暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | aiB组（7组） | | | | | | |
| 姓 名 | 高奕辰 | 专业 | 人工智能（吴健雄学院） | 班级 | 615182 | 学号 | 17 |
| 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**  高考是我国人才选拔的重要途径，每年的考生数量可达900万以上，而在高考之后的填报志愿环节则是学生选择高校的至关重要的一步。而高中生和家长对于高校及高考志愿填报并没有太多经验与工具，巨大的需求催生出此次项目——通过对对2017-2019三年的各大高校录取分数和各省市的分数段数据的分析，尝试开发一套基于历史数据的高考志愿填报助手的系统。 | | | | | | | |
| **二、小组任务和个人任务**  本小组为AI算法B组，负责的是聚类算法的设计与实现，并完成功能的实现。  我们组运用聚类算法最终实现的功能被确定为两个，首先是聚类生成各省每年的分段分布模型，其次是得出各大学各专业的等级分布模型，然后根据分段分布模型对大学进行推荐。（两个功能都是基于三年高考的录取分数的。）  作为组长，我的任务主要在三方面。首先是确定功能，即在实验开头以及过程中保持与组员和其他组的讨论，确定功能的内容以及功能的具体实现方法，并且进行合理的调整。其次是对小组代码进行管理，将小组成员的代码片段进行整合，实现整个应用的功能运行。再次是将小组代码进行对接，按照其他组的需求将小组代码进行修改，满足对接要求。 | | | | | | | |
| 1. **个人任务需求分析**   对于第一个任务，确定功能，是一个在实验前期非常重要的环节。我们组拿到的题目并没有特别明确，大概方向即为使用聚类算法分出分段信息。但是，我和组员们都感到仅仅分出分段无法对软件整体功能进行完善。我需要与组员讨论，对我们的工作进行细化。  其次，在小组代码管理上，小组在码云平台上实现代码的合作，我通过issue等方式向组员发送任务，调动小组成员，将功能变为任务分配下去。小组成员完成的结果一般都是独立的，我需要通过小组成员代码的输入输出接口将工作拼接到一起。同时，对于小组组员的一些跨组别的需求，如数据库接入等方面，我也需要写好对接代码。最后，需要将整个app函数化，便于对接。  最后进行小组间对接，我需要与其他组对接人员交流，按照他们的要求修改代码，同时进一步了解其他组的工作，按照其他组的工作及时调整工作方向。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  对于第一个任务，我在每次小组合作时都会和组员讨论，来修改接下来的任务和功能实现方向。首先，基于聚类算法，我们提出聚类出分段和聚类出专业评级，如开题报告中的描述。但在后期实现时，组员提出加入基于分段的学校推荐功能，从而使我们的app更加贴近用户，由此更加贴合软件的志愿填报功能。除此之外，由于小组成员提出算法深度不足的问题，我又提出组员使用类似损失函数的思想对算法参数进行优化的想法。  第二个任务是我的工作重点。首先我联系了数据库组确定了数据库结构，然后使用sqlite3模块写了读取数据库的模块。  1  之后将读取数据库封装到get\_dataset的函数中，之后修改函数中的sql语句即可读取对应信息。在此之后我还向组员介绍了数据库结构和具体用法，方便功能实现。  QQ截图20201005204011  其次，在具体功能实现时，我将组员的代码进行整合。对于，组员的代码，我进行审查后进行合并和测试，找出bug并进行解决。在仓库中我们可以看到，组员上交的代码多为小样，我对小样进行测试与合并，组合成一整个app项目。在此之后，我将代码整合为多个部分，成为项目框架中的标准格式。  **QQ截图20201005210340**  对于最后一个任务，我对我们的应用添加了可视化，使用matplotlab等模块画出统计图来输出算法分析的结果，并且按照对接要求输出图片。  **QQ截图20201005205908QQ截图20201005210106** | | | | | | | |
| **五、实验结果与分析**  输出示例：（三个示例图不属于同一个实例）  555_pic2  commandtestmajorRank  如图为第一个功能和第二个功能的输出结果示例，综合来看功能实现完成。  但是在总体功能的设计上，我们最终实现的功能和开题时有些不同。与开题时设计的功能相比，  我们多了基于分段的学校推荐，而少了一个同专业不同学校的评级。前者的添加已经在报告前  叙述，而后者的减少是由于在数据库中学校的的学科名称与数据库一级学科二级学科名称没有  进行关联，我们的组员在实现时在进行语义消歧等工作上消耗了过多时间，在结题时，最终决  定不完全实现此功能。 | | | | | | | | |
| **六、实验总结与心得体会**  综上，此次实验完成了ai算法组的任务，为高考志愿app添加了一个实用的辅助功能。  在本次实验中，我担任了人数最少的组的组长。我在实验过程中遇到了不少困难，但也获得了很多经验。首先，软件开发过程中，我体会到组长是整个组的方向指挥，组长除了将任务拆分并分配下去之外，更重要的是在此之前确定，并不断地调整功能，使我们做出的软件不仅仅符合整个大组的要求，还能让组员理解自己做的东西，否则组员没有目标，更无动力实现。  其次，在组内和组间合作中，确定从小到大每一个模块的输入和输出极其重要。在一次合作中，我将一些组员的输入定为csv文件，另一些组员的输入设为直接从数据库读取，这造成了极大的不便，平白添加了很多工作量。（我之后在整合时需要再将它们统一起来。）  最后，此次实验我体会到git开发模式的优势：线上开发不仅仅能够方便地给组员分配任务，组内的合作也可以通过仓库清楚了解，除此，以为基础PR的从老师到组长到组员的管理模式也极大地方便了整个大项目的开发。 | | | | | | | | |
|  | | | | | | | | |

2020年9月制