Session 5. Constrained Mining with Data Anti-Monotonicity

Data Space Pruning with Data Anti-Monotonicity

- A constraint c is *data anti-monotone*: In the mining process, if a data entry *t* cannot satisfy a pattern *p* under *c*, *t* cannot satisfy *p*'s superset either
  - Data space pruning*:* Data entry *t* can be pruned
- Ex. 1: c_1: *sum(S.Price)* $\geq v$ is data anti-monotone
- Ex. 2: c_2: *min(S.Price)* $\leq v$ is data anti-monotone
- Ex. 3: c_3: *range(S.profit)* ≥ 25 is data anti-monotone
  - Attention: Exploring recursive data space pruning

TID	Transaction
10	a, b, c, d, f, h
20	b, c, d, f, g, h
30	b, c, d, f, g
40	a, c, e, f, g

min_sup = 2

price(item)>0

Item	Profit
a	40
b	0
c	-20
d	-15
e	-30
f	-10
g	20
h	-5

Data space pruning in the FP-growth mining process:

b's-proj. DB

TID	Transaction
~~10~~	~~a, c, d, f, h~~
20	c, d, f, g, h
30	c, d, f, g

Recursive Data Pruning

Constraint: range{S.profit} ≥ 25

b's FP-tree

single branch: cdfg: 2