**2024年第十一届中国可视化与可视分析大会**

**数据可视化竞赛 赛道1**

**（ChinaVis Data Challenge 2024 - mini challenge 1）**

**答 卷**

参赛队名称：上海交通大学-白骐硕-赛道1

团队成员： 白骐硕，上海交通大学，qishuo-bai@sjtu.edu.cn，队长

吴治远，上海交通大学，wzy605399@sjtu.edu.cn

刘一诺，上海交通大学，lyn15806636972@sjtu.edu.cn

杨雨彤，上海交通大学，flora20@sjtu.edu.cn

曹俊翔，上海交通大学，cjxqaq@sjtu.edu.cn

董笑菊，上海交通大学，xjdong@sjtu.edu.cn，指导老师

团队成员是否与报名表一致（是或否）：是

是否学生队（是或否）： 是

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）：D3，Python

例：D3，Excel，MySQL，Qt，CVASTer（天津大学xxx中心开发的数据可视分析工具）

共计耗费时间（人天）： 40人天

本次比赛结束后，我们是否可以在网络上公布该答卷与视频（是或否）：是

**1、【知识掌握程度】分析学习者答题行为日志记录，从答题分数、答题状态等多维度属性量化评估知识点掌握程度，并识别其知识体系中存在的薄弱环节。**（建议参赛者回答此题文字不多于800字，图片不多于5张）

**1.1 系统概览**

**图形用户界面, 图示

描述已自动生成**

**图1：**可视分析系统概览图（系统完成后截一个更好的，标上ABCD，把聚类设置的浮窗也截上）

为了完成「析数启智」时序多变量教育数据可视分析挑战赛，我们基于赛事提供数据设计了可视分析系统LearnClusterVis，旨在帮助教育机构全面了解和优化教学效果。该系统通过多视图展示学习者的学习行为和知识掌握情况，包括：

* 聚类视图（Scatter View）：展示学习者的聚类情况，便于识别不同群体的学习模式。
* 画像视图（Portait View）：通过外圈圆环图详细展示群体或个体对不同知识点及子知识点的掌握情况；通过内圈雷达图展示学习者在不同特征上的表现，从而帮助用户深入分析学习行为模式。
* 题目视图（Title View）：展示各个题目的答题情况，包括每题的作答次数、分数分布、提交时间分布等信息。
* 答题周日志（Week View）：按周展示学习者对不同知识点和子知识点的掌握程度，便于跟踪学习者在不同时间段的表现。
* 学生视图（Student View）：详细展示每个学生的详细答题信息，包括提交次数和最高得分。

**1.2知识点掌握程度特征选取**

为了量化评估知识点掌握情况，我们首先在原始数据的基础上进行了初步分析和处理，最终选取了6个特征用来反映某位学习者对于某道题的掌握情况，然后根据题目的分值为每道题的掌握情况赋予权重，最后通过加权平均来计算学习者对于某个知识点或者子知识点的掌握程度。选取的特征及其解释如**表1**所示：（要不要加上计算的公式？）

**表1:** 知识点掌握程度

|  |  |  |
| --- | --- | --- |
| **特征字段** | **含义** | **解释** |
| score\_bonus | 答题得分加成 | 题目最终得分越高，该值越高 |
| tc\_bonus | 时间复杂度加成 | 题目时间复杂度越低，该值越高（如题目回答错误，则该值为0） |
| mem\_bonus | 空间复杂度加成 | 题目空间复杂度越低，该值越高（如题目回答错误，则该值为0） |
| \_error\_type\_penalty | 错误类型扣减 | 题目错误类型越少，该值越高 |
| \_test\_num\_penalty | 尝试次数扣减 | 题目尝试次数越少，该值越高 |
| rank\_bonus | 排名加成 | 题目排名越高，该值越高（OJ网站通常会设立用户排名，以用户的提交答案通过数多少或某个题目执行时间快慢为排名依据） |

**1.3 群体知识点掌握程度分析（scatter, portait）**

由于数据的高维度特性，我们希望通过聚类的方式来进行初始分析。用户可以首先点击右上角的设置按钮打开Cluster Configuration，选择聚类所使用的维度，并筛选数据范围（**图1-xxx**）。例如，通过选择

**1.4 个体知识点掌握程度分析（portait，week）**

**2、【学习模式】结合学习者的特征挖掘个性化学习行为模式，从多角度设计并展示学习者画像，如答题高峰时段、偏好题型、正确答题率等**（建议参赛者回答此题文字不多于800字，图片不多于5张）

**2.1 分析方法**

**2.2 学习者画像**

|  |  |  |  |
| --- | --- | --- | --- |
| **分析角度** | **特征字段** | **含义** | **解释** |
| 知识掌握程度 | score\_bonus | 答题得分加成 | 题目最终得分越高，该值越高 |
| tc\_bonus | 时间复杂度加成 | 题目时间复杂度越低，该值越高（如题目回答错误，则该值为0） |
| mem\_bonus | 空间复杂度加成 | 题目空间复杂度越低，该值越高（如题目回答错误，则该值为0） |
| \_error\_type\_penalty | 错误类型扣减 | 题目错误类型越少，该值越高 |
| \_test\_num\_penalty | 尝试次数扣减 | 题目尝试次数越少，该值越高 |
| rank\_bonus | 排名加成 | 题目排名越高，该值越高（OJ网站通常会设立用户排名，以用户的提交答案通过数多少或某个题目执行时间快慢为排名依据） |
| 学习习惯 | explore\_bonus | 探索加成 | 题目回答正确后，仍然尝试探索，探索次数越多该值越高 |
| enthusiasm\_bonus | 热情加成 | 题目发布后，全部提交次数的时间平均值越早，说明该生完成题目的热情越高，则该值越高 |

**3、【知识掌握程度and 学习模式】不同的学习模式直接影响到学习者对知识的吸收、整合及应用能力，高效的学习模式能够促进知识的深度理解和长期记忆。请对学习模式与知识掌握程度之间的潜在关系进行建模，利用图表的形式呈现结果并简要分析。**（建议参赛者回答此题文字不多于800字，图片不多于5张）

**2.1 分析方法**

**2.2 分析结果**

**图表？**

**4、【题目本身】合理的题目难度应当与学习者的知识掌握程度相匹配，当学习者知识掌握水平很高但答题正确率较低时，意味着题目难度超出了其能力范围。请试着利用可视分析方法找出这些不合理的题目。**（建议参赛者回答此题文字不多于800字，图片不多于5张）

**2.1 过难题目**

**2.2**

**5、结合上述分析结果，请你为题目设计者和课程管理人员提供一些宝贵的建议，以优化题库内容设置和改善教学质量，并简要说明理由。**（建议参赛者回答此题文字不多于500字，图片不多于3张）

**2.1 学习者画像分析设计**