

# 数据挖掘第三次作业

闫子健 2020E8013382014 沈阳计算所

## Part 1: written part

### 1. (a)

$\min\_sup = 0.4$

	a	b	c	d	e
sup(x)	0.7	0.6	0.5	0.5	0.8

故a, b, c, d, e均为频繁集

	ab	ac	ad	ae
sup(x)	0.4	0.3	0.4	0.6

	bc	bd	be
sup(x)	0.3	0.2	0.5

	cd	ce	de
sup(x)	0.2	0.3	0.5

故频繁集有ab, ad, ae, be, de

	abe	ade
sup(x)	0.3	0.4

故ade为频繁项集。

综上，频繁项集有{a}, {b}, {c}, {d}, {e}, {ab}, {ad}, {ae}, {be}, {de}, {ade}

### (b)

$\text{confidence}(\{a,b\} \Rightarrow \{c\}) = \text{count}(a,b,c) / \text{count}(a,b) = 2/4 = 0.5$

$\text{confidence}(\{c\} \Rightarrow \{a,b\}) = \text{count}(a,b,c) / \text{count}(c) = 2/5 = 0.4$

故不是对称测量。

### (c)

有频繁项集{ade}

对于频繁项集ade，有频繁子集{a},{d},{e},{ad},{ae},{de}

那么可以得到的规则如下：

	$a \wedge d \Rightarrow e$	$a \wedge e \Rightarrow d$	$d \wedge e \Rightarrow a$
confidence	1	0.667	0.8

故有  $\text{buys}(X, a) \wedge \text{buys}(X, d) \Rightarrow \text{buys}(X, e)$     0.4    1

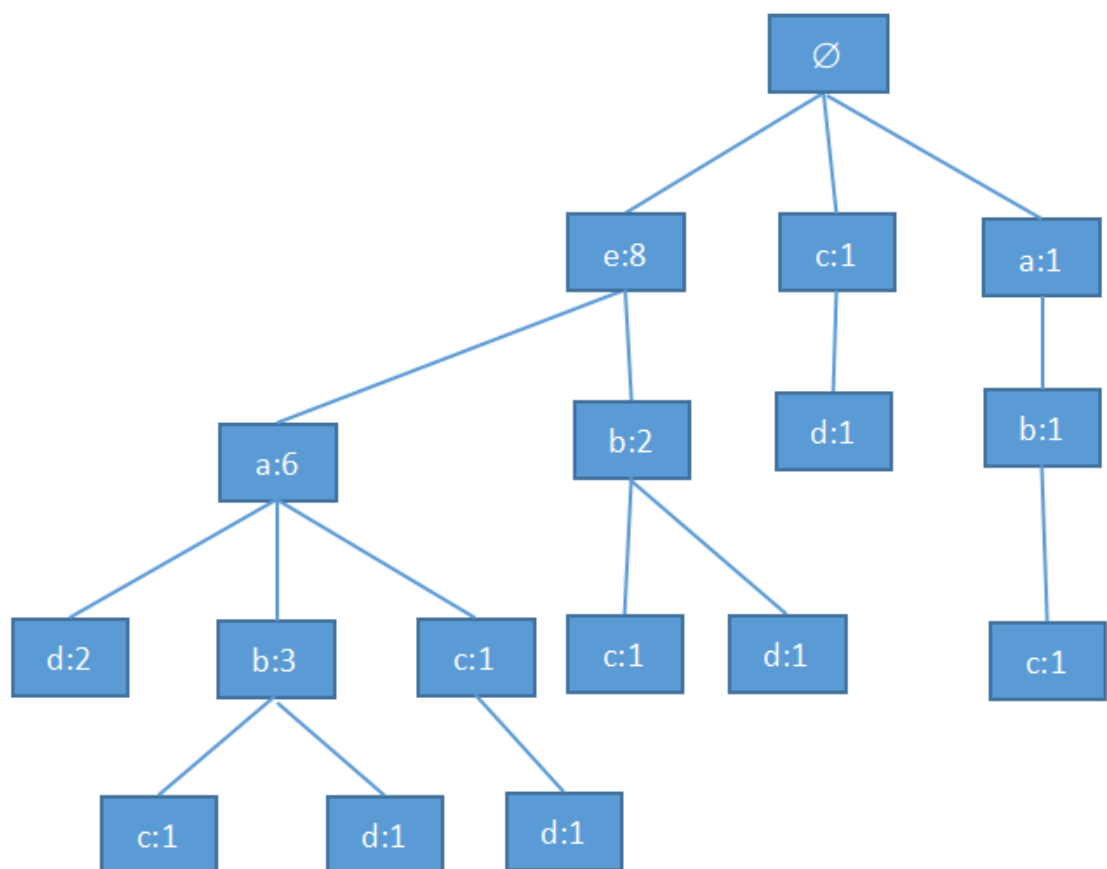
$\text{buys}(X, d) \wedge \text{buys}(X, e) \Rightarrow \text{buys}(X, a)$     0.4    0.8

## 2.(a)

Header Table	
<u>Item frequency head</u>	
e	8
a	7
b	6
c	5
d	5

Items Bought	(Ordered)frequent items
{a,d,e}	{e, a, d}
{a,b,c,e}	{e, a, b, c}
{a,b,d,e}	{e, a, b, d}
{a,c,d,e}	{e, a, c, d}
{b,c,e}	{e, b, c}
{b,d,e}	{e, b, d}
{c,d}	{c, d}
{a,b,c}	{a, b, c}
{a,d,e}	{e, a, d}
{a,b,e}	{e, a, b}

FP-tree如下:



### Conditional pattern bases

item	cond. pattern base
a	e:6
b	ea:3, e:2, a:1
c	eab:1, ea:1, eb:1, ab:1
d	ea:2, eab:1, eac:1, c:1,

(b)

fpgrowth比apriori更快一些

### 3. (a)

- 计算平均值

$$\bar{r}_{User1} = \frac{1 + 1 + 5 + 3}{4} = 2.5$$

$$\bar{r}_{User2} = \frac{4 + 5 + 3}{3} = 4$$

$$\bar{r}_{User3} = \frac{1 + 3 + 1 + 1}{4} = 1.5$$

$$\bar{r}_{User4} = \frac{4 + 3 + 2 + 1}{4} = 2.5$$

$$\bar{r}_{User5} = \frac{2 + 2 + 2 + 4}{4} = 2.5$$

- 计算item-item similarity

$$\cos(User2, User1) = \frac{1 \times 3 + 5 \times 5 + 3 \times 4}{\sqrt{35}\sqrt{50}} = 0.956$$

$$\cos(User2, User3) = \frac{1 \times 3 + 1 \times 5 + 1 \times 4}{\sqrt{3}\sqrt{50}} = 0.980$$

$$\cos(User2, User4) = \frac{4 \times 3 + 2 \times 5 + 1 \times 4}{\sqrt{21}\sqrt{50}} = 0.802$$

$$\cos(User2, User5) = \frac{2 \times 3 + 2 \times 5 + 4 \times 4}{\sqrt{24}\sqrt{50}} = 0.924$$

- Predict User 2's rating for Product 2. Assume that neighborhood size = 2

$$r_{User2, Product2} = 4 + \frac{0.980(3 - 1.5) + 0.956(1 - 2.5)}{0.980 + 0.956} = 4.02$$

## Part 2: Lab part

1.

The list of association rules generated by the model:

20 fields						
File Generate						
Sort by: Confidence 16						
Instances	Support	Confidence	Lift	Consequent	Antecedent 1	Antecedent 2
3590	7.800	57.900	1.254	milk	biscuits	pasta
4417	9.600	56.000	1.214	milk	water	pasta
3818	8.300	53.300	1.155	milk	juices	
5363	11.600	53.200	1.520	pasta	tomato sou...	
7045	15.200	52.200	1.131	milk	yoghurt	
9468	20.500	51.500	1.117	milk	biscuits	
5363	11.600	51.300	1.112	milk	tomato sou...	
6949	15.000	49.900	1.081	milk	coffee	
7084	15.300	49.700	1.077	milk	brioches	
4951	10.700	47.300	1.025	milk	coke	
12879	27.900	46.700	1.012	milk	water	
4803	10.400	46.400	1.006	milk	tunny	
46243	100.000	46.100	1.000	milk		
16201	35.000	45.900	0.994	milk	pasta	
3466	7.500	45.500	1.300	pasta	coffee	milk
5050	10.900	45.400	0.985	milk	beer	

- Sort the rules by lift

support	confidence	lift	Consequent	Antecedent1	Antecedent1
11.6	53.2	1.52	pasta	tomato souce	
7.5	45.5	1.3	pasta	coffee	milk
8.3	53.2	1.155	milk	juices	
15.2	52.2	1.131	milk	yoghurt	
20.5	51.5	1.117	milk	biscuits	
15	49.9	1.081	milk	coffee	

本来有一个milk和tomato souce关联的规则，不过不知道买牛奶的为啥还想买番茄酱，感觉逻辑上没有实际关联 就去掉了。

- Sort the rules by support

support	confidence	lift	Consequent	Antecedent1	Antecedent1
35	45.9	0.994	milk	pasta	
20.5	51.5	1.117	milk	biscuits	
15.3	49.7	1.077	milk	brioche	
15.2	52.2	1.131	milk	yoghurt	
15	49.9	1.081	milk	coffee	
11.6	53.2	1.520	pasta	tomato souce	

单纯从支持度来看，很多人都很喜欢在买牛奶时买一些固态的物品来吃，比如说意大利面，饼干，奶油卷之类的。同时，一部分人在购买牛奶时，尝尝也会购买其他的饮品，比如说酸奶，咖啡等。

- Sort the rules by confidence

support	confidence	lift	Consequent	Antecedent1	Antecedent1
11.6	53.2	1.5	pasta	tomato souce	
15.2	52.2	1.131	milk	yoghurt	
20.5	51.5	1.117	milk	biscuits	
15	59.9	1.081	milk	coffee	
15.3	49.7	1.077	milk	brioche	
10.7	47.3	1.025	milk	coke	

从置信度来说，有很高的可能性人们在买意大利面的时候会买番茄酱拌入其中一起吃。