasy zapewniają najbardziej istotne funkcjonalności dla podsumowań własności. Pierwszą z nich jest metoda o nazwie *getHeight()*, która zwraca wysokość zbioru rozmytego, tj. maksymalną wartość jaką przyjmuje funkcja przynależności w całej przestrzeni rozważań. Kolejna metoda to *getSupport()*, której wynikiem wykonania są wszystkie te elementy zbioru rozmytego, które przynależą do zbioru rozmytego w stopniu większym niż 0. Bardzo podobną metodą jest *getAlphaCount(double y)*, która jako argument przyjmuje wartość zbioru rozmytego. Argument ten stanowi próg dla selekcji elementów, dla których wartość funkcji przynależności jest powyżej niego. Kolejną w kolejności jest metoda *getComplement(double x)*, której zadaniem jest zwrócenie dopełnienia zbioru rozmytego dla argumentu x, czyli liczby, którą trzeba dodać w celu uzyskania wartości 1. Następne dwie metody, czyli *getAndResult(List<FuzzySet> fuzzySet)* i *getOrResult(List<FuzzySet> fuzzySet)* od-

powiadają za operacje sumy (OR) i iloczynu (AND) zbiorów rozmytych (wynikiem tych operacji również jest zbiór rozmyty). Suma zbiorów rozmytych jest obliczana jako maksimum (wykorzystanie T-normy) ze wszystkich wartości zbiorów rozmytych przekazanych w argumencie dla elementu x . Iloczyn zbiorów rozmytych jest obliczany jako minimum (wykorzystanie S-normy) ze wszystkich wartości zbiorów rozmytych przekazanych w argumencie dla wszystkich elementów w przestrzeni. Następną metodą jest metoda *getDegreeOfMembership(double x)*, która dla elementu x zwraca stopień przynależności do zbioru rozmytego (w metodzie tej następuje de facto wywołanie metody *countMembership(double x)* z klasy *MembershipFunction*. Kolejną metody to *isConvex()* oraz *isConcave()*, które zwracają wartość logiczną *true* w momencie, gdy zbiór rozmyty jest wypukły oraz wklęsły. Ostatnią z metod, którą planujemy zaimplementować jest metoda *getCardinality()*, dzięki której otrzymujemy moc zbioru rozmytego tj. sumę wartości funkcji charakterystycznej z wszystkich elementów zbioru rozmytego. [3]

MembershipFunction - klasa abstrakcyjna, która reprezentuje funkcje przynależności do zbioru rozmytego. Klasa ta zawiera jedną metodę abstrakcyjną o nazwie *countMembership(double x)*. Dzięki temu, w klasach dziedziczących można zaimplementować obliczanie stopnia przynależności w zależności od tego, jakiego rodzaju funkcji przynależności chcemy użyć. [5] Z klasy *MembershipFunction* dziedziczy pięć klas:

- **TrapezoidalMembershipFunction** - reprezentuje trapezoidalną funkcję przynależności. W przypadku klasy *LeftTrapezoidalMembershipFunction* (funkcja malejąca) wykorzystane zostają 3 argumenty:

- x1 - początek przedziału od którego funkcja przyjmuje wartość 1
- x2 - koniec przedziału dla którego funkcja przyjmuje wartość 1 i początek przedziału dla funkcji malejącej
- x3 - koniec przedziału dla funkcji malejącej i początek przedziału dla funkcji malejącej

W przypadku klasy *RightTrapezoidalMembershipFunction* również występują 3 argumenty:

- x1 - początek przedziału od którego funkcja jest rosnąca
- x2 - koniec przedziału dla którego funkcja jest rosnąca i początek przedziału dla funkcji stałej o wartości 1
- x3 - koniec przedziału dla funkcji stałej o wartości równej 1

Do funkcjonowania klasy *BothSidesTrapezoidalMembershipFunction* niezbędne są natomiast 4 argumenty:

- x1 - początek przedziału od którego funkcja zaczyna rosnąć
- x2 - koniec przedziału dla funkcji rosnącej i początek przedziału dla funkcji stałej o wartości 1
- x3 - koniec przedziału dla funkcji stałej o wartości 1 i początek przedziału dla funkcji malejącej
- x4 - koniec przedziału dla funkcji malejącej

- **TriangularMembershipFunction** - reprezentuje trójkątną funkcję przynależności. Jako, że funkcja ta może być malejąca, rosnąca lub też składać się z obu tego rodzaju funkcji, w tym celu z klasy *GaussianMembershipFunction* dziedziczą trzy funkcje, które reprezentują wymienione wyżej rodzaje monotoniczności. W przypadku klasy *LeftTriangularMembershipFunction*

4.2. Diagram UML generatora podsumowań. Krótka instrukcja użytkownika

Drugim pakietem użytym przez nas w aplikacji jest pakiet generatora podsumowań. Składa się on z następujących elementów: **Label** - klasa reprezentująca pojedynczą etykietę. Etykieta zawiera swoją nazwę oraz pole odpowiadające za klasę *FuzzySet*. Dzięki temu w etykiecie mamy dostęp do metod klasy *FuzzySet*, w tym *getDegreeOfMembership(double x)*, czyli stopień przynależności dla zadanego argumentu.

LinguisticVariable - klasa odpowiadająca za zmienną lingwistyczną. Posiada ona swoją nazwę oraz pole *rangeSet*, które odpowiada za przestrzeń rozważań dla podanej zmiennej lingwistycznej. Poza tym klasa ta posiada pole typu *Label*, które odpowiada za wyżej wymienioną etykietę.

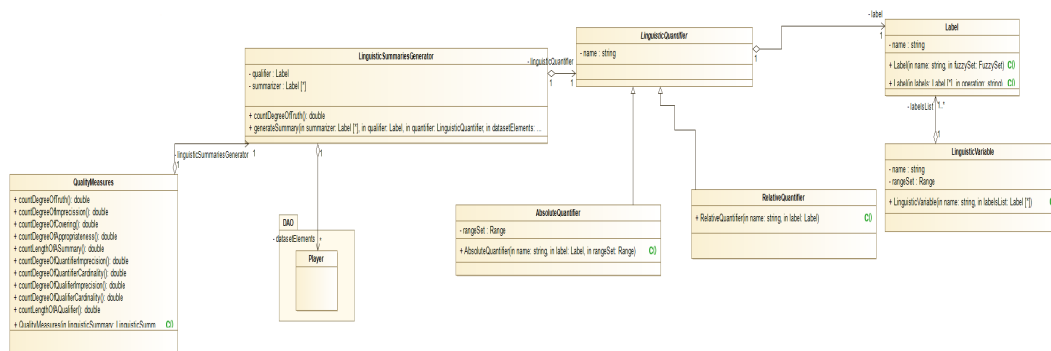
LinguisticQuantifier - klasa abstrakcyjna reprezentująca kwantyfikator lingwistyczny. Kwantyfikator lingwistyczny posiada etykietę. Z klasy abstrakcyjnej *LinguisticQuantifier* dziedziczą dwie następujące klasy:

- **AbsoluteQuantifier** - kwantyfikator absolutny posiadający dodatkowo poza polami, które dziedziczy przedział rozważań.
- **RelativeQuantifier** - kwantyfikator względny.

Player - klasa z modułu DAO odpowiadającego za wczytywanie danych. Klasa ta odpowiada za pojedynczego zawodnika wczytanego z bazy danych.

LinguisticSummariesGenerator - klasa odpowiadająca za generowanie podsumowań, a w zasadzie pojedynczego podsumowania. Zawiera ona kwantyfikator lingwistyczny, pole typu *Player*, co odpowiada podmiotowi podsumowania. Klasa zawiera także pola: *qualifier* (kwalifikator) oraz *summarizer* (sumaryzator). Oba pola są Listami typów *Label*. Klasa ta pozwala także oczywiście na wygenerowanie podsumowania.

QualityMeasures - klasa zawierająca jedno pole - *LinguisticSummariesGenerator*. Klasa ta pozwala na obliczenie miar jakości od T_1 do T_{11} [1] dla wygenerowanego podsumowania.



Rysunek 16. Diagram UML generatora podsumowań.

Diagram UML generatora podsumowań (warstwy obliczeniowej oraz interfejsu użytkownika). Krótki ilustrowany opis jak użytkownik może korzystać z aplikacji, w szczególności wprowadzać parametry podsumowań, odczytywać wyniki oraz definiować własne etykiety i kwantyfikatory. Wersja JRE

i inne wymogi niezbędne do uruchomienia aplikacji przez użytkownika na własnym komputerze.

Sekcja uzupełniona jako efekt zadania Tydzień 11 wg Harmonogramu Zajęć na WIKAMP KSR.

5. Jednopoimotowe podsumowania lingwistyczne. Miary jakości, podsumowanie optymalne

We wszystkich przedstawionych poniżej eksperymentach kolumna z nazwą "średnia" oznacza miarę jakości podsumowania optymalnego.

5.1. Jednopoimotowe podsumowania lingwistyczne w formie pierwszej

Dla jednopoimotowych podsumowań lingwistycznych w formie pierwszej przeprowadziliśmy dwa eksperymenty, w każdym z nich przyjmując inne wagi. Wyniki dla eksperymentu nr 1 prezentują się następująco:

The screenshot shows the 'Podsumowania lingwistyczne bazy danych' application. At the top, there are settings for 'Forma' (Pierwsza), 'Kwantyfikator' (Prawie 11144), 'Kwalifikator(y)' (empty), and 'Sumaryzator(y)' (przeciętnej wagi). Below these are input fields for 'Prawie 1...' and buttons for 'Wyczyść zaznaczenie'. The main section displays 'Wagi miar jakości' for 11 terms (T1 to T11) with values: T1=0,25, T2=0,14, T3=0,14, T4=0,08, T5=0,08, T6=0,14, T7=0,08, T8=0,08, T9=0, T10=0, T11=0. Below this is a table with 12 columns: 'tekst', T1, T2, T3, T4, T5, T6, T7, T8, T9, T10, T11, and 'średnia'. The table contains 15 rows of linguistic descriptions and their corresponding scores. At the bottom, there are buttons for 'Generuj' and 'Podsumowania Wielopoimotowe'.

tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	średnia
Około 2000 zawodników jest w wieku juniorskim,	0.49	0.78	0.22	0.0	1.0	0.73	0.87	0.89	0.0	0.0	1.0	0.59
Około 1/4 zawodników jest średniego wzrostu,	0.9	0.44	0.56	0.0	1.0	0.0	0.82	0.72	0.0	0.0	1.0	0.57
Więcej niż 0 zawodników jest idealny procent asyst,	0.23	0.84	0.16	0.0	1.0	0.87	0.93	0.9	0.0	0.0	1.0	0.55
Około 2000 zawodników jest idealny wpływ na drużynę,	0.24	0.83	0.17	0.0	1.0	0.73	0.87	0.92	0.0	0.0	1.0	0.53
Około 2000 zawodników jest pozytywny wpływ na drużynę,	0.26	0.6	0.4	0.0	1.0	0.73	0.87	0.72	0.0	0.0	1.0	0.51
Około 6000 zawodników jest średniego wzrostu,	0.0	0.44	0.56	0.0	1.0	0.73	0.87	0.72	0.0	0.0	1.0	0.45
Prawie żaden zawodników jest idealny procent asyst,	0.34	0.84	0.16	0.0	1.0	0.0	0.91	0.9	0.0	0.0	1.0	0.45
Prawie 11144 zawodników jest przeciętnej wagi,	0.0	0.46	0.54	0.0	1.0	0.76	0.88	0.72	0.0	0.0	1.0	0.45
Około 2000 zawodników jest w wieku juniorskim, średniego wzrostu,	0.0	0.65	0.12	0.0	0.5	0.73	0.87	0.82	0.0	0.0	1.0	0.39
Około 1/4 zawodników jest w wieku juniorskim,	0.14	0.78	0.22	0.0	1.0	0.0	0.82	0.89	0.0	0.0	1.0	0.39
Około 1/2 zawodników jest średniego wzrostu,	0.01	0.44	0.56	0.0	1.0	0.0	0.82	0.72	0.0	0.0	1.0	0.35
Około 3/4 zawodników jest dużo asyst na mecz,	0.0	0.69	0.31	0.0	1.0	0.0	0.82	0.79	0.0	0.0	1.0	0.35
Około 1/4 zawodników jest w wieku juniorskim, średniego wzrostu,	0.01	0.65	0.12	0.0	0.5	0.0	0.82	0.82	0.0	0.0	1.0	0.28

Rysunek 17. Wyniki eksperymentu nr 1 dla jednopoimotowych podsumowań lingwistycznych w formie pierwszej

Podsumowanie "Około 2000 zawodników jest w wieku juniorskim" osiągnęło najwyższą wartość miary jakości podsumowania optymalnego porównując z innymi podsumowaniami wygenerowanymi w tym eksperymencie. Wyniosła ona 0.59. Warto zauważyć, że w tym wypadku dodając kolejny suma-

ryzator jakość podsumowania maleje ("Około 1/4 zawodników jest średniego wzrostu" znajduje się na liście na pozycji drugiej, "Około 1/4 zawodników jest w wieku juniorskim" na pozycji 10, natomiast podsumowanie "Około 1/4 zawodników jest w wieku juniorskim, średniego wzrostu" dopiero na pozycji 13. Najwyższą wartość miary T1 osiągnęło podsumowanie "Około 1/4 zawodników jest średniego wzrostu", jednak biorąc pod uwagę miarę podsumowania optymalnego podsumowanie to znajduje się na 2 miejscu. Warto zwrócić uwagę na miarę T6, która lekko zakłamuje nam obraz. Dla kwantyfikatora względnego miara ta w każdym przypadku pokazuje 0. Wynika to z tego, że ten typ kwantyfikatorów wyraża się w tym wypadku funkcją Gaussowską, która to dąży do zera (jednak wartości 0 nie przyjmuje). W kolejnym eksperymencie zmodyfikowaliśmy nieco wagi średnich, a jego wyniki prezentują się następująco:

Podsumowania lingwistyczne bazy danych
Podsumowania jednopodmiotowe

Forma: Pierwsza Kwantyfikator: Prawie żaden Edytuj Kwalifikator(y): Edytuj Sumaryzator(y): fatalny wpływ na d... Edytuj

Prawie ż...

Wyczyść zaznaczenie Wyczyść zaznaczenie

Wagi miar jakości:

T1: 0,45 T2: 0,1 T3: 0,1 T4: 0,02 T5: 0,1 T6: 0,02

T7: 0,11 T8: 0,1 T9: 0 T10: 0 T11: 0

tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	średnia
Prawie żaden zawodników jest błyskawicznie wybrani w drafcie, mało rozegranych gier w sezonie,	0.98	0.79	0.03	0.01	0.5	0.0	0.91	0.87	0.0	0.0	1.0	0.76
Około 6000 zawodników jest dostateczna liczba zdobytych punktów na mecz,	0.88	0.16	0.84	0.0	1.0	0.73	0.87	0.45	0.0	0.0	1.0	0.75
Około 1/4 zawodników jest bardzo ciężkiej wagi,	0.81	0.63	0.37	0.0	1.0	0.0	0.82	0.8	0.0	0.0	1.0	0.73
Około 1/2 zawodników jest dostateczna liczba zdobytych punktów na mecz,	0.75	0.16	0.84	0.0	1.0	0.0	0.82	0.45	0.0	0.0	1.0	0.67
Więcej niż 0 zawodników jest idealną skuteczność rzutów,	0.46	0.86	0.14	0.0	1.0	0.87	0.93	0.93	0.0	0.0	1.0	0.62
Około 2000 zawodników jest idealną skuteczność rzutów,	0.21	0.86	0.14	0.0	1.0	0.73	0.87	0.93	0.0	0.0	1.0	0.5
Prawie żaden zawodników jest fatalny wpływ na drużynę,	0.1	0.7	0.3	0.0	1.0	0.0	0.91	0.85	0.0	0.0	1.0	0.43
Około 4000 zawodników jest idealną skuteczność rzutów,	0.0	0.86	0.14	0.0	1.0	0.73	0.87	0.93	0.0	0.0	1.0	0.4
Około 6000 zawodników jest normalnie wybrani w drafcie,	0.0	0.75	0.25	0.0	1.0	0.73	0.87	0.87	0.0	0.0	1.0	0.4
Więcej niż 0 zawodników jest fatalny wpływ na drużynę,	0.0	0.7	0.3	0.0	1.0	0.87	0.93	0.85	0.0	0.0	1.0	0.4
Prawie żaden zawodników jest bardzo ciężkiej wagi,	0.02	0.63	0.37	0.0	1.0	0.0	0.91	0.8	0.0	0.0	1.0	0.39
Około 1/2 zawodników jest bardzo ciężkiej wagi,	0.0	0.63	0.37	0.0	1.0	0.0	0.82	0.8	0.0	0.0	1.0	0.37
Prawie wszystkie zawodników jest błyskawicznie wybrani w drafcie, mało rozegranych gier w sez...	0.0	0.79	0.03	0.01	0.5	0.0	0.91	0.87	0.0	0.0	1.0	0.32

Generuj Podsumowania Wielopodmiotowe

Rysunek 18. Wyniki eksperymentu nr 2 dla jednopodmiotowych podsumowań lingwistycznych w formie pierwszej

W tym eksperymencie zmieniliśmy wagi, w szczególności zwiększając wagę dla miary T1 oraz zmniejszając wagę dla miary T6. Najlepszym podsumowaniem w tym eksperymencie okazało się "Prawie żaden zawodnik jest błyskawicznie wybrany w drafcie i ma mało rozegranych gier w sezonie [0.76]". Jest to jak najbardziej logiczne, ponieważ kluby wybierają zawodników w drafcie szybko, aby ci stanowili o sile ich drużyny. Podsumowanie to osiągnęło wartość T1 aż na poziomie 0.98. Jak widzimy, przy takim dobraniu wag jak powyżej byliśmy w stanie zminimalizować negatywny wpływ funkcji

gaussowskiej na ostateczny wynik miary jakości podsumowania optymalnego. Zarówno w tym jak i w poprzednim eksperymencie przyjęliśmy wagi dla T9, T10 i T11 na 0, ponieważ miary te dotyczą kwalifikatorów, których w formie pierwszej podsumowań jednopodmiotowych nie ma.

5.2. Jednopodmiotowe podsumowania lingwistyczne w formie drugiej

Dla formy drugiej jednopodmiotowych podsumowań lingwistycznych przeprowadziliśmy 1 eksperyment, którego wyniki przedstawiają się następująco:

Podsumowania lingwistyczne bazy danych
Podsumowania jednopodmiotowe

Forma: Druga
Kwantyfikator: Prawie wszystkie
Kwalifikator(y): idealną skuteczno...
Sumaryzator(y): pozytywny wpływ ...

Wagi miar jakości:

T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	średnia
0,10	0,09	0,09	0,09	0,09	0,09						
0,09	0,09	0,09	0,09	0,09							

tekst	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	średnia
Prawie żaden zawodników będących bardzo dużo zbiórek na mecz, jest bardzo niskiego wzrostu...	0.99	0.75	0.06	0.18	1.0	0.0	0.91	0.89	0.67	0.82	1.0	0.66
Około 1/2 zawodników będących bardzo ciężkiej wagi, jest dużo zbiórek na mecz,	0.72	0.67	0.53	0.2	1.0	0.0	0.82	0.79	0.63	0.8	1.0	0.65
Prawie wszystkie zawodników będących idealną skuteczność rzutów, jest idealny wpływ na dru...	0.0	0.83	0.38	0.2	1.0	0.0	0.91	0.92	0.86	0.93	1.0	0.63
Około 1/4 zawodników będących średnia liczba rozegranych gier w sezonie, jest doświadczony...	0.64	0.81	0.2	0.02	1.0	0.0	0.82	0.85	0.69	0.84	1.0	0.62
Około 3/4 zawodników będących idealną skuteczność rzutów, jest pozytywny wpływ na drużyn...	0.0	0.6	0.6	0.2	1.0	0.0	0.82	0.72	0.86	0.93	1.0	0.61
Prawie wszystkie zawodników będących idealną skuteczność rzutów, jest pozytywny wpływ na ...	0.0	0.6	0.6	0.2	1.0	0.0	0.91	0.72	0.86	0.93	1.0	0.61
Prawie żaden zawodników będących idealny procent asyst, jest nieskuteczni z rzutów,	0.13	0.0	1.0	0.0	1.0	0.0	0.91	0.88	0.84	0.9	1.0	0.6
Około 3/4 zawodników będących bardzo dużo zbiórek na mecz, jest bardzo wysokiego wzrostu,	0.18	0.4	0.85	0.25	1.0	0.0	0.82	0.7	0.67	0.82	1.0	0.6
Około 1/2 zawodników będących szybko wybrani w drafcie, jest w młodym wieku,	0.78	0.45	0.55	0.0	1.0	0.0	0.82	0.57	0.67	0.75	1.0	0.6
Około 1/2 zawodników będących bardzo dużo asyst na mecz, jest pozytywny wpływ na drużynę,	0.41	0.6	0.5	0.11	1.0	0.0	0.82	0.72	0.69	0.82	1.0	0.6
Około 1/4 zawodników będących bardzo mało zdobytych punktów na mecz, jest w średnim wi...	0.99	0.68	0.3	0.02	1.0	0.0	0.82	0.74	0.27	0.69	1.0	0.6
Prawie wszystkie zawodników będących idealny procent asyst, jest idealną skuteczność rzutów,	0.0	0.86	0.1	0.04	1.0	0.0	0.91	0.93	0.84	0.9	1.0	0.59
Prawie wszystkie zawodników będących idealny wpływ na drużynę, jest idealny procent asyst,	0.0	0.84	0.18	0.02	1.0	0.0	0.91	0.9	0.83	0.92	1.0	0.59

Generuj Podsumowania Wielopodmiotowe

Rysunek 19. Wyniki eksperymentu nr 1 dla jednopodmiotowych podsumowań lingwistycznych w formie drugiej

Tym razem wagi rozłożyliśmy równomiernie. Najlepszym podsumowaniem sortując po mierze podsumowania optymalnego okazało się "Prawie żaden zawodnik mający bardzo dużo zbiórek na mecz, jest bardzo niskiego wzrostu [0.66]" Podsumowanie to osiągnęło wartość miary T1 (stopień prawdziwości) aż 0,99. Stopień prawdziwości 0,99 osiągnęło także podsumowanie "Około 1/4 zawodników mających bardzo mało zdobytych punktów na mecz, jest w średnim wieku [0.60]". Końcowa miara dla tego podsumowania wyniosła 0.60. Wysoką wartość miary podsumowania optymalnego (0.65) uzyskało także podsumowanie "Około 1/2 zawodników będących bardzo ciężkiej wagi, ma dużo zbiórek na mecz. . Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Listy podsumowań jednopodmiotowych i tabele/rankingi podsumowań dla danych atrybutów obowiązkowe i dokładnie

opisane w „captions” (tytułach), konieczny opis kolumn i wierszy tabel. Dla każdego podsumowania podane miary jakości oraz miara jakości podsumowania optymalnego.

Sekcja uzupełniona jako efekt zadania Tydzień 11 wg Harmonogramu Zajęć na WIKAMP KSR.

6. Wielopodmiotowe podsumowania lingwistyczne i ich miary jakości

Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Uzasadnienie i metoda podziału zbioru danych na rozłączne podmioty. Listy podsumowań wielopodmiotowych i tabele/rankingi podsumowań dla danych atrybutów obowiązkowe i dokładnie opisane w „captions” (tytułach), konieczny opis kolumn i wierszy tabel. Konieczne uwzględnienie wszystkich 4-ch form podsumowań wielopodmiotowych.

****** Możliwe sformułowanie zagadnienia wielopodmiotowego podsumowania optymalnego ******.

******Ewentualne wyniki realizacji punktu „na ocenę 5.0” wg opisu Projektu 2. i ich porównanie do wyników z części obowiązkowej******.

Sekcja uzupełniona jako efekt zadania Tydzień 12 wg Harmonogramu Zajęć na WIKAMP KSR.

7. Dyskusja, wnioski

7.1. Jednopodmiotowe podsumowania lingwistyczne w formie pierwszej

W mierze T6, w której pod uwagę bierzemy nośnik kwantyfikatora niezwykle ważny jest typ funkcji przynależności dla kwantyfikatora. Dla funkcji Gaussowskiej miara ta zawsze przyjmowała będzie wartość 0, ponieważ wykres funkcji Gaussa na całej przestrzeni rozważań nie przyjmuje wartości 0. Łącząc sumaryzatory, których połączenie wydaje nam się logiczne i które wydają się w jakiś sposób skorelowane ze sobą (jak np. ”błyskawicznie wybrany w drafcie” i ”mało rozegranych gier w sezonie”) otrzymujemy podsumowania o bardzo wysokiej mierze T1, a więc stopniu prawdziwości. Warto zauważyć, że dla tego samego kwantyfikatora zmieniając tylko sumaryzatory, bądź dla tych samych sumaryzatorów zmieniając tylko kwantyfikatory tylko niektóre z miar zmieniają swoje wartości. Oznacza to po prostu, że część miar zależy tylko i wyłącznie od sumaryzatora/kwantyfikatora, nie łącząc tych dwóch zbiorów rozmytych ze sobą, stąd tak ważna jest analiza wszystkich miar przy wyciąganiu wniosków.

7.2. Jednopođmiotowe podsumowania lingwistyczne w formie drugiej

Z podsumowań jednopođmiotowych w formie drugiej pĹynie wiĹcej wniosków dotyczĄcych zbioru danych zawodników koszykówki. Możemy bowiem wywnioskowaĹ chociaŹby, Źe zawodnicy bardzo niskiego wzrostu nie przewoĹdzą w statystykach jeŹli chodzi o zbiórki. Natomiast zawodnicy waŹący bardzo duŹo czĹĹsto zaliczają duŹo zbiorów na mecz. Poza tym możemy wywnioskowaĹ, Źe okoĹo poĹowy zawodników wybranych szybko w drafcie sĄ aktualnie w mĹodym wieku. Możemy z tego zdania wysnuĹ wniosek, Źe zawodnicy wybrani w drafcie szybko poczĄtkowe lata grają w NBA, jednak potem ze rŹoĹnych wzglĹdów odchodzą z najbardziej prestiŹowej ligi koszykówki Źwiata. JeŹli chodzi o wnioski dotyczĄce miar pĹynĄce z tej czĹĹci eksperymentów to zdecydowanie byĹo w nich widaĹ, Źe im wiĹcej nie powiĄzanych ze sobĄ kwalifikatorów/sumaryzatorów uŹyjemy tym podsumowanie jest gorszej jakoŹci. Można takŹe zauwaŹyĹ, Źe miara T7 jest podobna dla wszystkich podsumowań. Źwiadczy to o tym, Źe kwantyfikatory sĄ rŹownomiernie rozĹoŹone - ich pola pod wykresami sĄ podobne. ZauwaŹamy takŹe, Źe miary T9, T10 i T11, czyli miary zaleŹĄce od kwalifikatorów przyjmują w wiĹkszoŹci przypadków duŹe wartoŹci. Źwiadczy to o tym, Źe kwalifikatory w formie drugiej podsumowań jednopođmiotowych zostaĹy dobrane odpowiednio.

DokĹadne interpretacje uzyskanych wyników w zaleŹnoŹci od parametrŹw klasyfikacji opisanych w punktach 3.-4 opisu Projektu 2. SzczegŹlnie istotne sĄ wnioski o charakterze uniwersalnym, istotne dla podobnych zadań. OmŹwiĹ i wyjaŹniĹ napotkane problemy (jeŹli byĹy). KaŹdy wniosek/problem powinien mieĹ poparcie w przeprowadzonych eksperymentach (odwoĹania do konkretnych wyników: tabel i miar jakoŹci). Ocena ktŹre wybrane kwantyfikatory, summaryzatory, kwalifikatory i/lub ich miary jakoŹci majĄ maĹe albo duŹe znaczenie dla wiarygodnoŹci i jakoŹci otrzymanych agregacji/podsumowań. Dla końcowej oceny jest to najwaŹniejsza sekcja sprawozdania, gĹdyŹ prezentuje poziom zrozumienia rozwiĄzywanego problemu.

** MoŹliwoŹci kontynuacji prac w obszarze logiki rozmytej i wnioskowania rozmytego, zwĹszcza w kontekŹcie pracy inŹynierskiej, magisterskiej, naukowej, itp. **

Sekcja uzupeĹniona jako efekt zadań Tydzień 11 i Tydzień 12 wg Harmonogramu ZajĹć na WIKAMP KSR.

8. Braki w realizacji projektu 2.

WymieniĹ wg opisu Projektu 2. wszystkie niezrealizowane obowiĄzkowe elementy projektu, ewentualnie podaĹ merytoryczne (ale nie czasowe) przyczyny tych braków.

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] Baza danych zawierająca statystyki koszykarzy z ligi NBA z lat 1996-2020 [przełączany 04.05.2021] Dostępna w: <https://www.kaggle.com/justinas/nba-players-data>
- [5] Dokumentacja diagramów UML [przełączana 24.05.2021] Dostępna w: <https://www.uml-diagrams.org/>

Literatura zawiera wyłącznie źródła recenzowane i/lub o potwierdzonej wiarygodności, możliwe do weryfikacji i cytowane w sprawozdaniu.