暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | AI算法应用C组 | | | | | | |
| 姓 名 | 汪铮沁 | 专业 | 人工智能 | 班级 | 091181 | 学号 | 09118105 |
| 实验时间 | 2020.8.31-2020.9.23 | | 指导教师 | 孔祥龙 | | 成绩 |  |
| **一、实验背景和目的**  高考是人生当中非常重要的一件事，除非少部分同学被提前录取或者排名分数特别理想，否则我们都会经历痛苦的志愿填报的阶段。那么，我们该如何针对自己的分数以及个人的兴趣爱好选择学校和专业，同时还要考虑到以后的就业情况和未来发展情况，还得结合自身能力来看，如果有条件还要考虑学校的实力、专业的实力、老师的实力，甚至考虑到学校的环境问题，如此看来，志愿填报真的是一件很重要也很艰难的事情。  我们现在的这个项目就是想帮助高考完的考生可以更好地也更方便更准确地选择志愿，使考生的分数不被浪费，也尽量使他能选择自己擅长且热爱的专业，也拥有光明的未来发展。 | | | | | | | |
| **二、小组任务和个人任务**  算法应用小组的任务是根据前端传来的数据等推荐大学及专业，分为学校推荐、聚类算法、专业推荐，我们是C组负责专业推荐。  但是专业的推荐也需要解决各学校专业名字不同的问题，这就需要语义消歧，推荐专业也分为两个方面，根据高考成绩或者其他能力推荐，于是我们又分为了三个组，我所在的组负责基于能力的专业推荐。  基于能力又可以分为三个方面，高考单科成绩、竞赛成绩、问卷调查结果，我负责和另外两位同学一起完成根据问卷对考生的能力进行评估的任务。 | | | | | | | |
| **三、个人任务需求分析**  说实话刚开始做这个任务还有点担心，根据调查问卷对考生的能力进行评估，听起来有点  难度，但好在老师给我们提供了学群和所需能力和包含专业和主要职业发展之间对应关系  的表格，让我们的任务有了思路。我们的任务简单分为两步，完成调查问卷、完成对应的  代码。因为有十八个能力，所以简单来说我们准备问十八个问题，根据用户回答的选项，  获得不同分数，然后根据不同学群所需的不同能力，计算学群所得分数，又由于每个学群  需要的能力个数是不一样的，不能单纯根据学群分数分析推荐度，所以还得数一下学群所  需能力个数，得分除以能力个数，就是我们最终得到的推荐度。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  １.完成调查问卷  　根据老师提供的资料（截取部分如下图）我们设计了调查问卷。  1  问卷设计的题目大致如下图几类：  （1）考生选择自身客观情况判断能力：    （2）考生选择个人主观情况判断能力：    （3）考生根据题目做出选择判断能力：    每个选项后面括号里的内容只是给我们待会儿写代码的时候用的，方便我们计算分数，  对考生的能力有一个可以直接观察的数据。  　２.完成代码  　　首先是两个列表，分别是学群名称、能力名称。    接下来这两个列表是分数对照、能力参照，分数对照是按照之前问卷每个选项后面括号  里的给出的，优良中差分别对应３２１０分，严格按照能力列表给出列表。能力参照则  是按照老师所给列表里学群与能力的对应关系，按照学群列表顺序给出所需能力的列表。  因为担心有考生故意选择不符合自己实际情况的选项来获得高分，我们的选项对应的分  数并不完全统一，３２１０是随机分配的。      接下来的是推荐专业的代码了，分为两块，一块是根据选项获得分数，一块是根据学群  所得分数得出推荐值。每个学群对应的能力个数并不相同，所以我们还需要将得分除以  能力个数，获得平均得分，即最后的学群推荐度。 | | | | | | | |
| **五、实验结果与分析**    上图就是我们完成的问卷，一共十八题，对应十八个能力，每个问题都有四个选项，每个选项  后都有不同分数，因为篇幅原因，只以小图的样子放出来。  但是这份问卷也有一些问题，首先就是这些问题都是我们三个人写的，专业方面肯定不值得肯  定，其次就是可能很多选项对应的分数不一定合理，比如选优的可以是４分，放大比例，但是  我们也不清楚这个比例该怎么确定，所以暂时定了３２１０，然后还有一种能力对应的题目只  有一个，感觉太过绝对，如果适当增加问题可能会更好一点，最后还有就是很多题目能得高分  的答案非常明显，如果某考生单纯选最优项，那就很容易得高分，就没有意义了。  下面四张图里我演示了两组数据。  第一组数据是随机输入的，第二组数据是我自己做了问卷以后得出的结果，结果显示我并不适  合目前这个专业（推荐度倒数第二），的确我选择这个专业的时候考虑不够成熟。          我们写的这个推荐的算法其实也有一点问题，每个学群需要的能力有好几个，但是实际上它们  的地位并不平等，有些能力非常重要，有些能力可能可有可无，但是我们简单的把它们平均以  后算推荐度，可能会导致结果不一定是最优解。 | | | | | | | | |
| **六、实验总结与心得体会**  这个项目，我们有超过一百位同学参与，每个人都有不同的分工，需要完成不同的任务。通过  这个项目我个人任务的部分，我了解了能力与学群的关系，写完了考生能力的调查问卷，完成  了根据问卷推测能力进而推荐学群的任务，通过了解其他同学的任务，我也了解了聚类算法、  数据挖掘、数据清洗、ｄｊａｎｇｏ开发、网页制作等等一系列不同的内容。而我们每个组之  间也都必须常常交流，互相传输数据，比如我们需要前端组的数据，也需要传输给后端组问卷  等内容，以前我们写这种项目从来没有那么多人过，这次也是一个非常不同的体验，而且由于  有同学在等着我们的数据，我们也不敢拖延，尽量提早完成，但是说实话，效率依然不行，这  可能是由于学校以及专业数据太多，两个组之间交流不够导致的。 | | | | | | | | |

2020年9月制