

Homework 1 Pt 1

Singles	Doubles	Triples
M 3 60%	MO 1	MKO 1
O 3 60%	ME 2	KOE 3 60%
K 5 100%	MY 2	OKY 2
E 4 80%	OE 3 60%	EKY 2
Y 3 60%	OE 2	MKY 2
N 2	KE 4 80%	
A 1	KY 3 60%	
C 1	EY 2	
D 1		
U 1		
I 1		

Below minsup at 3/5 = 60%

Frequent Itemsets

M	60%
O	60%
K	100%
E	80%
Y	60%
MK	60%
OK	60%
OE	60%
KE	80%
KY	60%
KOE	60%

(d) Association rules

M → K	3/3 = 100%
K → M	3/5 = 60%
O → K	3/3 = 100%
K → O	3/5 = 60%
O → E	3/3 = 100%
E → O	3/4 = 75%
K → E	4/5 = 80%
E → K	4/4 = 100%
K → Y	3/5 = 60%
Y → K	3/3 = 100%
K → OE	3/5 = 60%
O → KE	3/3 = 100%
E → KO	3/4 = 75%
OE → K	3/3 = 100%
KE → O	3/4 = 75%
KO → E	3/3 = 100%

min conf = 80%

Strikethru = doesn't meet min-conf

Strong assoc rules: conf 1. ft

M → K	100%	1
O → K	100%	1
O → E	100%	1.25
K → E	80%	1
E → K	100%	1
Y → K	100%	1
O → KE	100%	1.25
OE → K	100%	1
KO → E	100%	1.25

All doubles and triples not included were eliminated by Apriori property

Strikethru = does not mean minsup

All letter groups are considered sets

(b) A closed frequent itemset is an itemset that meets the min support and has no superset containing it that has a higher frequency/support.

{K}, {M, K}, {K, E}, {K, Y}, {K, O, E}

100%, 60%, 80%, 60%, 60%

(c) A max frequent itemset is an itemset that has no superset that meets the minsup.

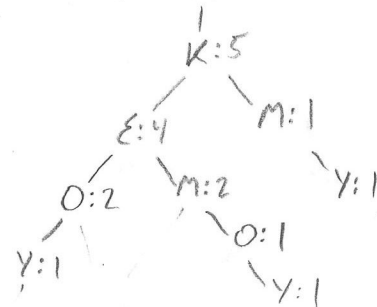
{M, K}, {K, Y}, {K, O, E}

(e)

Ex (2) (b) tree₃₃ = {}

F-List

K	5
E	4
M	3
O	3
Y	3



(c) FPgrowth(tree₃₃, null)

No support for null
FPgrowth(tree₃₃, Y) Tree:
CPB: {Y}
{K, E, O:13}
{K, E, M, O:13}
{M, K:13}

FPgrowth(tree₃₃, M)

CPB: {K:13}, {K, E:23}

Tree: {M}, {K:33}

(d) you only need to do two passes using FPgrowth

Freq Patterns

{M, K}, {M}, {Y, K}, {O}, {K, E}, {E}, {O, E}, {K, O}, {K, O, E}

FPgrowth(tree₃₃, O)

CPB: {K, E:23}, {K, E, M:13}

Tree: {O}, K:3, E:3

order
KEMOY
KEOY
KEM
KMY
KEO

LOE: {T100, T200, T500} → KOE

MKO: {T700}
OKY: {T100, T200}
EKY: {T100, T200}
MKY: {T100, T400}

MK: {T100, T300, T400} → OK
OK: {T100, T500} → OE
OE: {T100, T500} → KE
KE: {T100, T300} → KY
KY: {T100, T400} → KY

MO: {T100}
ME: {T100, T300}
MY: {T100, T400}
DY: {T100, T200}
EY: {T100, T200}

(b) LOI = 3/5
if set size ≥ 3 the fragment

M
O
K
E
Y

→ M: {T100, T200, T500}
→ O: {T100, T200, T500}
→ K: {T100, T200, T300, T400, T500}
→ E: {T100, T200, T300, T500}
→ Y: {T100, T200, T400}

EX(3)(a) M: {T100, T300, T400}

N: {T100, T200}
A: {T300}
C: {T400}
D: {T200}
U: {T400}
I: {T500}

EX(4)(a)

A	NA	B	LB	RB	Total
105	40	65	35	10	150
45	10	35	35	10	150

A, B support = 65/150 = .43 > .4 = min-sup

A → B confidence = 100/65 = .65 > .6 = min-conf

It is a strong rule b/c support and confidence are above minimum

(b) Lift = confidence / support of consequent = .65 / (105/150) = .7 = .929

Lift compares frequency of B happening due to A vs B just happening. Helps us determine whether A → B or B is just common. .929 < 1 so there is a loss of a chance of seeing B if we see A than if we don't

Know if we see A. They are almost independent b/c lift is close to 1. The rule is not very significant.

(c)

A → B = 105/150 = .70
A → B = 45/100 = .30
A → B = 50/105 = .35
A → B = 50/45 = .15

A	NA	B	LB	RB	Total
105	35	70	15	30	150
45	15	30	15	30	150

A → B support = 35/150 = .23 < .4 = min-sup

Lift = 23 / (100/150 * 45/150) = 1.16

Conf = 35/100 = .35 < .6 = min-conf

(e) A → NB

They seem to be fairly independent considering that the p value is only slightly above .05.

p-value = .0588

X² = 3.571 df=1

(d) X² = (65-70)² / 70 + (40-35)² / 35 + (35-30)² / 30 + (10-15)² / 15

4) (f) $\text{Not } B \rightarrow A$ $\text{conf} = \frac{35}{45} = .\overline{7}$ $\sim B \rightarrow A$ is stronger b/c it has a higher confidence.
 $\text{lift} = \frac{.23}{.2} = 1.\overline{16}$

(g) $\frac{1}{2}(.7 + .35) = .564$

(h)
$$\text{IR} = \frac{\text{Sup}(A) - \text{Sup}(\sim B)}{\text{Sup}(A \cup \sim B)} = \frac{100 - 45}{100 + 45 - 35} = \frac{55}{110} = \frac{1}{2}$$

5) FP growth