

Data Mining:

Concepts and Techniques

(3rd ed.)

— Tutorial 2 —

Slides Courtesy of Textbook

Tutorial 2: Assignment 2

- Important Issues
- Learning Resources
- Sample Questions
- Q & A

Important Issues

- Assignment 2 covers Chapter 4 and Chapter 5
- Deadline : Thursday (10/08) 11:59 PM
- No Late Submission !!!

Learning Resources

- Attend Lectures !!!!
- Lecture Video !!!
- Textbook + Slides !!
- Piazza & Office Hour !

About Assignment 2

- 5 Written Questions
- 1 Mini MP

Sample Question 1

- Cuboid ?
- Cell ?
- Measure: ?

Sample Question 1

Sample Question 1

Example:

Year vs. Dept.

	CS	ECE	EDU	LAW
Senior	30	40	20	25
Junior	20	50	25	30

Cuboid:

(Year, Dept.)

Cell: measure

(Senior, CS): 30, (Junior, CS): 20

Aggregated:

(* , CS): 50

Sample Question 2

- A: (a1, a2, a3, a4) : 1 and B: (b1, b2, b3, b4) : 1
 - Count of C: (*, a2, a3, a4) ?
1
 - Count of D: (b1, b2, b3, *) ?
1
 - Count of E: (*, *, *, *) ?
2

Sample Question 2

- A: (a1, a2, a3, a4) : 10 and B: (b1, b2, a3, a4) : 5
 - Count of C: (*, a2, a3, a4) ?
10
 - Count of D: (b1, *, a3, *) ?
5
 - Count of E: (*, *, a3, a4) ?
15

Sample Question 3

- We have 2 Base Cells:

- A: (a1, a2, a3, a4) : 10

- B: (b1, b2, a3, a4) : 10

List all the closed cells.

- Closed Cell: A cell c is a closed cell if there exists no cell, d, such that d is a **descendant** of cell c and d has the **same measure** value as c.

Sample Question 3

- We have 2 Base Cells:

- A: (a1, a2, a3, a4) : 10

- B: (b1, b2, a3, a4) : 10

List all the closed cells.

1) Is C: (*,*,*,*): 20 a closed cell? Why? NO!

2) Is D: (*,*,a3,a4): 20 a closed cell? Why? YES!

3) Is E: (a1,*,a3,a4): 10 a closed cell? Why? NO!

4) Closed cells:

(a1,a2,a3,a4), (b1,b2,a3,a4) and (*,*, a3,a4)

Sample Question 4

- We have 2 Base Cells:
 - A: (a1, a2, a3, a4,a5) : 1
 - B: (b1, b2, a3, a4,a5) : 1

How many distinct aggregated (non-base) cells?

Sample Question 4

- We have 2 Base Cells:
 - A: (a1, a2, a3, a4,a5) : 1
 - B: (b1, b2, a3, a4,a5) : 1

How many distinct aggregated (non-base) cells?

Before we start working on this, I have a question:

Is C: (*,a2,a3,a4,a5): 1 an aggregated cell? Yes!

Sample Question 4

Sample Question 4

- We have 2 Base Cells:
 - A: (a1, a2, a3, a4,a5) : 1
 - B: (b1, b2, a3, a4,a5) : 1

How many distinct aggregated (non-base) cells?

Divide into 2 parts based on first 2 dimension:

- 1) First 2 dimensions are both stars
- 2) In first 2 dimensions, there is at least 1 non-star dimension

Sample Question 4

- First 2 dimensions are both stars

A -> (*,*, a3,a4,a5)

B -> (*,*, a3,a4,a5)

Aggregated cell for A and B are same,
we have $2^3 = 8$ cells

Sample Question 4

- In first 2 dimensions, there is at least 1 non-star dimension

For cell A, it could be $(a_1, *, a_3, a_4, a_5) \dots$

For the first 2 dimensions of cell A: we have $2^2 - 1$

For the last 3 dimensions of cell A: we have 2^3

For cell A, we have $(2^2 - 1) * 2^3 = 24$

In this case, the aggregated cell for A and B are different, so totally we have $2 * 24 = 48$

Sample Question 4

- We have 2 Base Cells:
 - A: (a1, a2, a3, a4,a5) : 1
 - B: (b1, b2, a3, a4,a5) : 1

How many distinct aggregated (non-base) cells?

$$8 + 48 - 2 = 54$$

Sample Question 5

- We have the array data:

(a0,b0,c0):1	(a0,b0,c1):1	(a0,b0,c2):1
(a0,b1,c0):1	(a0,b1,c1):1	(a0,b1,c2):1

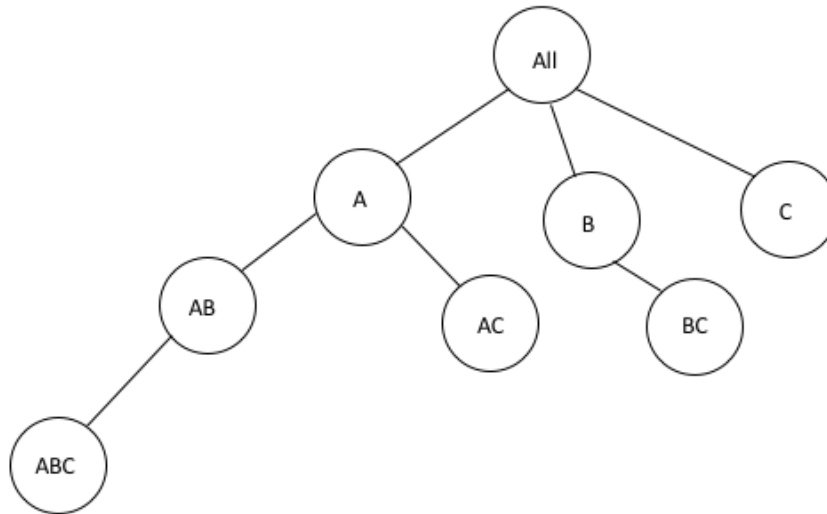
- Suppose we construct the iceberg cube using BUC for dimension A, B, C with the order of A->B->C. Mini_Support = 3
 - 1) Draw the trace tree
 - 2) List cells need to be computed

Sample Question 5

- The order is $A \rightarrow B \rightarrow C$. Draw the trace tree.

Sample Question 5

- The order is $A \rightarrow B \rightarrow C$. Draw the trace tree.



Sample Question 5

List cells being computed:

All: $(*,*,*) : 6$ – expansion

A: $(a0,*,*) : 6$ – expansion

AB: $(a0,b0,*) : 3$ – expansion

$(a0,b1,*) : 3$ – expansion

ABC: $(a0,b0,c0) : 1$ $(a0,b1,c0) : 1$

$(a0,b0,c1) : 1$ $(a0,b1,c1) : 1$

$(a0,b0,c2) : 1$ $(a0,b1,c2) : 1$

AC: $(a0,*,c0) : 2$ $(a0,*,c1) : 2$ $(a0,*,c2) : 2$

Sample Question 5

List cells being computed:

B: $(*,b0,*) : 3$ – expansion

$(*,b1,*) : 3$ – expansion

BC: $(*,b0,c0) : 1$ $(*,b1,c0) : 1$

$(*,b0,c1) : 1$ $(*,b1,c1) : 1$

$(*,b0,c2) : 1$ $(*,b1,c2) : 1$

C: $(*,*,c0) : 2$ $(*,*,c1) : 2$ $(*,*,c2) : 2$

Totally, there are 24 cells need to be computed.

Sample Question 5

What if we set the `mini_support` = 4? Which cells would be removed from the list?

Sample Question 5

List cells being computed:

All: $(*,*,*) : 6$ – expansion

A: $(a0,*,*) : 6$ – expansion

AB: $(a0,b0,*) : 3$

$(a0,b1,*) : 3$

~~ABC: $(a0,b0,c0) : 1$ $(a0,b1,c0) : 1$~~

~~$(a0,b0,c1) : 1$ $(a0,b1,c1) : 1$~~

~~$(a0,b0,c2) : 1$ $(a0,b1,c2) : 1$~~

AC: $(a0,*,c0) : 2$ $(a0,*,c1) : 2$ $(a0,*,c2) : 2$

Sample Question 5

List cells being computed:

B: (*,b0,*) : 3

(*,b1,*) : 3

~~BC: (*,b0,c0) : 1 (*,b1,c0) : 1~~

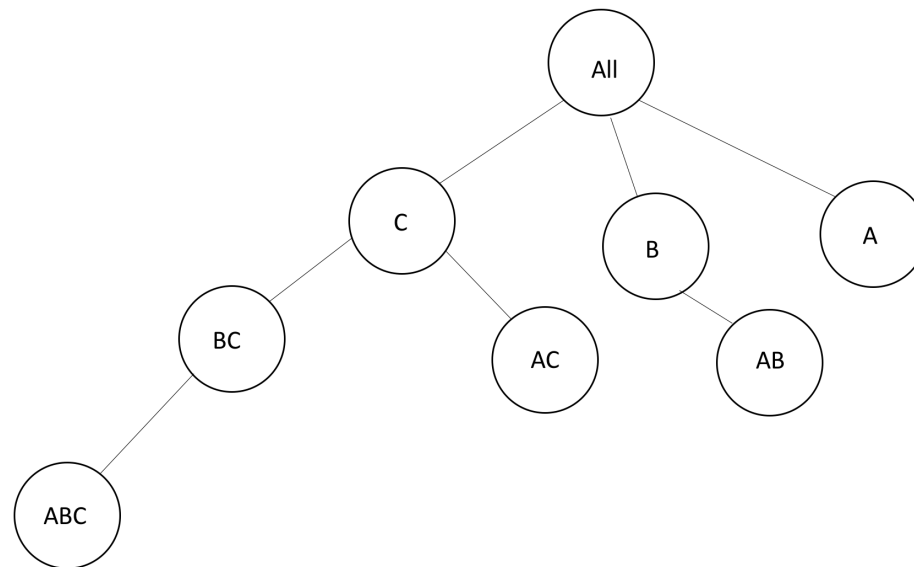
~~(*,b0,c1) : 1 (*,b1,c1) : 1~~

~~(*,b0,c2) : 1 (*,b1,c2) : 1~~

C: (*,*,c0): 2 (*,*,c1) : 2 (*,*,c2) : 2

Sample Question 5

- If we traverse in the order C->B->A. Draw the trace tree.



Q & A

- Thank you