Week 2: Frequent Itemsets (Basic)

Help Center

Warning: The hard deadline has passed. You can attempt it, but **you will not get credit for it**. You are welcome to try it as a learning exercise.

☐ **In accordance with the Coursera Honor Code, I (Manuel Bordés Rguez.) certify that the answers here are my own work.**

Question 1

Suppose we have transactions that satisfy the following assumptions:

- *s*, the support threshold, is 10,000.
- There are one million items, which are represented by the integers 0,1,...,999999.
- There are *N* frequent items, that is, items that occur 10,000 times or more.
- There are one million pairs that occur 10,000 times or more.
- There are 2*M* pairs that occur exactly once. *M* of these pairs consist of two frequent items, the other *M* each have at least one nonfrequent item.
- No other pairs occur at all.
- Integers are always represented by 4 bytes.

Suppose we run the a-priori algorithm to find frequent pairs and can choose on the second pass between the triangular-matrix method for counting candidate pairs (a triangular array count[i][j] that holds an integer count for each pair of items (i, j) where i < j) and a hash table of item-item-count triples. Neglect in the first case the space needed to translate between original item numbers and numbers for the frequent items, and in the second case neglect the space needed for the hash table. Assume that item numbers and counts are always 4-byte integers.

As a function of N and M, what is the minimum number of bytes of main memory needed to execute the a-priori algorithm on this data? Demonstrate that you have the correct formula by selecting, from the choices below, the triple consisting of values for N, M, and the (approximate, i.e., to within 10%) minumum number of bytes of main memory, S, needed for the a-priori algorithm to execute with this data.

◯ N = 30,000; M = 200,000,000; S = 1,800,000,000

- ◯ N = 50,000; M = 80,000,000; S = 1,500,000,000
- ◯ N = 20,000; M = 80,000,000; S = 1,100,000,000
- ◯ N = 30,000; M = 100,000,000; S = 500,000,000

Question 2

Imagine there are 100 baskets, numbered 1,2,...,100, and 100 items, similarly numbered. Item i is in basket j if and only if i divides j evenly. For example, basket 24 is the set of items {1,2,3,4,6,8,12,24}. Describe all the association rules that have 100% confidence. Which of the following rules has 100% confidence?

- ◯ {8,12} → 96
- ◯ {12,18} → 36
- ◯ {2,3,5} → 45
- ◯ {2,4} → 8

Question 3

Suppose ABC is a frequent itemset and BCDE is NOT a frequent itemset. Given this information, we can be sure that certain other itemsets are frequent and sure that certain itemsets are NOT frequent. Other itemsets may be either frequent or not. Which of the following is a correct classification of an itemset?

- ◯ ABCDE can be either frequent or not frequent.
- ◯ ABCDEF can be either frequent or not frequent.
- ◯ BCE is frequent.
- ◯ BCD can be either frequent or not frequent.

☐ **In accordance with the Coursera Honor Code, I (Manuel Bordés Rguez.) certify that the answers here are my own work.**

Submit Answers Save Answers

You cannot submit your work until you agree to the Honor Code. Thanks!