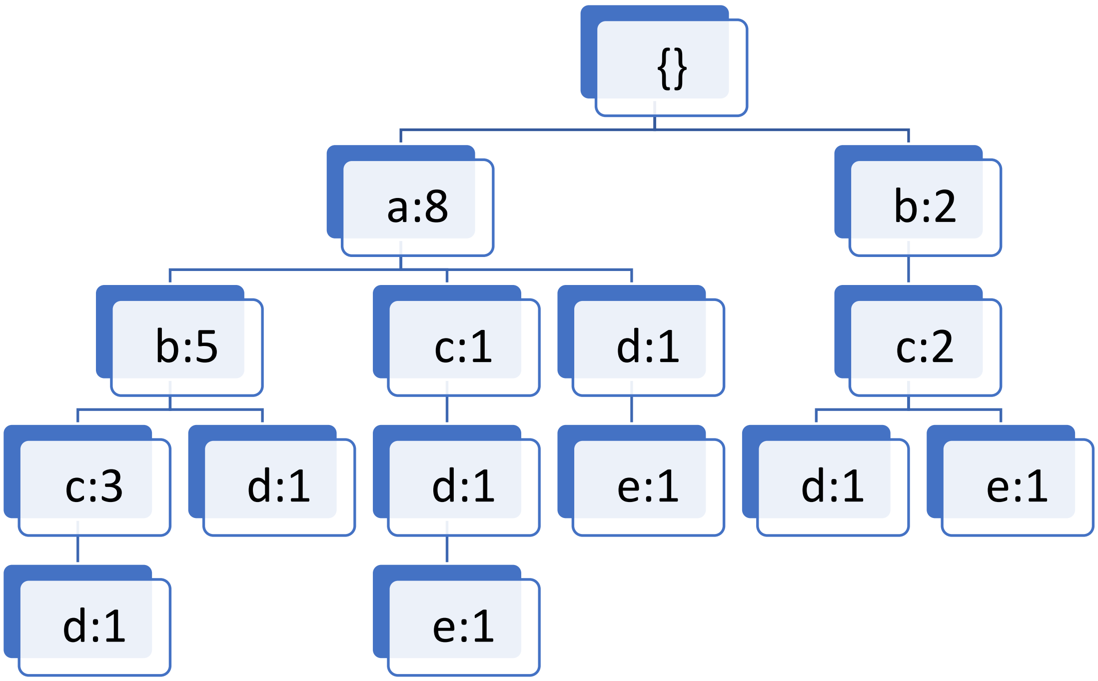
Ryan Karl
CSE-40647 Data Science
Dr. Meng Jiang
17 November 2019

Homework 4 Solutions

Question 1 Solution:



Frequent Patterns List with support:

{[a]:8}
{[b]:7}
{[c]:6}
{[d]:5}
{[e]:3}

{[a,b]:5}
{[a,c]:4}
{[a,d]:4}
{[a,e]:2}
{[b,c]:5}
{[b,d]:3}
{[c,d]:3}
{[c,e]:2}
{[d,e]:2}

{[a,d,e]:2}
{[a,b,d]:2}
{[a,b,c]:3}

{[a,c,d]:2}
{[b,c,d]:2}

Question 2 Solution:

Please note that Null Invariance is when the measure value does not change with the number of null-transactions. This means that a measure that is null invariant will not contain a term that represents the overall number of the transactions in its formula (for the proof this is represented as the term T). Below we prove that Lift is not null invariant but Cosine is null invariant. Notice that will use the shorthand:

$$s(X) = \frac{number\ of\ transactions\ containing\ X}{total\ number\ of\ transactions} = \frac{x}{T}$$

for the proofs, where the abbreviation x to represents the number of transactions containing X and similarly T is the total number of transactions. If we abbreviate the number of transactions containing both A and B as c, this means that

$$Lift(A,B) = \frac{s(A \cup B)}{s(A) \times s(B)} = \frac{\frac{c}{T}}{\frac{a}{T} \times \frac{b}{T}} = \frac{\frac{c}{T}}{\frac{a \times b}{T^2}} = \frac{c \times T \times T}{a \times b \times T} = \frac{c \times T}{a \times b}$$

and clearly we cannot remove the T term from the equation. Thus Lift is not null invariant. However, we see that

$$Cosine(A,B) = \frac{s(A \cup B)}{\sqrt{s(A) \times s(B)}} = \frac{\frac{c}{T}}{\sqrt{\frac{a}{T} \times \frac{b}{T}}} = \frac{\frac{c}{T}}{\frac{\sqrt{a \times b}}{\sqrt{T^2}}} = \frac{\frac{c}{T}}{\frac{\sqrt{a \times b}}{T}} = \frac{c \times T}{\sqrt{a \times b} \times T} = \frac{c}{\sqrt{a \times b}}$$

and clearly we are able to successfully remove the T from the equation. Thus Cosine is null invariant.

Question 3 Solution:

Pattern 1: {D}

Not a closed pattern. Notice that the min_sup for {D} is 2 (from transactions 3 and 4) but the min_sup for {B,D} is also 2 (again from transactions 3 and 4) and {B, D} is a superset of {D}.

Pattern 2: {A, B, C, F}

Is a closed pattern. Notice that the min_sup for {A, B, C, F} is 2 (from transactions 2 and 3). Any superset of {A, B, C, F} must have at least 5 five elements, but only one transaction contains 5 elements (transaction 3). This means it is impossible for a superset to exist that has a min_sup of at least 2.

Pattern 3: {B, F}

Not a closed pattern. Notice that the min_sup for {B, F} is 2 (transactions 2 and 3) but the min_sup for {B, C, F} is also 2 (again from transactions 2 and 3) and {B, C, F} is a superset of {B, F}.

Pattern 4: {B, D}

Is a closed pattern. Notice that the min_sup for {B, D} is 2 (transactions 3 and 4). Now transactions 1 and 2 do not contain D so we can ignore them. The only remaining element in transaction 4 besides B and D is E, but E is not contained in transaction 3. Therefore there is no superset of {B, D} that has a min_sup of at least 2.

Pattern 5: {A, C, F}

Is a closed pattern. Notice that the min_sup for {A, C, F} is 3 (transactions 1, 2 and 3). Now transaction 4 contains none of these elements so we can ignore it. Now besides A, C, and F, the other elements contained by transactions 1, 2 and 3 are G, B, and D, so for {A, C, F} to not be a closed pattern we would have to find at least of G, B, or D present in transactions 1, 2, and 3. Note that B and D are not present in transaction 1, D and G are not present in transaction 2, and G is not present in transaction 3. Therefore, we know there is no superset of {A, C, F} that has a min_sup of at least 3.

Question 4 Solution:

Pattern 1: {ACF}

Is a sequential pattern. Notice that {ACF} appears in both sequences 1 and 2, and can be constructed from events 1, 2, and 3 of both. Thus its min_sup = 2.

Pattern 2: {(FG)B}

Is not a sequential pattern. Notice that {(FG)B} cannot be constructed from any sequence. Clearly sequence 2 does not contain (FG) in any of its 3 events. Also, although (FG) does appear in event 3 of sequence 1 and event 2 of sequence 3, there is no B in any event following event 3 of sequence 1 and sequence 3 has no events following event 2.

Pattern 3: {(FG)}

Is a sequential pattern. Notice that {(FG)} appears in both event 3 of sequence 1 and event 2 of sequence 3, and thus its min_sup = 2.

Pattern 4: {B(FG)}

Is a sequential pattern. Notice that {B(FG)} can be constructed from events 1 and 3 of sequence 1 and events 1 and 2 of sequence 3, and thus its min_sup = 2.

Pattern 5: {GF}

Is not a sequential pattern. Notice that {GF} cannot be constructed from any sequence. Clearly sequence 2 does not contain G in any of its 3 events. Also, although G does appear in event 3 of sequence 1 and event 2 of sequence 3, there is no F in any event following event 3 of sequence 1 and sequence 3 has no events following event 2.

Ryan Karl
CSE-40647 Data Science
Dr. Meng Jiang
17 November 2019

Homework 4 Solutions

Question 5 Solution:

Frequent Patterns List with support:

{[a]:8}
{[b]:7}
{[c]:6}
{[d]:5}
{[e]:3}

{[a,b]:5}
{[a,c]:4}
{[a,d]:4}
{[a,e]:2}
{[b,c]:5}
{[b,d]:3}
{[c,d]:3}
{[c,e]:2}
{[d,e]:2}

{[a,d,e]:2}
{[a,b,d]:2}
{[a,b,c]:3}
{[a,c,d]:2}
{[b,c,d]:2}