

Povezovalna pravila (*Association rules*)

Povezovalna pravila (*association rules*) so pravila oblike $X \rightarrow Y$, kjer sta X in Y množici postavk (*items*). Namen učenja povezovalnih pravil na bazi podatkov je iskanje zanimivih relacij med atributi, ki lahko doprinesejo zanimiva (nova) spoznanja o samih podatkih oziroma o zakonitostih v dani problemski domeni. Naloga učnega algoritma je poiskati **vsa** povezovalna pravila, ki so dovolj pogosta, kar izračunamo s podporo (*support, coverage*) in zanesljiva, kar izračunamo z zaupanjem (*confidence, rule accuracy*).

$$Support(X \rightarrow Y) = \frac{\text{število}(XinY)}{\text{število_transakcij}} \quad Confidence(X \rightarrow Y) = \frac{\text{število}(XinY)}{\text{število}(X)}$$

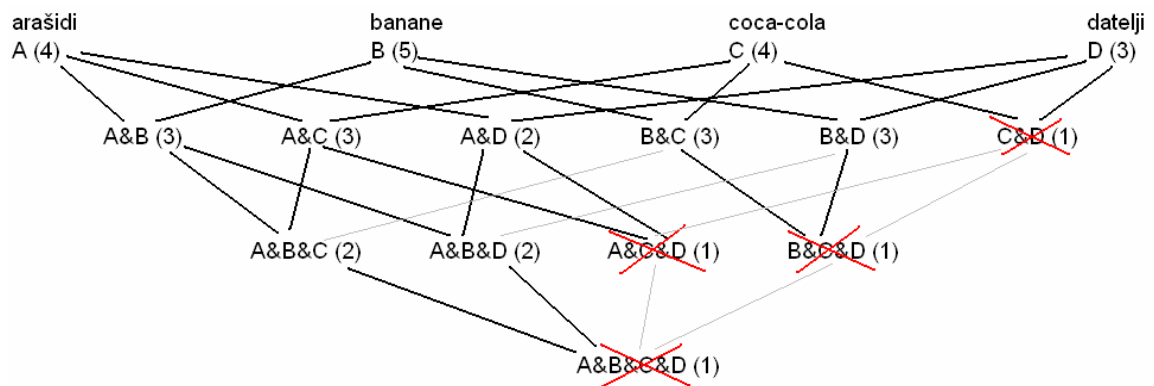
Klasifikacijska pravila (*classification rules*) so podobna povezovalnim pravilom, s to razliko, da v sklepnem delu nastopa samo ena postavka, ki je razred, saj so klasifikacijska pravila so namenjena klasifikaciji.

Primer

Zgradi povezovalna pravila (*association rules*) z minimalno podporo (*support, coverage*) 2/6 in minimalnim zaupanjem (*confidence, rule accuracy*) 70 %.

arašidi	banane	coca-cola	datelji
1	1	1	1
	1	1	
	1		1
1		1	
1	1		1
1	1	1	

1. Zgradimo velike množice (*large itemsets*) s podporo najmanj 2/6



2. Iz velikih množic (*itemsetov*), ki imajo podporo vsaj 2/6 zgradimo vsa možna povezovalna pravila (*association rules*).

A&B	Support = 3/6
A → B	Confidence = 2/4 = 50%
B → A	Confidence = 2/5 = 40%
A&C	Support = 3/6
A → C	Confidence = 3/4 = 75%
C → A	Confidence = 3/4 = 75%
A&D	Support = 2/6
A → D	Confidence = 2/4 = 50%
D → A	Confidence = 2/3 = 67%
B&C	Support = 3/6
B → C	Confidence = 3/5 = 60%
C → B	Confidence = 3/4 = 75%
B&D	Support = 3/6
B → D	Confidence = 3/5 = 60%
D → B	Confidence = 3/3 = 100%
A&B&C	Support = 2/6
A → B&C	Confidence = 2/4 = 50%
A&B → C	Confidence = 2/3 = 67%
A&C → B	Confidence = 2/3 = 67%
B → A&C	Confidence = 2/5 = 40%
B&C → A	Confidence = 2/3 = 67%
C → A&B	Confidence = 2/4 = 50%
A&B&D	Support = 2/6
A → B&D	Confidence = 2/4 = 50%
A&B → D	Confidence = 2/3 = 67%
A&D → B	Confidence = 2/2 = 100%
B → A&D	Confidence = 2/5 = 40%
B&D → A	Confidence = 2/3 = 67%
D → A&B	Confidence = 2/3 = 67%

3. Zberemo povezovalna pravila (*association rules*) z zaupanjem (*confidence*) vsaj 70%:

Pravilo	Conf.	Supp.
Arašidi → Coca-Cola	75%	50%
Coca-Cola → Arašidi	75%	50%
Coca-Cola → Banane	75%	50%
Datelji → Banane	100%	50%
Arašidi & Datelji → Banane	100%	33%