

初步分析文档

AI 算法应用 C 组

1 基于成绩的专业推荐

需求分析 志愿填报不仅需要选择学校，而且需要选择专业。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。在不考虑特别爱好的前提下，成绩依然是学生选择专业的重要依据。因此，在给使用者推荐学校之后，我们还需要依据成绩给他们推荐合适的专业。

初步实现思路 综合考虑历年来各个高校各个专业的录取分数（名次）区间以及热门程度，在给定高校的前提下为使用者推荐符合条件的专业。因为高校已经由第六组给定，所以基于成绩的专业推荐较为简单。

所需数据 近三年各个高校各个专业在各个省份的录取分数（名次）区间。

2 基于兴趣的专业推荐

需求分析 在根据成绩推荐可以或合适的专业的基础上，兴趣往往是专业填报时的关键因素，所以我们期望在原有基础上添加“根据兴趣进行专业推荐”这一功能。我们首先需要了解学生个人在初高中时期较为感兴趣的方面，如建筑构造、计算机技术等等，根据这些因素进行专业推荐，其次，在学生没有与专业相关的明确兴趣时，通过一些侧面的兴趣，如运动、乐器等来推断什么专业适合该学生，最后在学生没有明确的想法时，通过提供一些选项，让他们了解自己可能感兴趣的方面。

初步实现思路 首先建立起兴趣与专业的关系网和权重，在一些主成分达到一定阈值时直接推荐专业。对于侧面的一些兴趣因素，可以通过贝叶斯公式来推断合适的专业。最后，对于没有明确目标的学生，可以设计一个类似的答题问卷系统，来帮助其推断合适的感兴趣的专业。

所需数据 一些大方向上的兴趣走向及其与专业的契合度、每个院系中学生的兴趣占比（方便之后贝叶斯公式的运用）、一份有助于推断兴趣的问卷。

3 基于能力的专业推荐

需求分析 虽然在高考填报志愿时，不同专业对学生有理科和文科的要求，但是实际上各专业对学生能力要求都是非常具体的，单单以理科和文科不能完全表达。为了更好地让用户能够按照自己相应的能力来选择专业，我们需要统计两样指标。在用户这方面，我们需要获得他们在高中或初中的履历（如竞赛获奖，学生会经历，班委经历，社团经历）。在大学专业上，我们需要得到各专业的具体要求（如计算机专业可以要求用户有一定的计算机基础）

初步实现思路 首先需要把用户能力分成相应的类别。然后统计出各个专业的具体要求，作为加分项（人工智能专业可以把计算机以及数学的高考分数和竞赛获奖作为加分项目），然后根据用户的履历，对所有专业按照加分项加权求和得出各专业推荐系数。

所需数据 不同专业对具体能力的要求。

4 基于未来发展的专业推荐

需求分析 因为外界的各种因素，现在的学生在选择专业时都会考虑未来的发展状况，比如该专业的就业率、该专业的薪资状况，或者是该专业的工作环境（室内，室外）、工作强度。所以说我们有必要根据学生对未来发展的需求来给学生推荐专业。我们首先需要知道学生的理想专业的属性。比如想要就业率高，薪资高，室内工作为主，而对工作强度没有什么要求的学生就可以推荐计算机类；而想要室外工作的就不应该给这类学生推荐计算机类。同时学生可以对需求有主要次要的区分，比如优先看中就业率，其次看中薪水，对工作强度和工作环境没有什么要求。

初步实现思路 我们初步决定将所有学校所有专业关于未来发展的每一项属性全部分类，类中有符合该标签的该专业与该专业对应的学校。比如东南大学计算机类应该同时出现在就业率中就业率高的标签、薪资中的薪资高、工作环境中的室内标签下。当学生提出需求时，在首要条件的对应标签中找到推荐学校的专业，再去次要条件中将这些选出的专业进行比对，从而进行推荐。

所需数据 各专业的就业率与薪资水平。

5 专业大类和专业的语义确定

需求分析 考虑到目前的高校招生时越来越倾向于大类招生，且专业名称也常在原来的基础上加入“智能”、“智造”等关键词，因此导致专业名称五花八门，且容易造成歧义。另一个严重的问题是，不同的学校对于同一个专业有着不同的称呼，这样会导致难以将其看作一个专业来进行评估和推荐，缺乏横向比较，而且难以适用收集的大量数据。为解决这一问题，我们需要获得各学校各专业的基本含义，并将相同的归为一类，以便后

续的使用。在获取了专业的语义后，还可以通过聚类将其聚为几个大类，如计算机类、设计类等等，为后期根据用户兴趣、能力推荐专业提供了很大的方便。

初步实现思路 我们初步决定使用字符表示的距离相似度（Jaccard 距离和 Levenshtein 距离）以及语义表示的 word2vec 词向量来判断某几个名称是否代表相同的含义。对于一些少量难以通过语义理解的专业，后续也许会使用人工进行处理。另一种实现方法为爬取学校的相关文件，即某大类中含有的专业，此方法工作量巨大，可能难以实现，因此还是以语义为主要方向。

所需数据 这里需要的数据即为训练词向量所需的文本，我们初步打算使用一些中文百科词条以及相关的新闻素材来训练。

6 各专业的学校排名

需求分析 为了给用户推荐到一个合适的专业，需要即满足其兴趣要求，又要考虑到未来发展的前景。一个专业的高校排名很大程度上决定了这个专业内高校未来的就业情况和薪资情况等，排名靠前的高校在这个专业内就业情况较好，薪资空间较大。为了考量一个专业的高校排名情况，我们从学科评估等级、学科世界 ESI 排名、师资力量、科研经费四个方面来综合评估。

初步实现思路 把学科评估等级、学科世界 ESI 排名、师资力量、科研经费匹配到相应的评分数值，并合理的加权计算出体现专业高校排名的特征值。

所需数据 不同专业所有学校的学科评估等级、学科世界 ESI 排名、师资力量、科研经费。