

基于人格分析和数据驱动的班级学习小组划分模型

2025 级学生：李佳蓬 孙宇辰 杨子墨 指导教师：张磊 张翼

辽宁省葫芦岛市第一高级中学

摘要

在高中阶段，科学构建学习互助小组对提高班级整体学业水平具有关键价值。然而，传统随机分组模式通常未充分考虑学生在性格特质、学科专长及学习发展潜力等方面的个体差异，致使小组内部协作效能低下，甚至对学习气氛产生负面影响。针对此问题，本研究以融合 DISC 人格理论、学业能力互补分析及学习趋势预测方法，设计了一套适用于绝大多数高中班级的学习小组划分模型，致力于将学习能力与个体特质相契合的学生有机组合，以充分发挥小组内部的协同优势。

本文从三个维度对学生间的互助潜力进行量化评估：

首先，对于学生人格的分析，我们借助 DISC 行为风格测试，将每位学生的学习、性格特征量化为数据进行分析；其次，对于学生学科互补情况的分析，我们基于学生历次考试纯理六大科目的平均表现，根据一些数据的差值情况分析互补情况，以确保小组内双方能在各自的薄弱科目上获得有效支持；然后，对未来进步趋势的预测，我们通过对每位学生在各科历次考试排名建立线性回归模型。用拟合出的成绩变化的斜率，代表其学习进步的速率。结合学习幂定律与边际递减效应原理，构建对数形式的学习进步预测模型，量化小组合作中受助者成绩提升的幅度。

基于上述三项指标，将其相乘得到两位学生之间的“互助效能”，作为评估两人组合效果的综合标准。

在模型求解阶段，本研究以全班所有小组互助效能总和最大化为优化目标，采用贪心算法进行迭代匹配。为验证模型的有效性，本研究将模型分组结果与多种随机分组方案进行对比分析。结果表明，模型分组下的互助效能总和显著优于随机分组，证实该模型能够有效提升小组内部的协作质量。本研究还探讨了模型的局限性，包括考试成绩样本量有限、人格问卷可能存在主观偏差、贪心算法易陷入局部最优解以及时间参数设定依赖经验判断等问题。针对这些不足，提出了引入社交网络分析、模型简化、增加动态调整机制等改进方向。

本模型不仅适用于高三班级的学习小组配置，还可拓展至大学项目团队组建、企业新员工培训小组等多种场景，为基于个体差异的团队科学构建提供了一套系统化、可量化的方法论。

关键词：学习小组划分；人格分析；回归预测；特征工程；贪心算法；数据驱动

1 绪论

1.1 研究背景

在当前国家对高素质创新人才培养需求日益迫切的背景下，高中阶段作为人才储备的关键期，其教学质量的提升显得尤为重要。传统的教学模式正面临着如何更精细化地促进学生全面发展的新挑战，因此，探索一种能够兼顾学生个体差异、最大化发挥互助效能的分组机制，对于提升班级整体学业水平具有重要的现实意义。

当前，广大教育工作者为提升教学成效，促进学生全面发展，普遍采用组建学习互助小组的教学模式。然而，考虑到学生在各学科领域的能力差异与个性化学习风格特点，若采用简单随机分组策略，易导致资源分配失衡，如高水平学生过度集中引发组内恶性内卷，造成不必要的学习压力；或因性格特质不匹配导致合作障碍，使学习小组效能大打折扣，进而干扰学生的学业进度。为了防止随机分组造成的不合理问题以及学生自主学习造成的成绩极端分化现象。对于学生本身，我们简单量化学生性格特征；对于学生成绩，我们根据历次考试数据进行特征分析和回归分析。统筹多方建立模型，得出班级内学习小组划分的最优解。

1.2 问题提出

基于以上背景，我们需要建立数学模型解决以下问题：

- 如何通过设计问卷调查，收集学生在性格、兴趣、学习习惯等方面的数据，运用统计学方法进行分析，进而量化学生间的人格相容程度？
- 如何利用历史成绩数据，通过计算学生不同科目的成绩差异，评估其在哪些领域可以相互协助，从而量化成绩的优势互补程度？
- 如何通过分析组内成员的成绩趋势，简单预测小组互助对学生学业进步的影响？
- 如何结合学生的人格相容程度、成绩互补性、进步预测结果以及班级人数、性别比例等因素，运用优化算法求解班级学习小组划分的最优方案？

2 问题分析

2.1 人格相容程度的量化分析

为量化学生间的人格相容程度，本研究基于 DISC 人格理论设计问卷，收集学生在性格、兴趣、学习习惯等方面的数据。通过统计各人格类型中接纳其他类型的比例，定义“接纳度”；进而将两人彼此接纳度的乘积定义为“相容度”。该方法将非结构化的人格特征转化为可计算的数值，为后续分组提供了基础量化指标。

2.2 优势互补程度的量化评估

利用学生历次考试的各科（只考虑纯理选科，即选择语文，数学、英语、物理、化学、生物）的成绩期望值，**对每一科进行单独分析**，计算两人在每一科目上的成绩差与满分之比，我们将其定义为单科优势互补度。再取六科整体的期望值，定义为整体成绩的“优势互补度”。该指标反映了学生在学科知识结构上的差异性与互助潜力，值越大说明两人越能在学习上取长补短。

2.3 互助进步趋势的预测

为预估小组互助对学业的影响，首先基于学生历次考试排名进行线性回归，拟合出各科成绩趋势斜率。然后借鉴边际递减效应，引入对数模型： $\text{被帮助者的新斜率} = \text{原斜率} + (\text{帮助者斜率} - \text{原斜率}) \times (\ln t + 1) / (\ln T + 1)$ 。据此定义“趋势促进因子”，用以量化互助后学业进步的预期幅度

2.4 最优分组方案的求解

综合人格相容度、优势互补度、趋势促进因子，将三者的乘积定义为小组“帮助效能”。以全班所有小组帮助效能之和最大化为目标，采用贪心算法进行迭代求解。该方法避免了穷举的巨大计算量，能够快速得到科学分组的局部最优解，实现数据驱动的分组决策。

3 模型假设

在现实中，每位同学的学习生活与习惯等特征各不相同，为降低模型复杂度并便于分析，特做出如下假设：

- 1) 假定班级内同学们的学业均正常进行，无辍学、长期旷课等现象
- 2) 假定班级内同学们个人的学习习惯和方法短期内不会发生显著改变
- 3) 排除不可控因素对成绩数据的影响（如学生身体问题导致成绩不佳、学生违纪等情况）
- 4) 忽略班级内同学们人际关系对分组的影响
- 5) 规定统计所得数据均为真实数据，无谎报、恶意作答情况

4 符号说明

本文所使用到的一些符号说明如下：

表 1

	名称	出现位置
C_{ij}	人格相容度	6.1 人格相容度分析
$d_{Euclidean}$	闵可夫斯基距离	6.1 人格相容度分析
$S_{ij}^{subject}$	学科互补度	6.2 学生学科互补情况分析
S_{ij}	优势互补度	6.2 学生学科互补情况分析
$\widehat{g_i^{subject}}$	单科成绩回归模型	6.3 学生成绩趋势情况分析
$\beta_i^{subject}$	直线 $\widehat{g_i^{subject}}$ 的斜率	6.3 学生成绩趋势情况分析
p_{ij}	趋势促进因子	6.3 学生成绩趋势情况分析
E	帮助能效矩阵	6.4 小组内合作效果评价指标

5 模型准备

在正式建立学习小组划分模型之前，我们需要做好充分的准备工作，包括数据的收集、理论的选择和分析框架的构建。本次建模主要基于心理学和管理学中的相关理论，旨在科学地量化学生之间的相容性与互补性。

首先，在人格分析方面，我们引入了 DISC 个性测验作为理论基础。该理论由美国心理学家马斯顿提出，将人的行为风格分为支配型（D）、影响型（I）、支持型（S）和服从型（C）四种类型。我们通过自编问卷对全班同学进行测试，得到每位学生在四个维度上的得分，以此为基础量化人格特征，提出人格相容度的计算公式。

其次，在学业互补方面，我们以成绩优势互补度为指标，判断两人在同一学科上的互补潜力，通过期望值 i 、

j 来衡量互补程度。

此外，我们还引入了学习趋势分析，利用线性回归模型拟合学生历次成绩排名的变化趋势，并结合“学习幂定律”与“边际递减效应”，构建了学习进步预测模型。通过对时间参数的设定，我们尝试量化小组合作对学生成绩斜率的正向影响。

综上所述，模型准备阶段以 DISC 人格理论、Pavlov（2018）分析生产规划中学习效应的幂函数形式曲线模型为基础，结合成绩数据和排名趋势，为后续构建“帮助效能”指标和分组算法奠定了坚实的理论与数据基础。

6 模型的建立与求解

6.1 最优分组方案的求解

6.1.1 人格相容度分析

想要搭配出一个优秀的小组，需要考虑诸多因素。而组员的人格个性与精神内核的相容程度便是重要因素之一。直觉上，让两个“志同道合”的人成为搭档、建立小组会使工作完成的更好。事实上的确如此，Zahra Abbasi 团队对“深层次相似性”的研究¹ 证明了这一结论：人们更容易被与自己相似的人所吸引，而这种相似性会带来积极的工作成果。

我们搜索了网络上主流的行为、个性以及人格测试体系，结合本文研究需求与受众群体，选择使用“DISC 个性测验”作为人格相容度的量化体系。

DISC 个性测验是以美国心理学家威廉·莫尔顿·马斯顿（William Moulton Marston）的 DISC 行为模式理论为理论基础的一种人格测验，能够展现同学们的人格特点，并据此把人格类型分为支配型“D”、影响型“I”、支持型“S”、服从型“C”。但是 DISC 个性测验并不是把人们简单划分为四种类型，我们需要关注每个个体在这四个行为模式上的典型程度（即问卷得分）。

为了得到全班同学的 DISC 个性测验结果，我们设计了有关测试问卷，并在班级内进行下发、测验。DISC 问卷由 20 道判断题组成，每个人格类型有五道题，每一题对应的人格分值为 2 分，单个人格满分为 10 分。每道题各有 4 个描述个性特质的形容词，这些描述性的词语都采用中性词义，没有明显的褒贬意义显示出来，长度基本上一致，完成的时间统一规定为十分钟以内。

完成调查后，我们会得到反应每个同学人格类型典型程度的表格。结构形式如下（数据仅为实例）：

表 2

人格类型	D 支配型得分	I 影响型得分	S 支持型得分	C 服从型得分
学生				
张三	8	2	4	6
李四	2	2	8	4
.....	...	...	...	...

在计算两人之间的人格相容度时，需关注两人的 DISC 人格测试得分差。将 i 、 j 两人的问卷分数与满分作比后，投影在四维坐标系中，并计算他们的欧几里得距离

¹ Abbasi, Z., Billsberry, J., & Todres, M. (2023). Empirical studies of the "similarity leads to attraction" hypothesis in workplace interactions: a systematic review. *Management Review Quarterly*, *74*(2), 661 – 709.

$$d_{\text{Euclidean}} = \sqrt{\sum_{p \in \{D, I, S, C\}} \left(\frac{S_i^p}{10} - \frac{S_j^p}{10} \right)^2}, \quad (i \neq j).$$

欧几里得距离越大，两学生的人格差异越大。而后，遵循“深层相似性”的理论，整合定义 i 、 j 两人的人格相容度 (Compatibility) 具体形式如下：

$$C_{ij} = 1 - \frac{1}{4} d_{\text{Euclidean}}$$

其中， C_{ij} 的取值范围在 $[0, 1]$ ，取值越大，代表着两学生人格差异越小，那么其人格相容程度就越好。

6.1.2 学科互补情况分析

在同一学习小组中，为了确保能够组员互助提升，成员之间需要实现学科优势互补。比如有 i 、 j 两位同学： i 同学的语文、英语成绩较强，但物理、化学、数学薄弱； j 同学的语文英语成绩薄弱，但理科成绩较强。那么从单从**成绩分数**这一方面看， i 同学与 j 同学适合分一组，他们可以实现优势互补。在这一例子中， i 同学在文科方面帮助 j 同学， j 同学在理科方面帮助 i 同学。

实质上，判断两学生在某一学科上是否互补的方式就是判断他们在这一学科上的成绩差，并取二者**成绩差值与单科满分的比**作为量化标准， i 、 j 学生在某一学科上的学科**优势互补度**为

$$S_{ij}^{\text{subject}} = \left| \frac{i \text{ 单科成绩} - j \text{ 单科成绩}}{\text{单科满分}} \right|, \quad (i \neq j).$$

其中， i 、 j 两人分数取历次考试本科目成绩的**期望值**。

统筹整合语数外物化生六科，定义两学生**整体成绩上的优势互补度**为 i 、 j 学生六科**优势互补度**的期望，即：

$$S_{ij} = \frac{1}{6} \times \sum_{\text{subject}} S_{ij}^{\text{subject}}, \quad (i \neq j).$$

S_{ij} 的取值范围是 $[0, 1]$ ， S_{ij} 的数值越大，两人的优势互补度越高。其中 subject 属于语数外物化生六科所构成的集合。

6.1.3 学生成绩趋势情况分析

如何评价一个学生的进步？想必采用只观察预测学生的成绩分数高低的方式评价学生的进步是较为片面的。学习的进步是动态的过程，不是静态的节点；小组内的互助是有条不紊的，不是一蹴而就的。所以，我们对每个学生的历次考试成绩数据进行线性回归分析，通过拟合直线的斜率判断学生的**成绩走势**，并通过对组内二人的互助影响关系进行分析，预测出学生在学习小组建立后成绩斜率的提升量，以此作为学生进步程度的评估，并进行具体量化。

每次考试的题目难度不同，只看成绩无法确定学生的确切进步情况。所以我们根据每位同学的历次考试单科成绩的**校排名**数据，建立一次线性回归模型，通过最小二乘法，求解出回归方程，确定每个学生的单科成绩趋势（斜率）。以考试次序号为自变量，对**每个学生的每个单科成绩**建立一次线性回归模型：

$$\widehat{g_i^{\text{subject}}} = \beta_i^{\text{subject}} \bar{x}_i + \varepsilon$$

其中， β_i^{subject} 是拟合后一次直线的斜率，可以反映学生 i 的成绩进步趋势。

接下来，我们开始对学生在学习小组建立后成绩斜率的提升进行预测。根据**边际递减效应**，易得**随着学习时间的增长，学习者的学习斜率向知识传授者不断靠近**。通过这个关系，我们可以简单预测出在小组合作后被帮助者的回归斜率变化（进步趋势变化），进而得出这一小组的趋势促进因子。

在开始计算趋促因子之前，首先需要判断 i 、 j 两学生在某一科目之间的帮助关系，确定谁是提供帮助者，谁

是被帮助者。只需要通过比较二者在这一科目上分数的期望值大小即可完成判断——期望值大的是提供帮助者，期望值小的是被帮助者。

然后，根据上文计算出的二者的成绩斜率，以及边际递减法的模型，进行预测。其中， β_i 是被帮助者的成绩回归斜率， β_j 是提供帮助者的成绩回归斜率。则经小组合作后的 $\widehat{\beta}_i^{subject_t}$ 为：

$$\widehat{\beta}_i^{subject_t} = \beta_i + (\beta_j - \beta_i) \cdot \frac{\ln t + 1}{\ln T + 1}, \quad (i \neq j).$$

其中， t 为小组的合作时间， T 是学科弱势者完全接近掌握学科优势者知识的时间周期。根据实际情况， T 取值为 6 个月，约 180 日。 t 取值为 4 个月（约一学期）即 120 日。该公式基于认知心理学中的“学习的幂定律”²（练习曲线呈负加速），以及经济学中的“边际收益递减规律”³。通过引入时间参数 T （达到近似极限的总时间），将经典的对数函数学习模型⁴应用于预测辅导效果，同时体现了“初期进步快，后期趋近上限”的特点。

最后统筹整合两人在六个科目上的趋势预测，我们将趋势促进因子（process）定义为：

$$p_{ij} = \frac{1}{6} \sum_{subject} \left| \frac{\widehat{\beta}_i^{subject_t}}{\beta_j} \right|, \quad (i \neq j).$$

p_{ij} 越大，二者在未来的学习进步效果越好。

6.1.4 学生成绩趋势情况分析

为了量化小组两成员的综合互助效果，我们将 i 、 j 之间的人格相容度、优势互补度和趋势促进因子的乘积定义为 (i, j) 小组的帮助效能（Effectiveness），具体形式为：

$$E = c_{ij} \cdot s_{ij} \cdot p_{ij}, \quad (i \neq j).$$

其中， c_{ij} ， s_{ij} ， p_{ij} 均为上文的定义量。 $E[i][j]$ 则为 i 同学和 j 同学之间的帮助效能。其中， $E[i][j] = E[j][i]$ 恒成立。通过这一矩阵，我们可以便捷直观的获得小组们的互助效果数据，帮助效能也是评价小组优劣的重要指标。

6.2 模型求解

根据上文建立的模型框架，我们只需要向模型中代入班级数据。这个数据集可以来自于任何满足假设的班级，本模型对情境中的每个班级都适用。为了达到本文的应用目的，发挥实际意义，我们代入论文作者本班级的数据进行求解。

6.2.1 求解思路

数据获取完成后，首先启动数据集的特征工程，对数据进行清洗、整理。对每个学生建立各自的特征向量为：

$$Student_i = [Personality_i, Grades_i, Ranks_i, Trends_i]$$

在编程语言中， $Student_i$ 是一个多层嵌套的列表。其中：

- $Personality_i$ 是该同学的 DISC 人格类型得分，用于在未来计算两同学之间的人格相容度。
- $Grades_i$ 是一个一维列表，包含该学生历次考试各科的平均成绩数据。
- $Ranks_i$ 是一个二维列表，包含该学生历次考试六个学科的排名数据。
- $Trends_i$ 是一个一维列表，包含根据 $Ranks_i$ 计算而得的每个学科的成绩回归斜率。

然后，构建帮助能效矩阵，计算两两学生间的帮助效能。全班一共有五十个同学，若想要通过穷举的方式求解，则需要的计算量巨大，实际无法计算。所以，我们选择采用贪心算法求解。当全班所有小组的帮助能效和达

² 在《认知心理学及其启示》一书中，作者引用了相关研究，指出描述练习效果的曲线通常是幂函数，被称为学习的幂定律。

³ 本模型的数学格式效仿了 Pavlov (2018) 在分析生产规划中的学习效应时所使用的幂函数形式的曲线模型。

⁴ 一篇发表在《北京理工大学学报》的研究《学习投入与产出的经济学分析》通过实验验证了这一规律。

到极大值，则此时的分组方案就是科学分组的局部最优解。在理想状态下，当全班的帮助效能和达到最大值，则可取全局最优解。

$$\max \sum_{\text{所有组}(i,j)} E[i][j], \quad (i \neq j).$$

接下来，将帮助能效的因子的计算公式算法写入代码后，代入数据、整理结果，得到班级内学生们的特征向量。然后编写求取 $\sum_{(i,j)} E[i][j]$ 最大值的代码，使用贪心算法，找到当前步骤中帮助效能取值最大的组员进行配组。像这样遍历、迭代，直到程序完成组的划分后，输出结果。

6.2.2 代入数据，求出结果。

我们将学生的相关数据集导入程序脚本，进行简单地整理与计算，得出了学生们的有关参数数据。并且，对每个矩阵绘制了热力图如进行可视化，展示如下：

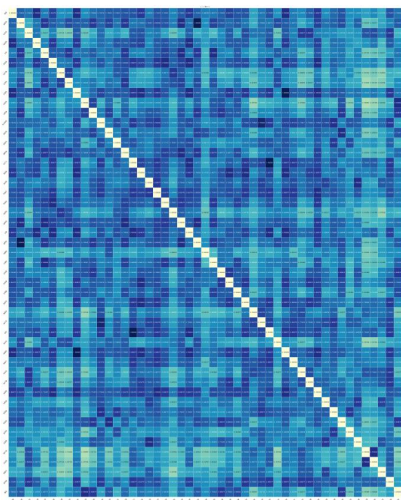


图 1 学科互补度矩阵热力图

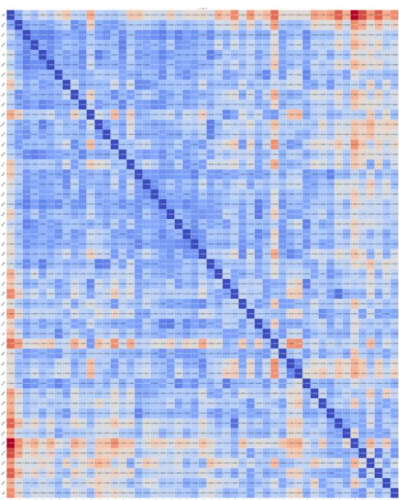


图 2 人格相容度热力图

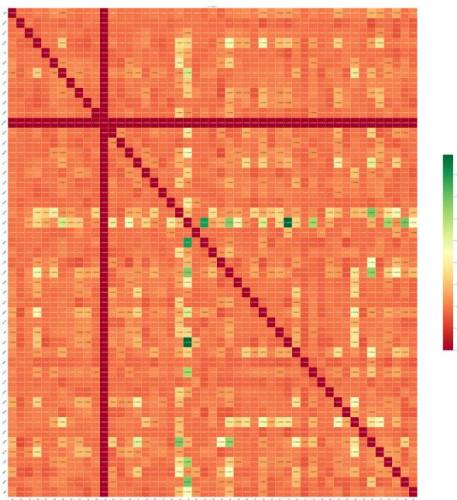


图 3 趋势促进因子热力图

注意，由于数据来源的班级中有一位同学成绩缺失，于是出现了图三中“红十字”的现象，实际上该同学的数据并未列入模型求解的数据集。

然后，将得出的特征向量数据导入贪心算法模型，计算求解。经过 27 秒后，程序输出了结果：

赵梓琪,阎沫含	张子川,王奥博	郭昊,李佳妮	赵泓翔,汪祉霖
何祉剑,杨嘉乐	魏子涵,王冠尧	赵跃策,刘思睿	张津维,龙帅谕
陈佳禾,魏昊林	任孟航,张佳睿	李佳蓬,张函铭	黄天朔,刘思源
王思蒙,杨潇然	才湛韬,曹韵	吴尚修,姬圣哲	郭瀚午,李思源
杨子墨,王雨荷	李鑫宇,陆矜	刘骏麒,赵泽铭	廉芮涵,金熙哲
杨子昂,赵子墨	吴楚越,王尊	王粤,李佳岐	孙宇辰,何欣研

表 3 模型最优分组结果

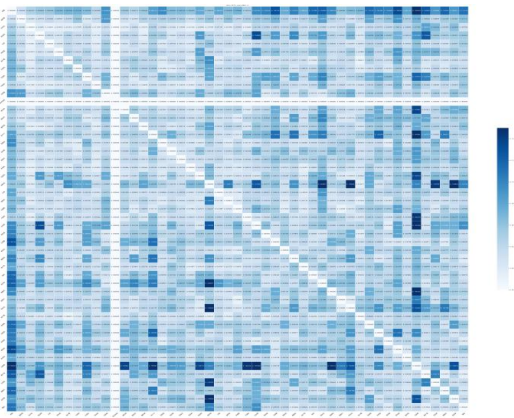


图 4 帮助能效矩阵热力图

根据这个方案进行学生小组划分，可以使得班级小组的帮助效能总和最大化，最大化后得到的班级帮助能效

总和约为 **2.03274**。

7 模型的分析与检验

7.1 最优程度检验

现在对上文求出的最优分组方案进行分析。为了检验这一结果是否是“最优解”，我们使用程序**随机**生成了 5 个结果，按照上文的定义计算后，他们的帮助能效总和分别约为 **0.71378**、**0.41572**、**1.17834**、**1.29765**、**1.49867**，这些随机方案的数据效果均不及上文求出的最优解，且部分随机结果与“最优解”相差甚远，这也进一步否定了现实中班级内随机分组的科学性合理性。通过与随机分组方案的数据结果比对，不难发现上文求出的“最优解”是可以做到**最优或局部最优**的。

7.2 误差分析

在数据层面上，所求结果所以依赖的数据集样本容量较小，仅有三次考试的成绩数据。这可能会导致一次回归模型的拟合效果差强人意，测试误差较大。同样，很难保证所有同学问卷填写的真实性，虽然在问卷中卷头部分有提示语，但是不排除问卷填写主观倾向主导的可能，因此人格分析可能存在误差。

模型与算法上，真实的学习效果在模型中的参数（如 $\widehat{\beta_i^{subject_t}} = \beta_i + (\beta_j - \beta_i) \cdot \frac{\ln t + 1}{\ln T + 1}$ 中的 $\frac{\ln t + 1}{\ln T + 1}$ ）可能与模型中不同。我们根据实际情况，依据经验对 **T**、**t** 进行了取值。但实际上经验主义并不科学，也许真实的取值与我们的取值有出入，遂可能造成误差。并且，贪心算法存在自身的局限性，可能导致结果仅为局部最优解，不排除有更优方案的可能。

8 模型的评价、改进及推广

8.1 模型的评价

我们将得出的分组结果在班级内进行了公示，并询问了同学们对小组划分的满意程度，经过统计，班级内学生对模型的满意度调查结果如下图所示：

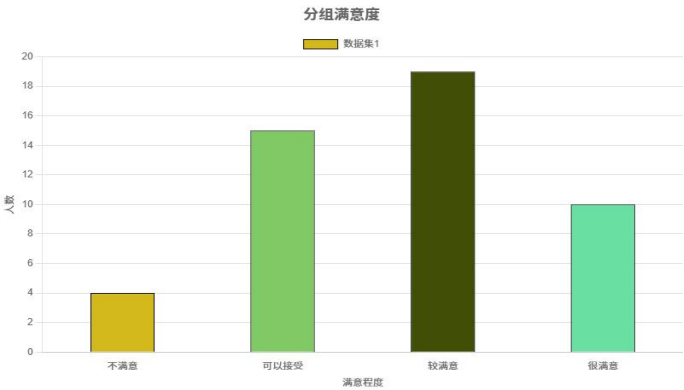


图 5 学生对模型结果的满意度

根据图表可见，有约 **8.33%**的同学对分组情况不满意，而对分组很满意的同学占比约 **20.8%**，相对而言，模型还有很大的改进空间。

8.2 改进方向

在现有模型的基础上，还可以考虑以下几点改进方向，让模型更加准确、贴合实际

- 引入**社交网络分析**，考虑学生之间的友谊关系进行分组
- 在优化模型的同时对模型进行简化，根据**奥卡姆剃刀原则（Occam's Razor）**，应该尽量优化模型的细枝末节，减少复杂度以提高鲁棒性、防止过拟合等。
- 增加**动态调整机制**，根据学习过程中的实际效果优化再分组

8.3 应用推广价值

经过适当的情景改进后，本模型不仅适用于高中班级的学习小组分组，还可推广应用于：

- 大学课程项目小组的组建
- 企业新员工培训小组的分配
- 社区志愿服务团队的搭配

9 参考文献

本文有关心理学、社会学等其他专业领域的理论与定律大多来源于网络搜寻。

- Abbasi, Z., Billsberry, J., & Todres, M. (2023). Empirical studies of the "similarity leads to attraction" hypothesis in workplace interactions: a systematic review. *Management Review Quarterly*, *74*(2), 661 – 709.
- Pavlov, O. V. (2018). "Dynamic models of production planning with continuous time in projects of new products development". *Journal of Physics: Conference Series*, 1096, 012180.
- 《认知心理学及其启示》——约翰·R.安德森（John R. Anderson）
- 《学习投入与产出的经济学分析》——《北京理工大学学报》
- DISC 个性测验 <https://baike.baidu.com/item/DISC个性测验/7778753> ——百度百科

10 附录说明

本论文中编写的相关代码与附件内容已经在 Github 开源：<https://github.com/shaidens/Model1>

- 附件 1-1：考试成绩数据（一）
- 附件 1-2：考试成绩数据（二）
- 附件 1-3：考试成绩数据（三）
- 附件 1-4：DISC 人格测试数据
- 附件 1-5：DISC 人格测试调查问卷
- 附件 2：数据整理、分析、模型求解 Python 脚本