

Data oddania: \_\_\_\_\_

Ocena: \_\_\_\_\_

Justyna Hubert 210200

Karol Podlewski 210294

## Zadanie 2: Podsumowania lingwistyczne\*

### 1. Cel

Celem zadania było zaprojektowanie aplikacji desktopowej, która posiada charakter doradczy, generujący pewną ilość podsumowań lingwistycznych dla podanej bazy, a następnie przedstawia użytkownikowi wybrane - według zastosowanych miar jakości wyniki, czyli podsumowania lingwistyczne.

### 2. Wprowadzenie

Zagadnieniem jakim zajmowaliśmy się w ramach projektu była analiza działania lingwistycznych podsumowań baz danych na zbiorach rozmytych. Zbiór rozmyty jest podstawowym pojęciem wykorzystywanym przy naszym zadaniu, zatem przytoczmy jego definicję:

**Definicja 1.** Niech  $\mathcal{X}$  będzie zbiorem, którego elementy interesują nas w sposób bezpośredni, czyli jest zbiorem klasycznym znanym z teorii mnogości (dany element przynależy do zbioru lub nie przynależy). Wówczas *zbiorem rozmytym opisanym w przestrzeni rozważań*  $\mathcal{X}$  nazywamy każdy zbiór  $A$  postaci:

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

gdzie  $\mu_A(x) : \mathcal{X} \rightarrow [0, 1]$  nazywamy *funkcją przynależności do zbioru rozmytego*  $A$ .

---

\* GitHub: <https://github.com/hubjust/KSR>

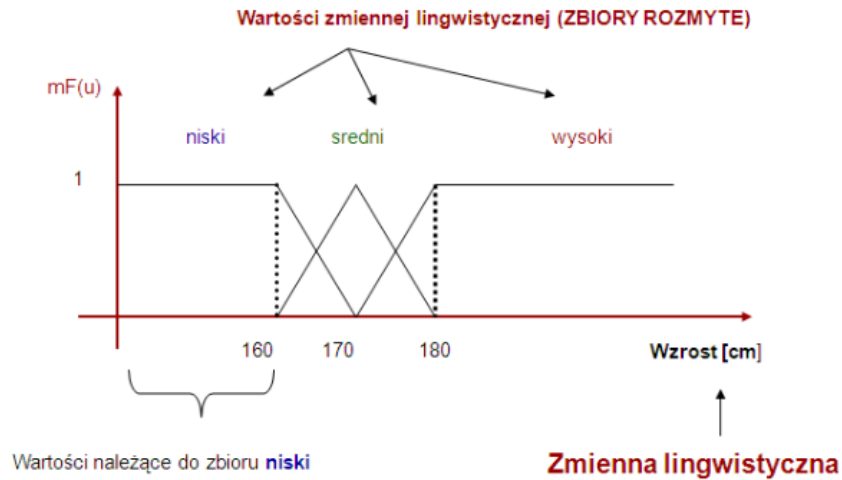
Funkcja przynależności określa w jakim stopniu dany element przynależy do zbioru. W zbiorach rozmytych zakres wartości jakie może ona przyjmować jest rozszerzony do przedziału  $[0,1]$ . W naszym projekcie skorzystaliśmy z funkcji przynależności trójkątnej oraz trapezoidalnej. Przytoczmy ich definicje:

**Definicja 2** (Zbiór rozmyty o trójkątnej funkcji przynależności). Zbiór rozmyty  $A$  typu I na uniwersum  $\mathbb{R}$  jest *liczbą rozmytą trójkątną o parametrach*  $a, b, c$  wtedy i tylko wtedy, gdy  $a \leq b \leq c$  oraz:

$$\mu_A(x) = \begin{cases} 0 & \text{gdy } x \in (-\infty, a], \\ (x - a)/(b - a) & \text{gdy } x \in (a, b), \\ 1 & \text{gdy } x = b, \\ (c - x)/(c - b) & \text{gdy } x \in (b, c), \\ 0 & \text{gdy } x \in [c, +\infty). \end{cases}$$

**Definicja 3** (Zbiór rozmyty o trapezoidalnej funkcji przynależności). Zbiór rozmyty  $A$  typu I na uniwersum  $\mathbb{R}$  jest *liczbą rozmytą trapezoidalną o parametrach*  $a, b, c, d$  wtedy i tylko wtedy, gdy  $a \leq b \leq c \leq d$  oraz:

$$\mu_A(x) = \begin{cases} 0 & \text{gdy } x \in (-\infty, a], \\ (x - a)/(b - a) & \text{gdy } x \in (a, b), \\ 1 & \text{gdy } x \in [b, c], \\ (d - x)/(d - c) & \text{gdy } x \in (c, d), \\ 0 & \text{gdy } x \in [d, +\infty). \end{cases}$$



Rysunek 1. Przykład funkcji przynależności - trójkątnej oraz trapezoidalnej [3]

Wyjaśnijmy także, czym jest lingwistyczne podsumowanie. Niech  $\mathcal{D}$  będzie bazą danych składającą się z  $m$  krotek opisujących poszczególne rekordy. Przyjmijmy, że każda kolumna opisuje cechę pewnego typu. Taką cechę możemy nazwać *zmienną lingwistyczną*. Może ona przyjmować konkretne wartości liczbowe lub rozmyte (np. mało/trochę/dużo/sporo). Zdefiniujmy także  $P$ . Niech  $P$  będzie podmiotem podsumowania lingwistycznego (np. mężczyźni, kobiety, samochody, zawodnicy). Bardzo ważnym elementem, wykorzystywanym we wszystkich rodzajach podsumowań lingwistycznych, jest kwantyfikator oznaczany jako  $Q$ . Przykładami kwantyfikatorów mogą być: "około 10", "ponad 70" (kwantyfikatory absolutne - zbiory rozmyte na uniwersum  $\mathbb{R}$ ) lub "większość", "znikoma część" (kwantyfikatory relatywne - zbiory rozmyte na uniwersum  $[0, 1]$ ). Istotny dla nas będzie stopień przynależności  $P$  do  $Q$ . Zdefiniujmy także sumaryzator  $S_j$ . Jest to zbiór rozmyty na zbiorze wartości przyjmowanych przez  $j$ -tą kolumnę bazy danych. Np. gdyby krotki dotyczyły różnych pojazdów, a jedną ze zmiennych lingwistycznych była ich prędkość, to sumaryzatory mogłyby mieć postać "jeździ szybko", "jeździ ponad 200km/h" itp.

Wykorzystując powyższe elementy można skonstruować **lingwistyczne podsumowanie bazy danych**, czyli:

$$Q \text{ } P \text{ jest/są } S_j \text{ } [T] ,$$

gdzie  $T$  to stopień prawdziwości podsumowania.

Przykład : *Dużo studentów zarabia średnią krajową [0.64]*, gdzie: "dużo" to kwantyfikator, "studentów" to podmiot lingwistyczny, "zarabia średnią krajową" to sumaryzator, a "[0,64]" to stopień prawdziwości podsumowania.

W celu rozszerzenie podsumowania lingwistycznego należy skorzystać ze złożonego sumaryzatora. Sumę sumaryzatorów można w podsumowaniu lingwistycznym zapisać za pomocą słowa "lub", zaś iloczyn za pomocą słowa "i". W rezultacie **podsumowanie ze złożonym sumaryzatorem** może mieć postać:

$$Q \text{ } P \text{ jest/są } S_1 \text{ i/lub } S_2 \text{ i/lub } \dots \text{ i/lub } S_n \text{ } [T] .$$

Przykład: *Dużo studentów zarabia średnią krajową i/lub nosi okulary [0.44]*.

Innym sposobem rozszerzenia pojęcia podsumowań jest zastosowanie kwalifikatora. Kwalifikator  $W$  jest zbiorem rozmytym na  $\mathcal{D}$ , który opisuje jakąś dodatkową właściwość. Typowe przykłady to "[osoby] które są bezrobotne", "[osoby] które są dziećmi". **Podsumowanie z kwalifikatorem** ma postać:

$$Q \text{ } P \text{ mających własność } W \text{ ma własność } S_j \text{ } [T] .$$

Przykład: *Studenci, którzy mają blond włosy zarabiają średnią krajową [0.28]*.

Aby określić jakość naszych podsumowań zaimplementowaliśmy poniższe miary jakości:

### 2.1. $T_1$ – stopień prawdziwości

Stopień prawdziwości jest najbardziej naturalną miarą jakości podsumowania. Określa ona sumę przynależności wszystkich rozważanych krotek do sumaryzatora  $S_j$ :

$$r = \sum_{i=1}^m \mu_{ce(S_j)}(d_i) ,$$

gdzie  $ce(S_j)$  jest rozszerzeniem cylindrycznym sumaryzatora  $S_j$ ,  $m$  liczba wszystkich krotek, a  $d_i$ . Dla kwantyfikatorów relatywnych stopień prawdziwości możemy zapisać jako

$$T_1 = \mu_Q\left(\frac{r}{m}\right),$$

zaś dla kwantyfikatorów absolutnych jako

$$T_1 = \mu_Q(r),$$

gdzie  $r$  jest kardynalnością.

### 2.2. $T_2$ – stopień nieprecyzyjności

Dla podsumowania z  $n$  sumaryzatorami  $S_1 \dots S_n$  możemy określić stopień nieprecyzyjności, definiowany następującym wzorem:

$$T_2 = 1 - \left( \prod_{j=1}^n \text{in}(S_j) \right)^{1/n} .$$

Wyrażenie  $\left( \prod_{j=1}^n \text{in}(S_j) \right)^{1/n}$  to określa średnią geometryczną ze stopni rozmycia wykorzystanych sumaryzatorów, czyli w jakim stopniu precyzyjny jest sumaryzator. Im mniejszy nośnik zbioru rozmytego, tym wyższa jest jego precyzja.

### 2.3. $T_3$ – stopień pokrycia

Stopień pokrycia  $T_3$  jest zdefiniowany dla podsumowań z kwalifikatorami. Stopień pokrycia  $T_3$  Dla każdego  $i = 1 \dots m$  (związanego z krotką  $d_i$  z bazy danych) możemy zdefiniować (z kwalifikatorem):

$$t_i = \begin{cases} 1 & \text{gdy } \mu_{ce(S_j)}(d_i) > 0 \wedge \mu_W(d_i) > 0 \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

$$h_i = \begin{cases} 1 & \text{gdy } \mu_W(d_i) > 0 \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

Bez kwalifikatora:

$$t_i = \begin{cases} 1 & \text{gdy } \mu_{ce(S_j)}(d_i) > 0 \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

$$h_1 = 1$$

Przy powyższych oznaczeniach:

$$T_3 = \frac{\sum_{i=1}^m t_i}{\sum_{i=1}^m h_i} .$$

Reprezentuje stopień w jakim nośnik sumaryzatora pokrywa się z nośnikiem kwalifikatora.

#### 2.4. $T_4$ – stopień trafności

Dla podsumowania z  $n$  sumaryzatorami  $S_1 \dots S_n$  oraz  $m$  krotkami w bazie danych możemy wprowadzić oznaczenia:

$$g_{ij} = \begin{cases} 1 & \text{gdy } \mu_{ce(S_j)}(d_i) > 0 \\ 0 & \text{w przeciwnym wypadku.} \end{cases}$$

oraz

$$r_j = \frac{\sum_{i=1}^m g_{ij}}{m} .$$

Wówczas możemy zapisać:

$$T_4 = \left| \prod_{j=1}^n r_j - T_3 \right| .$$

Określa jak wiele krotek przynależy do sumaryzatora, czyli czy dane podsumowanie jest właściwe dla zestawu danych.

#### 2.5. $T_5$ – długość podsumowania

Dla podsumowania z  $n$  sumaryzatorami  $S_1 \dots S_n$  miarę długości podsumowania definiujemy jako:

$$T_5 = 2 \left( \frac{1}{2} \right)^{|s|} .$$

Gdzie  $|s|$  jest ilością zbiorów rozmytych, z których skomponowany jest sumaryzator. Określa jakość podsumowania na podstawie złożoności sumaryzatora, czyli im więcej składowych sumaryzatora złożonego, tym niższa wartość tej miary.

#### 2.6. $T_6$ – stopień nieprecyzyjności kwantyfikatora

$T_6$ , czyli stopień nieprecyzyjności kwantyfikatora możemy zdefiniować jako:

$$T_6 = 1 - \ln(Q) .$$

Reprezentuje w jakim stopniu precyzyjny jest kwantyfikator. Im mniejszy nośnik zbioru rozmytego tym wyższa jest jego precyzja.

### 2.7. $T_7$ – stopień liczności kwantyfikatora

W przeciwieństwie do  $T_6$ , zamiast zliczać elementy z nośnika  $Q$ , policzymy moc zbioru rozmytego:

$$T_7 = 1 - \frac{|Q|}{|\mathcal{X}_Q|}.$$

Opisuje stopień precyzji kwantyfikatora, im mniejsza kardynalność kwantyfikatora tym jest on bardziej precyzyjny.

### 2.8. $T_8$ – stopień liczności sumaryzatora

W przypadku zastosowania sumaryzatora złożonego, podobnie jak przy poprzednich miarach, stosujemy średnią geometryczną. Dla podsumowania z  $n$  sumaryzatorami  $S_1 \dots S_n$ :

$$T_8 = 1 - \left( \prod_{j=1}^n \frac{|S_j|}{|\mathcal{X}_j|} \right)^{\frac{1}{n}}.$$

Opisuje stopień precyzji sumaryzatora, im mniejsza kardynalność kwantyfikatora tym jest on bardziej precyzyjny.

### 2.9. $T_9$ – stopień nieprecyzyjności kwalifikatora

Stopień precyzji kwalifikatora  $T_9$  jest oparty na drugiej formie podsumowań tzn.:  $Q$  obiektów będących/mających  $W$  jest/ma  $S$ , gdzie  $W$  jest reprezentowane przez zbiór rozmyty i jest kwalifikatorem. Definicja tej miary jest następująca:

$$T_9 = 1 - \ln(W).$$

Określa w jakim stopniu precyzyjny jest kwalifikator. Im szerszy nośnik zbioru rozmytego tym niższa jest jego precyzja, gdyż bierze pod uwagę większy zakres wartości.

### 2.10. $T_{10}$ – stopień liczności kwalifikatora

Stopień kardynalności kwalifikatora  $T_{10}$  definiujemy jako:

$$T_{10} = 1 - \frac{|W|}{|\mathcal{X}_g|}.$$

Opisuje stopień precyzji kwalifikatora, im większa jest kardynalność kwalifikator tym jest on mniej precyzyjny.

### 2.11. $T_{11}$ – długość kwalifikatora

Długość kwalifikatora  $T_{11}$  definiujemy następująco:

$$T_{11} = 2 \left( \frac{1}{2} \right)^{|W|}.$$

Wyznacza jakość podsumowania na podstawie złożoności kwalifikatora, Im bardziej złożony kwalifikator tym jakość podsumowania gorsza.

### 3. Opis implementacji

Program został stworzony w języku C#. Graficzny interfejs użytkownika został stworzony przy wykorzystaniu Windows Presentation Foundation. Logika aplikacji została odseparowana od GUI. W związku z tym, zaimplementowaliśmy trzy projekty: Logic, ViewModel oraz GUI.

#### 3.1. Logic

W tym projekcie zawarta została cała logika aplikacji. Odzworowany został model naszej bazy danych (*FifaPlayer.cs*), zaimplementowane zostały: funkcje przynależności trójkątna (*TriangularFunction.cs*) oraz trapezoidalna (*TrapezoidFunction*), zmienna lingwistyczna (*LinguisticVariable.cs*), kwantyfikator (*Quantifier.cs*), zmienna, która "na sztywno" określa nasz kwantyfikator (np. słaby, przeciętny, dobry) w zależności od podanych danych (*Variable.cs*), sumaryzator "i" (*And.cs*), a także sumaryzator "lub" (*Or.cs*). W projekcie logic znajduje się także klasa (*Measures.cs*), gdzie zawarliśmy wszystkie 11 miar jakości podsumowań.

#### 3.2. ViewModel

Klasa MainViewModel przyjmuje dane wejściowe od użytkownika i reaguje na jego poczynania wywołując wybrane akcje z logiki programu oraz odpowiada za odświeżanie widoków w interfejsie graficznym.

#### 3.3. GUI

Projekt GUI (graphical user interface) implementuje przejrzysty oraz łatwy w obsłudze graficzny interfejs użytkownika.

## 4. Materiały i metody

#### 4.1. Baza danych

Do przeprowadzenia badań i generowania konkretnych podsumowań wykorzystaliśmy bazę danych dotyczącą przechowującą statystyki piłkarzy z gry Fifa 2019 [4]. Składa się ona z 15397 krotek znajdujących się w tabeli z 20 różnymi kolumnami - w ramach naszego projektu skorzystaliśmy z 13. Przedstawiamy je poniżej:

- Wiek - wartości z przedziału [17-45]
- Wzrost (cm) - wartości z przedziału [155-205]
- Waga (kg) - wartości z przedziału [50-11]
- Tempo - wartości z przedziału [0-97]
- Przyspieszenie - wartości z przedziału [13-98]

- Prędkość - wartości z przedziału [12-97]
- Dribbling - wartości z przedziału [0-97]
- Zręczność - wartości z przedziału [14-98]
- Balans - wartości z przedziału [16-99]
- Reakcje - wartości z przedziału [30-96]
- Kontrola piłki - wartości z przedziału [3-97]
- Opanowanie - wartości z przedziału [3-97]
- Precyzja - wartości z przedziału [0-93]
- Ustawienie się - wartości z przedziału [2-95]

Każda z ww. kolumn jest typem całkowitym.

## 4.2. Sumaryzatory i kwalifikatory

Poniżej zaprezentowaliśmy poszczególne sumaryzatory oraz kwalifikatory wykorzystane w naszym programie.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Bardzo młody	17	17	18	20
Młody	19	21	24	29
Dorosły	28	30	35	37
Dojrzały	36	40	45	45

Tabela 1. Przyporządkowane parametry funkcji trapezoidalnej dla wieku.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Bardzo niski	155	155	160	162
Niski	161	165	168	170
Przeciętny	169	172	176	180
Wysoki	179	182	186	192
Bardzo wysoki	191	196	205	205

Tabela 2. Przyporządkowane parametry funkcji trapezoidalnej dla wzrostu.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Niska	50	59	66	73
Standardowa	72	77	79	84
Postawna	83	86	89	91
Ciężka	90	98	110	110

Tabela 3. Przyporządkowane parametry funkcji trapezoidalnej dla wzrostu.



<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Niskie	0	15	27	35
Średnie	33	46	58	66
Wysokie	65	79	88	97

Tabela 4. Przyporządkowane parametry funkcji trapezoidalnej dla tempa.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słabe	13	25	29	35
Przeciętne	33	46	58	66
Dobre	65	79	88	98

Tabela 5. Przyporządkowane parametry funkcji trapezoidalnej dla przyspieszenia.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słaba	12	25	29	35
Przeciętna	33	46	58	69
Dobra	65	79	88	97

Tabela 6. Przyporządkowane parametry funkcji trapezoidalnej dla prędkości.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słaby	0	15	27	37
Przeciętny	36	46	58	66
Dobry	67	79	88	97

Tabela 7. Przyporządkowane parametry funkcji trapezoidalnej dla dribblingu.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słaba	14	21	27	35
Przeciętna	33	46	58	69
Dobra	67	79	88	98

Tabela 8. Przyporządkowane parametry funkcji trapezoidalnej dla zręczności.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słaby	16	21	29	35
Przeciętny	38	46	58	72
Dobry	71	79	88	99

Tabela 9. Przyporządkowane parametry funkcji trapezoidalnej dla balansu.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słabe	30	37	43	47
Przeciętne	46	59	66	72
Szybkie	71	79	88	96

Tabela 10. Przyporządkowane parametry funkcji trapezoidalnej dla reakcji.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słaba	3	14	23	26
Przeciętna	25	39	48	57
Dobra	56	63	69	75
Bardzo dobra	74	81	88	99

Tabela 11. Przyporządkowane parametry funkcji trapezoidalnej dla kontroli piłki.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słabe	3	15	24	32
Zadowalające	31	44	57	66
Bardzo dobre	65	75	88	97

Tabela 12. Przyporządkowane parametry funkcji trapezoidalnej dla opanowania.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słaba	3	15	24	32
Przeciętna	31	44	57	66
Dobra	65	75	88	97

Tabela 13. Przyporządkowane parametry funkcji trapezoidalnej dla celności.

<b>Etykieta</b>	<b>a</b>	<b>b</b>	<b>c</b>	<b>d</b>
Słabe	3	15	24	32
Przeciętne	31	44	57	66
Dobre	65	75	88	97

Tabela 14. Przyporządkowane parametry funkcji trapezoidalnej dla ustawiania się.

### 4.3. Kwantyfikatory

Kwantyfikatory podzieliliśmy na względne i absolutne . Przedstawiamy je poniżej:

Etykieta	Funkcja przynależności	a	b	c	d
Żaden	Trójkątna	0	0	0.1	-
Mniej niż ćwierć	Trapezoidalna	0	0	0.25	0.35
Około jedna trzecia	Trójkątna	0.23	0.33	0.43	-
Około połowa	Trójkątna	0.4	0.5	0.6	-
Około dwie trzecie	Trójkątna	0.56	0.66	0.76	-
Większość	Trójkątna	0.73	0.83	0.93	-
Prawie każdy	Trapezoidalna	0.85	0.9	1	1.05

Tabela 15. Przyporządkowane parametry dla kwantyfikatora względnego.

Etykieta	Funkcja przynależności	a	b	c	d
Mniej niż 100	Trapezoidalna	0	0	99	150
Około 250	Trójkątna	150	250	350	-
Około 500	Trójkątna	400	500	600	-
Około 750	Trójkątna	650	750	850	-
Więcej niż 100	Trapezoidalna	950	1000	15397	15397

Tabela 16. Przyporządkowane parametry dla kwantyfikatora absolutnego.

## 5. Badania

Postanowiliśmy podzielić nasze badania na 3 części:

1. W pierwszym badaniu sprawdzaliśmy jakie wartości przybiorą miary podsumowań dla różnych kwantyfikatorów.
2. W drugim teście porównywaliśmy podsumowania z oraz bez kwalifikatora.
3. W trzecim badaniu naszym celem było porównanie podsumowań z jednym sumaryzatorem oraz ich połączeń spójnikami ORAZ i LUB.

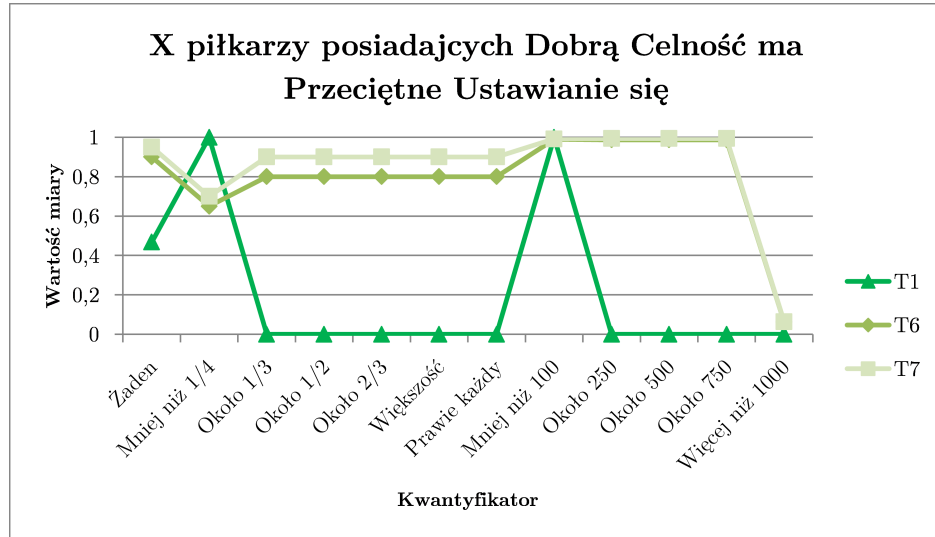
### 5.1. Pierwszy eksperyment

W tym badaniu wygenerowaliśmy 3 komunikaty, dotyczące odpowiednio zależności pomiędzy:

1. Dobrą Celnością a Przeciętym Ustawianiem się,
2. Słabą Prędkością a Zadowalającym Opanowaniem,
3. Młodym Wiekem a Dobrą Prędkością.

Miary  $T_2 - T_5$  oraz  $T_8 - T_{10}$  były stałe dla różnych kwantyfikatorów, dlatego nie umieszczaliśmy ich w tabeli oraz na wykresie, gdyż kwantyfikator nie miał wpływu na ich wartość. Zamieściliśmy ich wartości pod tabelą porównującą wartości miar  $T_1$ ,  $T_6$  i  $T_7$ .

### 5.1.1. Eksperyment 1.1



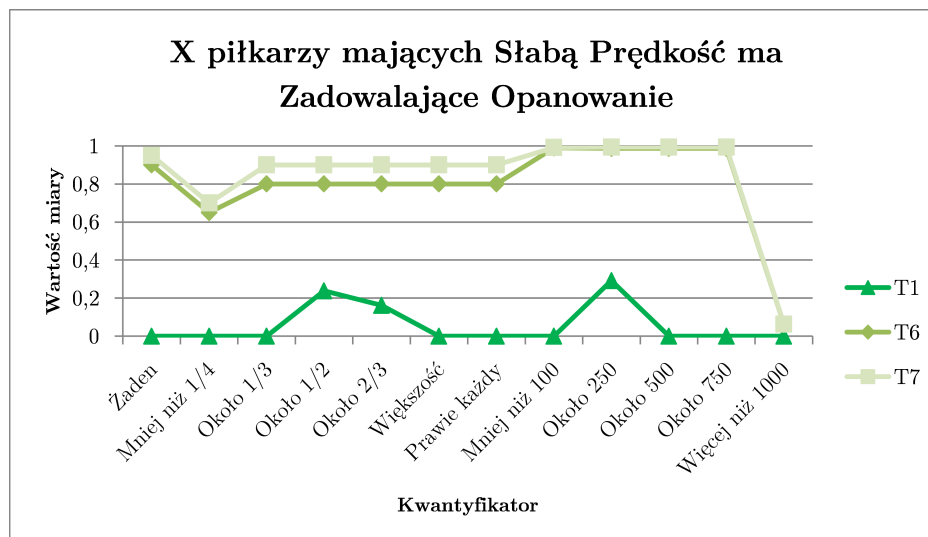
Rysunek 2. Wykres przedstawiający wyniki Eksperymentu 1.1

Kwantyfikator	$T_1$	$T_6$	$T_7$
Żaden	0.468	0.9	0.95
Mniej niż 1/4	1	0.65	0.7
Około 1/3	0	0.8	0.9
Około 1/2	0	0.8	0.9
Około 2/3	0	0.8	0.9
Większość	0	0.8	0.9
Prawie każdy	0	0.8	0.9
Mniej niż 100	1	0.99	0.992
Około 250	0	0.987	0.994
Około 500	0	0.987	0.994
Około 750	0	0.987	0.994
Więcej niż 1000	0	0.987	0.063

Tabela 17. Tabela przedstawiający wyniki Eksperymentu 1.1

Uzyskane miary, które były jednakowe dla każdego kwantyfikatora:  $T_2 = 0.431$ ,  $T_3 = 0.151$ ,  $T_4 = 0.263$ ,  $T_5 = 1$ ,  $T_8 = 0.998$ ,  $T_9 = 0.827$ ,  $T_{10} = 0.999$ ,  $T_{11} = 1$ .

### 5.1.2. Eksperyment 1.2



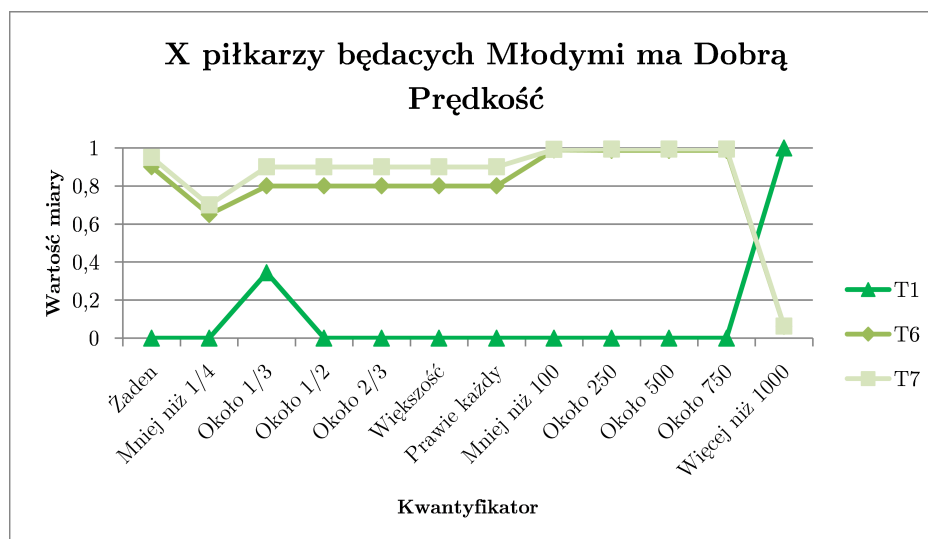
Rysunek 3. Wykres przedstawiający wyniki Eksperymentu 1.2

Kwantyfikator	$T_1$	$T_6$	$T_7$
Żaden	0	0.9	0.95
Mniej niż 1/4	0	0.65	0.7
Około 1/3	0	0.8	0.9
Około 1/2	0.238	0.8	0.9
Około 2/3	0.162	0.8	0.9
Większość	0	0.8	0.9
Prawie każdy	0	0.8	0.9
Mniej niż 100	0	0.99	0.992
Około 250	0.292	0.987	0.994
Około 500	0	0.987	0.994
Około 750	0	0.987	0.994
Więcej niż 1000	0	0.062	0.063

Tabela 18. Tabela przedstawiający wyniki Eksperymentu 1.2

Uzyskane miary, które były jednakowe dla każdego kwantyfikatora:  $T_2 = 0.318$ ,  $T_3 = 0.718$ ,  $T_4 = 0.211$ ,  $T_6 = 1$ ,  $T_8 = 0.998$ ,  $T_9 = 0.939$ ,  $T_{10} = 0.999$ ,  $T_{11} = 1$ .

### 5.1.3. Eksperyment 1.3



Rysunek 4. Wykres przedstawiający wyniki Eksperymentu 1.3

Kwantyfikator	$T_1$	$T_6$	$T_7$
Żaden	0	0.9	0.95
Mniej niż 1/4	0	0.65	0.7
Około 1/3	0.344	0.8	0.9
Około 1/2	0	0.8	0.9
Około 2/3	0	0.8	0.9
Większość	0	0.8	0.9
Prawie każdy	0	0.8	0.9
Mniej niż 100	0	0.99	0.992
Około 250	0	0.987	0.994
Około 500	0	0.987	0.994
Około 750	0	0.987	0.994
Więcej niż 1000	1	0.987	0.063

Tabela 19. Tabela przedstawiająca wyniki Eksperymentu 1.3

Uzyskane miary, które były jednakowe dla każdego kwantyfikatora:  $T_2 = 0.499$ ,  $T_3 = 0.565$ ,  $T_4 = 0.254$ ,  $T_6 = 1$ ,  $T_8 = 0.999$ ,  $T_9 = 0.347$ ,  $T_{10} = 1$ ,  $T_{11} = 1$ .

## 5.2. Drugi eksperyment

W tym badaniu wygenerowaliśmy 2 komunikaty, które skupiały się na:

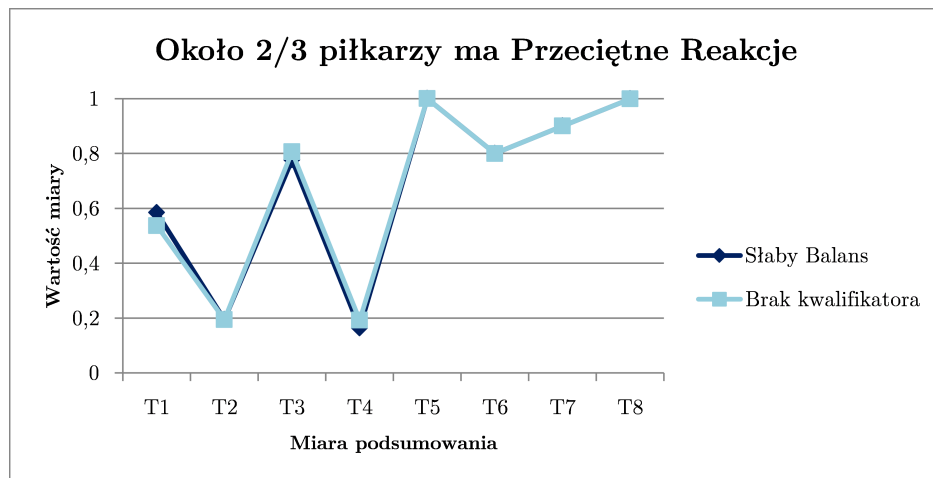
1. Przeciętnych Reakcjach,
2. Standardowej Wadze.

Porównywaliśmy wpływ obecności kwalifikatora (odpowiednio Słabego Bilansu oraz Niskiego Wzrostu) lub jego braku na wartości miar. Porównywaliśmy

zarówno jednakowe jak i różniące się kwantyfikatory, tak, by miara  $T_1$  nie wynosiła 0.

Miary  $T_9 - T_{11}$  zależały od kwalifikatora - nie dało się ich obliczyć bez niego, dlatego nie były porównywane w tym badaniu.

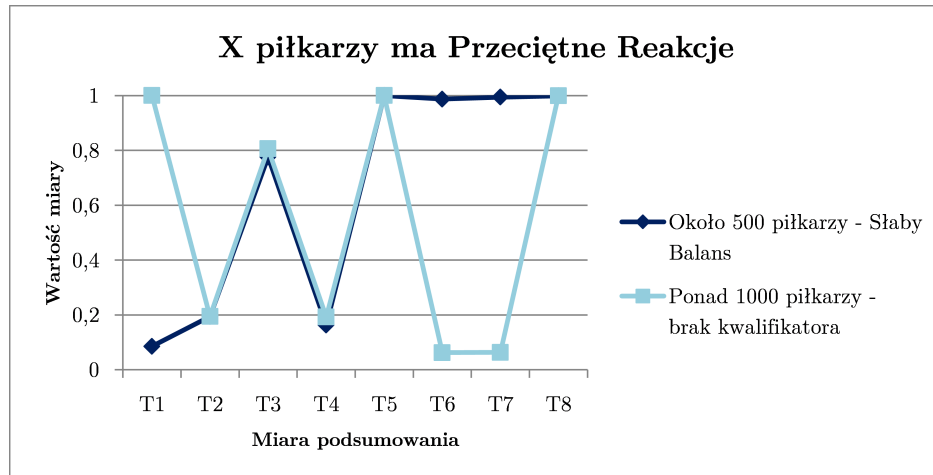
### 5.2.1. Eksperyment 2.1



Rysunek 5. Wykres przedstawiający wyniki Eksperymentu 2.1 - kwantyfikator względny

Miara	Słaby balans	Brak kwalifikatora
$T_1$	0.584	0.537
$T_2$	0.194	0.194
$T_3$	0.776	0.806
$T_4$	0.163	0.192
$T_5$	1	1
$T_6$	0.8	0.8
$T_7$	0.9	0.9
$T_8$	0.999	0.999

Tabela 20. Tabela przedstawiająca wyniki Eksperymentu 2.1 - kwantyfikator względny (około 2/3 piłkarzy)



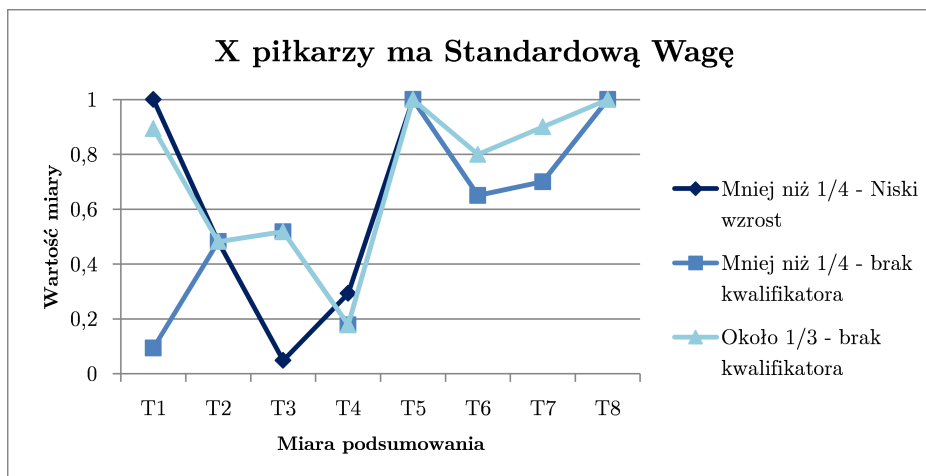
Rysunek 6. Wykres przedstawiający wyniki Eksperymentu 2.1 - kwantyfikator bezwzględny

Miara	Słaby balans	Brak kwalifikatora
	Okolo 500	Więcej niż 1000
$T_1$	0.537	1
$T_2$	0.194	0.194
$T_3$	0.776	0.806
$T_4$	0.163	0.192
$T_5$	1	1
$T_6$	0.987	0.062
$T_7$	0.994	0.063
$T_8$	0.999	0.999

Tabela 21. Tabela przedstawiająca wyniki Eksperymentu 2.1 - kwantyfikator bezwzględny



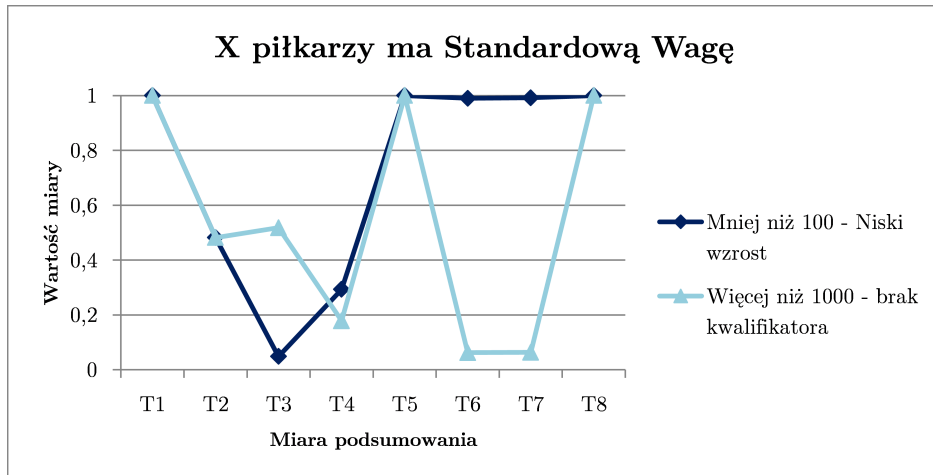
### 5.2.2. Eksperyment 2.2



Rysunek 7. Wykres przedstawiający wyniki Eksperymentu 2.2 - kwantyfikator względny

Miara	Niski Wzrost	Brak kwalifikatora	Brak kwalifikatora
	Mniej niż 1/4	Mniej niż 1/4	1/3
$T_1$	1	0.094	0.894
$T_2$	0.482	0.482	0.482
$T_3$	0.048	0.518	0.518
$T_4$	0.293	0.178	0.178
$T_5$	1	1	1
$T_6$	0.65	0.65	0.8
$T_7$	0.7	0.7	0.9
$T_8$	1	1	1

Tabela 22. Tabela przedstawiająca wyniki Eksperymentu 2.2 - kwantyfikator względny



Rysunek 8. Wykres przedstawiający wyniki Eksperymentu 2.2 - kwantyfikador bezwzględny

	Niski Wzrost	Brak kwalifikatora
Miara	Mniej niż 100	Więcej niż 1000
$T_1$	1	1
$T_2$	0.482	0.482
$T_3$	0.048	0.518
$T_4$	0.293	0.178
$T_5$	1	1
$T_6$	0.99	0.062
$T_7$	0.992	0.063
$T_8$	1	1

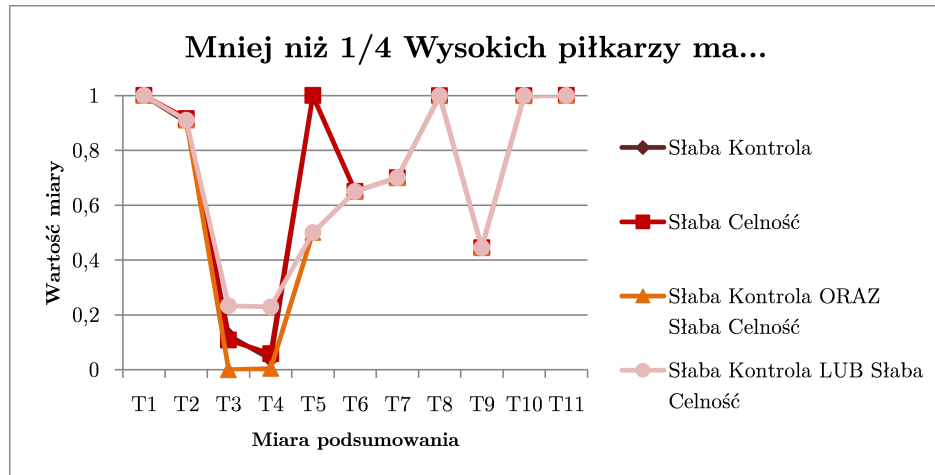
Tabela 23. Tabela przedstawiająca wyniki Eksperymentu 2.2 - kwantyfikador bezwzględny

### 5.3. Trzeci eksperyment

W tym badaniu wygenerowaliśmy komunikaty dotyczących wysokich piłkarzy, którzy mają:

1. Słabą Kontrolę,
2. Słabą Celność,
3. Słabą Kontrolę ORAZ Słabą Celność,
4. Słabą Kontrolę LUB Słabą Celność.

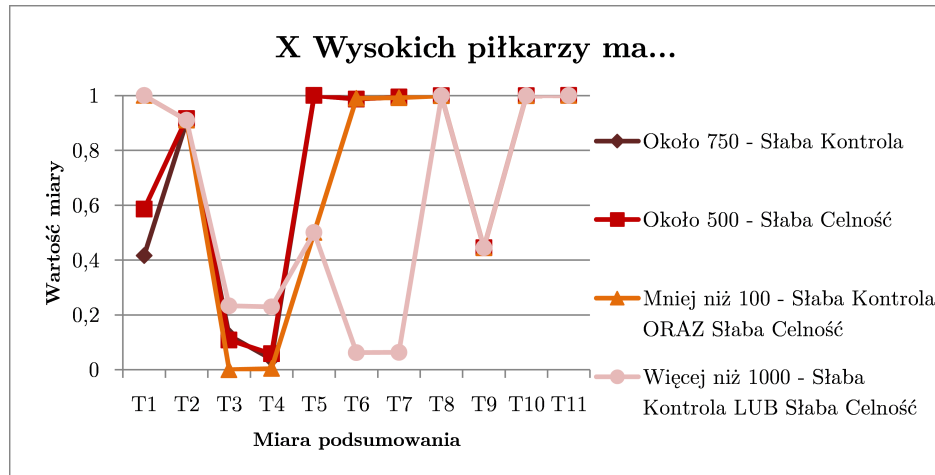
Porównywaliśmy zarówno jednakowe, jak i różniące się kwantyfikatory, dla których miara  $T_1$  była różna od 0.



Rysunek 9. Wykres przedstawiający wyniki Eksperymentu 3 - kwantyfikator względny

Miara	Słaba Kontrola	Słaba Celność	ORAZ	LUB
$T_1$	1	1	1	1
$T_2$	0.904	0.916	0.91	0.91
$T_3$	0.125	0.108	0	0.233
$T_4$	0.038	0.058	0.004	0,229
$T_5$	1	1	0.5	0.5
$T_6$	0.065	0.065	0.65	0.65
$T_7$	0.7	0.7	0.7	0.7
$T_8$	0.999	0.999	0.999	0.999
$T_9$	0.445	0.445	0.445	0.445
$T_{10}$	0.999	0.999	0.999	0.999
$T_{11}$	1	1	1	1

Tabela 24. Tabela przedstawiająca wyniki Eksperymentu 3 - kwantyfikator względny (mniej niż 1/4 piłkarzy)



Rysunek 10. Wykres przedstawiający wyniki Eksperymentu 3 - kwantyfikator bezwzględny

Miara	Slaba Kontrola	Slaba Celność	ORAZ	LUB
	Okolo 700	Okolo 500	Mniej niż 100	Więcej niż 1000
$T_1$	0.416	0.586	1	1
$T_2$	0.904	0.916	0.91	0.91
$T_3$	0.125	0.108	0	0.233
$T_4$	0.038	0.058	0.004	0.229
$T_5$	1	1	0.5	0.5
$T_6$	0.987	0.987	0.99	0.062
$T_7$	0.994	0.994	0.992	0.063
$T_8$	0.999	0.999	0.999	0.999
$T_9$	0.445	0.445	0.445	0.445
$T_{10}$	0.999	0.999	0.999	0.999
$T_{11}$	1	1	1	1

Tabela 25. Tabela przedstawiająca wyniki Eksperymentu 3 - kwantyfikator bezwzględny

## 6. Dyskusja

### 6.1. Wpływ kwantyfikatora na miary

Miara  $T_1$  obrazująca prawdziwość danego podsumowania, zmieniała się wraz z rozważanym kwantyfikatorem - dla większości będąc równa 0. Obrazuje ona nie tylko jak dużo piłkarzy spełnia dane kryterium, ale pozwala też szacować jak duża jest grupa, która podsumowanej cechy nie spełnia. Dla eksperymentu 1.1 widać jak mało jest piłkarzy którzy posiadając Dobrą Celność, cechują się też Przeciętnym Ustawianiem się. Piłkarze cechujący się Słabą Prędkością w dużej części posiadają Zadowalające Opanowanie -  $T_1$  jest niskie dla miar około  $1/2$  oraz około  $2/3$ , dzięki czemu wiemy, że właściwa grupa byłaby opisana kwantyfikatorem zawierającym się w przedziale  $(\frac{1}{2}, \frac{2}{3})$ .

Liczba ta oscyluje wokół 250 piłkarzy. Sprawa wygląda inaczej dla Młodych piłkarzy - około 1/3 z nich ma Dobrą Prędkość, jednak tym razem jest to ponad 1000 piłkarzy.

Miary  $T_6$  i  $T_7$  opisują sam kwantyfikator, a dokładnie jego precyzyjność. Wyraźnie widać, że miara mniej niż 1/4 opisuje większy zbiór niż reszta zaproponowanych przez nas kwantyfikatorów względnych. Bardzo nieprecyzyjny jest kwantyfikator bezwzględny więcej niż 1000. Wartości tych miar są identyczne dla wszystkich podsumowań i zależą tylko od kwantyfikatora.

Miary  $T_2 - T_5$  oraz  $T_8 - T_{11}$  nie zmieniały swoich wartości w zależności od użytego kwantyfikatora, co nie jest zaskakujące, mając na uwadze braku wykorzystania kwantyfikatora w tych wzorach.

## 6.2. Wpływ kwalifikatora na miary

Cieężko jest porównać miarę  $T_1$  dla kwalifikatora lub jego braku - pracujemy wtedy na innych zbiorach. Wartości dla kwantyfikatorów względnych mogą się zmienić nieznacznie bądź drastycznie. Prawie na pewno zmieniają się wartości dla kwantyfikatorów bezwzględnych (zakładając, że są odpowiednio zdefiniowane). Porównując wyniki skupiliśmy się tylko na tych, gdzie wartości miary  $T_1$  były większe od zera, przez co miary  $T_6$  i  $T_7$  potrafią się różnić, co jednak nie ma nic wspólnego z kwalifikatorem.

Miara  $T_3$  mówi nam o tym, jak bardzo nośnik kwalifikatora pokrywają się z nośnikiem sumaryzatora. Im wyższa wartość jest tej cechy, tym kwalifikator i sumaryzator są ze sobą mniej powiązane. Zarówno Słaby Balans jak i brak kwalifikatora są mocno powiązane z Przeciętnymi Reakcjami, nie można jednak tego powiedzieć o Standardowej Wadze oraz braku kwalifikatora. Prawie w ogóle nie są ze sobą powiązane Niski Wzrost oraz Standardowa Waga (o czym także mówi kwantyfikator mniej niż 100, dla którego miara  $T_1$  jest równa 1). Im wyższa jest ta wartość, tym większy jest kwantyfikator.

Miara  $T_4$  mówi o tym, jak wiele krotek przynależy do sumaryzatora. Ze względu na  $T_3$  we wzorze, jej wartość się zmienia, jednak wpływie tej wartości przyjrzymy się w następnej części omówienia eksperymentów.

Miary  $T_2, T_5$  oraz  $T_8$  nie zmieniały swoich wartości w obrębie jednego eksperymentu, gdyż zależą od sumaryzatora, a nie kwalifikatora.

## 6.3. Wpływ sumaryzatora na miary

Miara  $T_2$  która opisuje stopień nieprecyzyjności sumaryzatora. Nie zależy ona od typu połączenia, przez co dla obu złożonych podsumowań wartość miary jest średnią dla wartości miar z pojedynczym sumaryzatorem. Mając na uwadze nie tylko ten eksperyment, ale i wszystkie poprzednie, można powiedzieć, że miara ta jest tym większa, im bardziej precyzyjny jest użyty sumaryzator. Na przykład Standardowa Waga nie jest tak precyzyjnym su-

maryzatorom jak Słaba Kontrola czy Słaba Celność.

Miara  $T_3$  pokazuje, że dla połączonych spójnikiem "lub" sumaryzatorów powiązanie z kwalifikatorem jest większa, zaś dla spójnika "i" wartość ta w przypadku tego podsumowania spada do 0.

Miara  $T_4$  mówi o tym ile krotek należy do sumaryzatora. Określa jak bardzo podsumowane wartości pasują do zestawu danych. Dla stwierdzenia "Mniej niż 100 Wysokich piłkarzy ma Słabą Kontrolę oraz Słabą Celność" miara  $T_4$  jest bardzo mała, co wyraźnie mówi, że podsumowanie to nie pasuje do zestawu danych.

Miara  $T_5$  zależy od ilości sumaryzatorów. Dla jednego sumaryzatora osiąga wartość 1, zaś dla dwóch - wartość 0.5. Wartość tej miary wraz z zwiększeniem liczby sumaryzatorów wciąż by malała.

Miara  $T_8$  opisująca precyzyjność sumaryzatora dla wszystkich przeprowadzonych przez nas eksperymentów uzyskiwała wartości bardzo bliskie bądź równe 1.

#### 6.4. Inne spostrzeżenia

Miara  $T_9$  opisująca precyzyjność kwalifikatora mówi o tym jak precyzyjny jest kwalifikator. Dla wysokich wartości tej miary, kwantyfikatory są niskie (zakładając, że kwantyfikatory te osiągają wartość większą niż 0 dla  $T_1$ ).

Miara  $T_{10}$  opisująca stopień precyzji kwalifikatora dla wszystkich przeprowadzonych przez nas eksperymentów uzyskiwała wartości bardzo bliskie bądź równe 1, podobnie jak miara  $T_8$ .

Z racji zastosowania pojedynczego kwalifikatora, miara  $T_{11}$  zawsze osiągała wartość 1. W przypadku naszego programu, nie wnosi ona żadnej informacji. Miary  $T_5, T_8$  i  $T_{10}$  także nie wnoszą istotnych informacji.

### 7. Wnioski

- Należy dobierać wąskie wartości przy funkcjach przynależności, aby uzyskać miarodajne wyniki.
- Parametry dla kwantifikatora absolutnego muszą być dobrze dobrane do typu i rodzaju bazy - baza składająca się z dużej liczby elementów i dużej liczby atrybutów może potrzebować małych wartości bezwzględnych.
- Miara  $T_1$  jest niezbędna. Dopiero gdy ona osiąga wartość większą od 0, powinniśmy analizować pozostałe miary.
- Dla bardzo precyzyjnych podsumowań, wartości miar  $T_2, T_6, T_7$  oraz  $T_9$  powinny być jak najwyższe.
- Miary  $T_5$  oraz  $T_{10}$  są bardzo proste w analizie - im więcej sumaryzatorów/kwalifikatorów, tym wartość takiej miary jest mniejsza. Sprawia to, że nie zawsze są miarodajne.

- Miary  $T_8$  i  $T_{10}$  nie sprawdzają się przy analizie dla zapytań z małą liczbą sumaryzatorów oraz kwalifikatorów.

## Literatura

- [1] Methods for the linguistic summarization of data - applications of fuzzy sets and their extensions, Adam Niewiadomski, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2008
- [2] <http://www.cs.put.poznan.pl/amichalski/si.dziennie/AI7.new.fuzzy.b&w.pdf>
- [3] [http://home.agh.edu.pl/~mrzyglod/iw/iw\\_pliki/iw-is-L2-2017-2018.pdf](http://home.agh.edu.pl/~mrzyglod/iw/iw_pliki/iw-is-L2-2017-2018.pdf)
- [4] <https://www.kaggle.com/aishwarya1992/fifa-19-player-database>