**HW2 BLOOM FILTER**

**一、实验概述**

本人编写了bloom\_filter类进行测试，其中选择了输入集为0—999999，测试集为1000000—2999999。数据集取得非常大是为了尽可能获得趋于理论值的报错率。哈希函数采用了助教给出的Murmurhash类，生成k个不同哈希函数的方案是把哈希函数内置的seed参数设置为1—k。

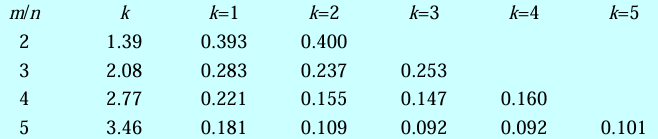
**二、实验结果**

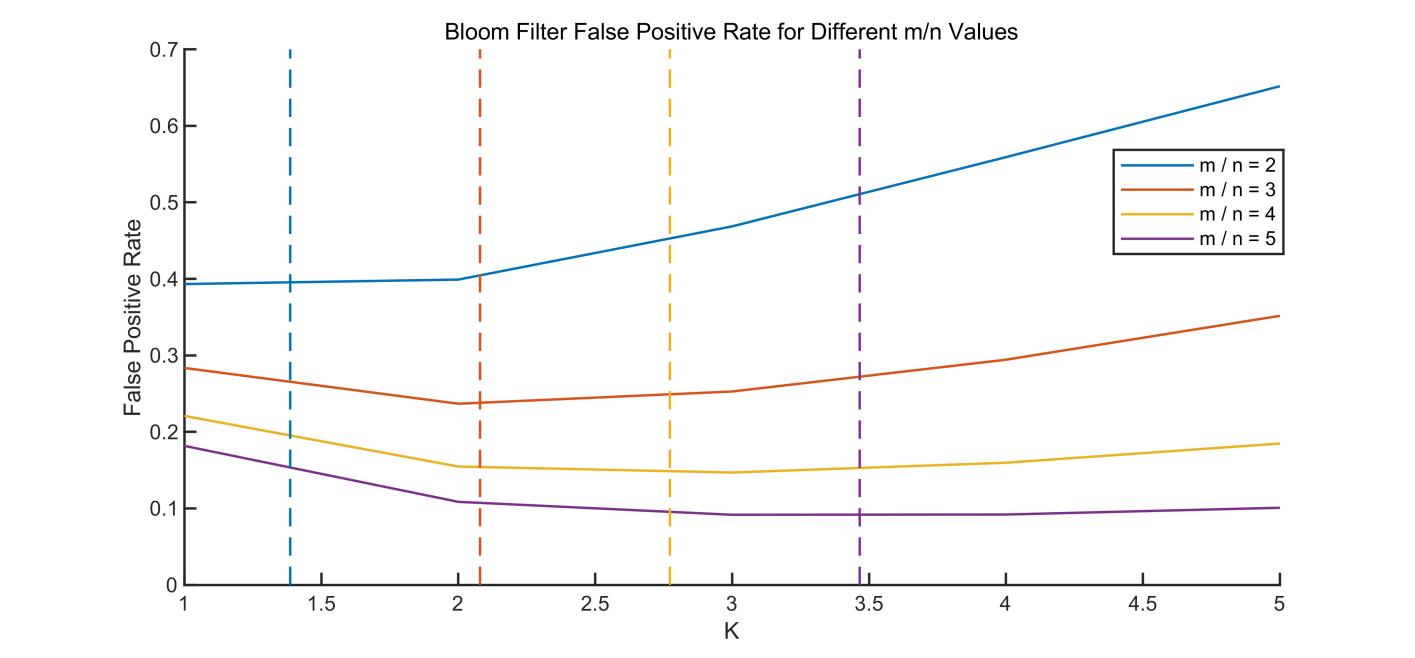
下表与图是实验得到结果。其中图中虚线标注了由理论值公式

进行计算得到的k值。

|  |  |  |
| --- | --- | --- |
| m/n | k | 报错率 |
| 2 | 1 | 0.393233 |
| 2 | 2 | 0.399058 |
| 2 | 3 | 0.46862 |
| 2 | 4 | 0.559119 |
| 2 | 5 | 0.651818 |
| 3 | 1 | 0.283552 |
| 3 | 2 | 0.236801 |
| 3 | 3 | 0.252763 |
| 3 | 4 | 0.29435 |
| 3 | 5 | 0.351687 |
| 4 | 1 | 0.220889 |
| 4 | 2 | 0.154701 |
| 4 | 3 | 0.146875 |
| 4 | 4 | 0.159671 |
| 4 | 5 | 0.184731 |
| 5 | 1 | 0.181719 |
| 5 | 2 | 0.108541 |
| 5 | 3 | 0.09165 |
| 5 | 4 | 0.091965 |
| 5 | 5 | 0.100801 |

**表1. 测试结果表**

**图1.理论k值**

**图2. 测试结果曲线图**

**三、结果分析**

可以发现表格中的数据与课件给出的理论值几乎一致，且图中的虚线表示的理论k值也与实验得出的最优k值符合。