Bloom过滤器的false positive概率计算

Bloom过滤器使用bit数组作为基础数据结构，bit位数为m。假设hash函数选择数组中每个bit位的概率相同，那么其中某一个bit位在添加数据时不被某个hash函数设置为1的概率是：

(1)

如果选择个hash函数，那么完成一次添加数据，某个bit位仍旧没有被设置为1的概率是：

(2)

根据

(3)

则有

(4)

添加n个数据后，某个bit位仍旧为0的概率是

(5)

其设置为1的概率为

(6)

现在加入一个新的数据，那么其对应的k个bit均设置为1的概率为，

(6)

令

(7)

则(6)式转换为

(8)

求解就是求解的最小值，对u求解导数得

(9)

再令

(10)

(11)

继续化简

(12)

=0 (13)

当时（看出来的）等式成立，得到

即

，继而得到

(13)

[1] https://andybui01.github.io/bloom-filter/#implementation-and-benchmarks