暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | kgInference(赵基藤组) | | | | | | |
| 姓 名 | 廖滔 | 专业 | 人工智能 | 班级 | 091182 | 学号 | 09118236 |
| 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**  当今时代，高考已经是全国众多考生的必经之路，然而高考后大家都需要进行志愿填报，由于各个大学招生情况相去甚远，每一年，每一个省份的情况也一直是处在不断变化的状态，这给考生们带来了不便。于是我们尝试做一个志愿推荐系统，提前收集各学校的招生信息，依据考生实际情况给予志愿填报的建议。 | | | | | | | |
| **二、小组任务和个人任务**  小组任务： 1、 依据高校分数线进行统计分析，生成相应图表，展示变化趋势； 2、根据用户输入的动态个人信息进行相应的可视化； 3、以地图为依据生成可视化界面，展示高校信息； 4、基于知识图谱的推理和问答； 5、对应的django框架编写。  个人任务：知识图谱问答  基于知识图谱的问答，尝试理解考生询问的问题，比如某省的某个排名区间能上什么大学或者某省考生想读某个专业，尝试去寻找相关专业强劲的学校。并给出我们的回答。 | | | | | | | |
| **三、个人任务需求分析**  每个考生都会有先选学校还是先选专业的考虑，部分人可能优先选学校，部分人会优先选择自己钟爱的专业。无论如何，志愿填报推荐系统既然要推荐，就肯定是以考生的意愿为前提。既然以考生意愿为前提，就必须理解考生的问题。我主要书写一个程序，把问题当中的关键信息抽取出来，也就是“理解”问题的关键所在。只有明确了考生的意向，比如想冲985，想冲211；或者想读数学专业，有哪些大学数学专业不错；又或者某考生想在某个省份上学，优先选择那个省份的大学等等。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**    先设定特征词路径，在项目下有一个叫做dict的文件夹，里面用txt文件存着一系列信息。score 里面存放着0-750的分数，year则是高考年份， university 对应了我们爬虫所得的几十所大学的录取分数数据，major 是大学的专业，province是我国的省份。从里面读取了一系列的特征词。    随后是特征词的加载。  qwds尾缀相当于是帮助我们切割语句的一个定点，比如一句话“我能不能上北京大学”，检测到有“能不能上”这个词语，那这个词语后面极大概率是一所学校的名称。同理，“我想读数学系”  检测到这个“系”那么说明前后很可能就是某个专业的名称。    上图是应用之前的切割词来帮助我们判断这个信息是哪一种是类型的，是一本线，二本线还是一些别的东西。    此处用一个循环，判断词语的类型，然后添加到一个列表里面。    **以上是**过滤的加速以及我们的问句过滤。    **这个是上面用到的一个函数**，判断句子里面有没有某个词语出现，从而返回一个True False。 | | | | | | | |
| **五、实验结果与分析**    举了一个简单的例子，在广东省的一个考生报出了分数，询问是否能考上东南大学。输出的结果是对问题的信息进行了抽取，得出三个重要信息，省份，分数，想报的大学。 | | | | | | | | |
| **六、实验总结与心得体会**  本次实验是全体同学合力做一个高考志愿填报系统，我个人虽然只是做一个不大的程序，帮助系统理解一些志愿填报基础问题的语义，但是这个程序也是后续工作的一个基础，因为只有理解了问题问什么东西，才能进一步匹配后续相关大学或者相关专业的数据，以便进行推荐。另外在本次作业中，我们逐渐认识到靠单打独斗是解决不了问题的，必须通过及时的沟通和团队的协同作战，才能较好地完成任务。另外，合作也不仅仅是小组之内的合作，和别的小组也是合作关系，协调关系。所以在这次作业中，既提高了编程能力，也培养了在一个团队中如何通过协调合作完成一个大项目的能力。 | | | | | | | | |

2020年9月制