

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 wyznaczania reduktu
- Alternatywne metody wyznaczania reduktu

Znaczenie i kierunki rozwoju 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

<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 2. Tablica decyzyjna przykładowego zbioru.

U – pacjenci

$C = \{\text{ból głowy, ból mięśni, temperatura}\}$, 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 3. Sprzeczne informacje w zbiorze – przypadki, których nie można jednoznacznie sklasyfikować.

Relacja nierozróżnialności

Relację nierozróżnialności zdefiniowana jest jako

$$IND(B) = \{(x, y) \in U \times U : \forall_{a \in B} a(x) = a(y)\},$$

gdzie $B \subseteq A$.

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

W oparciu o posiadane dane, można stwierdzić, że:

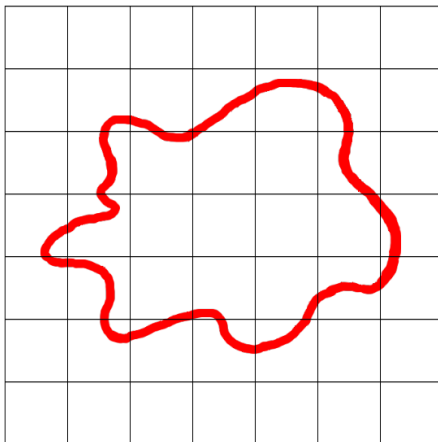
- ▶ $\{1, 3, 6\}$ to zbiór przypadków, które (na podstawie atrybutów warunkowych) możemy *jednoznacznie* zaklasyfikować do grupy pacjentów chorych na gripę;
- ▶ $\{1, 2, 3, 5, 6\}$ to zbiór przypadków, które *mogą* być zakwalifikowane jako pacjenci chorzy na gripę;
- ▶ $\{2, 5\}$ to zbiór przypadków,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

Dolne przybliżenie

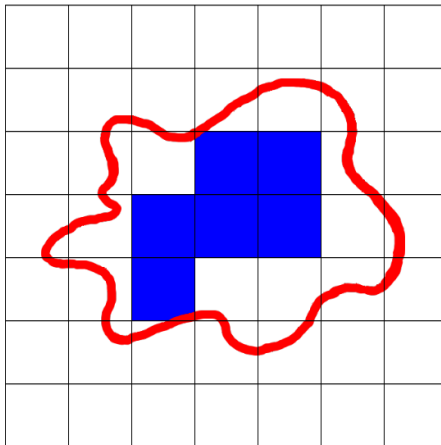
Wszystkie te elementy, które można jednoznacznie zaklasyfikować do danego zbioru, według posiadanej wiedzy na ich temat.

Górne przybliżenie

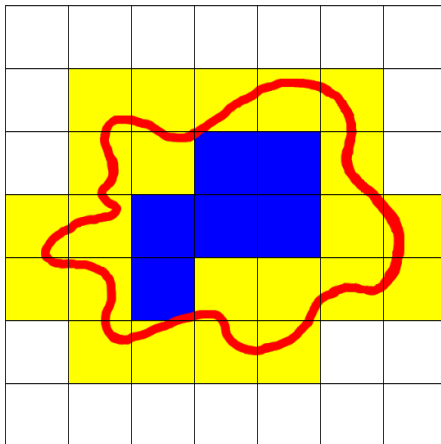
Wszystkie te elementy, których przynależności do danego zbioru nie można wykluczyć.



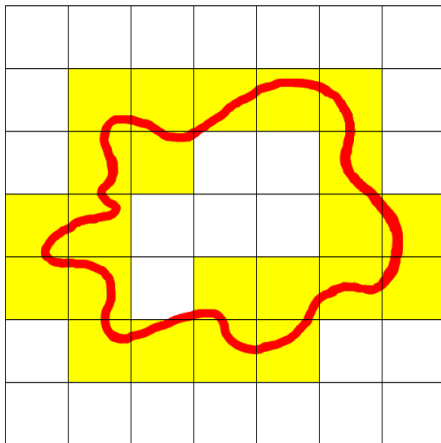
Rysunek 1. Przykładowy zbiór.



Rysunek 2. Dolne przybliżenie zbioru.



Rysunek 3. Górne przybliżenie zbioru.



Rysunek 4. Obszar brzegowy zbioru.

- ▶ Zbiory przybliżone oparte są o logikę trójwartościową;
- ▶ zbiór przybliżony jest zbiorem niedefiniowalnym – nie można go jednoznacznie scharakteryzować na podstawie własności jego elementów;
- ▶ **zbiór przybliżony można jednak scharakteryzować za pomocą dwóch zbiorów definiowalnych – dolnego i górnego przybliżenia.**

Problem redukcji

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

Zbiór niezależny

Zbiór atrybutów $B_1 \subset A$ jest *niezależny* w danym systemie informacyjnym, jeśli dla każdego $B_2 \subset B_1$ zachodzi $IND(B_1) \neq IND(B_2)$.

Redukt

Reduktem zbioru atrybutów $B_1 \subseteq A$ nazywamy każdy niezależny zbiór $B_2 \subseteq B_1$, dla którego $IND(B_1) = IND(B_2)$, przy czym B_2 powinien być jak najmniej liczny. Może istnieć wiele reduktów.

- ▶ W każdej komórce macierzy rozróżnialności umieszczane są atrybuty rozróżniające każde dwa obiekty,
- ▶ rozróżniamy jedynie obiekty należące do różnych klas decyzyjnych,
- ▶ wystarczające jest wypełnienie jedynie połowy macierzy (dzielonej według głównej przekątnej), ponieważ jeśli obiekt x_1 jest rozróżnialny od atrybutu x_2 przez zbiór atrybutów K , to obiekt x_2 jest rozróżnialny od atrybutu x_1 przez ten sam zbiór atrybutów.

	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 4. 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
5	tak	nie	podwyższona	nie

Tabela 5. Fragment tablicy decyzyjnej.

	1	2	3	4	5	6
5	g, m	?	?	∅	∅	—

Tabela 6. 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
5	tak	nie	podwyższona	nie

Tabela 7. Fragment tablicy decyzyjnej.

	1	2	3	4	5	6
5	g, m	∅	?	∅	∅	—

Tabela 8. 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
5	tak	nie	podwyższona	nie

Tabela 9. Fragment tablicy decyzyjnej.

	1	2	3	4	5	6
5	g, m	∅	m, t	∅	∅	–

Tabela 10. 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 11. 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 12. Macierz rozróżnialności po redukcji.

1. Zliczenie wystąpień atrybutów w macierzy rozróżnialności.
2. Wybór atrybutu występującego najliczniej w macierzy rozróżnialności; dodanie wybranego atrybutu do wynikowego zbioru atrybutów *Red*.
3. Wykreślenie komórek zawierających wybrany atrybut.
4. Jeśli wszystkie komórki zostały wykreślone, wynikiem jest uzyskany zbiór *Red*, w przeciwnym razie powrót do kroku 1.

	1	2	3	4	5	6
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 13. Fragment macierzy rozróżnialności zawierający istotne dane.

g – 4

m – 4

t – 6

	1	2	3	4	5	6
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 14. Fragment macierzy rozróżnialności zawierający istotne dane.

$$Red = \{t\}$$

$$g - 1$$

$$m - 1$$

	1	2	3	4	5	6
4	t	g , m, t	g , t	∅	–	–
5	g , m	∅	m , t	∅	∅	–
6	∅	∅	∅	m , t	g , t	∅

Tabela 15. Fragment macierzy rozróżnialności zawierający istotne dane.

$$Red = \{t, g\} \vee Red = \{t, m\}$$

	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 16. Macierz rozróżnialności oryginalnego zbioru.

$$Red = \{t, g\} \vee Red = \{t, m\}$$

Wyznaczanie reduktu w zbiorze przybliżonym jest problemem NP-zupełnym – nie jest możliwe znalezienie rozwiązania w czasie wielomianowym.

Rozwiązania sprzętowe:

- ▶ specjalizowane układy programowalne.

Rozwiązania przybliżone – wykorzystanie innych metod sztucznej inteligencji:

- ▶ algorytmy ewolucyjne,
- ▶ algorytmy mrówkowe,
- ▶ inteligencja roju,
- ▶ metody połączone.

- ▶ Opis zbiorów przybliżonych z wykorzystaniem sieci komórkowych (*cellular networks*; na tej podstawie w 1994 r. opracowana została koncepcja PRSComp (*Parallel Rough Sets COMPUter*) – Rybiński, Muraszkiewicz (1984);
- ▶ idea samouczącego się systemu, wykorzystującego sieci komórkowe – uczenie systemu w warstwie oprogramowania, na podstawie wyników tworzona jest warstwa sprzętowa – Lewis, Perkowski, Jóźwiak (1999);
- ▶ implementacja systemu opartego o sieci komórkowe – Dharmadhikari, Ngo, Lewisa (1999);
- ▶ koncepcja RSP (*Roug Set Processor*) – Pawlak (2004);

- ▶ idea systemu wspierającego w sposób sprzętowy minimalizację funkcji logicznych, tworzonych na podstawie macierzy rozróżnialności – Kanasugi, Yokoyama (2001);
- ▶ implementacja wyżej wymienionego rozwiązania w FPGA, 700-krotne skrócenie czasu obliczeń w porównaniu do tradycyjnych rozwiązań – Kanasugi, Matsumoto (2007);
- ▶ koprocesor wspomagający sprzętowo niektóre operacje na ZP – Tiwari, Kothari (2015);
- ▶ implementacje sprzętowe na FPGA i CPLD metod ZP od operacji podstawowych (jak np. obliczanie aproksymacji), przez wyznaczanie rdzeni i reduktów, po generowanie reguł decyzyjnych – Kopczyński (2016).

- ▶ Algorytmy genetyczne, oparte o sprowadzenie problemu redukcji do problemu pokrycia kolumnowego macierzy, w której kolumny odpowiadają atrybutom, a wiersze parom obiektów, które wymagają rozróżnienia – Wróblewski (2001);
- ▶ redukcja ZP przy pomocy algorytmów mrówkowych – Jensen, Shen (2003);
- ▶ algorytm hybrydowy, łączący algorytmy mrówkowe i inteligencję roju – Pratiwi, Choo, Muda (2011);

- ▶ propozycja algorytmu genetycznego opartego o kodowanie chromosomów jako zbioru atrybutów w potencjalnym redukcje – Zhengjiang, Jingmin, Yan (2012);
- ▶ propozycja ulepszenia wyżej wymienionego algorytmu – atrybuty należące do rdzenia chronione są przed mutacją, selekcja osobników na podstawie znormalizowanego współczynnika istotności – Chen, Liu, Wan (2014).

- ▶ 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,
- ▶ ogólniej – przetwarzanie dużych zbiorów 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
- [3] Jakub Wróblewski
Adaptacyjne metody klasyfikacji obiektów
- [4] Maciej Kopczyński
Wspomaganie decyzji oparte na sprzętowej realizacji metod zbiorów przybliżonych

Pytania

