Recitation 03 –Association Rule Mining – Jan 25, 2017

Teaching Assistant: Anatoli Shein

Assume the following market basket transactions for a fictional super-market that carries the following products: **A** (angel hair pasta), **B** (bread), **C** (cereal), **D** (diapers), **E** (eggs), and **F** (flour).

#1	A	B		D		F
#2		B	C		E	
#3	A		C	D		
#4	A	B		D		F
#5		B		D		F
#6	A		C		E	
#7	A	B	C	D		F
#8			C		E	F
#9	A	B	C	D		F
#10			C	D	E	F

(Q1) Compute the support counts for each of the size-1 itemsets:

	A	B	C	D	E	F
Count:	6	6	7	7	4	7

(Q2) Compute the support counts for each doubleton (i.e., size-2 itemsets):

	F	E	D	C	B
A	AF:4	AE:1	AD:5	AC:4	AB:4
B	BF:5	BE:1	BD:5	BC:3	
C	CF:4	CE:4	CD:4		
D	DF:6	DE:1			
E	EF:2				

(Q3) Assuming a minimum support count threshold of 3, identify which of the above doubletons would be frequent itemsets.

AF AD AC AB BF BD BC CF CE CD DF

(Q4) For the frequent size-2 itemsets identified in Q3, generate all possible association rules and compute their confidence. Assuming a minimum confidence threshold of 50% identify which association rules are good.

C	=>	E	0.5714285714
E	=>	C	1
A	=>	C	0.6666666667
C	=>	A	0.5714285714
B	=>	D	0.8333333333
D	=>	B	0.7142857143
D	=>	F	0.8571428571
F	=>	D	0.8571428571
B	=>	F	0.8333333333
F	=>	B	0.7142857143
B	=>	C	0.5
C	=>	D	0.5714285714
D	=>	C	0.5714285714
A	=>	D	0.8333333333
D	=>	A	0.7142857143
C	=>	F	0.5714285714
F	=>	C	0.5714285714
A	=>	F	0.6666666667
F	=>	A	0.5714285714
A	=>	B	0.6666666667
B	=>	A	0.6666666667

(Q5) Using the frequent size-2 itemsets from Q3, generate all possible candidate frequent size-3 itemsets. Discard those size-3 itemsets that are not possible to be frequent (frequent means to have support count of at least 3) using the subset property.

CDF

ABD

BCF

ABF

ADF

ACD

ABC

BDF

ACF

BCD

(Q6) Compute the actual support counts of the size-3 itemsets identified in Q5, and select those with support count of at least 3.

CDF 3

ABD 4

ABF 4

ADF 4

ACD 3

BDF 5

(Q7) For the frequent size-3 itemsets identified in Q6, generate all possible association rules and compute their confidence. Assuming a minimum confidence threshold of 50%, identify which association rules are good.

A	=>	D	F	0.6666666667
D	=>	A	F	0.5714285714
F	=>	D	A	0.5714285714
A	D	=>	F	0.8
A	F	=>	D	1
D	F	=>	A	0.6666666667
A	=>	B	F	0.6666666667
B	=>	A	F	0.6666666667
F	=>	B	A	0.5714285714
A	B	=>	F	1
A	F	=>	B	1
B	F	=>	A	0.8
C	D	=>	F	0.75
C	F	=>	D	0.75
D	F	=>	C	0.5
B	=>	D	F	0.8333333333
D	=>	B	F	0.7142857143
F	=>	D	B	0.7142857143
B	D	=>	F	1
B	F	=>	D	1
D	F	=>	B	0.8333333333
A	=>	C	D	0.5
A	C	=>	D	0.75
A	D	=>	C	0.6
C	D	=>	A	0.75
A	=>	B	D	0.6666666667
B	=>	A	D	0.6666666667
D	=>	B	A	0.5714285714
A	B	=>	D	1
A	D	=>	B	0.8
B	D	=>	A	0.8