

Exercise

Trace the results of using the Apriori algorithm on the grocery store example with support threshold $s=33.33\%$ and confidence threshold $c=60\%$. Show the candidate and frequent itemsets for each database scan. Enumerate all the final frequent itemsets. Also indicate the association rules that are generated and highlight the strong ones, sort them by confidence.

Transaction ID	Items
T1	HotDogs, Buns, Ketchup
T2	HotDogs, Buns
T3	HotDogs, Coke, Chips
T4	Chips, Coke
T5	Chips, Ketchup
T6	HotDogs, Coke, Chips

Apriori Algorithm

Minimum support = $(33.33/100) * 6 = 2$

1st scan

C1

Itemset	Sup-count
HotDogs	4
Buns	2
Ketchup	2
Coke	3
Chips	4

L1

Itemset	Sup-count
HotDogs	4
Buns	2
Ketchup	2

Coke	3
Chips	4

2nd scan

C2

Itemset	Sup-count
HotDogs, Buns	2
HotDogs, Ketchups	1
HotDogs, Coke	2
HotDogs, Chips	2
Buns, Ketchup	1
Buns, Coke	0
Buns, Chips	0
Ketchup, Coke	0
Ketchup, Chips	1
Coke, Chips	3

L2

Itemset	Sup-count
HotDogs, Buns	2
HotDogs, Coke	2
HotDogs, Chips	2
Coke, Chips	3

3rd scan

C3

Itemset	Sup-count
HotDogs, Buns, Coke	0
HotDogs, Buns, Chips	0
HotDogs, Coke, Chips	2

L3

Itemset	Sup-count
HotDogs, Coke, Chips	2

All frequent item sets.

$L1 \cup L2 \cup L3 = \{\{\text{HotDogs}\}, \{\text{Buns}\}, \{\text{Ketchup}\}, \{\text{Coke}\}, \{\text{Chips}\}, \{\text{HotDogs}, \text{Buns}\}, \{\text{HotDogs}, \text{Coke}\}, \{\text{HotDogs}, \text{Chips}\}, \{\text{Coke}, \text{Chips}\}, \{\text{HotDogs}, \text{Coke}, \text{Chips}\}\}$

Strong association rules

HotDogs -> Buns	Support (2/6) *100 = 33.33%	Confidence = (2/4) *100 = 50%	
Buns -> HotDogs	Support (2/6) *100 = 33.33%	Confidence = (2/2) *100 = 100%	Selected
HotDogs -> Coke	Support (2/6) *100 = 33.33%	Confidence = (2/4) *100 = 50%	
Coke -> HotDogs	Support (2/6) *100 = 33.33%	Confidence = (2/3) *100 = 66.67%	Selected
HotDogs -> Chips	Support (2/6) *100 = 33.33%	Confidence = (2/4) *100 = 50%	
Chips -> HotDogs	Support (2/6) *100 = 33.33%	Confidence = (2/4) *100 = 50%	
Coke -> Chips	Support (3/6) *100 = 50%	Confidence = (3/3) *100 = 100%	Selected
Chips -> Coke	Support (3/6) *100 = 50%	Confidence = (3/4) *100 = 75%	Selected

HotDogs -> Coke \wedge Chips	Support (2/6) *100 = 33.33%	Confidence = (2/4) *100 = 50%	
Coke -> HotDogs \wedge Chips	Support (2/6) *100 = 33.33%	Confidence = (2/3) *100 = 66.67%	Selected
Chips -> HotDogs \wedge Coke	Support (2/6) *100 = 33.33%	Confidence = (2/4) *100 = 50%	
HotDogs \wedge Coke -> Chips	Support (2/6) *100 = 33.33%	Confidence = (2/2) *100 = 100%	Selected
HotDogs \wedge Chips -> Coke	Support (2/6) *100 = 33.33%	Confidence = (2/2) *100 = 100%	Selected
Coke \wedge Chips => HotDogs	Support (2/6) *100 = 33.33%	Confidence = (2/3) *100 = 66.67%	Selected

Strong rules are rules that have confidence = 50%, 66.67%, 75% and 100%

Sort from maximum to minimum by confidence

Buns -> HotDogs **Confidence** = $(2/2) * 100 = 100\%$
Coke -> Chips **Confidence** = $(3/3) * 100 = 100\%$
HotDogs \wedge Coke -> Chips **Confidence** = $(2/2) * 100 = 100\%$
HotDogs \wedge Chips -> Coke **Confidence** = $(2/2) * 100 = 100\%$
Chips -> Coke **Confidence** = $(3/4) * 100 = 75\%$
Coke -> HotDogs **Confidence** = $(2/3) * 100 = 66.67\%$
Coke -> HotDogs \wedge Chips **Confidence** = $(2/3) * 100 = 66.67\%$
Coke \wedge Chips => HotDogs **Confidence** = $(2/3) * 100 = 66.67\%$

Frequent Pattern-Growth Algorithm

Minimum support = $(33.33/100) * 6 = 2$

1st scan

Itemset	Sup-count
HotDogs	4
Buns	2
Ketchup	2
Coke	3
Chips	4

Sort frequent

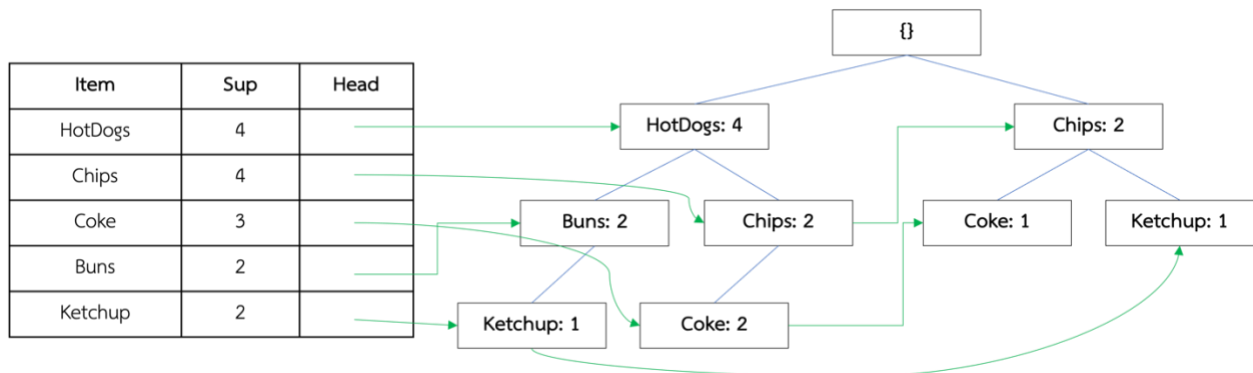
F-list = HotDogs, Chips, Cokes, Buns, Ketchup

Itemset	Sup-count
HotDogs	4
Chips	4
Coke	3
Buns	2
Ketchup	2

Construct FP-tree

Transaction ID	Items	(ordered) frequent items
T1	HotDogs, Buns, Ketchup	HotDogs, Buns, Ketchup
T2	HotDogs, Buns	HotDogs, Buns
T3	HotDogs, Coke, Chips	HotDogs, Chips, Coke
T4	Chips, Coke	Chips, Coke
T5	Chips, Ketchup	Chips, Ketchup
T6	HotDogs, Coke, Chips	HotDogs, Chips, Coke

Header Table



Conditional pattern bases

Item	Conditional pattern bases
HotDogs	Empty
Chips	{{HotDogs: 2}}
Coke	{{HotDogs, Chips: 2}, (Chips: 1)}
Buns	{{HotDogs: 2}}
Ketchup	{{HotDogs, Buns: 1}, (Chips: 1)}

Conditional FP-tree

Item	Conditional pattern bases	Conditional FP-tree
HotDogs	Empty	Empty
Chips	{{HotDogs: 2}}	{{HotDogs: 2}} Chips
Coke	{{HotDogs, Chips: 2}, (Chips: 1)}	{{Chips: 3}} Coke
Buns	{{HotDogs: 2}}	{{HotDogs: 2}} Buns
Ketchup	{{HotDogs, Buns: 1}, (Chips: 1)}	Empty

Frequent Pattern rules

Item	Frequent Pattern Generated
HotDogs	Empty
Chips	{{HotDogs, Chips: 2}}
Coke	{{Chips, Coke: 3}}
Buns	{{HotDogs, Buns: 2}}
Ketchup	Empty