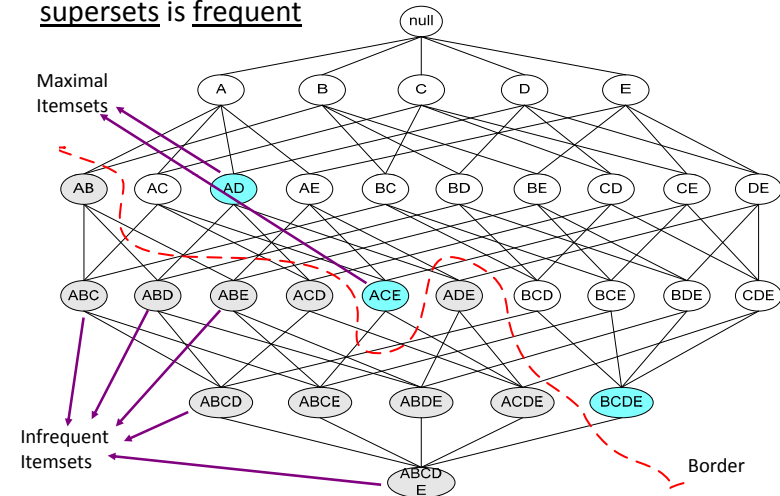


Exercises

Frequent Itemsets
Maximum Frequent Itemsets
Closed frequent Itemsets

Maximal Frequent Itemset

- An itemset is maximal frequent if none of its immediate supersets is frequent

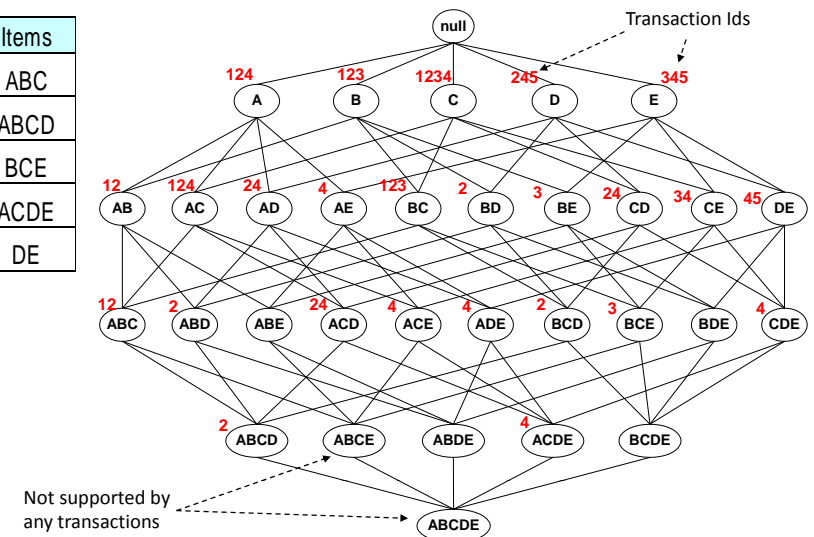


Closed Itemset

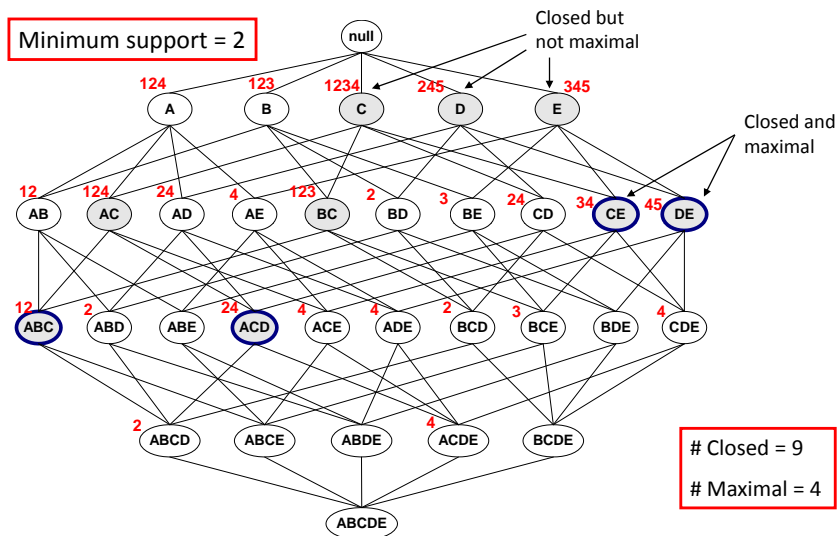
- An itemset is closed if none of its immediate supersets has the **same** support count as the itemset

Maximal vs Closed Itemsets

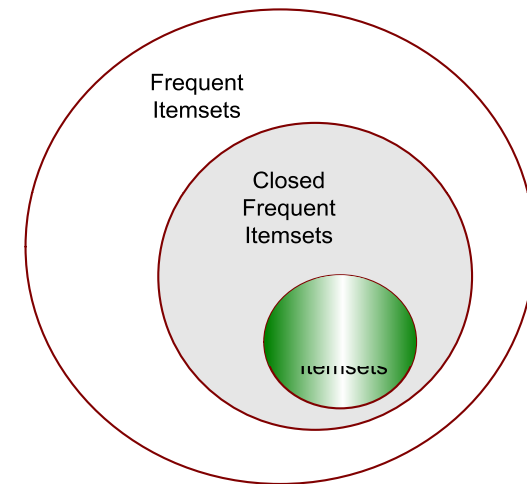
TID	Items
1	ABC
2	ABCD
3	BCE
4	ACDE
5	DE



Maximal vs Closed Frequent Itemsets



Maximal vs Closed Itemsets



TID	Items
1	1, 3, 4
2	2, 3, 5
3	1, 2, 3, 5
4	2, 5

- Find support count for itemset $X = \{2, 5\}$.
- Ans : 3
(as it appears in transactions 2, 3 and 4)

TID	Items
1	1, 3, 4
2	2, 3, 5
3	1, 2, 3, 5
4	2, 5

- If minimum support is 2, find all frequent itemsets.
- Ans :
 $\{1\}, \{2\}, \{3\}, \{5\}, \{1,3\}, \{2,3\}, \{2,5\}, \{3,5\}$ and $\{2,3,5\}$

- If minimum support is 2,
find all maximum frequent itemsets:

• Ans :

freq itemsets : {1}, {2}, {3}, {5}, {1,3}, {2,3}, {2,5},
{3,5} and {2,3,5}

~~{1}, {2}, {3}, {5}, {1,3}, {2,3}, {2,5}, {3,5}~~ and {2,3,5}

{1,3} and {2,3,5}

TID	Items
1	1, 3, 4
2	2, 3, 5
3	1, 2, 3, 5
4	2, 5

- If minimum support is 2,
find all closed frequent itemsets.

• Ans :

freq itemsets : {1:2}, {2:3}, {3:3}, {5:3}, {1,3:2},
{2,3:2}, {2,5:3}, {3,5:2} and {2,3,5:2}

~~{1:2}, {2:3}, {3:3}, {5:3}, {1,3:2}, {2,3:2}, {2,5:3},~~
~~{3,5:2}~~ and {2,3,5:2}

{3}, {1,3}, {2,5} and {2,3,5}

TID	Items
1	1, 3, 4
2	2, 3, 5
3	1, 2, 3, 5
4	2, 5

Exercises

Freq Itemsets / Assoc Rules

● Association Rule Evaluation Metrics

—Support (s)

=Fraction of transactions that contain both X and Y

—Confidence (c)

=Measures how often items in Y appear in transactions that contain X

● Example:

{Milk, Diaper} \Rightarrow Beer

$$s = \frac{\sigma(\text{Milk, Diaper, Beer})}{|T|} = \frac{2}{5} = 0.4$$

$$c = \frac{\sigma(\text{Milk, Diaper, Beer})}{\sigma(\text{Milk, Diaper})} = \frac{2}{3} = 0.67$$

TID	Items
1	Bread, Milk
2	Bread, Diaper, Beer, Eggs
3	Milk, Diaper, Beer, Coke
4	Bread, Milk, Diaper, Beer
5	Bread, Milk, Diaper, Coke

- Compute the support for itemsets {a}, {b, d}, and {a,b,d} by treating each transaction ID as a market basket.

Customer ID	<u>Transaction ID</u>	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Compute the support for itemsets {a}, {b, d}, and {a,b,d} by treating each transaction ID as a market basket.

- Ans : support

{a} =

{b, d} =

{a,b,d} =

Customer ID	<u>Transaction ID</u>	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Use the results in the previous problem to compute the confidence for the association rules {b, d} → {a} and {a} → {b, d}.
State what these values mean in plain English.

Customer ID	<u>Transaction ID</u>	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Use the results in the previous problem to compute the confidence for the association rules {b, d} → {a} and {a} → {b, d}.
State what these values mean in plain English.

- Ans : confidence

{b, d} → {a} =

{a} → {b, d} =

Customer ID	<u>Transaction ID</u>	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Compute the support for itemsets {a}, {b, d}, and {a,b,d} by treating each customer ID as a market basket.

<u>Customer ID</u>	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Compute the support for itemsets {a}, {b, d}, and {a,b,d} by treating each customer ID as a market basket.

- Ans : support

{a} =

{b, d} =

{a,b,d} =

<u>Customer ID</u>	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Use the results in the previous problem to compute the confidence for the association rules {b, d} → {a} and {a} → {b, d}.
State what these values mean in plain English.

<u>Customer ID</u>	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

- Use the results in the previous problem to compute the confidence for the association rules {b, d} → {a} and {a} → {b, d}.
State what these values mean in plain English.

- Ans : confidence

{b, d} → {a} =

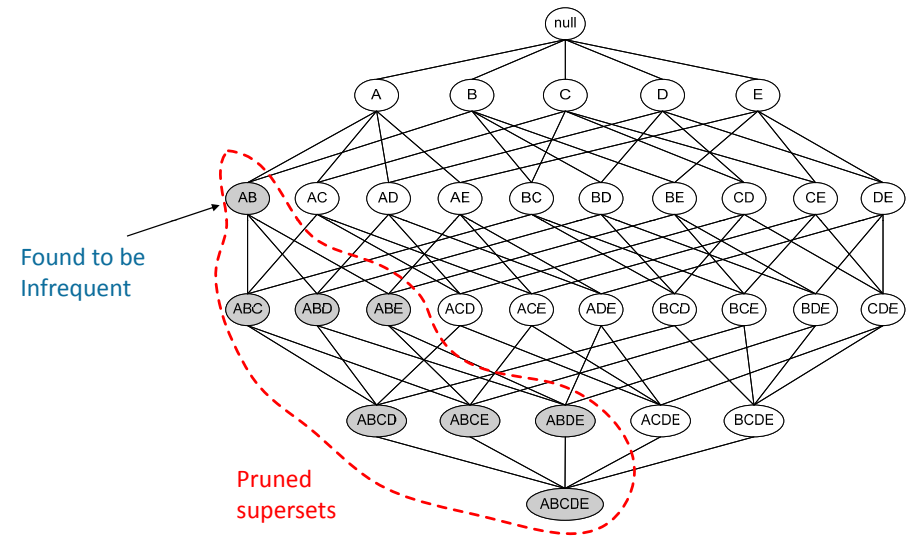
{a} → {b, d} =

<u>Customer ID</u>	Transaction ID	Items Bought
1	0001	{a, d, e}
1	0024	{a, b, c, e}
2	0012	{a, b, d, e}
2	0031	{a, c, d, e}
3	0015	{b, c, e}
3	0022	{b, d, e}
4	0029	{c, d}
4	0040	{a, b, c}
5	0033	{a, d, e}
5	0038	{a, b, e}

Exercises

Apriori

Illustrating Apriori Principle



- Use the two step approach to generate all rules having support $\geq .4$ and confidence $\geq .6$ and support $\geq .3$ and confidence $\geq .6$ for the transactions below.

Transaction ID	Items Bought
1	{Milk, Beer, Diapers}
2	{Bread, Butter, Milk}
3	{Milk, Diapers, Cookies}
4	{Bread, Butter, Cookies}
5	{Beer, Cookies, Diapers}
6	{Milk, Diapers, Bread, Butter}
7	{Bread, Butter, Diapers}
8	{Beer, Diapers}
9	{Milk, Diapers, Bread, Butter}
10	{Beer, Cookies}

Apriori
=====

Minimum support: 0.4 (4 instances)
Minimum metric <confidence>: 0.6
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Large Itemsets L(1):
Milk=TRUE 5
Beer=TRUE 4
Diapers=TRUE 7
Bread=TRUE 5
Butter=TRUE 5
Cookies=TRUE 4

Size of set of large itemsets L(2): 2

Large Itemsets L(2):
Milk=TRUE Diapers=TRUE 4
Bread=TRUE Butter=TRUE 5

Best rules found:

- Butter=TRUE 5 ==> Bread=TRUE 5 conf:(1)
- Bread=TRUE 5 ==> Butter=TRUE 5 conf:(1)
- Milk=TRUE 5 ==> Diapers=TRUE 4 conf:(0.8)

Apriori

Minimum support: 0.35 (4 instances) Minimum support: 0.3 (3 instances)

Minimum metric <confidence>: 0.6

Number of cycles performed: 13

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6

Large Itemsets L(1):

Milk=TRUE 5

Beer=TRUE 4

Diapers=TRUE 7

Bread=TRUE 5

Butter=TRUE 5

Cookies=TRUE 4

Size of set of large itemsets L(2): 7

Large Itemsets L(2):

Milk=TRUE Diapers=TRUE 4

Milk=TRUE Bread=TRUE 3

Milk=TRUE Butter=TRUE 3

Beer=TRUE Diapers=TRUE 3

Diapers=TRUE Bread=TRUE 3

Diapers=TRUE Butter=TRUE 3

Bread=TRUE Butter=TRUE 5

Size of set of large itemsets L(3): 2

Large Itemsets L(3):

Milk=TRUE Bread=TRUE Butter=TRUE 3

Diapers=TRUE Bread=TRUE Butter=TRUE 3

Best rules found:

1. Butter=TRUE 5 ==> Bread=TRUE 5 conf:(1)

2. Bread=TRUE 5 ==> Butter=TRUE 5 conf:(1)

3. Milk=TRUE Butter=TRUE 3 ==> Bread=TRUE 3 conf:(1)

4. Milk=TRUE Bread=TRUE 3 ==> Butter=TRUE 3 conf:(1)

5. Diapers=TRUE Butter=TRUE 3 ==> Bread=TRUE 3 conf:(1)

6. Diapers=TRUE Bread=TRUE 3 ==> Butter=TRUE 3 conf:(1)

7. Milk=TRUE 5 ==> Diapers=TRUE 4 conf:(0.8)

8. Beer=TRUE 4 ==> Diapers=TRUE 3 conf:(0.75)

9. Bread=TRUE 5 ==> Milk=TRUE 3 conf:(0.6)

10. Milk=TRUE 5 ==> Bread=TRUE 3 conf:(0.6)