暑期学校实验项目：高考志愿填报助手

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 小组名称 | AI算法B组(第七组) | | | | | | |
| 姓 名 | 徐志修 | 专业 | 吴健雄学院 | 班级 | 615182 | 学号 | 61518205 |
| 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**  高考志愿的填报对于每位高考考生都是至关重要的大事。我们希望根据往年高考志愿填报的规律，进行信息的整合和汇总，得出高考填报的普遍规律，进而对各位填写志愿的高考考生进行指导。 | | | | | | | |
| **二、小组任务和个人任务**  小组任务:  通过对各分段的特征信息进行聚类，需要对多年的分数，分数的排名，专业进行分析，得出恰当的分段分布模型和专业分布模型，辅助推荐系统的构建。  个人任务:  （1）根据得到的分数聚类，划分考生的分数所属区间，将考生分成不同的分段。再结合用户输入的分数进行判断，得出用户的分数所属。再根据用户的分数，历年高考分数聚类以及各个大学的录取分数，综合得出给推荐用户的大学。 | | | | | | | |
| **三、个人任务需求分析**  (1)用户所属分数区间的判断:  我们根据往年高考分数分布的规律，可以大概得出高考考生的分数分布可分为五个类别，即高分段，中高分段，中分段，中低分段以及低分段。我们希望能够根据用户输入的分数与分数聚类进行对比，得出用户所属的分数区间，从而更好地为用户推荐适合的大学。  （2）大学综合分数的计算和用户推荐大学的选取：  由于大学录取是按照各个专业采用不同的录取分数，而我们想为用户推荐的是大学而非具体的专业，因此我们需要通过大学的各个专业的分数得到大学的综合分数。此外，我们需要根据用户的具体分数，用户所属的分数段以及各个大学的综合分数进行判断，得到最适合推荐给用户的前三所大学。 | | | | | | | |
| **四、实验过程（需附上关键代码及相关说明）**  (1)用户所属区间判断:  首先根据同组同学的工作，我们可以根据往年的分数分布得到分数的聚类，如下图所示：    上图为江苏省2019高考分数的数据分布，可以看出高考分数基本满足正态分布。  我们对用户输入的分数进行判断，根据用户输入的分数与各个聚类中心的距离相比较，得出用户所属的聚类中心。    由于我们之前进行聚类运算时，是使用各个数据点与中心的绝对距离之间进行比较，并进行迭代计算，从而得到各个聚类的中心点，所以我们使用用户输入数据点与各个聚类中心点的绝对距离进行比较，从而得到用户所属的聚类编号，再通过聚类的编号得到用户具体属于哪个分段。我们得到了用户编号之后，可以更好地得到用户填写志愿的普遍规律，从而从宏观的角度对用户的高考填报进行指导  (2)大学综合分数的计算:  我们根据各个大学的每个专业的录取分数进行计算，达到每个大学的综合分数，从而得到用户适合的大学。    首先我们从数据库中得到大学的专业录取分数，对其进行数据的读入，并进行记录。    对各个大学的综合分数进行计算，最后得到的结果如图所示:    其中第一个数据代表省份，第二个数据代表文理科，第三个数据为大学的名称，第四个数据则为各个大学计算出的综合分数。  (3)大学推荐:  我们根据大学的综合分数和用户输入的分数，用户所属的聚类，综合得出为用户推荐的三所大学。 | | | | | | | |
| **五、实验结果与分析**  聚类结果:    可以看出数据分布比较平均，大致服从正态分布，从而证明了我们使用聚类方法的正确性。  同时,对各个省的聚类数据而言,处理后会获得一个距离中心的列表，借由这个列表可以完成学校推荐的功能，列表的部分数据形式大致如下所示。前面的编号为”省份+年份+科类”，后面的5个数据为5个聚类中心点。 | | | | | | | |
| 通过上面获得的数据,结合推荐程序,最终所输出的推荐学校如下: | | | | | | | |
| **六、实验总结与心得体会**  在本次实验中，我主要有以下的几点体会:   1. 人工智能相关知识在实际工程的应用：   之前，我们在各种课堂中学习了很多人工智能的相关的知识和算法，而本次软件实践课程则让我们将人工智能中的一个算法：聚类算法在实际工程中进行应用。本次我们只在一维维度上进行聚类算法的实践，但仍然也遇到了一些问题，如聚类中数据点产生缺失，一些聚类的情况不符合朴素的分类观点，对于这些问题，我们进行了小组内的讨论，并进行了一些资料的查询，从而更好地解决问题。这使得我们意识到，一般的人工智能理论在实际应用中，仍需要一定的调整，以适合实际工程的需要。   1. 现代编程工作的合作模式：   我们在本次编程项目中，采取了小组合作的方式完成，同时不同小组之间也有专门的同学负责小组的对接工作。由于一些交流上的问题，我们与其他小组的对接出现了一些问题，并花了一些时间协调解决。这使我意识到在现代编程的项目中，一个人只按照自己的习惯低头写代码是行不通的。在大型项目中，交流和沟通是非常重要的。 | | | | | | | |