软件实践第7组（ai2组）初步项目设计

一 项目概述

1. 工作内容

本项目的主要内容是对高考志愿填报推荐系统进行ai算法的设计与实现。本小组的具体工作是进行聚类算法的设计与实现，并和其他小组进行对接。我们的聚类算法需要对多年的分数、排名及高校进行分析，得出分段分布模型，以及专业的分布模型。

1. 开发条件：

项目开发平台：PyCharm 2018 等python编程平台

Pytorch等AI开发平台

项目开发语言：python

开发人员：61518207 张政 61518217 高奕辰 61518219 张欣然

61518229 游家伟 61518205 徐志修 61518424 王贵涛

61518426 周之遥 61518210 吉中旭

项目预计完成日期：2020 年 9月

二 功能需求分析

本小组在和其他小组讨论之后，确定了我们的工作重点，并在之后会追求与其他小组良好的对接合作。

1. 分段分布聚类

本功能来源于客户对自身分数定位的宏观理解需求。客户在填报志愿时首先要对自己的分数有一个比较宏观的理解，比如“高、中、低”分段。因为高考的背景是全省甚至是全国的，仅仅根据自己的排名数据很难确定自己在全省的位置（其原因可能有对全省参与人数难以了解、直白的数字数据不易于接受），进而难以决定志愿填报。所以我们希望用户在使用系统时可以立刻了解到自己的分数大致定位，这样用户可以立刻对自己是去开挖掘机还是上东大有了大致的想法。

其次，本功能还用于提取数据的宏观特征。对于每一个查询者，其输入数据为分数、地域，及可能的专业意向、偏爱的地理位置；而经过本功能的算法处理，其输入可以增添“分段”这一维度。这样能够使数据分析更具深度，提高预测学校等等其他功能的准确度

1. 高校专业分布聚类

考虑到存在用户想对学校和专业有进一步的了解，可以推出排名系统，根据各主流排名、教资力量等指标对学校专业进行排名，从而让用户对学校及专业有直观的了解。而由于高校的专业实力评估指标较为模糊，我们只能将各个高校的专业实力进行大致的聚类，形成A、B、C以及新设立等等的大致排名。

该功能的主要目的为辅助用户进行专业与学校选择的综合决策，在实际实行时，还可以结合用户自身分数、当地招生等情况进行结果的筛选等操作。

三 算法实现条件与目标设计

本部分以上述两种功能进行分别阐述。

1. 分段分布聚类
   1. 实现条件

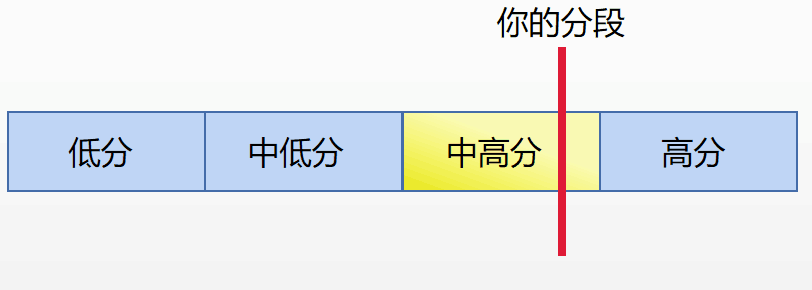
本功能的初始数据集为各省的分数对应人数的分布。而由于高考是拥有极大的参加人数的，所以分数分布有一定的概率分布模型，模型拥有一定的可行性。

模型有可能根据知识图谱的构建进行改进。

* 1. 目标设计

构建一个分数分段分布模型，其分段可由3个及以上组成（具体数值由模型和准确性综合决定），如“低，中，高”，或者“低，中低，中高，高”等等。

其次，进行可视化设计，加深用户对于自身定位的宏观理解，概念设计图如下图：



图一

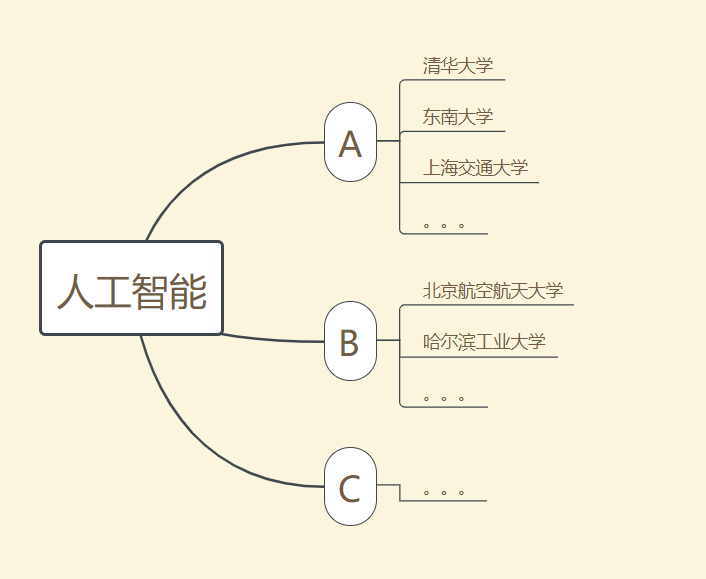
1. 高校专业分布聚类
   1. 实现条件

本功能的初始数据集为各个高校各个专业的录取分数线，同时酌情加入当年权威机构对于各个高校各个专业的排名状况，并为二者赋予一定权重。由于高校都有文理、地域等等分别，同时专业实力并不是快速改变的，模型具有一定的可行性。

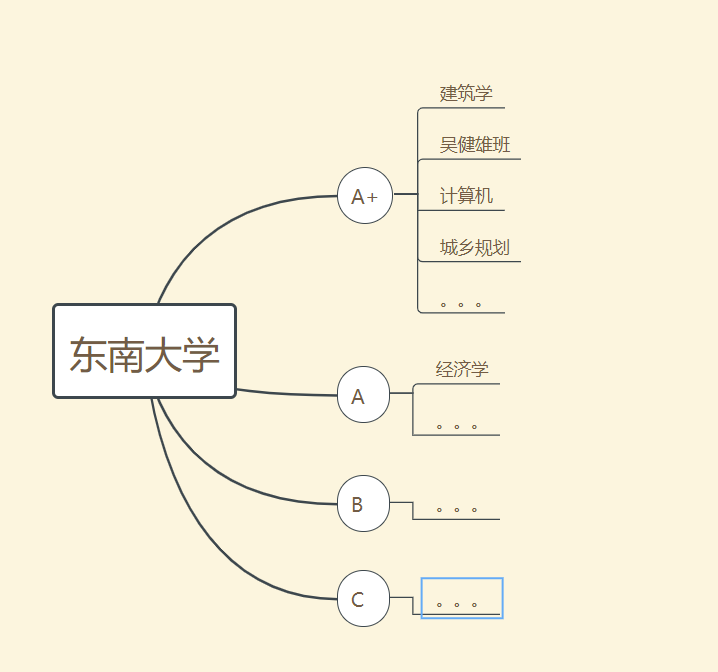
模型有可能根据知识图谱的构建进行改进。

* 1. 目标设计

构建一系列高校专业评级表，可以由专业作为主维度，高校作为副维度，也可以由高校作为主维度，专业作为副维度，其结果概念设计如下：



图二



图三

2020.9.4