Uniwersytet
Ekonomiczny
w Katowicach

blisko

międzynarodowo

przez całe życie

LOGIKA DLA INFORMATYKÓW

Zbiory przybliżone

dr hab. Ewa Michalska

Zbiory przybliżone

Zdzisław Pawlak (1926-2006) polski matematyk, informatyk twórca teorii zbiorów przybliżonych sformułowanej w 1982 roku (stanowiącej rozwinięcie klasycznej teorii zbiorów). Teoria zbiorów przybliżonych jest metodą analizy danych.

- Porównując dwa dowolne obiekty (nawet jeśli są bardzo podobne) zawsze możemy znaleźć między nimi różnice, zwłaszcza jeśli uwzględnimy dostatecznie dużo cech (atrybutów), z dostatecznie dużą dokładnością.
- Jeśli zmniejszymy dokładność opisu (ograniczymy się np. tylko do cech istotnych dla danego zastosowania), wówczas obiekty wcześniej rozróżnialne staną się nierozróżnialne.
- Zbiory przybliżone powstały (po odrzuceniu wymogu istnienia ściśle określonych granic) jako zbiory definiowane za pomocą dolnego i górnego przybliżenia zbioru.

Teorie zbiorów

- W klasycznej teorii zbiorów (teorii mnogości), zbiór jest definiowany przez swoje elementy, nie jest potrzebna żadna dodatkowa wiedza o elementach uniwersum, z których tworzymy zbiór.
- W teorii zbiorów przybliżonych (*rough set theory*), zakładamy, że mamy pewne informacje o elementach uniwersum i dane te są wykorzystywane do tworzenia zbiorów.
- W teorii zbiorów rozmytych (*fuzzy set theory*), elementy mogą należeć do zbioru w pewnym stopniu, a nie definitywnie jak ma to miejsce w klasycznej teorii zbiorów.

Teoria zbiorów przybliżonych

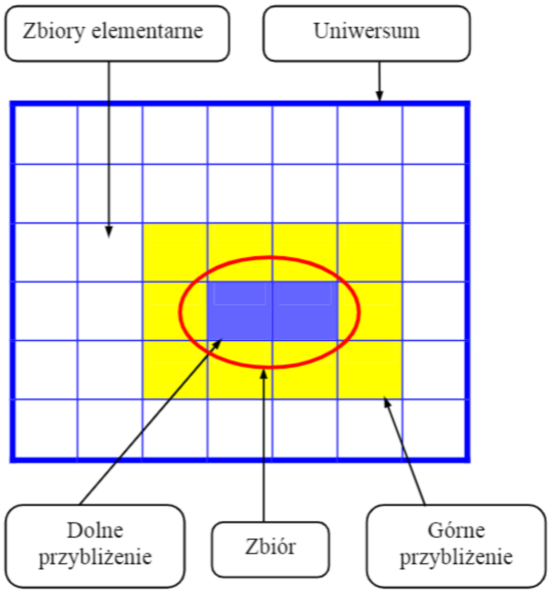
- Elementy, o których mamy identyczną informację są podobne i tworzą tzw. **zbiory elementarne**, stanowiące podstawę rozumowań w teorii zbiorów przybliżonych.
- Suma dowolnych zbiorów elementarnych jest nazywana **zbiorem definiowalnym**.
- Zbiory, które nie są zbiorami definiowalnymi nazywamy **zbiorami przybliżonymi**.

UWAGA:
Zbiorów przybliżonych nie można jednoznacznie scharakteryzować przez własności ich elementów. Dlatego w teorii zbiorów przybliżonych wprowadza się pojęcia dolnego i górnego przybliżenia zbioru, które pozwalają każdy zbiór niedefiniowalny (przybliżony) scharakteryzować za pomocą dwóch zbiorów definiowalnych – dolnego i górnego przybliżenia.

Teoria zbiorów przybliżonych

- **Dolnym przybliżeniem zbioru** są wszystkie elementy, które (w świetle posiadanej wiedzy) mogą być jednoznacznie zaklasyfikowane do rozważanego zbioru.
- **Górnym przybliżeniem zbioru** są wszystkie elementy, które (w świetle posiadanej wiedzy) nie mogą być wykluczone jako elementy rozważanego zbioru.
- Różnica między górnym a dolnym przybliżeniem jest nazywana **obszarem brzegowym** (brzegiem) zbioru.

Dany zbiór jest zbiorem przybliżonym wtedy i tylko wtedy, gdy jego obszar brzegowy jest niepusty.



Główne rodzaje wnioskowania

Wnioskowanie dedukcyjne – wnioskowanie dedukcyjne daje narzędzia służące do wyprowadzania zdań prawdziwych z innych zdań prawdziwych. Wnioskowanie dedukcyjne prowadzi zawsze do konkluzji prawdziwych. Wnioskowanie to jest stosowane w rozumowaniach na gruncie matematyki czy logiki. Rozstrzygania prawdziwości hipotez dokonuje się drogą formalnego rozumowania.

Wnioskowanie indukcyjne – punktem wyjścia do tego typu rozumowań są pewne fakty częściowe o badanej rzeczywistości (przykłady), które są następnie uogólniane, tworząc wiedzę o szerszym świecie, niż ten który stanowił punkt wyjścia wnioskowania. Wnioskowanie indukcyjne nie prowadzi do wniosków prawdziwych, a jedynie do wniosków prawdopodobnych (możliwych). Rozstrzygania prawdziwości hipotez dokonuje się na podstawie eksperymentu (np. w fizyce).

Główne cechy wnioskowania

Wnioskowanie dedukcyjne:
- zastosowania: matematyka, logika
- pełna teoria
- wnioskowanie zawsze prawdziwe
- weryfikacja hipotez – dowód

Wnioskowanie indukcyjne:
- zastosowania: nauki przyrodnicze i techniczne
- częściowe teorie
- wnioski prawdopodobne (możliwe)
- weryfikacja hipotez – eksperyment

Teoria zbiorów przybliżonych a praktyka

Wykorzystanie zbiorów przybliżonych w analizie danych umożliwia:

- szukanie zależności między danymi
- redukcję danych
- określenie wag danych
- generowanie reguł decyzyjnych z danych

Zastosowania:

- teoria decyzji
- bazy danych
- medycyna
- farmakologia
- bankowość
- lingwistyka
- ekonomia
- teoria sterowania
- teoria konfliktów
- ochrona środowiska

Zbiory przybliżone w analizie danych

System informacyjny jest sposobem przedstawiania informacji o obiektach charakteryzowanych przez ten sam zbiór cech

Systemem informacyjnym nazywamy uporządkowaną czwórkę $SI=\langle U, Q, V, f\rangle$

gdzie:

U – niepusty, skończony zbiór obiektów,

Q – zbiór wybranych cech (atrybutów) obiektów ze zbioru U

V – zbiór wszystkich możliwych wartości rozważanych cech

f – funkcja informacyjna, $f: U \times Q \rightarrow V$

Ponadto:

x – obiekt zbioru U

q – cecha obiektu x

$v_q^x = f(x,q)$ to wartość cechy q obiektu x

V_q – zbiór wszystkich możliwych wartości cechy q

$f(x,q) \in V_q$

$$V = \bigcup_{q \in Q} V_q$$

Tablica decyzyjna

Tablicą decyzyjną nazywamy uporządkowaną piątkę $DT=\langle U, C, D, V, f\rangle$

gdzie:

U – niepusty, skończony zbiór obiektów,

C – zbiór cech (atrybutów) warunkowych

D – zbiór cech (atrybutów) decyzyjnych

V – zbiór wszystkich możliwych wartości rozważanych cech

f – funkcja informacyjna

$C \cup D = Q$

Przyład 1: Przykład systemu informacyjnego

Wiersze opisują pacjentów, a kolumny tabeli reprezentują symptomy choroby (gorączka, kaszel, zmęczenie, utrata smaku lub węchu). Ostatnia kolumna definiuje podział pacjentów na dwie klasy: chorujących na COVID i niechorujących na COVID.

Pacjent (U)	gorączka (q_1)	kaszel (q_2)	zmęczenie (q_3)	utrata smaku lub węchu (q_4)	COVID (q_5)
1	Nie	Tak	Tak	Tak	Tak
2	Tak	Nie	Tak	Nie	Tak
3	Nie	Nie	Tak	Nie	Nie
4	Nie	Tak	Nie	Tak	Tak
5	Tak	Nie	Tak	Nie	Nie
6	Tak	Tak	Nie	Tak	Tak

Przyład 1 c.d.:

Problem: Znaleźć zależność pomiędzy występowaniem COVID (lub niewystępowaniem) – atrybut decyzyjny, a symptomami opisującymi pacjentów – atrybuty warunkowe. Opisać zbiór przypadków {1,2,4,6} (lub zbiór {3,5}) w kategoriach wartości atrybutów warunkowych.

Pacjent (U)	gorączka ($c_1=q_1$)	kaszel ($c_2=q_2$)	zmęczenie ($c_3=q_3$)	utrata smaku lub węchu ($c_4=q_4$)	COVID ($d_1=q_5$)
1	Nie	Tak	Tak	Tak	Tak
2	Tak	Nie	Tak	Nie	Tak
3	Nie	Nie	Tak	Nie	Nie
4	Nie	Tak	Nie	Tak	Tak
5	Tak	Nie	Tak	Nie	Nie
6	Tak	Tak	Nie	Tak	Tak

Podział zbioru $Q=\{q_1,q_2,q_3,q_4,q_5\}$ na:
- atrybuty warunkowe $C=\{c_1,c_2,c_3,c_4\}=\{q_1,q_2,q_3,q_4\}$
- atrybuty decyzyjne $D=\{d_1\}=\{q_5\}$

Przyład 1 c.d.:

Analizowane dane nie są spójne ponieważ przypadki 2 i 5 dostarczają sprzecznych informacji, obaj pacjenci są opisani tymi samymi wartościami atrybutów warunkowych, ale są przydzieleni do różnych klas decyzyjnych. Zatem zbiór pacjentów chorujących na COVID nie może być jednoznacznie opisany w kategoriach {gorączka, kaszel, zmęczenie, utrata smaku lub węchu}.

Można jednak opisać ten zbiór w sposób przybliżony:

- {1,4,6} jest maksymalnym zbiorem przypadków, które są **z pewnością** zaklasyfikowane do klasy pacjentów chorujących na COVID – dolne przybliżenie
- {1,2,4,5,6} jest zbiorem przypadków, co do których **możliwe** jest ich przydzielenie do klasy pacjentów chorujących na COVID (nie można ich wykluczyć) – górne przybliżenie
- {2,5} jest zbiorem przypadków, które nie mogą być jednoznacznie przydzielone do klasy COVID lub brak COVID, ze względu na sprzeczny opis atrybutów warunkowych – obszar brzegowy.

Obiekty P-nierozróżnialne

Jeśli pewne dwa obiekty $x_a, x_b \in U$ mają takie same wartości wszystkich cech q należących do zbioru $P \subseteq Q$, co możemy zapisać:

$$\forall_{q \in P}\ f(x_a, q) = f(x_b, q) \quad (\text{lub inaczej } \forall_{q \in P}\ f_{x_a}(q) = f_{x_b}(q))$$

to mówimy, że obiekty te są **P-nierozróżnialne** lub, że są ze sobą w relacji **P-nierozróżnialności** ($x_a \tilde{P} x_b$)

gdzie:

U – niepusty, skończony zbiór obiektów,

x – obiekt ze zbioru U

Q – zbiór cech (atrybutów) obiektów ze zbioru U

q – cecha obiektu x

P – podzbiór zbioru cech (atrybutów) obiektów ze zbioru U, $P \subseteq Q$

f – funkcja informacyjna, $f: U \times Q \rightarrow V$

V – zbiór wszystkich możliwych wartości cech

Relacja P-nierozróżnialności

Relacją P-nierozróżnialności nazywamy relację $\tilde{P}$ określoną

na przestrzeni U×U, spełniającą zależność:

$$x_a \tilde{P} x_b \leftrightarrow \forall_{q \in P}\ f_{x_a}(q) = f_{x_b}(q)$$

w której x_a, $x_b \in U$ oraz $P \subseteq Q$.

<u>UWAGA:</u>

Relacja $\tilde{P}$ jest zwrotna, symetryczna i przechodnia, jest więc relacją równoważności. Relacja równoważności dzieli zbiór, w którym jest określona, na rodzinę rozłącznych podzbiorów - klas abstrakcji tej relacji.

Klasą abstrakcji relacji $\tilde{P}$ w przestrzeni U nazywamy zbiór wszystkich obiektów $x \in U$ będących w relacji $\tilde{P}$. Dla każdego $x_a \in U$ istnieje dokładnie jeden taki zbiór:

$$[x_a]_{\tilde{P}} = \{x \colon x_a \tilde{P}\ x\}.$$

Rodzinę wszystkich klas abstrakcji relacji $\tilde{P}$ w przestrzeni U, nazywamy **zbiorem ilorazowym** i oznaczamy symbolem $U/\tilde{P}$.

Aproksymacja (przybliżenie) zbioru

$\tilde{P}$ **- dolną aproksymacją zbioru** $X\subseteq U$ nazywamy zbiór $\underline{\tilde{P}}X$ zdefiniowany następująco:

$$\underline{\tilde{P}}X=\{x\in U: [x]_{\tilde{P}} \subseteq X\}$$

UWAGA:

Dolną aproksymacją zbioru X jest więc zbiór tych obiektów $x\in U$, o których na podstawie wartości cech P możemy z całą pewnością powiedzieć, że są elementami zbioru X. Inaczej, obiekt $x\in U$ jest elementem dolnej aproksymacji, jeżeli cała klasa abstrakcji, do której on należy, jest podzbiorem zbioru X.

$\tilde{P}$ **- górną aproksymacją zbioru** $X\subseteq U$ nazywamy zbiór $\overline{\tilde{P}}X$ zdefiniowany następująco:

$$\overline{\tilde{P}}X=\{x\in U: [x]_{\tilde{P}}\cap X\neq\emptyset\}$$

UWAGA:

Górną aproksymacją zbioru X jest więc zbiór tych obiektów $x\in U$, o których na podstawie wartości cech P nie możemy z całą pewnością powiedzieć, że nie są elementami zbioru X. Inaczej, obiekt $x\in U$ jest elementem górnej aproksymacji, jeżeli klasa abstrakcji, do której on należy, ma niepustą część wspólną ze zbiorem X.

Aproksymacja (przybliżenie) zbioru

$\tilde{P}$ **- pozytywny obszar zbioru** $X \subseteq U$ jest tożsamy z jego dolną aproksymacją

$$\mathrm{Pos}_{\tilde{P}}(X)=\underline{\tilde{P}}X$$

$\tilde{P}$ **- negatywny obszar zbioru** $X \subseteq U$ jest różnicą zbioru U i górnej aproksymacji zbioru X

$$\mathrm{Neg}_{\tilde{P}}(X)=U\setminus\overline{\tilde{P}}X$$

$\tilde{P}$ **- brzegowy obszar zbioru** $X \subseteq U$ jest różnicą górnej aproksymacji zbioru X i dolnej aproksymacji zbioru X

$$\mathrm{Bn}_{\tilde{P}}(X)=\overline{\tilde{P}}X\setminus\underline{\tilde{P}}X$$

Aproksymacja (przybliżenie) zbioru

Zbiór $X \subseteq U$ nazywamy $\widetilde{P}$ **- dokładnym** jeśli zachodzi równość jego dolnej i górnej aproksymacji

$$\overline{\widetilde{P}}X=\underline{\widetilde{P}}X$$

lub $\widetilde{P}$ **- przybliżonym** w przeciwnym razie

$$\overline{\widetilde{P}}X\neq\underline{\widetilde{P}}X$$

Wartość wyrażoną wzorem:

$$\mu_{\widetilde{P}}(X)=\frac{\overline{\overline{\underline{\widetilde{P}}X}}}{\overline{\overline{\overline{\widetilde{P}}X}}}$$

nazywamy $\widetilde{P}$ **- dokładnością** aproksymacji zbioru X.

<u>UWAGA:</u>

Symbolem $\overline{\overline{A}}$ oznaczamy miarę zbioru A. W przypadku zbiorów skończonych jako miarę możemy przyjąć moc zbioru, w przypadku zbiorów ciągłych – długość przedziału, pole powierzchni, objętość itd..

Aproksymacja (przybliżenie) zbioru

Zbiór $X \subseteq U$ nazywamy:

(a) w przybliżeniu $\widetilde{P}$ **- definiowalnym**, jeżeli

$$\underline{\widetilde{P}}X \neq \emptyset$$

$$\overline{\widetilde{P}}X \neq U$$

(b) w przybliżeniu $\widetilde{P}$ **- niedefiniowalnym**, jeżeli

$$\underline{\widetilde{P}}X = \emptyset$$

$$\overline{\widetilde{P}}X \neq U$$

(c) zewnętrznie $\widetilde{P}$ **- niedefiniowalnym**, jeżeli

$$\underline{\widetilde{P}}X \neq \emptyset$$

$$\overline{\widetilde{P}}X = U$$

(d) całkowicie $\widetilde{P}$ **- niedefiniowalnym**, jeżeli

$$\underline{\widetilde{P}}X = \emptyset$$

$$\overline{\widetilde{P}}X = U$$

Przyład 2:

Dana jest tablica zawierająca informacje odnotowane przez właściciela komisu samochodowego

obiekt (U)	**liczba drzwi (q_1)**	**moc silnika (q_2)**	**kolor (q_3)**	**marka (q_4)**
x_1	2	60	niebieski	Toyota
x_2	2	100	czarny	Nissan
x_3	2	200	czarny	Ferrari
x_4	2	200	czerwony	Ferrari
x_5	2	200	czerwony	Toyota
x_6	3	100	czerwony	Toyota
x_7	3	100	czerwony	Toyota
x_8	3	200	czarny	Ferrari
x_9	4	100	niebieski	Nissan
x_{10}	4	100	niebieski	Nissan

Dziedziny cech (zbiór wartości funkcji informacyjnych):

$V_{q_1}=\{2,3,4\}$, $V_{q_2}=\{60,100,200\}$, $V_{q_3}=\{czarny,niebieski,czerwony\}$, $V_{q_4}=\{Ferrari,Nissan,Toyota\}$

Przyład 2:

Wykorzystując tablice decyzyjną, stworzymy system ekspertowy, który na podstawie informacji o liczbie drzwi, mocy silnika i kolorze samochodu określi jego markę

Reguła (R)	liczba drzwi ($c_1=q_1$)	moc silnika ($c_2=q_2$)	kolor ($c_3=q_3$)	marka ($d_1=q_4$)
1	2	60	niebieski	Toyota
2	2	100	czarny	Nissan
3	2	200	czarny	Ferrari
4	2	200	czerwony	Ferrari
5	2	200	czerwony	Toyota
6	3	100	czerwony	Toyota
7	3	100	czerwony	Toyota
8	3	200	czarny	Ferrari
9	4	100	niebieski	Nissan
10	4	100	niebieski	Nissan

Przyład 2 c.d.:

Uniwersum $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$

Podział zbioru $Q = \{q_1, q_2, q_3, q_4\}$ na:

- atrybuty warunkowe
  $C = \{c_1, c_2, c_3\} = \{q_1, q_2, q_3\} = \{$liczba drzwi,moc silnika,kolor$\}$
- atrybuty decyzyjne
  $D = \{d_1\} = \{q_4\} = \{$marka$\}$

Treść zawartą w tablicy decyzyjnej można przedstawić w postaci reguł:

R1: JEŻLI c_1=2 I c_2=60 I c_3=niebieski TO d_1=Toyota
R2: JEŻLI c_1=2 I c_2=100 I c_3=czarny TO d_1=Nissan
R3: JEŻLI c_1=2 I c_2=200 I c_3=czarny TO d_1=Ferrari
R4: JEŻLI c_1=2 I c_2=200 I c_3=czerwony TO d_1=Ferrari
R5: JEŻLI c_1=2 I c_2=200 I c_3=czerwony TO d_1=Toyota
R6: JEŻLI c_1=3 I c_2=100 I c_3=czerwony TO d_1=Toyota
R7: JEŻLI c_1=3 I c_2=100 I c_3=czerwony TO d_1=Toyota
R8: JEŻLI c_1=3 I c_2=200 I c_3=czarny TO d_1=Ferrari
R9: JEŻLI c_1=4 I c_2=100 I c_3=niebieski TO d_1=Nissan
R10: JEŻLI c_1=4 I c_2=100 I c_3=niebieski TO d_1=Nissan

Przyład 2 c.d.:

Klasy abstrakcji relacji C-nierozróżnialności $\tilde{C}$ określonej przez zbiór cech C

$$[x_1]_{\tilde{C}}=\{x_1\}$$

$$[x_2]_{\tilde{C}}=\{x_2\}$$

$$[x_3]_{\tilde{C}}=\{x_3\}$$

$$[x_4]_{\tilde{C}}= [x_5]_{\tilde{C}}=\{x_4,x_5\}$$

$$[x_6]_{\tilde{C}}= [x_7]_{\tilde{C}}=\{x_6,x_7\}$$

$$[x_8]_{\tilde{C}}=\{x_8\}$$

$$[x_9]_{\tilde{C}}= [x_{10}]_{\tilde{C}}=\{x_9,x_{10}\}$$

pary obiektów w zbiorach $\{x_4,x_5\},\{x_6,x_7\},\{x_9,x_{10}\}$ są C-nierozróżnialne

Zbiór ilorazowy:

$$U/\tilde{C}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4,x_5\},\{x_6,x_7\},\{x_8\},\{x_9,x_{10}\}\}$$

Przyład 2 c.d.:

W przestrzeni U istnieją trzy zbiory samochodów:

$$X_F=\{x_3,x_4,x_8\}$$
$$X_N=\{x_2,x_9,x_{10}\}$$
$$X_T=\{x_1,x_5,x_6,x_7\}$$

Dolne aproksymacje zbiorów X_F, X_N, X_T:

$$\underline{\tilde{C}}X_F=\{x_3\}\cup\{x_8\}=\{x_3,x_8\}$$
$$\underline{\tilde{C}}X_N=\{x_2\}\cup\{x_9,x_{10}\}=\{x_2,x_9,x_{10}\}$$
$$\underline{\tilde{C}}X_T=\{x_1\}\cup\{x_6,x_7\}=\{x_1,x_6,x_7\}$$

Obiekt $x\in U$ jest elementem dolnej aproksymacji, jeżeli cała klasa abstrakcji, do której on należy, jest podzbiorem zbioru X

Górne aproksymacje zbiorów X_F, X_N, X_T:

$$\overline{\overline{\tilde{C}}}X_F=\{x_3\}\cup\{x_4,x_5\}\cup\{x_8\}=\{x_3,x_4,x_5,x_8\}$$
$$\overline{\overline{\tilde{C}}}X_N=\{x_2\}\cup\{x_9,x_{10}\}=\{x_2,x_9,x_{10}\}$$
$$\overline{\overline{\tilde{C}}}X_T=\{x_1\}\cup\{x_4,x_5\}\cup\{x_6,x_7\}=\{x_1,x_4,x_5,x_6,x_7\}$$

Obiekt $x\in U$ jest elementem górnej aproksymacji, jeżeli klasa abstrakcji, do której on należy, ma niepustą część wspólną ze zbiorem X

Przyład 2 c.d.:

Obszar brzegowy zbiorów X_F, X_N, X_T:

$$Bn_{\tilde{C}}(X_F)=\overline{\tilde{C}}X_F\backslash\underline{\tilde{C}}X_F=\{x_3,x_4,x_5,x_8\}\backslash\{x_3,x_8\}=\{x_4,x_5\}$$

$$Bn_{\tilde{C}}(X_N)=\overline{\tilde{C}}X_N\backslash\underline{\tilde{C}}X_N=\{x_2,x_9,x_{10}\}\backslash\{x_2,x_9,x_{10}\}=\varnothing$$

$$Bn_{\tilde{C}}(X_T)=\overline{\tilde{C}}X_T\backslash\underline{\tilde{C}}X_T=\{x_1,x_4,x_5,x_6,x_7\}\backslash\{x_1,x_6,x_7\}=\{x_4,x_5\}$$

Obszar pozytywny zbiorów X_F, X_N, X_T:

$$Pos_{\tilde{C}}(X_F)=\underline{\tilde{C}}X_F=\{x_3,x_8\}$$

$$Pos_{\tilde{C}}(X_N)=\underline{\tilde{C}}X_N=\{x_2,x_9,x_{10}\}$$

$$Pos_{\tilde{C}}(X_T)=\underline{\tilde{C}}X_T\}=\{x_1,x_6,x_7\}$$

Obszar negatywny zbiorów X_F, X_N, X_T:

$$Neg_{\tilde{C}}(X_F)=U\backslash\overline{\tilde{C}}X_F=\{x_1,x_2,x_6,x_7,x_9,x_{10}\}$$

$$Neg_{\tilde{C}}(X_N)=U\backslash\overline{\tilde{C}}X_N=\{x_1,x_3,x_4,x_5,x_6,x_7,x_8\}$$

$$Neg_{\tilde{C}}(X_T)=U\backslash\overline{\tilde{C}}X_T=\{x_2,x_3,x_8,x_9,x_{10}\}$$

Przyład 2 c.d.:

Dokładność aproksymacji zbiorów X_F, X_N, X_T:

$$\mu_{\tilde{C}}(X_F)=\frac{\overline{\overline{\tilde{C}X_F}}}{\overline{\overline{\overline{\tilde{C}}X_F}}}=\frac{2}{4}=0,5$$

$$\mu_{\tilde{C}}(X_N)=\frac{\overline{\overline{\tilde{C}X_N}}}{\overline{\overline{\overline{\tilde{C}}X_N}}}=\frac{3}{3}=1$$

$$\mu_{\tilde{C}}(X_T)=\frac{\overline{\overline{\tilde{C}X_T}}}{\overline{\overline{\overline{\tilde{C}}X_T}}}=\frac{3}{5}=0,6$$

Dokładność aproksymacji zbioru X_N wynosi 1, co potwierdza fakt, że jest on $\tilde{C}$ - dokładny, czyli jednoznacznie zdefiniowany przez cechy należące do zbioru C.

Przyład 3:

Uniwersum $U = [0;10)$, $x \in U$

Zbiór cech $Q=\{q_1,q_2\}$, gdzie:

q_1 – część całkowita liczby x (cecha, funkcja podłoga,całość liczby x: $\lfloor x \rfloor$)

q_2 – część ułamkowa liczby x (mantysa liczby x: $x - \lfloor x \rfloor$)

stąd $x=q_1+q_2$

Funkcje informacyjne:

$$f(x,q_1)=f_x(q_1)= \lfloor x \rfloor$$
$$f(x,q_2)=f_x(q_2)= x - \lfloor x \rfloor$$

Dziedziny cech (zbiór wartości funkcji informacyjnych):

$$V_{q_1}=\{0,1,2,3,4,5,6,7,8,9\},\ V_{q_2}=[0,1)$$

Przyład 3 c.d.:

Klasy abstrakcji:

a) w zbiorze cech $Q=\{q_1,q_2\}$ wszystkie obiekty są rozróżnialne, istnieje zatem nieskończenie wiele jednoelementowych klas abstrakcji relacji Q-nierozróżnialności i każdy element U tworzy własną klasę

b) w zbiorach $P1=\{q_1\}$, $P2=\{q_2\}$

- dla relacji P1-nierozróżnialności mamy 10 klas abstrakcji

$$[0]_{\widetilde{P1}}=[0;1)$$
$$[1]_{\widetilde{P1}}=[1;2)$$
$$[2]_{\widetilde{P1}}=[2;3)$$
$$\ldots$$
$$[9]_{\widetilde{P1}}=[9;10)$$

- dla relacji P2-nierozróżnialności mamy nieskończenie wiele dziesięcioelementowych klas abstrakcji, są to zbiory liczb z przestrzeni U o takiej samej części ułamkowej np.:

$$[0,6]_{\widetilde{P2}}=\{0,6;\ 1,6;\ 2,6;\ 3,6;\ 4,6;\ 5,6;\ 6,6;\ 7,6;\ 8,6;\ 9,6\}$$

Zbiory ilorazowe relacji $\widetilde{P1}$ i $\widetilde{P2}$:

$U/\widetilde{P1}=\{[0;1);[1;2);[2;3);[3;4);[4;5);[5;6);[6;7);[7;8);[8;9);[9;10)\}$

$U/\widetilde{P2}$ - zbiór wszystkich klas abstrakcji

Przyład 3 c.d.:

W przestrzeni U wyodrębnijmy pewien zbiór X w postaci przedziału:

$$X=[1,25;\ 6,78]$$

Dolne aproksymacje zbioru X względem relacji $\widetilde{P1}$ i $\widetilde{P2}$:

$$\underline{\widetilde{P1}}X=[2]_{P1}\cup[3]_{P1}\cup[4]_{P1}\cup[5]_{P1}=[2;6)$$

$$\underline{\widetilde{P2}}X=\varnothing$$

Obiekt $x\in U$ jest elementem dolnej aproksymacji, jeżeli cała klasa abstrakcji, do której on należy jest podzbiorem zbioru X

Górne aproksymacje zbioru X względem relacji $\widetilde{P1}$ i $\widetilde{P2}$:

$$\overline{\widetilde{P1}}X=[1]_{P1}\cup[2]_{P1}\cup[3]_{P1}\cup[4]_{P1}\cup[5]_{P1}\cup[6]_{P1}=[1;7)$$

$$\overline{\widetilde{P2}}X=U$$

Obiekt $x\in U$ jest elementem górnej aproksymacji, jeżeli klasa abstrakcji, do której on należy, ma niepustą część wspólną ze zbiorem X

Przyład 3 c.d.:

Obszar brzegowy zbioru X=[1,25; 6;78] dla zbioru cech P1 i P2

$$Bn_{\widetilde{P1}}(X)=\overline{\widetilde{P1}}X\setminus\underline{\widetilde{P1}}X=[1;7)\setminus[2;6)=[1;2)\cup[6;7)$$

$$Bn_{\widetilde{P2}}(X)=\overline{\widetilde{P2}}X\setminus\underline{\widetilde{P2}}X=U\setminus\varnothing=U$$

Obszar pozytywny zbioru X=[1,25; 6;78] dla zbioru cech P1 i P2

$$Pos_{\widetilde{P1}}(X)=\underline{\widetilde{P1}}X=[2;6)$$

$$Pos_{\widetilde{P2}}(X)=\underline{\widetilde{P2}}X=\varnothing$$

Obszar negatywny zbioru X=[1,25; 6;78] dla zbioru cech P1 i P2

$$Neg_{\widetilde{P1}}(X)=U\setminus\overline{\widetilde{P1}}X=U\setminus[1;7)=[0;1)\cup[7;10)$$

$$Neg_{\widetilde{P2}}(X)=U\setminus\overline{\widetilde{P2}}X=U\setminus U=\varnothing$$

Przyład 3 c.d.:

Dokładność aproksymacji zbioru X=[1,25; 6;78] dla zbioru cech P1 i P2

$$\mu_{\widetilde{P1}}(X_F)=\frac{\overline{\overline{\underline{\widetilde{P1}}X}}}{\overline{\overline{\overline{\widetilde{P1}}X}}}=\frac{4}{6}=\frac{2}{3}$$

$$\mu_{\widetilde{P2}}(X_N)=\frac{\overline{\overline{\underline{\widetilde{P2}}X}}}{\overline{\overline{\overline{\widetilde{P2}}X}}}=\frac{0}{10}=0$$

Dokładność aproksymacji zbioru X dla zbioru cech P2 wynosi 0, co oznacza, że jest on całkowicie $\widetilde{P2}$ - niedefiniowalny.

Aproksymacja (przybliżenie) rodziny zbiorów

$\widetilde{P}$ **- dolną aproksymacją rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ nazywamy zbiór $\underline{\widetilde{P}}X$ opisany następująco:

$$\underline{\widetilde{P}}X=\{\underline{\widetilde{P}}X_1, \underline{\widetilde{P}}X_2, \ldots, \underline{\widetilde{P}}X_n\}$$

<u>UWAGA:</u>

Dolną aproksymacją zbioru X jest więc zbiór dolnych aproksymacji zbiorów należących do zbioru X (rodziny zbiorów).

$\widetilde{P}$ **- górną aproksymacją rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ nazywamy zbiór $\overline{\widetilde{P}}X$ opisany następująco:

$$\overline{\widetilde{P}}X=\{\overline{\widetilde{P}}X_1, \overline{\widetilde{P}}X_2, \ldots, \overline{\widetilde{P}}X_n\}$$

<u>UWAGA:</u>

Górną aproksymacją zbioru X jest więc zbiór górnych aproksymacji zbiorów należących do zbioru X (rodziny zbiorów).

Aproksymacja (przybliżenie) rodziny zbiorów

$\widetilde{\mathrm{P}}$ **- pozytywny obszar rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ definiujemy jako sumę pozytywnych obszarów zbiorów X_i, dla $i=1,\ldots,n$

$$Pos_{\tilde{P}}(X) = \cup_{X_i \in X} Pos_{\tilde{P}}(Xi)$$

$\widetilde{\mathrm{P}}$ **- negatywny obszar rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ jest różnicą zbioru U i sumy negatywnych obszarów zbiorów X_i, dla $i=1,\ldots,n$

$$Neg_{\tilde{P}}(X) = U \setminus \cup_{X_i \in X} Neg_{\tilde{P}}(Xi)$$

$\widetilde{\mathrm{P}}$ **- brzegowy obszar rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ definiujemy jako sumę obszarów brzegowych zbiorów X_i, dla $i=1,\ldots,n$

$$Bn_{\tilde{P}}(X) = \cup_{X_i \in X} Bn_{\tilde{P}}(Xi)$$

Aproksymacja (przybliżenie) rodziny zbiorów

$\widetilde{P}$ **- jakość aproksymacji rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ wyraża się wzorem

$$\gamma_{\widetilde{P}}(X) = \frac{\overline{\overline{Pos_{\widetilde{P}}(X)}}}{\overline{\overline{U}}}$$

$\widetilde{P}$ **- dokładność aproksymacji rodziny zbiorów** $X=\{X_1, X_2, \ldots, X_n\}$, gdzie $X \subseteq U$ wyraża się wzorem

$$\beta_{\widetilde{P}}(X) = \frac{\overline{\overline{Pos_{\widetilde{P}}(X)}}}{\sum_{X_i \in X} \overline{\overline{\overline{\widetilde{P}}X_i}}}$$

Przyład 2 c.d.:

Rozważmy rodzinę zbiorów X, której elementami są trzy zbiory samochodów:

$$X_F=\{x_3,x_4,x_8\},\ X_N=\{x_2,x_9,x_{10}\},\ X_T=\{x_1,x_5,x_6,x_7\}$$

$\tilde{C}$-dolna aproksymacja rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$\underline{\tilde{C}}X=\{\underline{\tilde{C}}X_F,\ \underline{\tilde{C}}X_N,\ \underline{\tilde{C}}X_T\}=\{\{x_3,x_8\},\{x_2,x_9,x_{10}\},\{x_1,x_6,x_7\}\}$$

$\tilde{C}$-górna aproksymacja rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$\overline{\tilde{C}}X=\{\overline{\tilde{C}}X_F,\ \overline{\tilde{C}}X_N,\ \overline{\tilde{C}}X_T\}=\{\{x_3,x_4,x_5,x_8\},\{x_2,x_9,x_{10}\},\{x_1,x_4,x_5,x_6,x_7\}\}$$

Przyład 2 c.d.:

$\tilde{C}$-brzegowy obszar rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$Bn_{\tilde{C}}(X)= Bn_{\tilde{C}}(X_F)\cup Bn_{\tilde{C}}(X_N)\cup Bn_{\tilde{C}}(X_T)=\{x_4,x_5\}\cup\emptyset\cup\{x_4,x_5\}=\{x_4,x_5\}$$

$\tilde{C}$-pozytywny obszar rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$Pos_{\tilde{C}}(X)=Pos_{\tilde{C}}(X_F)\cup Pos_{\tilde{C}}(X_N)\cup Pos_{\tilde{C}}(X_T)=\{x_3,x_8\}\cup\{x_2,x_9,x_{10}\}\cup\{x_1,x_6,x_7\}=$$
$$=\{x_1,x_2,x_3,x_6,x_7,x_8,x_9,x_{10}\}$$

$\tilde{C}$-negatywny obszar rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$Neg_{\tilde{C}}(X)=U\backslash(Neg_{\tilde{C}}(X_F)\cup Neg_{\tilde{C}}(X_N)\cup Neg_{\tilde{C}}(X_T))=$$
$$=U\backslash\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_9,x_{10}\}=\emptyset$$

Przyład 2 c.d.:

$\tilde{C}$ - jakość aproksymacji rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$\gamma_{\tilde{C}}(X)=\frac{\overline{\overline{\mathrm{Pos}_{\tilde{C}}(X)}}}{\overline{\overline{U}}} = \frac{8}{10}$$

$\tilde{C}$ - dokładność aproksymacji rodziny zbiorów $X=\{X_F, X_N, X_T\}$:

$$\beta_{\tilde{C}}(X)=\frac{\overline{\overline{\mathrm{Pos}_{\tilde{C}}(X)}}}{\sum_{X_i \in X} \overline{\overline{\overline{\tilde{C}}X_i}}} = \frac{8}{4+3+5} = \frac{2}{3}$$

Analiza tablic decyzyjnych

Wykrycie zależności między cechami (atrybutami) systemu informacyjnego pozwala stwierdzać czy do jednoznacznego opisu obiektu należącego do zbioru U konieczna jest znajomość wartości wszystkich jego cech.

Stopień zależności zbioru atrybutów P2 od zbioru atrybutów P1, gdzie P1,P2$\subseteq$Q, jest określony następująco:

$$k=\gamma_{\widetilde{P1}}(P2^*)$$

gdzie $\gamma_{\widetilde{P1}}$ oznacza jakość aproksymacji

<u>UWAGA:</u>

Zapis P1$\xrightarrow{k}$P2 oznacza, że zbiór atrybutów P2 zależy od atrybutów P1

w stopniu k<1, gdy k=1, piszemy P1$\rightarrow$P2

Analiza tablic decyzyjnych

Reguły tablicy decyzyjnej nazywamy **deterministycznymi**, jeżeli dla każdej pary reguł $R_a \neq R_b$ z równości wartości wszystkich atrybutów warunkowych C wynika równość wartości atrybutów decyzyjnych D, tzn.:

$$\forall_{\substack{R_a, R_b = 1,\ldots,N \\ R_a \neq R_b}} : \forall_{c \in C}\, f_{R_a}(c) = f_{R_b}(c) \rightarrow \forall_{d \in D}\, f_{R_a}(d) = f_{R_b}(d)$$

Reguły tablicy decyzyjnej nazywamy **niedeterministycznymi**, jeżeli dla każdej pary reguł $R_a \neq R_b$ z równości wartości wszystkich atrybutów warunkowych C nie wynika równość wartości atrybutów decyzyjnych D, tzn.:

$$\exists_{\substack{R_a, R_b \\ R_a \neq R_b}} : \forall_{c \in C}\, f_{R_a}(c) = f_{R_b}(c) \rightarrow \exists_{d \in D}\, f_{R_a}(d) \neq f_{R_b}(d)$$

<u>UWAGA:</u>

Tablica decyzyjna jest **dobrze określona** jeśli jej wszystkie reguły są deterministyczne. W przeciwnym przypadku mówimy, że jest **źle określona**.

Analiza tablic decyzyjnych

Tablica decyzyjna posiadająca zbiór atrybutów warunkowych C oraz zbiór atrybutów decyzyjnych D jest **dobrze określona**, jeśli zbiór atrybutów decyzyjnych zależy od zbioru atrybutów warunkowych w stopniu równym 1 (C $\rightarrow$ D), czyli

$$\gamma_{\tilde{P1}}(D^*) = 1$$

<u>UWAGA:</u>
Źle określoną tablicę decyzyjną (tablicę zawierającą reguły niedeterministyczne) można przekształcić w tablicę dobrze określoną poprzez:

- usunięcie reguł niedeterministycznych
- rozszerzenie zbioru atrybutów warunkowych

Przyład 2 c.d.:

Zbadamy stopień zależności zbioru atrybutów warunkowych C od zbioru atrybutów decyzyjnych D

$$k=\gamma_{\tilde{C}}(D^*)=\frac{\overline{\overline{Pos_{\tilde{P}}(D^*)}}}{\overline{\overline{U}}}$$

Klasy abstrakcji relacji D-nierozróżnialności to zbiory X_F, X_N, X_T, zatem:

$$U/\tilde{D}=D^*=\{X_F, X_N, X_T\}$$

$\tilde{C}$-pozytywny obszar zbioru D*:

$$Pos_{\tilde{C}}(D^*)=Pos_{\tilde{C}}(X_F)\cup Pos_{\tilde{C}}(X_N)\cup Pos_{\tilde{C}}(X_T)=\{x_3,x_8\}\cup\{x_2,x_9,x_{10}\}\cup\{x_1,x_6,x_7\}=$$
$$=\{x_1,x_2,x_3,x_6,x_7,x_8,x_9,x_{10}\}$$

$$k=\gamma_{\tilde{C}}(D^*)=\frac{8}{10}$$

Stopień zależności zbioru atrybutów warunkowych C zależy od zbioru atrybutów decyzyjnych D w stopniu k=0,8, co zapisujemy

$$C \overset{0,8}{\longrightarrow} D$$

Otrzymana wartość k<1 oznacza, że tablica decyzyjna jest źle określona. Na podstawie zbioru atrybutów warunkowych nie można jednoznacznie wnioskować o przynależności obiektów przestrzeni U do poszczególnych zbiorów X_F, X_N, X_T, będących klasami abstrakcji relacji D-nierozróżnialności.

Przyład 2 c.d.:

Sposoby przekształcania tablicy decyzyjnej zawierającej reguły niedeterministyczne w tablicę dobrze określoną

- usuwanie reguł niedeterministycznych (reguły 4 i 5)

Reguła (R)	liczba drzwi ($c_1=q_1$)	moc silnika ($c_2=q_2$)	kolor ($c_3=q_3$)	marka ($d_1=q_4$)
1	2	60	niebieski	Toyota
2	2	100	czarny	Nissan
3	2	200	czarny	Ferrari
6	3	100	czerwony	Toyota
7	3	100	czerwony	Toyota
8	3	200	czarny	Ferrari
9	4	100	niebieski	Nissan
10	4	100	niebieski	Nissan

$k=\gamma_{\tilde{C}}(D^*) = \frac{\overline{\overline{\{x_1,x_2,x_3,x_6,x_7,x_8,x_9,x_{10}\}}}}{\overline{\overline{\{x_1,x_2,x_3,x_6,x_7,x_8,x_9,x_{10}\}}}} = \frac{8}{8} = 1$ - tablica jest dobrze określona

Przyład 2 c.d.:

- rozszerzenie zbioru atrybutów warunkowych
  (dodajemy kolejne atrybuty np.: c_4 - tapicerka, c_5 – felgi)

Reguła (R)	liczba drzwi (c_1)	moc silnika (c_2)	kolor (c_3)	tapicerka (c_4)	felgi (c_5)	marka (d_1)
1	2	60	niebieski	tkanina	stal	Toyota
2	2	100	czarny	tkanina	stal	Nissan
3	2	200	czarny	skóra	Al	Ferrari
4	2	200	czerwony	skóra	Al	Ferrari
5	2	200	czerwony	tkanina	stal	Toyota
6	3	100	czerwony	skóra	stal	Toyota
7	3	100	czerwony	tkanina	stal	Toyota
8	3	200	czarny	skóra	Al	Ferrari
9	4	100	niebieski	tkanina	stal	Nissan
10	4	100	niebieski	tkanina	Al	Nissan

Przyład 2 c.d.:

Nowy zbiór atrybutów warunkowych $C'=\{c_1,c_2,c_3,c_4,c_5\}$

Dziedziny atrybutów:

$V_{c_1}=\{2,3,4\}$, $V_{c_2}=\{60,100,200\}$, $V_{c_3}=\{czarny,niebieski,czerwony\}$, $V_{c_4}=\{tkanina,skóra\}$, $V_{c_5}=\{stal,Al\}$, $V_{d_1}=\{Ferrari,Nissan,Toyota\}$

Klasy abstrakcji relacji C′-nierozróżnialności $\widetilde{C'}$ określonej przez zbiór cech C′ to zbiory jednoelementowe, ponieważ każdy element przestrzeni U ma przynajmniej jedną różną wartość cechy

Zbiór ilorazowy:

$$U/\widetilde{C'}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6,\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$$

Obszar pozytywny zbiorów X_F, X_N, X_T:

$$Pos_{\widetilde{C'}}(X_F)=\underline{\widetilde{C'}}X_F=\{x_3,x_4,x_8\}=X_F$$

$$Pos_{\widetilde{C'}}(X_N)=\underline{\widetilde{C'}}X_N=\{x_2,x_9,x_{10}\}=X_N$$

$$Pos_{\widetilde{C'}}(X_T)=\underline{\widetilde{C'}}X_T\}=\{x_1,x_5,x_6,x_7\}=X_T$$

Przyład 2 c.d.:

Górne aproksymacje zbiorów X_F, X_N, X_T:

$$\overline{\overline{\tilde{C}'}}X_F=\{x_3\}\cup\{x_5\}\cup\{x_8\}=\{x_3,x_4,x_8\}=X_F$$
$$\overline{\overline{\tilde{C}'}}X_N=\{x_2\}\cup\{x_9\}\cup\{x_{10}\}=\{x_2,x_9,x_{10}\}=X_N$$
$$\overline{\overline{\tilde{C}'}}X_T=\{x_1\}\cup\{x_5\}\cup\{x_6\}\cup\{x_7\}=\{x_1,x_5,x_6,x_7\}=X_T$$

$\tilde{C}'$-pozytywny obszar rodziny zbiorów D^*:

$$Pos_{\tilde{C}'}(D^*)=\{x_3,x_4,x_8\}\cup\{x_2,x_9,x_{10}\}\cup\{x_1,x_5,x_6,x_7\}=$$
$$=\{x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_9,x_{10}\}=U$$

$\tilde{C}'$ - jakość aproksymacji rodziny zbiorów D^*:

$$\gamma_{\tilde{C}'}(D^*)=\overline{\overline{Pos_{\tilde{C}'}(D^*)}}/\overline{\overline{U}}=\frac{10}{10}=1$$

$\tilde{C}'$ - dokładność aproksymacji rodziny zbiorów D^*:

$$\beta_{\tilde{C}'}(D^*)=\overline{\overline{Pos_{\tilde{C}'}(D^*)}}/\sum_{X_i\in D^*}\overline{\overline{\overline{\overline{\tilde{C}'}}X_i}}=\frac{\overline{\overline{U}}}{\overline{\overline{U}}}=\frac{10}{10}=1$$

Obie wartości wynoszą 1, rozszerzona tablica decyzyjna jest zatem dobrze określona.

Analiza tablic decyzyjnych c.d.

Zbiór atrybutów $P1 \subseteq Q$, jest **niezależny** w danym systemie informacyjnym, jeżeli dla każdego $P2 \subset P1$ zachodzi $\widetilde{P1} \neq \widetilde{P2}$. W przeciwnym przypadku zbiór P1 jest zależny

Zbiór atrybutów $P1 \subseteq Q$, jest **niezależny** ze względu na zbiór atrybutów $P2 \subseteq Q$ (P2-niezależny), jeśli dla każdego $P3 \subset P1$ zachodzi $\widetilde{P1} \neq \widetilde{P2}$.

$$Pos_{\widetilde{P1}}(P2^*) \neq Pos_{\widetilde{P3}}(P2^*)$$

W przeciwnym przypadku zbiór P1 jest P2-**zależny**

Analiza tablic decyzyjnych c.d.

Reduktem zbioru atrybutów $P1 \subseteq Q$ nazywamy każdy niezależny zbiór $P2 \subset P1$, dla którego $\widetilde{P2} = \widetilde{P1}$

Reduktem względnym zbioru atrybutów $P1 \subseteq Q$ ze względu na P2 (tzw. P2-reduktem) nazywamy każdy P2-niezależny zbiór $P3 \subset P1$, dla którego $\widetilde{P3} = \widetilde{P1}$

Atrybut $p \in P1$ jest **nieusuwalny** ze zbioru P1, jeżeli dla $P2=P1 \setminus \{p\}$ zachodzi $\widetilde{P2} \neq \widetilde{P1}$. W przeciwnym przypadku atrybut p jest **zbędny**.

Zbiór wszystkich atrybutów nieusuwalnych ze zbioru P nazywa się **rdzeniem** P, co zapisujemy następująco:

$$CORE(P)=\{p \in P:\ \widetilde{P'} \neq \widetilde{P},\ P'=P \setminus \{p\}\}$$

Analiza tablic decyzyjnych c.d.

Znormalizowany współczynnik istotności podzbioru zbioru atrybutów warunkowych $C' \subseteq C$ ma postać

$$\sigma_{(C,D)}(C') = \frac{\gamma_{\tilde{C}}(D^*) - \gamma_{\tilde{C}''}(D^*)}{\gamma_{\tilde{C}}(D^*)}$$

gdzie $C'' = C \setminus C'$

UWAGA:
Wartość zerowa znormalizowanego współczynnika istotności otrzymana dla danego podzbioru atrybutów warunkowych C wskazuje, że podzbiór ten może być usunięty ze zbioru atrybutów warunkowych bez szkody dla aproksymacji rodziny zbiorów D*

Dowolny podzbiór atrybutów warunkowych $C' \subseteq C$ nazywamy przybliżonym D-reduktem zbioru atrybutów C, a **błąd przybliżenia** tego reduktu wyznaczamy ze wzoru:

$$\varepsilon_{(C,D)}(C') = \frac{\gamma_{\tilde{C}}(D^*) - \gamma_{\tilde{C}'}(D^*)}{\gamma_{\tilde{C}}(D^*)}$$

Przyład 2 c.d.:

Rozszerzony zbiór atrybutów warunkowych $C'=\{c_1,c_2,c_3,c_4,c_5\}$

Przykładowe podzbiory zbioru C':

$C1=\{c_1,c_3,c_4,c_5\}$, $C2=\{c_1,c_2,c_3\}$, $C3=\{c_1,c_2\}$, $C4=\{c_1,c_3\}$, $C5=\{c_2,c_3\}$, $C6=\{c_1\}$, $C7=\{c_2\}$, $C8=\{c_3\}$

- Zbiór C' jest zbiorem zależnym, ponieważ podzbiór $C1\subset C'$ generuje takie same klasy abstrakcji jak C'

  Zbiory ilorazowe:

$$U/\widetilde{C'}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6,\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$$
$$U/\widetilde{C1}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6,\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$$

- Zbiór C2 jest zbiorem niezależnym, ponieważ wszystkie podzbiory różne od C2 generują inne niż dla C2 klasy abstrakcji

  Zbiory ilorazowe:

$$U/\widetilde{C2}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4,x_5\},\{x_6,x_7\},\{x_8\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C3}=\{\{x_1\},\{x_2\},\{x_3,x_4,x_5\},\{x_6,x_7\},\{x_8\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C4}=\{\{x_1\},\{x_2,x_3\},\{x_4,x_5\},\{x_6,x_7\},\{x_8\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C5}=\{\{x_1\},\{x_2\},\{x_3,x_8\},\{x_4,x_5\},\{x_6,x_7\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C6}=\{\{x_1,x_2,x_3,x_4,x_5\},\{x_6,x_7,x_8\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C7}=\{\{x_1\},\{x_2,x_6,x_7,x_9,x_{10}\},\{x_3,x_4,x_5,x_8\}\}$$
$$U/\widetilde{C8}=\{\{x_1,x_9,x_{10}\},\{x_2,x_3,x_8\},\{x_4,x_5,x_6,x_7\}\}$$

Przyład 2 c.d.:

Rozszerzony zbiór atrybutów warunkowych $C'=\{c_1,c_2,c_3,c_4,c_5\}$
Przykładowe podzbiory zbioru $C1=\{c_1,c_3,c_4,c_5\}$:
$C9=\{c_1,c_3,c_4\}$, $C4=\{c_1,c_3\}$, $C6=\{c_1\}$, $C7=\{c_2\}$, $C8=\{c_3\}$,

- Zbiór C′ jest zbiorem zależnym, ponieważ podzbiór $C1 \subset C'$ generuje takie same klasy abstrakcji jak C′

  Zbiory ilorazowe:

$$U/\widetilde{C'}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6,\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$$
$$U/\widetilde{C1}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6,\},\{x_7\},\{x_8\},\{x_9\},\{x_{10}\}\}$$

- Zbiór C1 jest zbiorem niezależnym, ponieważ wszystkie podzbiory różne od C1 generują inne niż dla C1 klasy abstrakcji

  Zbiory ilorazowe:

$$U/\widetilde{C4}=\{\{x_1\},\{x_2,x_3\},\{x_4,x_5\},\{x_6,x_7\},\{x_8\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C6}=\{\{x_1,x_2,x_3,x_4,x_5\},\{x_6,x_7,x_8\},\{x_9,x_{10}\}\}$$
$$U/\widetilde{C7}=\{\{x_1\},\{x_2,x_6,x_7,x_9,x_{10}\},\{x_3,x_4,x_5,x_8\}\}$$
$$U/\widetilde{C8}=\{\{x_1,x_9,x_{10}\},\{x_2,x_3,x_8\},\{x_4,x_5,x_6,x_7\}\}$$
$$U/\widetilde{C9}=\{\{x_1\},\{x_2\},\{x_3\},\{x_4\},\{x_5\},\{x_6,\},\{x_7\},\{x_8\},\{x_9,x_{10}\}\}$$

(UWAGA: Sprawdzić klasy abstrakcji dla pozostałych podzbiorów zbioru C1!!!)

Przyład 2 c.d.:

Rozszerzony zbiór atrybutów warunkowych $C'=\{c_1,c_2,c_3,c_4,c_5\}$

- Zbiór $C1=\{c_1,c_3,c_4,c_5\}\subset C'$ jest reduktem zbioru C′, ponieważ generuje takie same klasy abstrakcji jak zbiór C′ i jest zbiorem niezależnym
- Atrybut c_2 jest zbędny ponieważ $C1=C'\setminus\{c_2\}$ oraz $U/\widetilde{C'}=U/\widetilde{C1}$
- Rdzeniem zbioru C′ są atrybuty $\{c_1,c_3,c_4,c_5\}$ tzn.:

$$CORE(C')=\{c_1,c_3,c_4,c_5\}$$

- Jakość aproksymacji rodziny zbiorów D* ze względu na zbór cech C1 ($\widetilde{C1}$-jakość):

$$\gamma_{\widetilde{C1}}(D^*)=\overline{\overline{Pos_{\widetilde{C1}}(D^*)}}/\overline{\overline{U}}=\frac{10}{10}=1$$

- Współczynnik istotności podzbioru $\{c_2\}$ zbioru atrybutów warunkowych C

$$\sigma_{(C',D)}(\{c_2\})=(\gamma_{\widetilde{C'}}(D^*)-\gamma_{\widetilde{C1}}(D^*))/\gamma_{\widetilde{C'}}(D^*)=(1-1)/1=0$$

Wartość 0 oznacza, że atrybut c_2 jest nieistotny, jego usunięcie nie wpłynie zatem na jakość aproksymacji rodziny zbiorów D*. Ponadto dla wyjściowego zbioru reguł warunkowych C, błąd przybliżenia wynosi 0,2