

1 Asociačné pravidlá

Naším cieľom je implementovať hľadanie asociačných pravidiel algoritmom APRI-ORI. Nech D je súbor transakcií. Pre jednoduchosť budeme predpokladať, že D je bunkové pole rozmerov $N \times 1$, ktorého položky sú jednoriadkové vektory obsahujúce čísla položiek v jednej transakcii. Základom je hľadanie častých množín položiek. Podpora množiny položiek X

$$Podpora(X) = \frac{|\{i \mid X \subseteq D\{i\}\}|}{N}$$

Kód v Matlabe:

```
function S=ItemSetSupport(X,D)
% Compute support of the set of items X
% in the set of transactions D

N=size(D,1)
count=0

for I =1:N
    T=D{I}
    if isempty(setdiff(X,T))
        count=count+1
    end
end
S=count/N
end
```

Podobne jednoducho sa dá implementovať výpočet podpory, spoľahlivosti a zlepšenia pravidla $A \rightarrow B$.

1.1 Príklad

Máme súbor 100 transakcií s položkami a, b, c . Zaujímajú nás pravidlá s minimálnou podporou 0.60, teda množina položiek obsiahnutých v pravidle sa musí vyskytnúť aspoň v 60 transakciách.

Množina položiek	Počet výskytov
$\{a\}$	65
$\{b\}$	50
$\{c\}$	80

čtetnosť množiny položiek $\{b\}$ je pod dolnou hranicou, tak žiadna dvojprvková množina položiek obsahujúca b nemôže mať dostatočnú podporu – zostáva jediná možnosť $\{a, c\}$.

1.2 Algoritmus Apriori (Agrawal & Srikant 1994)

Užívateľ zadá minimálnu podporu p . Na redukciu počtu uvažovaných množín položiek sa použije nasledujúca idea:

1. jedno prvkové časté množiny položiek (s podporou aspoň p) sa použijú na generovanie 2-prvkových množín častých položiek,
2. 2-prvkové časté množiny položiek sa použijú na generovanie 3-prvkových množín častých položiek,
3. ...

Generovanie pravidiel:

- (1) Vygenerujeme všetky množiny položiek s podporou aspoň p (nazývajú sa časté množiny položiek)
- (2) Pre každú častú množinu položiek X vygenerujeme všetky asociačné pravidlá tvaru $A \rightarrow B$ také, že

$$A \cup B = X \text{ a } A \cap B = \emptyset.$$

- (3) Vylúčime pravidlá, ktoré nespĺňajú ďalšie užívateľom zadané podmienky, napr. minimálnu spoľahlivosť.

1.3 Úloha

Naprogramujte funkcie

- Podpora(A, B, D) – vráti podporu pravidla $A \rightarrow B$ v súbore transakcií D ,
- Spol(A, B, D) – vráti spoľahlivosť pravidla $A \rightarrow B$ v súbore transakcií D ,
- Zlep(A, B, D) – vráti zlepšenie pravidla $A \rightarrow B$ v súbore transakcií D ,
- Apriori(P, S, D) – vráti (alebo len vypíše) všetky pravidlá s popodporou aspoň P a spoľahlivosťou aspoň S v súbore transakcií D ,

Volanie pre súbor transakcií

```
1 3 4
1 2 3 5
2 3 5
2 5
1 2 3 6
```

by mohlo vyzerat nasledovne

```
Apriori(0.25,0.58,D)
Found a total of 16 association rules:
=====
Rule Confidence Support
=====
1 2 ==> 3 1 0.4
3 5 ==> 2 1 0.4
1 ==> 2 3 0.666 0.4
1 3 ==> 2 0.666 0.4
2 3 ==> 1 0.666 0.4
```

5 ==> 2 3 0.666 0.4
2 3 ==> 5 0.666 0.4
2 5 ==> 3 0.666 0.4
1 ==> 3 1 0.6
5 ==> 2 1 0.6
3 ==> 1 0.75 0.6
2 ==> 3 0.75 0.6
3 ==> 2 0.75 0.6
2 ==> 5 0.75 0.6
5 ==> 3 0.666 0.4
1 ==> 2 0.666 0.4