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 | 曹思辰 | 专业 | 人工智能 | 班级 | 091181 | 学号 | 09118113 |
| 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**  据了解，高考志愿填报对高考录取结果，起到了关键性的作用，不仅关系着考生四年的大学经历，更与考生未来人生发展有着非常重要的关系。现今社会有很多不公平的现象，但高考却是最公平的制度，至少高考时按照分数来录取，大家都是凭自己的努力。马上又到了一年一度的高考了，考生考的好不好，他们的分数可以上什么样的院校，如何将每一分得到最合理的使用，是所有人最为关心的问题。因此志愿填报是高考完的重中之重。志愿填报是职业发展的关键起点，大学的选择影响个人高度和圈子，学生们还得根据自身选择专业。因此为了能更好的帮助考生选择适合自己的学校和专业，我们推出了一款高考志愿推荐系统。 | | | | | | | |
| **二、小组任务和个人任务**  **小组任务：**  在本次高考志愿推荐系统中，我们AI算法应用C组主要负责的是根据用户填入的信息以及AI算法应用A组给出的推荐学校来给出适合用户的专业（其中包括专业的推荐度与风险值）  **个人任务**：  我们组在内部一共分成了三个小组，分别负责：根据成绩推荐成绩；更具能力推荐专业；语义消歧。我担当的职务是根据成绩推荐专业的小组长，组员为曾家俊和丁自超。我  的任务主要是给组员分配任务，让他们帮我收集一些数据，而我则是编写我们小组功能所需要的代码以及编写整个AI算法应用C组的最终整合代码。 | | | | | | | |
| **三、个人任务需求分析**  我的个人任务除了给我们3人小组分配任务外，具体可以被分成3个：1、编写基于总成  绩推荐专业的代码；2、编写基于能力推荐专业中的根据用户具体单科成绩推出能力  这部分的代码；3、编写整个AI算法应用C组的最终的整合代码。  1、编写基于总成绩推荐专业的代码：  在这一任务中，我需要根据用户输入的总成绩以及AI算法应用A组给出的推荐学校给出  各专业的推荐度以及风险值。因为成绩高低的标准每年都不一样，所以学生总成绩的排名  才是我们需要的数据，为此我需要获得往三年各校各专业的录取最低排名和用户的输入  的排名。    2、编写基于能力推荐专业中的根据用户具体单科成绩推出能力这部分的代码：  在基于能力推荐专业中，我们整个大组一共整理以下内容:由具体专业到一级专业的映射；  由一级专业到学群的映射；各学群所需要的能力；能力列表。我通过三个部分计算出各学  群的推荐分，进而得出其学群中包含专业的推荐分。而我负责的就是根据用户具体单科成  绩推出用户的能力从而得出各学群的推荐分，然后与另一小组进行学群推荐分的整合。为  此，我们需要先根据用户输入的单科成绩筛选出用户的优势学科。因为不同地区的考试科  目不同，评分指标也不同，所以我需要获得不同地区的高考科目，以及根据地区的不同设  定不同的阈值来确定用户的优势学科。除此之外，我们还需要得到科目与能力间的对应关  系。以上这两样数据的设置和采集我交给了我的两个组员去完成。而我负责根据他们采集  到的数据完成功能的实现。  3、编写整个AI算法应用C组的最终的整合代码：  在这个任务中，我需要设置权重融合基于单科推荐学群、基于问卷推荐学群、基于竞赛推  荐学群三种学群推荐度，然后建立映射关系，来得到由AI算法应用A组给出的推荐大学  中各专业的推荐度。把这个推荐度与基于总成绩的推荐度进行融合来获得最终的推荐度。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**    1、编写基于总绩推荐专业的代码：  因为第一组迟迟没有把数据存入数据库中，所以我的工作大部分都是基于读取csv、txt和json  文件的。  因为我们需要的是最低录取排名，所以我根据高校录取分数线整合和各地一分一段表写了代码来获得高校录取排名文件：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\7W7[0A0FKQ6Y]59[{Q6R7}J.png  在这一文件中我在高校录取分数线整合文件的基础上加了排名这一列  以下代码是读取csv数据，以及根据省份、推荐学校、文理类别的筛选数据的代码：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\M}C24AV1$F1F(SD5QKVLG9Q.png  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\AZYJFTY%2~KZF%5X41NC8JG.png  从中我们筛选出某具体专业，三年的录取排名。  我通过比较用户输入的排名与过往三年某专业的最低录取排名计算出推荐度和风险值。  推荐度：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\FOTV]ZG95VKJY_`0]TK(7DB.png  风险值，我设置了一个分段函数来将其归一到[0.1]间  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\W31BWWC@0WVU({7A%NJ{C`P.png  计算出所有专业的推荐度与风险值，并将推荐度归一到[0,1]间  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\J57TN`NLF{J$@5R59U$YBR6.png  2、编写基于能力推荐专业中的根据用户具体单科成绩推出能力这部分的代码：  我利用其它组员收集与编写出的数据构建了四种字典，分别为：  学群对应能力：sg2abl  一级学科对应学群：maj12sg  科目对应能力：sub2abl  具体学科对应一级学科：major2major1  考虑到最后是从数据库中读取数据，前三个字典内容较小就直接把结果放到函数  里了，最后一个较多，准备录取数据库中。  然后，编写了获得优势学科的代码：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\OB~2FKB@1{RGY86(XE)$H@S.png  写了由单科成绩获得学群推荐度的代码：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\F1MXT[V9`_(35}@~VFHSA%Q.png  3、编写整个AI算法应用C组的最终的整合代码：  我先对基于单科推荐学群、基于问卷推荐学群、基于竞赛推荐学群三种学群推荐度进行了整合：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\706ITO2572_E(8PC@GR%A$5.png  然后编写了由学群得到具体专业推荐度的代码，获得基于能力推荐专业的结果：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\}8ZBT2{8R3FA6N38UXZ5ITG.png  最后把这一结果与基于总成绩推荐专业的结果整合：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\CIUJ8RHN$@`8CPC}{)J1C~U.png | | | | | | | |
| **五、实验结果与分析**  1、基于总成绩推荐专业：  输入如下：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\AD@6()NJSR4$_@{7M4N{83E.png  输出：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\KA6]JUF23`H9J)$Y~77IC}M.png  2、基于能力推荐专业中的根据用户具体单科成绩推出能力这部分：  输入：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\$$7WCMMQ]RDHYJL)HCU`HGB.png  输出：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\52%MVT)T4@_ZD@OC~{(GX_H.png  3、整个AI算法应用C组的最终的整合代码  输入：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\B~A_5`AEADWYVM4O]AV[3{W.png  输出：  C:\Users\csc\AppData\Roaming\Tencent\Users\641143099\QQ\WinTemp\RichOle\O5RZ$42P5P52DN(Q6J@$1GN.png | | | | | | | | |
| **六、实验总结与心得体会**  本次实验模仿了公司运转的模式，让我们在本科生就体验到这样的工作模式，我觉得是非常有意义的，老师的想法也是非常有创意的。在本次实验中，我第一次体验到我作为项目的一部分而工作，而不是所有任务的负责人。在这种模式下，我们组与组之间，成员与成员的沟通是必不可少的，不管是任务的衔接还是小组成员的分工，都是以前自己一个人写代码未曾体验到的。所以说，在这门课中，我收获了许多。  当然，本次实验中也充斥着问题。第一组迟迟没有完成数据的处理，数据中也存在着许多问题（格式不统一、数据错误等），这导致我们其它许多小组的任务都无法进行。虽然，大家或多或少都对第一组存在些意见，但他们也联合数据库组在结课前一天把较为正确的数据写到了数据库里（结课后也在继续更新数据）。他们的精神是值得肯定的，毕竟我们是在一起完成项目，任何一环出现问题，其实大家都得承担责任，毕竟我们是一个整体，应当互相帮助。  所以总体来说，我还是比较满意，希望以后有时间能继续来看看与完善这次大家一起努力的结晶。 | | | | | | | | |

2020年9月制