

Local Patterns in Data

Esther Galbrun

Spring 2023



UNIVERSITY OF
EASTERN FINLAND

Q1.1: Transactional dataset

Consider the following dataset and min. supp. threshold $\sigma = 2$.
Indicate the number of...

- i) ... transactions
- ii) ... items
- iii) ... transactions of size two
- iv) ... itemsets of size two
- v) ... frequent items

(0) { B, H, }
(1) { T }
(2) { T }
(3) { H, T }
(4) { B,D, R,S }
(5) { D, S }
(6) { B, R,S }
(7) { S }
(8) { D, S }
(9) { D, R }

Q1.2: Level-wise candidate generation

Considering the same dataset and min. supp. threshold $\sigma = 2$.

The frequent itemsets of size 2 are

$\{S, D\}, \{S, R\}, \{S, B\}, \{D, R\}$ and $\{R, B\}$.

What candidates of size 3 are generated in the next step, using extension by one item?

Which one can be pruned without support counting?

$\{S, D, R\}$

$\{S, D, B\}$

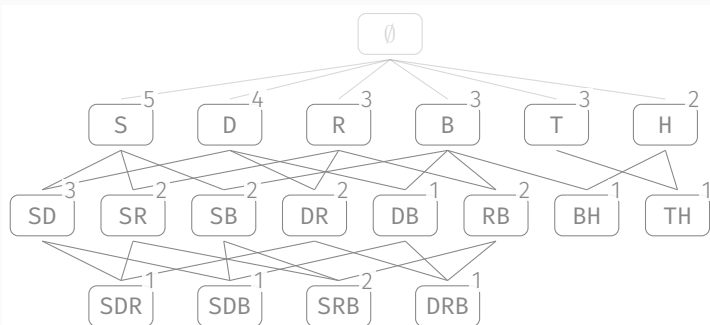
$\{S, R, B\}$

$\{D, R, B\}$

$\{R, B, H\}$

Q1.3: Itemset lattice

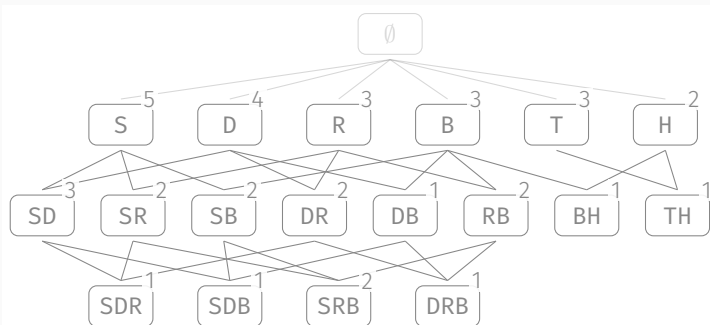
Considering the same dataset and min. supp. threshold $\sigma = 2$. Below is a part of the itemset lattice. Indicate the number of...



- ... frequent itemsets
- ... closed itemsets
- ... maximal itemsets

Q1.4: Association rules

Considering the same dataset and min. supp. threshold $\sigma = 2$. Below is a part of the itemset lattice.



Of the association rules that can be formed with $\{S, R, B\}$ which has the highest confidence?