Page No :
Date :

Data Mining
Assignment III

① **Apriori-based approach**

Apriori based frequent substructure mining algorithms share similar characteristics with Apriori based frequent itemset mining algorithms. The search for frequent graphs start with graph of small 'size' and proceeds in a bottom-up manner by generating candidates having an extra vertex, edge and path. The definition of the graph depends on the algorithm used.

The main design complexity of Apriori-based substructure mining algorithm is the candidate generation step. The candidate generation in frequent itemset mining is straightforward. The AGM algorithm uses a vertex-based candidate generation method that increases the substructure size by one vertex at each iteration of Apriori Graph.

② minimum support = 2.

C_1 =

Itemset	Support count.
I_1	5
I_2	7
I_3	6
I_4	2
I_5	2

L_1 =

Itemset	s.
I_1	5
I_2	7
I_3	6
I_4	2
I_5	2.

C_2 = I_1, I_2

~~I_1, I_3~~

~~I_1, I_4~~

~~I_4, I_5~~

~~I_5~~

C_2 =

Itemset	S.C.
I_1, I_2	4
I_1, I_3	4
I_1, I_4	1
I_1, I_5	2
I_2, I_3	4
I_2, I_4	2
I_2, I_5	2
I_3, I_4	0
I_3, I_5	1
I_4, I_5	0

L_2 =

(I_1, I_2)	4
(I_1, I_3)	4
(I_2, I_3)	4
(I_1, I_5)	2
(I_2, I_4)	2
(I_2, I_5)	2

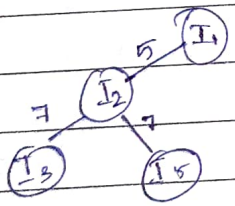
C_3 =

(I_1, I_2, I_3)	2
(I_1, I_2, I_4)	1
(I_1, I_2, I_5)	2
(I_2, I_3, I_5)	1
(I_2, I_3, I_4)	0

L_3 =

(I_1, I_2, I_3)	2
(I_1, I_2, I_5)	2

$C_3 = (I_1, I_2, I_3, I_5)$ — 1 ✗



(b) confidence = 70%

	Confidence
$(I_1, I_2) \Rightarrow I_3$	2/4 = 50%
$(I_1, I_3) \Rightarrow I_2$	2/4 = 50%
$(I_2, I_3) \Rightarrow I_1$	2/4 = 50%
$(I_1, I_2) \Rightarrow I_5$	2/4 = 50%

Page No. :
Date :

Confidence

$(I_2, I_5) \Rightarrow I_1$	$2/2 = 100\%$
$(I_1, I_5) \Rightarrow I_2$	$2/2 = 100\%$
$I_3 \Rightarrow (I_1, I_2)$	$2/6 = 33.3\%$
$I_2 \Rightarrow (I_1, I_3)$	$2/7 = 28.5\%$
$I_1 \Rightarrow (I_2, I_3)$	$2/5 = 40\%$
$I_2 \Rightarrow (I_1, I_5)$	$2/7 = 28.5\%$
$I_4 \Rightarrow (I_2, I_5)$	$2/5 = 40\%$
$I_5 \Rightarrow (I_1, I_2)$	$2/2 = 100\%$

Strong association rules $\Rightarrow$ $(I_1, I_5) \Rightarrow I_2$, $(I_2, I_5) \Rightarrow I_1$, $I_5 \Rightarrow (I_1, I_2)$ } 100%.