COMP4433 Data Mining and Data Warehousing

FAQ on Association Rule Mining II (with suggested answers)

1. Given the following transactional database for sequential association analysis. Let *min_sup*=25% and *min_conf*=50%.

Customer	Transaction Time	Items
David	2 Feb 2002	30,50
David	10 Feb 2002	50
David	21 Feb 2002	70
John	5 Feb 2002	10,30
John	7 Feb 2002	50
John	16 Feb 2002	40,60,70
John	27 Feb 2002	50,90
Peter	16 Feb 2002	30,50,70
Aaron	5 Feb 2002	30
Aaron	7 Feb 2002	30, 50
Aaron	17 Feb 2002	70
Aaron	27 Feb 2002	50
Leon	13 Feb 2002	30

Modify the AprioriAll-based sequential pattern mining algorithm so that it allows reoccurrence of itemsets in frequent sequences, e.g., a customer bought itemset *A* and then itemset *B* will buy itemset *A* again in the future.

(a) Find all frequent sequences using your modified algorithm.

(b) Random pick one frequent 4-sequence, if there exists one, and generate the corresponding strong sequential association rule(s).

(a) *Answer*

Sort Phase:

Customer ID	Customer Sequence
David	<(30 50),(50),(70)>
John	<(10 30),(50),(40 60 70),(50 90)>
Peter	<(30 50 70)>
Aaron	<(30),(30 50),(70),(50)>
Leon	<(30)>

Litemset Phase: min_sup = 25% (i.e. >= 2 customers)

Freq. Itemsets	Mapped to
(30)	1
(50)	2
(70)	3
(30 50)	4

Transformation Phase

Customer ID	Customer Sequence	Transformed Sequence	Mapping
David	<(30 50),(50),(70)>	<{(30),(50),(30 50)},{(50)},{(70)}>	<{1,2,4}{2}{3}>
John	<(10 30),(50),(40 60 70),(50 90)>	<{(30)},{(50)},{(70)},{(50)}>	<{1}{2}{3}{2}>
Peter	<(30 50 70)>	<{(30) (50) (70) (30 50)}>	<{1,2,3,4}>
Aaron	<(30),(30 50),(70),(50)>	<{(30)},{(30),(50),(30 50)},{(70)},{(50)}>	<{1}{1,2,4}{3}{2}>
Leon	<(30)>	<(30)>	<{1}>

Sequence Phase:

1-sequence	Count	2-sequence	Count	3-sequence	Count	4-sequence	Count
<1>	5	~~<1 1>~~	~~1~~	**<1 2 3>**	**3**	**<1 2 3 2>**	**2**
<2>	4	**<1 2>**	**3**	**<1 3 2>**	**2**	~~<1 2 2 3>~~	~~0~~
<3>	4	~~<2 1>~~	~~0~~	**<1 2 2>**	**2**	~~<1 2 3 3>~~	~~0~~
<4>	3	**<1 3>**	**3**	~~<1 3 3>~~	~~0~~	~~<1 3 2 2>~~	~~0~~
		~~<3 1>~~	~~0~~	~~<2 2 2>~~	~~0~~	~~<1 2 2 2>~~	~~0~~
		~~<1 4>~~	~~1~~	~~<2 2 3>~~	~~1~~	~~<2 3 2 2>~~	~~0~~
		~~<4 1>~~	~~0~~	**<2 3 2>**	**2**		
		<2 2>	**3**	~~<2 3 3>~~	~~0~~		
		<2 3>	**3**	~~<3 2 2>~~	~~0~~		
		<3 2>	**2**	~~<4 2 2>~~	~~0~~		
		~~<2 4>~~	~~0~~	~~<4 2 3>~~	~~1~~		
		<4 2>	**2**	~~<4 3 2>~~	~~1~~		
		~~<3 3>~~	~~0~~	~~<4 3 3>~~	~~0~~		
		~~<3 4>~~	~~0~~				
		<4 3>	**2**				
		~~<4 4>~~	~~0~~				

(b) *Answer*

Rule	Support	Confidence	Strong or not?
1 → 2 3 2	2 (40%)	2/5 (40%)	No
1 2 → 3 2	2 (40%)	2/3 (~66.7%)	Yes
1 2 3 → 2	2 (40%)	2/3 (~66.7%)	Yes