暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | AI算法c组 | | | | | | |
| 姓 名 | 佘瑾南 | 专业 | 人工智能 | 班级 | 091181 | 学号 | 09118115 |
| 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**  在全球疫情的背景下，高三毕业班学子面临着巨大的学业压力和择校困难。为了减轻他们的高考志愿难题，东南大学人工智能专业开设的由孔祥龙老师指导的软件实践课程，决定迎难而上，运用所学知识服务社会，报效祖国，并在此实践中扎实基础，磨炼专业技能，开发一款推荐高考志愿的程序，帮助高三毕业学生解决择校选专业的难题。此项目基于Python的Django框架，进行web开发。结合，数据收集与处理，数据库，知识图谱，AI算法，UIi设计等多方面知识，进行综合实践。此项目不仅培养同学们的专业素养，还讲软件实践应用到高校学习中，使同学们能接触到企业开发的过程，同时还培养了团队合作能力。 | | | | | | | |
| **二、小组任务和个人任务**  小组任务：   1. 语义消歧：确定各专业的语义表示，统一不同高校对于同一个专业的不同称呼，便于横向比较、提高数据利用率以及大类的推荐。列出所有高校的所有专业名称，找到与每个专业最接近的一个或多个一级学科，建立对照表。 2. 基于总成绩的专业推荐：在不考虑特别爱好的前提下，成绩是学生选择专业的重要依据。因此，在给使用者推荐学之后，还需要依据成绩给他们推荐合适的专业。不同的高校录取分数线不同，同一高校不同专业的录取分数线也不同。数据量不足。可以根据算法A组给出的学校和用户输入的成绩与排名，给出该学校各专业的推荐度与风险值。 3. 基于能力的专业推荐：如何评估用户的能力，以及根据其能力推荐专业，是十分关键的问题。这里我们参考了孔老师提供的数据，以下是其中一部分。在孔老师提供数据的基础上，我们引入了学群这一概念。最终总计有18个学群，19项能力。   个人任务：  语义消歧部分：确定各专业的语义表示，统一不同高校对于同一个专业的不同称呼，便于横向比较、提高数据利用率以及大类的推荐。列出所有高校的所有专业名称，找到与每个专业最接近的一个或多个一级学科，建立对照表。 | | | | | | | |
| **三、个人任务需求分析**  确定各专业的语义表示，统一不同高校对于同一个专业的不同称呼，便于横向比较、提高数据利用率以及大类的推荐。  1.在高校录取分数线整合表中获取所有高校所有专业的名称  2.在二级学科改进表中获取所有学科门类、一级学科和二级学科名称  3.使用Jaccard距离计算专业名称与学科名称的相似度，找出每个专业最接近的学科名称  4.人工检查并纠错  5.补充自定义的一级学科  6.将学科名称统一为一个或多个一级学科  7.删除特殊的专业名称  8.人工检查并纠错 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  将学科门类和二级学科统一为一个或多个一级学科，多个一级学科用’;’隔开    添加一级学科对应的ID：    删除特殊的专业名称：    人工检查并纠错 | | | | | | | |
| **五、实验结果与分析**  专业与学科对照表：    新增学科门类及一级学科：    专业与一级学科对照表： | | | | | | | | |
| **六、实验总结与心得体会**  本次项目实践是一次重要的理论与实际相结合的机会，本组成员共同合作，通过对文献的查询，对已有web网页制作的借鉴一起完成了这个项目，更深入的了解了Python库中包含的各种函数及其实际功能。使得我们对于程序整体的规划与设计有了新的认识与了解。  虽然我们组在本次项目制作中承担的是AI算法，但是在我们组内又分成了三个小组，分别承担着语义消歧，基于总成绩的专业推荐，基于能力的专业推荐三个部分。而我是语义消歧小组的一员。我们的任务主要有：在高校录取分数线整合表中获取所有高校所有专业的名称，在二级学科改进表中获取所有学科门类、一级学科和二级学科名称，使用Jaccard距离计算专业名称与学科名称的相似度，找出每个专业最接近的学科名称，人工检查并纠错，补充自定义的一级学科，将学科名称统一为一个或多个一级学科，删除特殊的专业名称，人工检查并纠错等。在整个实验过程中，虽然我们的任务无比艰巨，需要人工查验并修改几千条数据，但是小组内成员互相打气，互相帮助最终攻克一个个难关，我觉得这个过程是非常可贵的。所以我觉得这次项目实践，不仅学习的是动手操作能力，更是学习小组合作的一个非常好的机会。  虽然在本次课程设计中个人的贡献不可否认，但我觉得最重要的还是小组成员间的合作。通过这次课程设计，让我第一次体验到了以小组形式完成一个任务的过程，受益匪浅。在开始阶段，各人分开搜集资料，由于个体的差异，搜集的资料存在差异在所难免。所以在资料汇总的时候，思想火花碰撞的过程就尤为重要。听取他人的智慧，并加以评论，适时地提出自己的看法，最后达成意见的统一。在明确任务后，小组成员间进行任务分配，每个人完成各自的任务，在完成后进行统筹协调，整个过程都让我学习到了很多。 | | | | | | | | |

2020年9月制