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 | 张硕 | 专业 | 人工智能 | 班级 | 1班 | 学号 | 09118112 |
| 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**  高考作为我国选拔未来优秀人才的重要手段，已经成为国家、社会、家长考生关注的重点，除了高考成绩外，高考志愿填报也是高考后续阶段中最重要的一个环节，直接关系着考生的未来。然而，现阶段市场上的高考志愿填报系统存在许多不足，例如：系统功能单一，没有直观的、用户更容易接受的方式展现，仅仅根据考生分数进行推荐等。本次实验目的即为针对当前高考志愿填报系统存在的问题，设计并实现一个功能丰富、用户使用更方便的高考志愿填报系统。 | | | | | | | |
| **二、小组任务和个人任务**  小组任务：  基于成绩的专业推荐：成绩是选择专业时最为主要的因素，在第六组推荐学校的基础上，我们对每个学校推荐一定数量的专业，并用指标来衡量推荐程度  基于能力的专业推荐：不同的专业对学生的能力要求也不同，以此为依据来优先推荐专业可以一定程度上提高推荐的合理性  专业的语义确定：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。  个人任务：  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**  已有数据：  高校录取分数线整合（省份名字统一）.csv    一级学科改进.csv    二级学科改进.csv    任务背景：  各个高校有不同的专业命名规则，同一种专业在不同的高校常常有不同的名称，这给知识  图谱构建和查询某专业的高校排名带来了困难，因此有必要统一各个高校的专业名称，将  高校的专业映射到一个标准的一级学科。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  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    需要提取的数据是第五列的专业名称，去重后得到所有出现过的专业名称。  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    第一列为学科ID，1位表示学科门类，3位表示一级学科，5位表示二级学科，第二列为学科名称，需要读取并保存。同时建立ID-学科和学科-ID的字典。    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  与jieba等分词方式相比，直接拆分每个字效果更好；同时和bert、word2vec等预训练的模型相比，使用Jaccard距离匹配的准确率更高。    4.人工检查并纠错  有的匹配结果并不准确，需要人工检查并改为和专业名称最相近的学科  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，并添加学科ID        7.删除特殊的专业名称    8.人工检查并纠错  添加属于某专业的一级学科或者删除不属于某专业的一级学科。 | | | | | | | |
| **五、实验结果与分析**      实验结果为专业-一级学科对照表，第一列为数据集中所有专业的名称，第二列为与该专  业最接近的一个或多个一级学科，第三列为一级学科对应的学科ID | | | | | | | |
| **六、实验总结与心得体会**  这次的软件实践项目是我目前在大学参与的规模最大的项目，虽然整个项目我只参与了一小部分，但从项目分组到个人任务都让我收获良多。  在班级项目的小组中我选择了AI算法小组，因为我希望亲自参与一个系统的算法内核的构思与创建，与此同时算法又起到了一个承前启后的作用，和数据组、前端组都要有合作，因此是一个比较重要的部分。  我们小组的任务是专业推荐，其中又分成三个部分：基于成绩的专业推荐、基于能力的专业推荐、专业的语义确定，我选择的是专业的语义确定，需要把各个学校形形色色的专业名称统一为标准的一级学科，从而方便基于专业的各种检索。我们尝试了多种方法之后发现直接使用Jaccard距离计算专业名称与学科名称的相似度准确率更高，使用预训练模型的效果反而不太好，原因应该是模型需要使用专业相关的数据再次训练，而我们很难找到足够多的数据来训练，因此就选择了较为简单的Jaccard距离表示相似度。  在计算相似度时，如果直接和一级学科匹配，会出现原本和某个二级学科最接近的专业匹配到了另一个不包含该二级学科的一级学科的情况，因此我们先将专业与所有学科匹配，再统一为一级学科。  在实验的过程中，我体会到了分工合作的重要性，将一个较大的项目层层分工下去让每个人各司其职又相互配合，是一项困难而重要的工作。同时，明确的任务目标有时比做任务更重要，如果目标总是改动会使任务的工作难以开展，甚至前功尽弃。  总而言之，这次实验让我了解到了项目开发的一般过程，也体会到了开发大型项目时分工合作的重要性，在以后的小组任务中我会更加注重小组分工和目标，提高团队合作的效率。 | | | | | | | |

2020年9月制