

班级快速排序实验总结

庄逸

基本数据总览及分析

快速排序

冒泡排序

快速排序

数据数	比较次数	实验用时 (分钟)
8	13	1.37
9	18	2.5
12		
15	40	6.75
15	36	6.05
15	42	5.2
16	46	5.4
16	44	5.45
16	38	3.5
17	55	16
17	69	11
17	45	5
18	110	15
20	61	9.67
20	63	10.75
20		3.5

设数据数为 x

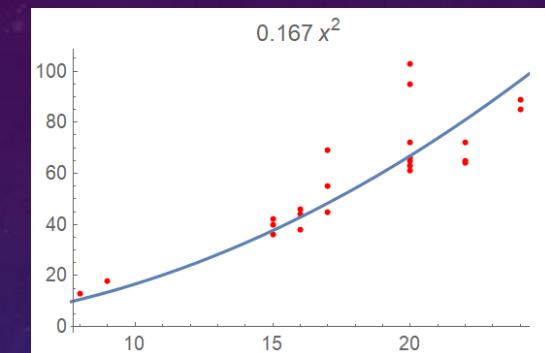
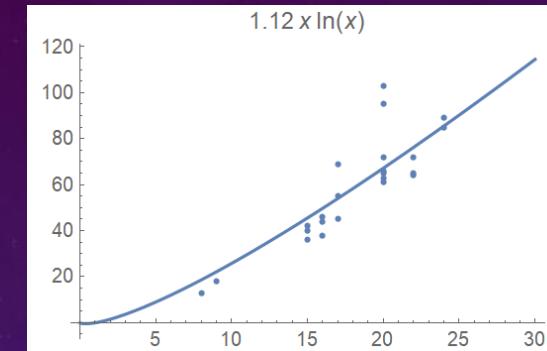
比较次数为 $f(x)$

去除明显的误差点

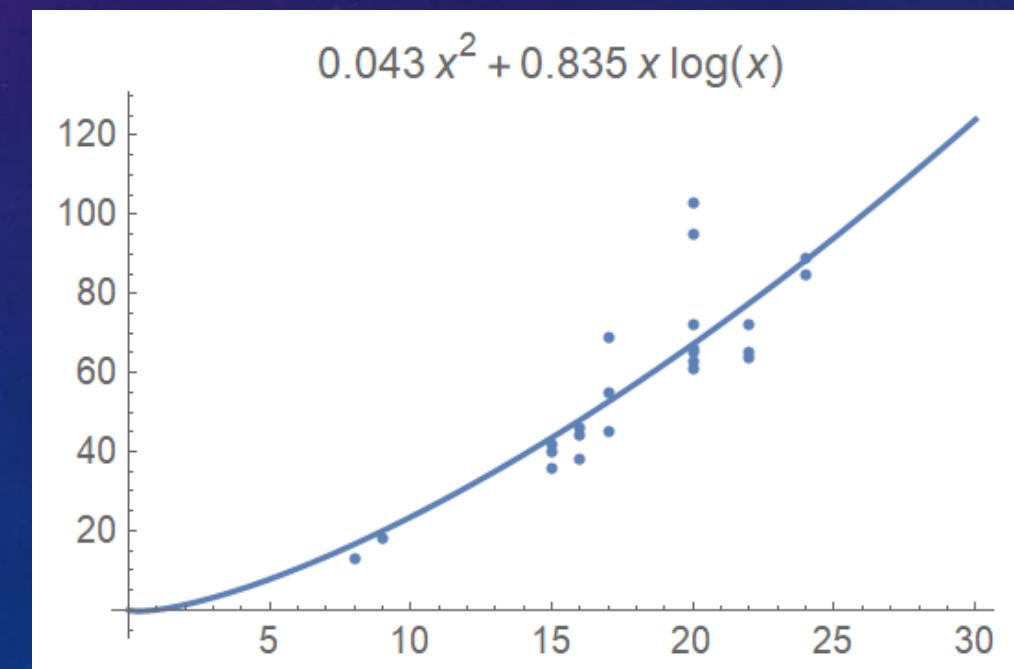
分别用 $x \ln x$ 和 x^2 和 $ax^2 + b \ln x$ 拟合，曲线如右图

时间影响因素较多，不作单变量拟合

结论： $x \ln x$ 成分相对偏大



20	66	5.6
20	65	8.1
20	72	9.2
20	95	
20	103	
22	65	14.7
22	72	9.7
22	64	8.7
23		5.1
23		
24	85	11.5
24	89	10.16
26	28	16.5
26	31	12.5
27		



冒泡排序

数据数	比较次数	理论次数	实验用时 (分钟)
15	108	105	8.4
16	113	120	5.88
17	145	136	3
18		153	8
20	55	190	8.25
20	190	190	8
21	210	210	5.5
25	190	300	12.7
26	109	325	10.67

特点：

比较次数明显比快速排序多
平均值140;60

所用时间略微比快速排序少
平均值7.8;8.4

为什么呢？

专题分析

- 误差及解决
- 数据记录
- 随机选取随机吗？
- 提高速度、稳定性
- 人体计算机VS电子计算机
- 冒泡排序的算法
- 创新
- 值得注意的现象

误差及解决

- 1702B: 草坪高低影响
- 1703A: 主要是缘于场地草坪的凹凸不平
- 1704A: 因为地势，站姿等因素，对身高的判断不是很精确；因为天气炎热，排序速度受到了一定影响。
- 1705: 在判断谁高上有一定误差
- 1709: 很多同学的身高处在同一水平线上，使得比较极为困难；地形凹凸不平，使得比较上会产生误差。可能因为“处理器”的视角问题和所站的位置问题使得同样两个“数据”进行比对后的结果不一样。
- 1710: 由于角度原因或是地面不平整，控制器无法正确判断同学的身高或判断失误。

- 1701: 视为身高相同处理
- 1702B: 多人共同裁定，按第一次比较为准
- 1703A: 让每名同学记住第一次比较时的结果
- 1704A: 两个人身高差异在一厘米左右即认为身高相同。
- 1710: 另选一名同学来定夺，或者让两名同学报出自己的具体身高，直接比较数据大小。
- 老师: 去年有同学搭了一个小平台。

数据记录

- 一个问题
- 1701: 有时控制器在发现两名同学身高相同时，便没有进行移动直接进入了下一组。其实是需要记录的。
- 一个技巧
- 1703A: 只需记录递归运算时每一组的人数。由于基准数据需要且只需与每个组内数据比较一次，比较数据即为该组人数减一。

随机选取随机吗？

- 1701: 控制器直接选取了中间位置的同学作为“随机”的快速排序基准点。这样做违背了随机选取的计算机原则。
- 1702B: 记录了第一轮选择的基准在排序完成后队列中的位置，预计会与所花步数有相关性，但数据并无反映。
- 1702B: 试图在每轮比较中依据先前经验选取身高最中央的一个，期望能加快速度。但是由于身高误差等因素，数据并无反映。
- 1709: 三次实验花费时间受随机选取的“数据”影响。
- 1709: 实际实验中，“处理器”会根据自己的下意识判断取中间的“数据”，所以在多次实验的时候会造成步数接近一致且较少。

提高速度、稳定性

- 数字化
- 1702A: 第一次完成快排后对每位同学身高进行了编码，使得在后续排列中对于身高相近的同学，只需各自报数进行比较。
- 预实验，多次实验
- 1702A: 数据能够“学习”
- 1709: 三次实验花费时间变短，也因为大家对这个实验越来越熟悉。
- 存储
- 17078: 不足在于我们没有存储模块，每次操作都要读取数据（询问同学的身高），这让场面一度十分尴尬。
- 1709: 大家对彼此的身高没有直观的印象，比对花去了很长时间。

人体计算机VS电子计算机—优点

- 1703B：教给同学们快速排序的执行方法之后，同学们便开始自动地进行排序。
- 1703B：前两次实验实现了多线程排序，各个数组之间的排序可以同时进行而且完全不会相互干扰。
- 1703B：电子计算机强大的原因是中央处理器运行频率极高，在单位时间内可以处理的指令数相当可观。但具体到其指挥的个体，其实非常简单，功能也很弱。而人体不一样，单个的人类个体拥有非常高的智能。人类个体完全可以发挥自己的智能从而实现数据之间自主的沟通与比较操作，这样将使排序速度大大快于照搬电子计算机的模式。
- 1702A：大大节省了比较和移动的时间，使得快排实验进行速度越来越快。
- 1704A：因为不符合人的正常思维，“快排”变成了“慢排”；相反地，冒泡排序法更符合人的思维，也更直观。
- 1706A：计算速度不同，计算机可以进行数据的备份，人体计算机排序对数据的操作相对来说更加随意。
- 1709：三次实验花费时间变短，也因为大家对这个实验越来越熟悉。

人体计算机VS电子计算机—缺点

- 1704B：人会听不懂命令。
- 1705：扮演识别符的同学有时会逻辑混乱。
- 1709：大家最开始还没有进入状态，没有产生数据的代入感。
- 1709：一方面“数据”的运动可能会导致排序的错误，另一方面“数据”运动不按次序，容易造成步数数错。

分布式并行人体计算机？

- LFZ: "分布式并行算法：你们自己找一下合适的位置，然后自己调整一下"
- 可能是最快的排序方式，但在计算机上难以实现。
- 难以清楚地解释排序原理
- 使用最普遍

冒泡排序的总体特点：次数多，速度快

- **1702A:** 虽然冒泡算法在步骤复杂度上远高于快速排序，但是每一步判定简单。
- **1703A:** 在计算机中，用时多少主要取决于运算次数的多少，也即比较次数的大小，而在人体计算机的实验中，主要耗时项目在于调用数据。因此数据组几乎不需进行移动的冒泡排序实验用时最短。
- **1704A:** 次数大大增加，但时间减少很多
- **1704A:** 冒泡排序法更符合人的思维，也更直观。
- **1704B:** 冒泡排序会做很多无用功，我们到一半已经知道了结果，但还是要重复排序步骤。

- **17078:** 相比于快速排序进程加快，因为我们无须读取数据就可以直接比较数据（身高），因为是对相邻的两名同学做比较。
- **1710:** 冒泡过程中，操作失误，可能忘记了冒泡过的同学。
- 解决方法：我们在冒泡过程中，让已经冒泡好的人后退一步，每次冒泡结束后，只能有一人后退一步。
- **1710:** 在输入值，输出值相同的情况下，冒泡排序的比较次数、交换次数、用时都多于快速排序。

创新

- **1703A:** 新算法节约时间的部分主要来源于“调用数据”的部分，也即让同学们走较短的路程。新算法是直接按次序比较数据并将它们分为两组。但这在计算机中是不可实现的。但该同学将全体同学排成一圈的创举无疑是值得鼓励的。
- 部分实验采用不同的输入输出
- **1701:** 输出年龄
- **1702B:** 输入身高，输出学号
- **1706B, 1710等:** 输入随机
- **1703A:** 在第二次的实验中，加入了“数据读取器”设定，直接得到身高数据而非相对大小。数据读取器向控制器汇报第一次排序的次序作为身高数据，而控制器只进行身高数据的判断，以模拟计算机的真实情境。
- 但是这项改进并不算很成功。失败部分源于小组同学没有肉眼读取身高的能力，同时也由于同学们不能很快地反馈上一次比较中的实验结果充当数据。

创新

- 1706B：堆排序
- 1、建堆：建立一个完全二叉树。每向堆中添加一个“数据”时要满足：堆的每一个父节点身高都小于其子节点；堆的每个左子树和右子树也是一个堆；
- 2、移出：移出堆顶的“数据”，排为第一位“数据”；
- 3、调整：将一叶节点的“数据”移至堆顶，根据建堆的原则向下调整“数据”；
- 4、重复：将新堆堆顶“数据”移出排为第二位“数据”，再将一叶节点数据移至堆顶，向下调整，接着不断重复移出和重复步骤；
- 5、完成：堆中所有“数据”被移出后，将得到按身高升序排列的一列“数据”。

值得注意的现象

- 1704A: 在计算机操作时, 次数与时间总是成正比, 而对人来说, 显然不是这样。三次实验中反而出现了次数与时间成反比的现象。
- 原因: 只在一定范围内; 学习; 操作简单
- 1702B&1704B: 冒泡排序中有时会出现后面很多人都已经排序完毕, 但是仍然要比到上一轮最后一个人之前, 做了不少无用功。
- 可以借此改进冒泡排序算法。
- 1701: 有时控制器在发现两名同学身高相同时, 便没有进行移动直接进入了下一组。其实是需要记录的。
- 事实上, 这一部分操作员可能没有特意比较, 记录员也可能没有数比较的次数, 这是做了冒泡排序实验的班级比较次数比理论值低的主要原因。

谢谢！