Suggested: Data Mining

ECLAT (Equivalence Class Transformation): Vertical Apriori

- Both Apriori and FP-growth use horizontal data format .
- ECLAT mines frequent itemsets using the Vertical Data Format.
- It is a depth first search based algorithm.
- In this, each item is stored together with its T_ID (Transaction ID).
- It uses intersection based approach to compute the support an itemset.

TID	List of Items
T1	I1, I2, I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

Min Support Count=2, Confidence =70%

Generate Association Rule using ECLAT Algorithm

Database is in Horizontal Data Format.

TID	List of Items
T1	I1, I2, I5
T2	I2,I4
T3	I2,I3
T4	I1,I2,I4
T5	I1,I3
T6	I2,I3
T7	I1,I3
T8	I1,I2,I3,I5
T9	I1,I2,I3

Step 1: Database is in Vertical Data Format

Itemset	List of Items
I1	T1, T4, T5, T7, T8, T9
I2	T1, T2, T3, T4, T6, T8, T9
I3	T3, T5, T6, T7, T8, T9
I4	T2, T4
I5	T1, T8

Step 2: Itemset generated by intersection of 1 itemset

Itemset	List of Items
I1	T1, T4, T5, T7, T8, T9
I2	T1, T2, T3, T4, T6, T8, T9
I3	T3, T5, T6, T7, T8, T9
I4	T2, T4
I5	T1, T8

Itemset	List of Items
I1, I2	T1,T4,T8,T9
I1, I3	T5,T7, T8, T9
I1, I4	T4
I1, I5	T1,T8
I2, I3	T3, T6, T8, T9
I2, I4	T2, T4
I2, I5	T1, T8
I3, I4	--
I3, I5	T8
I4, I5	--

Itemset	List of Items
I1, I2	T1,T4,T8,T9
I1, I3	T5,T7, T8, T9
I1, I4	T4
I1, I5	T1,T8
I2, I3	T3, T6, T8, T9
I2, I4	T2, T4
I2, I5	T1, T8
I3, I4	--
I3, I5	T8
I4, I5	--

Itemset	List of Items
I1, I2	T1,T4,T8,T9
I1, I3	T5,T7, T8, T9
I1, I5	T1,T8
I2, I3	T3, T6, T8, T9
I2, I4	T2, T4
I2, I5	T1, T8

Min. Support=2

Step 3: Itemset generated by intersection of 2 itemset

Itemset	List of Items
I1, I2	T1,T4,T8,T9
I1, I3	T5,T7, T8, T9
I1, I5	T1,T8
I2, I3	T3, T6, T8, T9
I2, I4	T2, T4
I2, I5	T1, T8

Itemset	List of Items
I1, I2,I3	T8, T9
I1, I2, I5	T1, T8
I1, I3, I5	T8
I2, I3, I4	--
I2, I3, I5	T8
I2, I4, I5	--

Itemset	List of Items
I1, I2,I3	T8, T9
I1, I2, I5	T1, T8

Step 3: Itemset generated by intersection of 3 itemset

Itemset	List of Items
I1, I2,I3	T8, T9
I1, I2, I5	T1, T8

Itemset	List of Items
I1, I2, I3, I5	T8

Table with four itemset is null

Rules can be formed from following three itemset

Itemset	List of Items
I1, I2,I3	T8, T9
I1, I2, I5	T1, T8

We can expand any rule.
For e.g. I1, I2 and I5

Confidence =70%

Association Rule	Confidence	Confidence (%)
I1^I2→I5	C(I1,I2,I5)/C(I1,I2)=2/4	50%
I1^I5→I2	C(I1,I2,I5)/C(I1,I5)=2/2	100%
I2^I5→I1	C(I1,I2,I5)/C(I2,I5)=2/2	100%
I1→I2^I5	C(I1,I2,I5)/C(I1)=2/6	33%
I2→I1^I5	C(I1,I2,I5)/C(I2)=2/7	29%
I5→I1^I2	C(I1,I2,I5)/C(I5)=2/2	100%

Advantages of ECLAT:

- It uses less memory than Apriori as its concept is based on depth first search.
- It requires less time for frequent pattern generation than Apriori. Because there is no repeated scanning of the data to compute individual support values.

Disadvantages of ECLAT:

- Suitable for small dataset.
- When T_ID list is large, then it takes more time to store candidate set. Also it requires more time for intersection.