#36: Counting and Cardinality

April 27, 2022 · *Brad Solomon*

Combining bloom filters

Given the following bloom filters, write a bloom filter which contains all the items present in both filters.

[0]	0	[0]	0	[0]	
[1]	1	[1]	1	[1]	
[2]	1	[2]	0	[2]	
[3]	0	[3]	0	[3]	
[4]	1	[4]	1	[4]	
[5]	0	[5]	1	[5]	
[6]	0	[6]	0	[6]	

Sequence Bloom Trees

Given the bit vectors (1010), (0010), (0001), and (0101), draw a sequence bloom tree that stores all vectors as leaves. Consider how the arrangement of leaves can affect the usefulness of the tree!

Counting Bloom Filter: Insertion

Construct a counting BF. **S = { 16, 8, 4, 13, 29, 11, 22 },**
h1(k) = k % 7, h2(k) = 2k+1 % 7

[0]	
[1]	
[2]	
[3]	
[4]	
[5]	
[6]	

Counting Bloom Filter: Deletion

Which of the following items **cannot** be deleted at least once?

S = { 0, 1, 2, 3, 4}, h(k) = k % 5,

[0]	5
[1]	2
[2]	1
[3]	3
[4]	7

What is the downside to allowing deletion?

CBF: Minimal Increase

If X hashes to indices 0, 2, and 3 what is our best estimate of X's current count? How can we adjust insertion to take that into account?

[0]	**3**
[1]	**4**
[2]	**2**
[3]	**1**
[4]	**4**
[5]	**6**
[6]	**8**

CBF: Recurring Minimum

In the above example, X only has a single minimum value. What does this mean and what can we do to improve our accuracy for counting the frequency of X?

Cardinality

Cardinality is a measure of:

Cardinality Estimation

If I randomly sampled values from 0 – 1000 (no repeats) and told you that the minimum value was 300, what is your best estimate for the cardinality in the random set?

What if the minimum value was 20?

K-minimum Estimation

Will the k-th minimum give me a better, worse, or the same estimation accuracy as the minimum? Why?

CS 225 – Things To Be Doing:

1. Continue working on mp_schedule
2. Either work on your final project or prepare for final exam