

CSG230 Fall'08 Quiz 1

Oct 9, 2008

You may keep the problem statement. Write your solutions on the answer sheet IN INCREASING ORDER to facilitate grading. Write your name on the first page.

- (15%) Given a transactional database (TDB): (tid=1: A,C,D), (tid=2: B,C,E), (tid=3: A,B,C,E), (tid=4: B,E), (tid=5: E,F). Let $min_support = 40\%$.
Find all frequent itemsets using FP-growth.
- (15%) Using the TDB and $min_support$ in problem 1, find all max frequent itemsets using MaxMiner.
- (15%) Using the TDB and $min_support$ in problem 1, find all closed frequent itemsets using CLOSET.
- (15%) Draw two lattices, both using “proper superset” as the partial-order operator. The first lattice is on the results of problem 1 (all frequent itemsets). The second lattice is on the results on problem 3 (all closed itemsets).
- (15%) Using the TDB and $min_support$ in problem 1, let $min_confidence = 100\%$, find all strong association rules.
- (10%) Suppose the Apriori algorithm has found the following itemsets to be the only frequent itemsets with four items: ABCD, ABDE, ABDF, BDEF. To generate frequent itemsets with five items, Apriori has three steps: (a) generate some candidates; (b) prune certain candidates before examining the database; (c) prune non-frequent candidates by database look-up.
What candidate(s) is/are generated in step (a)? Explain.
What candidate(s) is/are pruned in step (b)? Why?
- (15%) Consider the following taxonomy: an “item(I)” can be “fruit(F)” or “candy(C)”; a “fruit” can be “apple(A)” or “banana(B)”. Draw the complete lattice of all generalized itemsets.

CS G 230 Fall '08 Quiz 1 Solution

1.

A	C	D	
B	C	E	
A	B	C	E
B	E		
E	F		

△ frequent items:

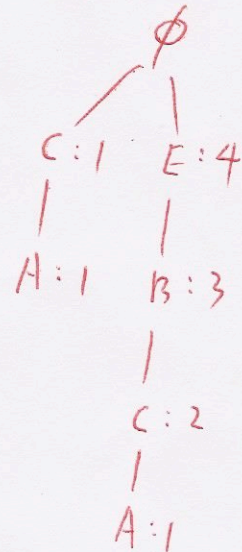
A	B	C	D	E	F
2	3	3	1	4	1
✓	✓	✓	✗	✓	✗

△ Reorder: \overrightarrow{EBCA}

△ transformed TDB:

- | | | | |
|---|---|---|---|
| C | A | | |
| E | B | C | |
| E | B | C | A |
| E | B | | |
| E | | | |

△ FP-tree



2 A-pattern base: C:1, EBC:1

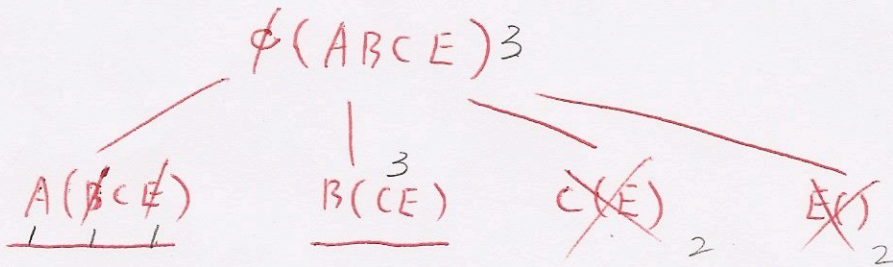
2 C-pattern base: EB:2

2 B-pattern base: E:3

2 E-pattern base: φ

- Result:
- A, AC
 - C, EC, BC, EBC
 - B, EB
 - E

2.



Max itemsets:

AC, BCE

2

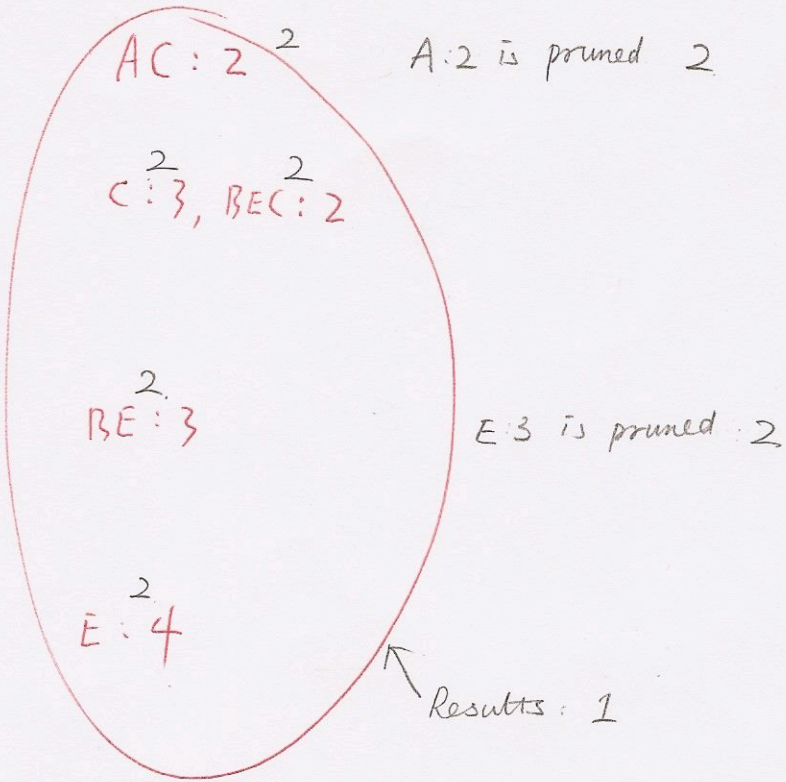
3. $DB|A = \begin{matrix} C \\ BCE \end{matrix}$

$DB-A|C = \begin{matrix} \phi \\ BE \\ BE \end{matrix}$

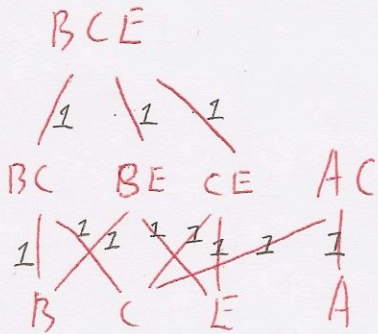
$DB-AC|B = \begin{matrix} E \\ E \\ E \end{matrix}$

$DB-ACB|E = \begin{matrix} \phi \\ \phi \\ \phi \\ \phi \end{matrix}$

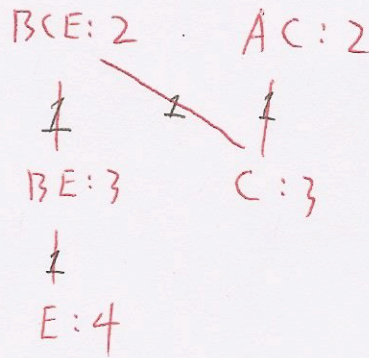
Candidates = result



4. Lattice of all FPs



Lattice of closed FPs



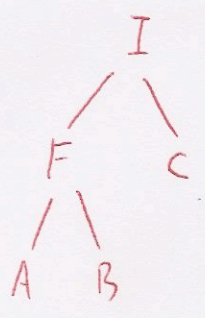
- 5.
- | | | | |
|------------------------------|---|-------------------------------|---|
| $A \rightarrow C \checkmark$ | 1 | $E \rightarrow BCX$ | I |
| $C \rightarrow AX$ | 1 | $B \rightarrow ECX$ | 1 |
| $E \rightarrow CX$ | 1 | $C \rightarrow BE X$ | 1 |
| $C \rightarrow EX$ | 1 | $BC \rightarrow E \checkmark$ | I |
| $B \rightarrow CX$ | 1 | $BE \rightarrow CX$ | 1 |
| $C \rightarrow BX$ | 1 | $CE \rightarrow B \checkmark$ | 1 |
| $E \rightarrow BX$ | 1 | | |
| $B \rightarrow E \checkmark$ | 1 | | |

So result =

$A \rightarrow C$
 $B \rightarrow E$
 $BC \rightarrow E$ 1
 $CE \rightarrow B$

6. (a) Aprior generates ABDEF 2
 because ABDE and ABDF differ only in the last item. 3
- (b) It is then pruned, because ABEF is not frequent 3

7. taxonomy



lattice

