Review Test Submission: Assignment 2

User	Peng Li
Course	Fall 2016-CS 412-Introduction to Data Mining-Sections A3, A4, NPV, P3, P4
Test	Assignment 2
Started	9/22/16 7:48 PM
Submitted	10/4/16 10:08 PM
Due Date	10/4/16 11:59 PM
Status	Completed
Attempt Score	0 out of 100 points
Time Elapsed	290 hours, 20 minutes
Results Displayed	Submitted Answers

Question 1

(24') Assume a base cuboid of 10 dimensions contains only two base cells:

- (a_1, a_2, a_3, a_4, b_5, ..., b_9, b_10), and
- (b_1, b_2, b_3, b_4, b_5, ..., b_9, b_10),

where a_i is not equal to b_i (for any i = 1, ..., 10). The measure of the cube is count. Answer the following questions.

a. How many nonempty aggregated (i.e., non-base) cells does a complete cube contain? **[a]**

b. How many nonempty aggregated cells does an iceberg cube contain, if the condition of the iceberg cube is count ≥ 2? **[b]**

c. How many non-star dimensions does the closed cell with count 2 have? **[c]**

d. How many closed cells are there in the full cube? **[d]**

Specified Answer for: a 1982

Specified Answer for: b 64

Specified Answer for: c 6

Specified Answer for: d 3

Question 2

(36') Given the dataset, it contains 50 rows, with each row representing a business. For each business, there are six fields (BusinessID, City, State, Category, Rating, Price). The fields are separated by tabs. We now want to construct a cube over four given dimensions (Location, Category, Rating, Price) with "Count"

as the measure. Note that in the Location dimension, there is a concept hierarchy, i.e., City and State. Based on the dataset, answer the following questions.

a. How many cuboids are there in this cube? **[a]**

b. How many nonempty cells are there in the cuboid (Location(City), Category, Rating, Price)? **[b]**

c. Drill up by climbing up in the Location dimension from City to State. How many nonempty cells are there in the cuboid (Location(State), Category, Rating, Price)? **[c]**

d. How many nonempty cells are there in the cuboid (*, Category, Rating, Price)? **[d]**

e. What is the count for the cell (Location(State) = "Illinois", *, Rating = "3", Price = "moderate")? **[e]**

f. What is the count for the cell (Location(City) = "Chicago", Category= "food", *, *)? **[f]**

Specified Answer for: a 24

Specified Answer for: b 46

Specified Answer for: c 36

Specified Answer for: d 22

Specified Answer for: e 5

Specified Answer for: f 3

Question 3

(40') Given dataset which contains 100 transactions, each line is a transaction and each transaction contains item(s) separated by spaces. Please do frequent pattern mining on this dataset and answer the following questions. If the answer is not an integer, please round the result to 3 decimal places.

(15') 1. Suppose the minimum support is 20. Then:

a. Count the number of frequent patterns. **[1a]**
b. Count the number of frequent patterns with length 3. **[1b]**
c. Count the number of max patterns. **[1c]**

(25') 2. Suppose we decrease the minimum support to 10.

a. Count the number of frequent patterns. **[2a]**
b. Count the number of frequent patterns with length 3. **[2b]**
c. Count the number of max patterns. **[2c]**
d. What is the confidence measure of the association rule: (C, E) → A? **[2d]**
e. What is the confidence measure of the association rule: (A, B, C) → E? **[2e]**

Specified Answer for: 1a 30

Specified Answer for: 1b 8

Specified Answer for: 1c 7

Specified Answer for: 2a 55

Specified Answer for: 2b 20

Specified Answer for: 2c 6

Specified Answer for: 2d 0.679

Specified Answer for: 2e 0.742

Tuesday, October 4, 2016 10:08:56 PM CDT

← OK