

# Traženje asocijacijskih pravila

Matej Mihelčič

Prirodoslovno-matematički fakultet, Sveučilište u Zagrebu

*matmih@math.hr*

28. travnja, 2026.



# Traženje asocijacijskih pravila

- Vidjeli smo da skupovi frekventnih objekata predstavljaju skupine objekata koje se često javljaju u bazi transakcija. Iz aspekta potrošačke košarice, to su artikli koje puno korisnika kupuje zajedno pri svojim kupnjama.
- Asocijacijska pravila su prirodno proširenje skupova frekventnih objekata, kod kojih ne želimo samo detektirati frekventne skupove objekata već i implikacijske odnose među njima. Promatrajući potrošačku košaricu, možemo se zapitati, ukoliko korisnik ima određeni skup artikala u košarici, za koje artikle bi moglo vrijediti da su mu dodatno potrebni do kraja kupnje?
- Odgovor na gornje pitanje može pomoći trgovačkom centru da optimizira razmještaj proizvoda na način da se takve skupine proizvoda nalaze prostorno blizu (što smanjuje vjerojatnost da će korisnik zaboraviti kupiti neki od potrebnih proizvoda) ili da potencijalno ponudi akcijske popuste ukoliko korisnik kupi dva ili više proizvoda zajedno.

# Traženje asocijacijskih pravila

- Jedan primjer asocijacijskog pravila može biti: 70% kupaca koji kupe vino i sir, također kupe i grožđe.
- Asocijacijska pravila se mogu koristiti i puno generalnije, kao metodologija koja omogućava traženje implikacijskih ovisnosti između atributa u skupovima podataka. Razumijevanje implikacijskih ovisnosti je bitno u raznim domenskim istraživanjima (npr. medicina, ekologija, ekonomija).
- Tradicionalni zadatak i algoritmi traženja asocijacija rade nad ekvivalentnim transakcijskim bazama kao i postupci traženja frekventnih skupova objekata. Novije adaptacije omogućuju rad nad generalnim tabličnim skupovima podataka koristeći binarne, kategorijske i numeričke attribute.
- Asocijacijsko pravilo  $\{A, B, C\} \rightarrow \{D, E\}$  označava da veliki dio transakcija za koje vrijedi  $A = 1$ ,  $B = 1$ ,  $C = 1$  također ima vrijednost  $D = 1$ ,  $E = 1$ .

Neka je  $\mathcal{I} = \{i_1, \dots, i_k\}$  skup objekata i  $D$  transakcijska baza.

## Definicija

Kažemo da je skup objekata  $Z \subseteq \mathcal{I}$  asocijacija ako frekvencija pojavljivanja od  $Z$  odstupa od očekivane frekvencije, uzimajući u obzir frekvencije individualnih objekata  $X \in Z$ .

Ukoliko je vjerojatnost kupnje purice 3%, a mlinaca 1%, očekujemo da će 0.03% kupaca te proizvode kupiti zajedno, ako su proizvodi međusobno nezavisni. Ukoliko iz podataka uočimo da 0.8% kupaca kupuje te proizvode zajedno, to odstupa od našeg očekivanja, stoga imamo asocijaciju između ta dva objekta. Postavlja se pitanje vrijedi li purica  $\rightarrow$  mlinci ili mlinci  $\rightarrow$  purica. Dakle, asocijacijska pravila imaju smjer i on je bitan za interpretaciju otkrivenog znanja.

# Traženje asocijacijskih pravila

## Definicija

Kažemo da je  $X \rightarrow Y$  asocijacijsko pravilo s antecedentom  $X$  i konzekventom  $Y$  ako  $Z = X \cup Y \subseteq I$  i  $X \cap Y = \emptyset$ .

## Definicija

Neka je  $X \rightarrow Y$  asocijacijsko pravilo. Njegova potpora se definira kao  $Pot(X \rightarrow Y) = Pot(X \cup Y)$ , gdje se potpora skupova objekata  $S$  definira kao  $Pot(S) = \frac{|\{T \in D | S \subseteq T\}|}{|D|}$ .

## Definicija

Neka je  $X \rightarrow Y$  asocijacijsko pravilo. **Pouzdanost** asocijacijskog pravila se definira kao  $Pouz(X \rightarrow Y) = P(Y|X) = \frac{Pot(X \cup Y)}{Pot(X)}$ .

# Traženje asocijacijskih pravila

- Pouzdanost asocijacijskog pravila  $X \rightarrow Y$  nam govori koji postotak transakcija koje zadovoljavaju  $X$  zadovoljavaju i  $Y$ . Ukoliko to svojstvo vrijedi za sve transakcije, tada je pouzdanost 1.0 i ne postoji kontraprimjer asocijacije. Najčešće je ipak u praksi pouzdanost asocijacijskog pravila manja od 1.0.

## Definicija

Za danu bazu transakcija  $D$  i dva praga: prag minimalne pouzdanosti  $\eta$  i prag minimalne potpore  $\sigma$ , **problem traženja asocijacijskih pravila** je generirati sva asocijacijska pravila  $X \rightarrow Y$  s  $Pot(X \rightarrow Y) \geq \sigma$  i  $Pouz(X \rightarrow Y) \geq \eta$  u  $D$ .

- Korištenje potpore i pouzdanosti kao jedinih mjera kvalitete asocijacijskih pravila, iako se često događa u praksi, može dovesti do otkrivanja asocijacijskih pravila koja ne sadrže korisno znanje. Npr. objekti koji su jako česti, kao što je kruh, se mogu slučajno javljati dosta često i s nekim potpuno nepovezanim proizvodom, npr. baterijama. Zbog toga asocijacija baterija → kruh može imati visoku pouzdanost, ali biti praktično neinteresantna.
- Često korištenje potpore i pouzdanosti je posljedica algoritamskih optimizacija koje omogućavaju učinkovito pretraživanje.
- Nedostaci navedenih mjera se često mogu kompenzirati u fazi postprocesiranja.

# Traženje asocijacijskih pravila - algoritam

- Prvo traži sve frekventne skupove objekata standardnim Apriori algoritmom.
- Iz svakog frekventnog skupa objekata se stvaraju konzekvente asocijacijskih pravila.
- Glavno opažanje koje omogućava učinkovito generiranje asocijacija je da za konzekvente asocijacijskih pravila, uz mjeru pouzdanosti, vrijedi **zatvorenje prema dolje** kao i za skupove objekata s obzirom na potporu.
- Za svaki frekventni skup objekata  $Z$  i njegove podskupove  $X$ , stvaramo pravila  $X \rightarrow Z \setminus X$ . Stvoreno asocijacijsko pravilo uključujemo u rezultat ako  $Pot(Z)/Pot(X) \geq \eta$  (ako asocijacijsko pravilo ima dovoljno visoku pouzdanost).
- Iz svojstava skupova objekata vrijedi da  $X_1 \subseteq X_2 \subseteq Z$  povlači  $Pot(X_1) \geq Pot(X_2) \geq Pot(Z)$ . Stoga,  $Pot(Z)/Pot(X_1) \leq Pot(Z)/Pot(X_2) \leq 1$ .

- Slijedi da je vrijednost pouzdanosti za asocijacijsko pravilo  $X_2 \rightarrow Z \setminus X_2$  viša od pravila  $X_1 \rightarrow Z \setminus X_1$ .
- Generalno će najveća vrijednost pouzdanosti biti postignuta za najveću antecedentu  $X_2$  (odnosno najmanju konzekventnu  $Z \setminus X_2$ ).
- Pretragu stoga možemo započeti generirajući asocijacijska pravila koja imaju konzekventu s jednim objektom. Ukoliko takva asocijacijska pravila ne prijeđu minimalni prag pouzdanosti, ne moramo evaluirati pravila s većom konzekventom.
- Stoga, da bi pravilo  $X_1 \rightarrow Y_1$  uz konzekventu  $Y_1 = Z \setminus X_1$  prešlo minimalni prag pouzdanosti, sva pravila  $X_2 \rightarrow Y_2$  uz  $Y_2 \subseteq Y_1 \subseteq Z$  moraju također prijeći minimalni prag pouzdanosti.
- Iz toga slijedi da za pravila  $X \rightarrow Y$  uz  $Z = X \cup Y$  imamo identično zatvorenje konzekvenata prema dolje (s obzirom na pouzdanost) kao i skupova objekata (s obzirom na potporu).

## Algoritam Apriori za traženje asocijacijskih pravila

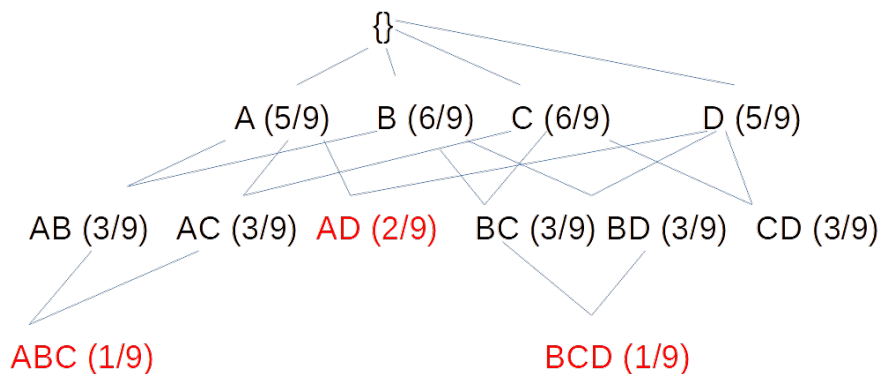
```
Ulaz:  $D, \sigma, \eta$ 
Izlaz:  $\mathcal{AR}(D, \sigma, \eta)$ 
 $C_1 := \{\{i\} \mid i \in \mathcal{I}\};$ 
 $k := 1;$ 
while  $C_k \neq \emptyset$  do
    racunajPotporu( $C_k, D$ );
     $F_k := \{S \in C_k \mid \text{supp}(S) \geq \sigma\};$ 
     $C_{k+1} := \text{generirajKandidate}(F_k);$ 
     $k := k + 1;$ 
end while
 $\mathcal{F} = F_1 \cup F_2 \cup \dots \cup F_{k-1};$  // svi frekventni skupovi objekata
 $\mathcal{AR} := \emptyset;$ 
for all  $Z \in \mathcal{F}$  do
     $C_1 := \{\{i\} \mid i \in Z\};$  // kand. 1-konzekvente
     $k := 1;$ 
    while  $C_k \neq \emptyset$  do
         $F_k := \{X \in C_k \mid \text{Pouz}(X \rightarrow Z - X) \geq \eta\};$ 
         $\mathcal{AR} := \mathcal{AR} \cup \{X \rightarrow Z - X \mid X \in F_k\};$ 
         $C_{k+1} := \text{generirajKandidate}(F_k);$ 
         $k := k + 1;$ 
    end while
end for
return  $\mathcal{AR}$ 
```

# Traženje asocijacijskih pravila - primjer

tid	A	B	C	D
1	1	1	0	1
2	0	1	1	0
3	0	1	0	1
4	0	0	1	1
5	1	1	0	0
6	1	0	1	0
7	0	1	1	1
8	1	0	1	1
9	1	1	1	0

- Pretpostavimo  $\sigma = 0.3, \eta = 0.6$ .

# Traženje asocijacijskih pravila - primjer



- Asocijacijska pravila stvaramo iz frekventnih skupova objekata:  $\{A\}, \{B\}, \{C\}, \{D\}, \{A, B\}, \{A, C\}, \{B, C\}, \{B, D\}, \{C, D\}$ .
- Jednočlani frekventni skupovi objekata **ne mogu** tvoriti asocijacijska pravila.
- Iz frekventnog skupa objekata  $\{A, B\}$  dobijemo dva kandidata za asocijacijska pravila: a)  $A \rightarrow B$  i  $B \rightarrow A$ .
  - $Pouz(A \rightarrow B) = \frac{3}{5} = 0.6$
  - $Pouz(B \rightarrow A) = \frac{3}{6} = 0.5$
  - Stoga prihvaćamo samo asocijacijsko pravilo  $A \rightarrow B$ .
- Iz frekventnog skupa objekata  $\{A, C\}$  dobijemo dva kandidata za asocijacijska pravila: a)  $A \rightarrow C$  i  $C \rightarrow A$ .
  - $Pouz(A \rightarrow C) = \frac{3}{5} = 0.6$
  - $Pouz(C \rightarrow A) = \frac{3}{6} = 0.5$
  - Stoga prihvaćamo samo asocijacijsko pravilo  $A \rightarrow C$ .

- Iz frekventnog skupa objekata  $\{B, C\}$  dobijemo dva kandidata za asocijacijska pravila: a)  $B \rightarrow C$  i  $C \rightarrow B$ .
  - $Pouz(B \rightarrow C) = \frac{3}{6} = 0.5$
  - $Pouz(C \rightarrow B) = \frac{3}{6} = 0.5$
  - Stoga ne prihvaćamo niti jedno od dva asocijacijska pravila.
- Iz frekventnog skupa objekata  $\{B, D\}$  dobijemo dva kandidata za asocijacijska pravila: a)  $B \rightarrow D$  i  $D \rightarrow B$ .
  - $Pouz(B \rightarrow D) = \frac{3}{6} = 0.5$
  - $Pouz(D \rightarrow B) = \frac{3}{5} = 0.6$
  - Stoga prihvaćamo samo asocijacijsko pravilo  $D \rightarrow B$ .
- Iz frekventnog skupa objekata  $\{C, D\}$  dobijemo dva kandidata za asocijacijska pravila: a)  $C \rightarrow D$  i  $D \rightarrow C$ .
  - $Pouz(C \rightarrow D) = \frac{3}{6} = 0.5$
  - $Pouz(D \rightarrow C) = \frac{3}{5} = 0.6$
  - Stoga prihvaćamo samo asocijacijsko pravilo  $D \rightarrow C$ .

Softver za kreiranje asocijacijskih pravila:

- Java - **WEKA** (Apriori, FP-growth).
- Java - **SPMF biblioteka** (Apriori, TopKRules, IGB, HGB itd).