

Odzyskiwanie wiedzy z tzw. koszyków

Klienci sklepu kupują różne zestawy produktów zwane koszykami ponumerowanymi od 1 do N (ich liczby) np.:

(1 2)

(1 2 3)

(1 2)

(1)

(1 2 3)

Jeśli dany koszyk występuje (też jako część składowa) więcej razy niż zakładana wartość progowa to jest istotny statystycznie np. $\frac{|P_{(1\ 2)}|}{N} > 20\%$, gdzie $|P_{(1\ 2)}|$ jest liczbą koszyków z produktami 1 i 2, a N liczbą wszystkich koszyków.

Odzyskiwanie wiedzy z tzw. koszyków

Najpierw wybiera się częste podgrupy produktów:

$$\frac{|P_{(1)}|}{N} = 1 > 20\%, \quad \frac{|P_{(1\ 2)}|}{N} = \frac{4}{5} > 20\%, \quad \frac{|P_{(1\ 2\ 3)}|}{N} = \frac{2}{5} > 20\%.$$

Następnie z częstych podgrup (1), (1 2), (1 2 3) można tworzyć reguły asocjacyjne czyli kojarzyć te podgrupy ze sobą np.:

$$(1) \rightarrow (1\ 2) - (1)$$

Do części warunkującej wybrano częsty podzbiór, a w części warunkowanej jest koszyk zawierający tamten podzbiór z części warunkującej (asocjacyjne skojarzenie), ale np. z reguły *modus ponens* ($\frac{a, a \Rightarrow b}{b}$) wynika, że konkluzja nie może zawierać przesłanki, stąd różnica w regule asocjacyjnej. W wyniku tych operacji powstaje reguła:

$$(1) \rightarrow (2)$$

Inna reguła uzyskana w ten sam sposób to:

$$(1\ 2) \rightarrow (1\ 2\ 3) - (1\ 2)$$

$$(1\ 2) \rightarrow (3)$$

Wiedza w bazach relacyjnych

W bazach relacyjnych wiedza jest zapisywana w tabelach. W kolumnach występują atrybuty mające zdyskretyzowane wartości np.: atrybut A_1 ma n_1 wartości od v_{11} do v_{1n_1} . W wybranej kolumnie atrybut C jest nazywany kategorią, a jej wartości d etykietami.

A_1	...	A_i	C
$(v_{11}, \dots, v_{1n_1})$...	$(v_{11}, \dots, v_{1n_i})$	(d_1, \dots, d_{n_C})
v_{11}	...	v_{23}	d_3
v_{12}	...	v_{25}	d_{n_C}
v_{1n_1}	...	v_{2n_i}	d_1
...

Pierwsza przykładowa tabela

x	wiek	samochód	ryzyko
1	18	maluch	duże
2	35	maluch	małe
3	50	sportowy	duże
4	66	minivan	duże
5	18	sportowy	duże
6	35	minivan	małe
7	60	maluch	małe
8	70	sportowy	duże
9	25	minivan	małe

Druga przykładowa tabela

x	aura	temperatura	wilgotność	wiatr	$c(x)$
1	słoneczna	ciepła	duża	słaby	0
2	słoneczna	ciepła	duża	silny	0
3	pochmurna	ciepła	duża	słaby	1
4	deszczowa	umiarkowana	duża	słaby	1
5	deszczowa	zimna	normalna	słaby	1
6	deszczowa	zimna	normalna	silny	0
7	pochmurna	zimna	normalna	silny	1
8	słoneczna	umiarkowana	duża	słaby	0
9	słoneczna	zimna	normalna	słaby	1
10	deszczowa	umiarkowana	normalna	słaby	1
11	słoneczna	umiarkowana	normalna	silny	1
12	pochmurna	umiarkowana	duża	silny	1
13	pochmurna	ciepła	normalna	słaby	1
14	deszczowa	umiarkowana	duża	silny	0