

Studenckie Koło Naukowe Math4You  
Wydział Informatyki Politechniki Białostockiej

# Zbiory przybliżone

## Polska szkoła sztucznej inteligencji

Jan Gromko

22 kwietnia 2017 r.



## Wprowadzenie

- Czym są zbiory przybliżone – historia i idea
- Podstawowe pojęcia

## Problem redukcji

- Istota problemu redukcji
- Prosty algorytm redukcji
- Problem złożoności wyznaczania reduktów
- Przybliżone metody wyznaczania reduktu

## Znaczenie zbiorów przybliżonych

- Możliwości i zalety
- Zastosowania

- ▶ Teoria zaproponowana w 1982 r. przez prof. Zdzisława Pawlaka.
- ▶ Wprowadzona jako nowe matematyczne podejście do pojęć nieostrych i metoda analizy danych.

- ▶ Zbiory przybliżone oparte są o logikę trójwartościową;
- ▶ zbiór przybliżony jest zbiorem niedefiniowalnym – nie można go jednoznacznie scharakteryzować na podstawie własności jego elementów.

<i>Pacjent</i>	<i>Ból głowy</i>	<i>Ból mięśni</i>	<i>Temperatura</i>	<i>Grypa</i>
1	nie	tak	podwyższona	tak
2	tak	nie	podwyższona	tak
3	tak	tak	wysoka	tak
4	nie	tak	normalna	nie
5	tak	nie	podwyższona	nie
6	nie	nie	wysoka	tak

Tabela 1. Tablica decyzyjna przykładowego zbioru.

$S = (U, A, V, f)$  – system informacyjny

$U$  – zbiór obiektów (*uniwersum*)

$A$  – zbiór atrybutów

$V = \bigcup_{a \in A} V_a$  – zbiór wszystkich możliwych wartości atrybutów

$V_a$  – dziedzina atrybutu  $a \in A$

$f : U \times A \rightarrow V$  – funkcja informacyjna

$DT = (U, C, D, V, f)$  – tablica decyzyjna

$C$  – zbiór atrybutów warunkowych

$D$  – zbiór atrybutów decyzyjnych

$A = C \cup D$  – zbiór atrybutów



$U$  – pacjenci

$C = \{\textit{ból głowy}, \textit{ból mięśni}, \textit{temperatura}\}$

$D$  – grypa



## **Problem:**

Znaleźć zależność między występowaniem/niewystępowaniem grypy a symptomami występującymi u pacjentów, czyli znaleźć zależność między atrybutem decyzyjnym a wartościami atrybutów warunkowych, opisujących poszczególne obiekty.

<i>Pacjent</i>	<i>Ból głowy</i>	<i>Ból mięśni</i>	<i>Temperatura</i>	<i>Grypa</i>
2	tak	nie	podwyższona	tak
5	tak	nie	podwyższona	nie

Tabela 2. Sprzeczne informacje w zbiorze – przypadki, których nie można jednoznacznie sklasyfikować.

## Relacja nierozróżnialności

Relację nierozróżnialności zdefiniowana jest jako

$$I(B) = \{(x, y) \in U \times U : \forall a \in B \ f(a, x) = f(a, y)\},$$

gdzie  $B \subseteq A$ .

Jeśli  $(x, y) \in I(B)$ , wówczas obiekty  $x$  i  $y$  są nierozróżnialne ze względu na podzbiór atrybutów  $B$ .

**W oparciu o posiadane dane, można stwierdzić, że:**

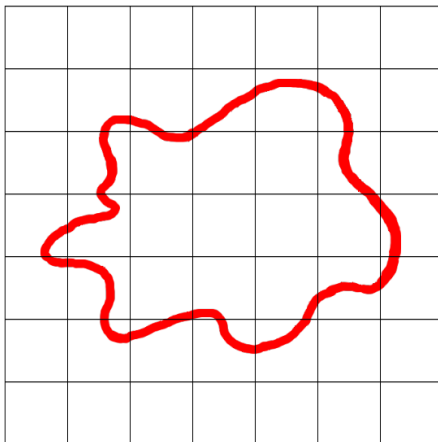
- ▶  $\{1, 3, 6\}$  to zbiór przypadków, które (na podstawie atrybutów warunkowych) możemy *jednoznacznie* zaklasyfikować do grupy pacjentów chorych na gripę;
- ▶  $\{1, 2, 3, 5, 6\}$  to zbiór przypadków, które *mogą* być zakwalifikowane jako pacjenci chorzy na gripę;
- ▶  $\{2, 5\}$  to zbiór przypadków, które nie mogą być jednoznacznie zaklasyfikowane jako pacjenci, którzy są lub nie są chorzy na gripę.

## Dolne przybliżenie

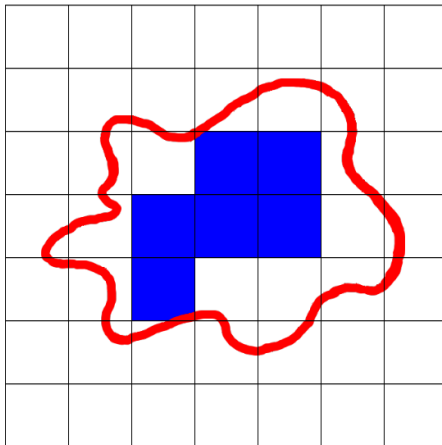
Wszystkie te elementy, które można jednoznacznie zaklasyfikować do danego zbioru, według posiadanej wiedzy na ich temat.

## Górne przybliżenie

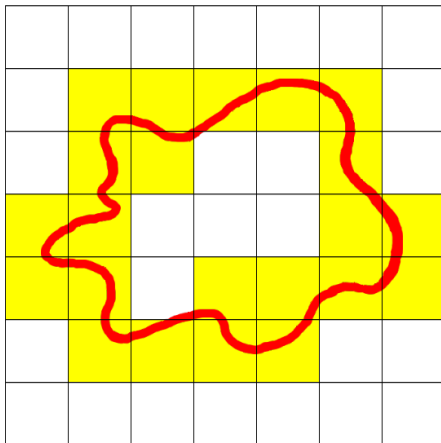
Wszystkie te elementy, których przynależności do danego zbioru nie można wykluczyć.



Rysunek 1. Przykładowy zbiór.

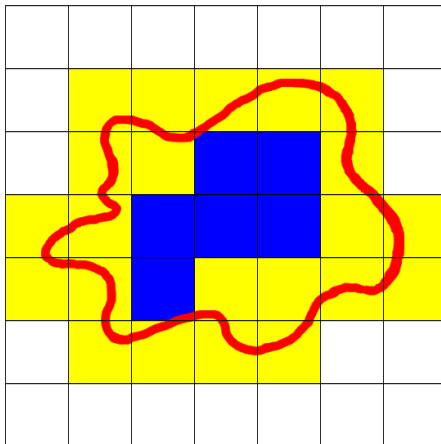


Rysunek 2. Dolne przybliżenie zbioru.



Rysunek 3. Obszar brzegowy zbioru.





Rysunek 4. Górne przybliżenie zbioru.

**Czy można zredukować zbiór pod względem atrybutów  
w ten sposób, by zachowana była rozróżnialność elementów  
z oryginalnego zbioru?**

	1	2	3	4	5	6
1	$\emptyset$	—	—	—	—	—
2	$\emptyset$	$\emptyset$	—	—	—	—
3	$\emptyset$	$\emptyset$	$\emptyset$	—	—	—
4	$t$	$g, m, t$	$g, t$	$\emptyset$	—	—
5	$g, m$	$\emptyset$	$m, t$	$\emptyset$	$\emptyset$	—
6	$\emptyset$	$\emptyset$	$\emptyset$	$m, t$	$g, t$	$\emptyset$

Tabela 3. Macierz rozróżnialności.

$g$  – ból głowy;  $m$  – ból mięśni;  $t$  – temperatura

<i>Pacjent</i>	<i>Ból głowy</i>	<i>Ból mięśni</i>	<i>Temperatura</i>	<i>Grypa</i>
1	nie	tak	podwyższona	tak
4	nie	tak	normalna	nie

Tabela 4. Fragment tablicy decyzyjnej.

	1	2	3	4	5	6
4	t	?	?	∅	—	—

Tabela 5. Fragment macierzy rozróżnialności.

<i>Pacjent</i>	<i>Ból głowy</i>	<i>Ból mięśni</i>	<i>Temperatura</i>	<i>Grypa</i>
2	tak	nie	podwyższona	tak
4	nie	tak	normalna	nie

Tabela 6. Fragment tablicy decyzyjnej.

	1	2	3	4	5	6
4	t	g, m, t	?	∅	—	—

Tabela 7. Fragment macierzy rozróżnialności.

<i>Pacjent</i>	<i>Ból głowy</i>	<i>Ból mięśni</i>	<i>Temperatura</i>	<i>Grypa</i>
3	tak	tak	wysoka	tak
4	nie	tak	normalna	nie

Tabela 8. Fragment tablicy decyzyjnej.

	1	2	3	4	5	6
4	t	g, m, t	g, t	∅	–	–

Tabela 9. Fragment macierzy rozróżnialności.

	1	2	3	4	5	6
1	$\emptyset$	—	—	—	—	—
2	$\emptyset$	$\emptyset$	—	—	—	—
3	$\emptyset$	$\emptyset$	$\emptyset$	—	—	—
4	t	g, m, t	g, t	$\emptyset$	—	—
5	g, m	$\emptyset$	m, t	$\emptyset$	$\emptyset$	—
6	$\emptyset$	$\emptyset$	$\emptyset$	m, t	g, t	$\emptyset$

Tabela 10. Macierz rozróżnialności.

	1	2	3	4	5	6
1	$\emptyset$	—	—	—	—	—
2	$\emptyset$	$\emptyset$	—	—	—	—
3	$\emptyset$	$\emptyset$	$\emptyset$	—	—	—
4	t	g, t	g, t	$\emptyset$	—	—
5	g	$\emptyset$	t	$\emptyset$	$\emptyset$	—
6	$\emptyset$	$\emptyset$	$\emptyset$	t	g, t	$\emptyset$

Tabela 11. Macierz rozróżnialności po redukcji.



- ▶ Szukanie zależności między danymi,
- ▶ redukcja zbiorów danych,
- ▶ określenie wagi danych,
- ▶ generowanie reguł decyzyjnych.

- ▶ Teoria ZP nie wymaga założeń na temat danych, takich jak prawdopodobieństwo czy rozmytość,
- ▶ szybkie algorytmy analizy danych,
- ▶ łatwa interpretacja wyników,
- ▶ matematyczna prostota.

- ▶ Medycyna,
- ▶ farmakologia,
- ▶ bankowość,
- ▶ lingwistyka,
- ▶ rozpoznawanie mowy,
- ▶ ochrona środowiska,
- ▶ bazy danych.

Ograniczenie liczby badań medycznych do jedynie tych, które są naprawdę konieczne do rozpoznania choroby.

- ▶ Zmniejszenie ryzyka powikłań u pacjenta,
- ▶ zmniejszenie kosztów badań.

- [1] Zdzisław Pawlak  
Zbiory przybliżone – nowa matematyczna metoda analizy danych
- [2] Leszek Rutkowski  
Metody i techniki sztucznej inteligencji

# Pytania

