

数据科学基础

大作业说明

陈振宇 南京大学智能软件工程实验室 www.iselab.cn

Python基础编程特征数据分析

- 提交一份开放性研究报告, 页数不限, 格式不限
- 采用一系列数据分析方法,回答 "Python基础编程特征" 的相关系列问题
- 所有分析代码开源,附上开源地址以方便助教核查
- 截止提交日期: 2020年6月6日, 提交方式另行通知
- 可由1-3人组队完成,但Python练习完成题目100道 以上的同学需要单独完成
- 占课程考核总成绩的30%

学生完成情况统计表

完成题量/题	分数/分	人数/人	比例
≥ 199	100	114	44%
179 – 198	90 – 99	54	21%
159 – 178	80 – 89	26	10%
139 – 158	70 – 79	11	4%
119 – 138	60 – 69	9	3%
≤ 118	0 - 59	48	18%

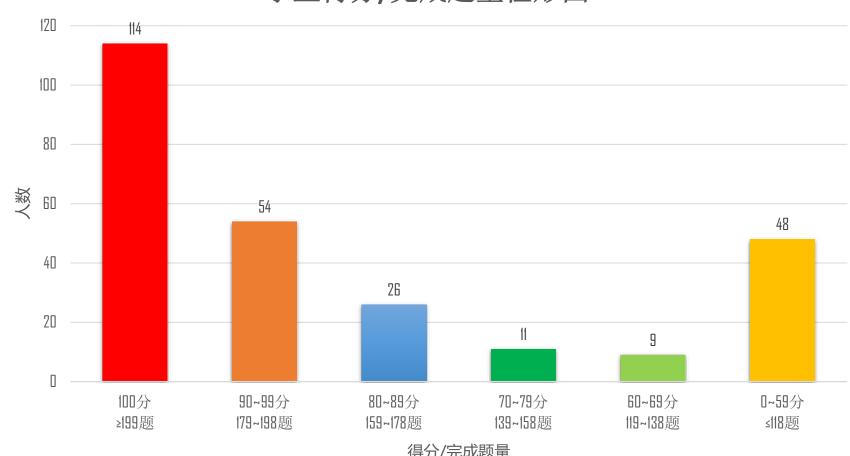
注: 1. 分数按照"学生通过题目量/总题量"计算2. 每个人的总题量: 200题,总人数: 262。

学生得分特征统计表

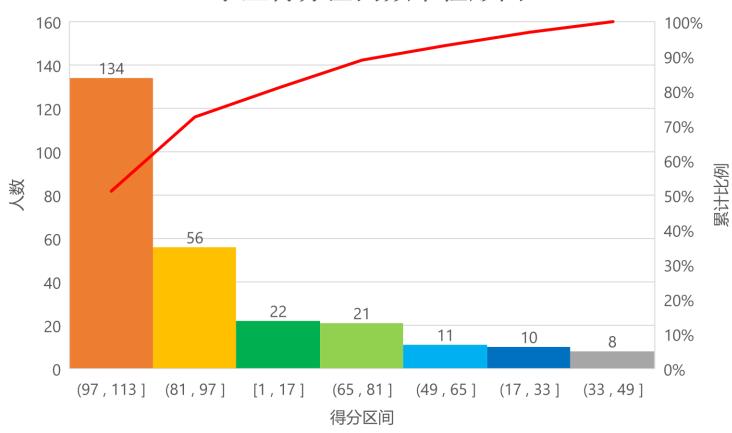
学生得分特征	分数/分	完成题量/题	人数/人
平均数	81	162	2
众数	100	200	114
中位数	98	196	7
标准差	30	无	无

注: 1. 分数按照"学生通过题目量/总题量"计算2. 每个人的总题量: 200题,总人数: 262。

学生得分/完成题量柱形图

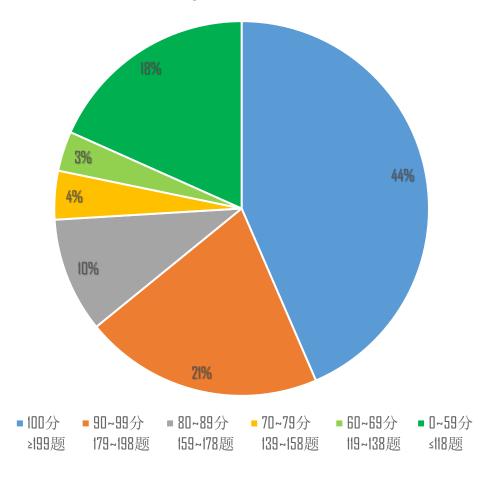


学生得分区间频率柱形图



横坐标是以16分为跨度的区间, 左侧纵坐标为人数, 右侧纵坐标为人数累计所占比例。

学生得分/完成题量饼状图



得到满分的人数比例最高,达到了44%; 60~69分的人数比例最低,只有3%; 得到90分以上的人数比例有65%。

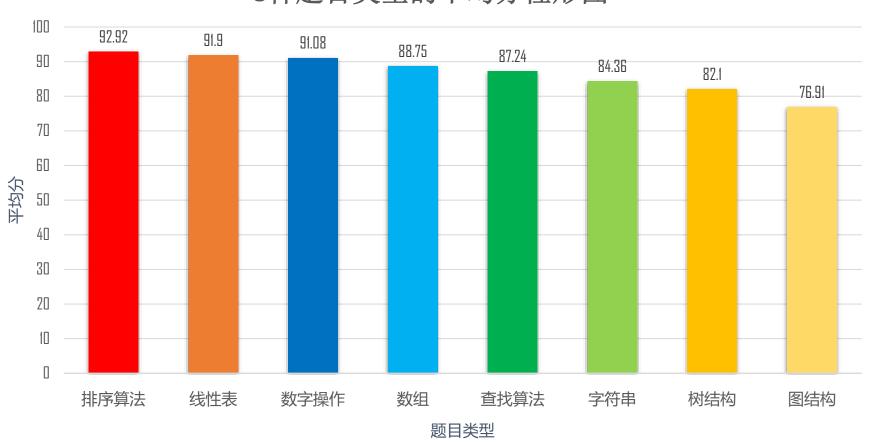
8类题目的平均分

题目类型	平均分
排序算法	92.92
查找算法	87.24
图结构	76.91
树结构	82.10
数字操作	91.08
字符串	84.36
线性表	91.90
数组	88.75

注: 这里的得分不是"学生通过题目量/总题量" 而是按照慕测上的赋分标准。

表中8种算法类别的平均分一定程度上 可以反映该类题型的难度。

8种题目类型的平均分柱形图



通过平均分这个维度推断, 题目难度由易到难的顺序: 排序算法、线性表、数字 操作、数组、查找算法、 字符串、树结构、图结构。

• 69%的同学未通过全部题目,未通过全部题目的原因?

1. 每题都做,但是未通过全部用例(题目较难): 45%

2. 未完成全部题目(有未提交题目): 55%

· 为什么未完成全部题目(有未提交题目)?

1. 个别题目难度大(只有个别题目未提交): 32%

2. 来不及(集中在截止日期前提交,且有大批未提交题目): **27%**

3. 不想做了(未集中在截止日期前提交,且有大批未提交题目):41%

・ 为什么不想做了(未集中在截止日期前提交)?

1. 题目太简单了(已做部分全通过): 22%

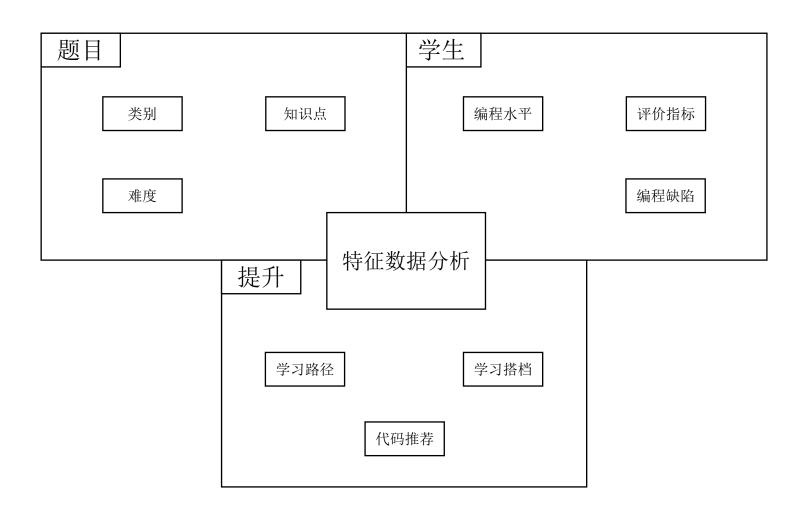
2. 题目太难了(已做部分未全通过): 78%

注:以上数据由抽样调查得到

- 根据上述未通过统计数据,可知同学未通过全部题目的原因分布如下:
- 1. 题目难 80%
- 2. 来不及 15%
- 3. 题目简单 5%

注: 以上数据由抽样调查得到

关于题目的研究问题



关于题目的研究问题示例

- 如何自动对题目进行分类?
 直接匹配关键字是一种思路,比如"排序",可以归类为算法类。
 能否逆向从实现代码的逻辑进行分类?
- 如何自动分析题目的知识点?直接分析题目的文字描述是一种思路,但是...假如我们提供很多编程答案给你作为数据分析基础呢?
- 如何自动定义题目的难度?
 例如,我们直观理解越多人错的题目越难,直接计算平均分作为难度的反向评分。但这样容易有****问题。

关于学生的研究问题示例

- 如何选择学生的编程能力评价指标?直接按答对题目的数量计算就好了,但是很多都全对的...假如我们提供很多编程答案给你作为数据分析基础呢?
- 如何自动评价学生的编程水平?直接针对能力评价指标进行相应的计算?假如对于排序很擅长,但是在查找上的表现很糟糕我们该怎么评价?
- 如何自动挖掘学生的编程缺陷?这位同学编程都答对了,但是...这位同学编程有困难,我们可以给他/她...

关于更多的研究问题示例

- 如何自动生成编程学习路径?陈老师出题太笨太懒了,应该先给这道题然后那道题…我觉得嘛,每个人都应该有自己的编程学习路径?
- 如何自动挖掘最佳编程CP? 他编程这不行,他编程那不行,但他们俩在一起就行… CP可以更多人…
- 如何自动推荐代码片段?
 这个题目我大体思路是明确的,就是这个点不知道怎么实现?
 如果给我推的代码可以直接运行成功,我又不想去研究它了...

为什么要认真做大作业

- 陈老师容易发神经, 我不想惹他...
- 我想让学弟学妹们《一小时精通Python编程》...
- 我数学分析没有学好,先争取个满分30分...
- 优秀作品搭配博士生辅助完成学术论文, 迈上学术巅峰...

