

# 一、被试者能力评价体系的建立

## 0. 被试者分组的建立

- 已知Mooctest平台提供的数据，是基于将近300名被测者分成五组，且五组题目各不相同、组间题目重合度低、但是组内题目相同、组间题目难度相似。
- 又由于Mooctest平台未提供分组数据，我们根据 **同分组题目组相同**、**每组均有同学完成全部题目** 这两个已知条件，设计代码进行分组，代码见仓库；
- 分组结果：

- 各分组人数

组别	G1	G2	G3	G4	G5
分组人数	68	53	44	54	52

- 各分组题目数目

组别	G1	G2	G3	G4	G5
题目数量	200	200	200	200	206

- 具体分组数据（被试者属于哪一组、每一组有哪些题目）见仓库 `group-tests.json`、`group-result.json`

## 1. 评价指标的选取

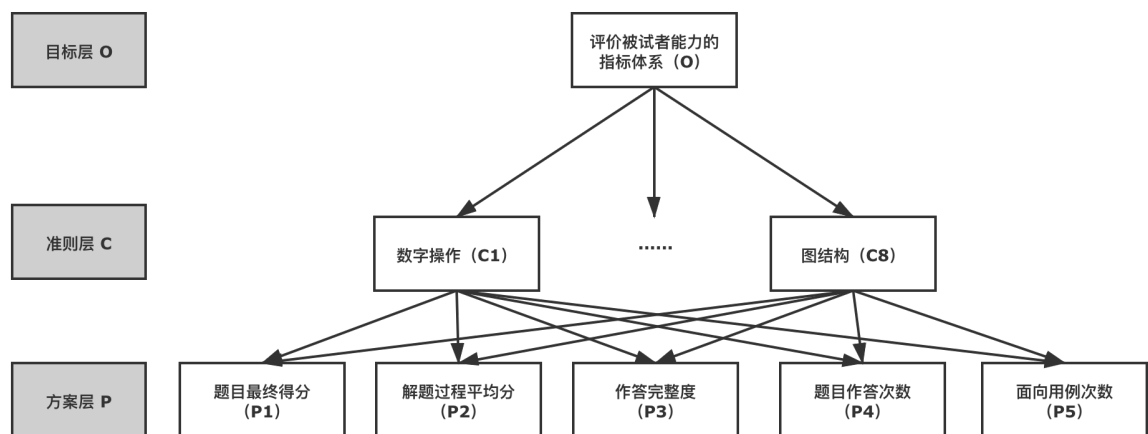
- 基于已有数据对被试者的能力进行评价，最重要的是根据被试者的提交次数、每次提交成绩、最终提交成绩与提交代码质量进行分析。根据上述 **1.0** 部分我们已经对总计300余名被试者的分组情况进行确定。
- 显然，最终提交得分越高、提交题目占为被试者提供题目比例越大，则被试者在此方面的能力越强。然而，多次提交等刷分行为、直接面向测试用例编程的作弊行为等，也均会影响被试者能力评价体系的建立。
- 当被试者存在面向用例的作弊行为时，在最终提交得分处获得的高评价一定会被否决。但是，由于慕测平台提供的数据仅限于 **提交记录**、而非 **测试用例运行记录**，所以对题均提交次数的研究价值不大。（1）
- 我们在理论分析的基础性上，结合了已知数据的可靠性，确定了以下指标及处理方法：
  - 题目需要分类处理，不同类型间题目不具有可比较性；
  - 被试者需要根据所做题目不同划分为不同组别，不同组别被试者不具有可比较性；
  - 提交成绩选用两组数据统计量：
    - 对于每道题目，我们均将最高提交得分作为最终得分指标，最高得分越高，说明对该题目的解答正确性越高；
    - 对于每次提交，我们将某题目的所有提交计算算术平均分，该项分值越高，说明此被

- 试者在本道题目中表现越好；
4. 作答完整度选用两组数据统计量：
    1. 作答题目数与全部题目数比：该项比例越高，说明答题越完整；
    2. 提交次数与作答题目数比：该项比例越低，说明本题对于此被试者耗费的精力越少。由于（1）的原因，该项权重不应过大；
  5. 诚信指标主要采用一组数据：
    1. 诚信指标：被试者面向用例等不诚信行为次数多，最终评分一定差。
- 我们采用如图所示的方法对两两指标的重要性进行检验

等级	同样重要	优秀	明显重要	强烈重要	很差
量化指标	1	3	5	7	9

## 2. 评价指标的检验

- 为确保上述指标的适用性，我们使用层次分析法进行检验



- 在选择不同准则时，我们考虑到每道题目均有自然的分类，对不同分类的题目都进行一次指标重要性调查是很有必要的
- 重要性填表： TODO

	C1 数字操作				
	P1	P2	P3	P4	P5
P1	1				
P2		1			
P3			1		
P4				1	
P5					1

- 可得方案层对目标层的权重为
  - 通过一致性检验

## 二、TOPSIS 法与评价指标模型的建立

### 1. TOPSIS模型背景

TOPSIS（Technique for Order Preference by Similarity to Ideal Solution, 逼近理想解排序法 或 优劣解分析法），是一种常用的综合评价方法，充分利用原始数据信息，精确的反应各个评价方案之间的差距

- 选用 TOPSIS 模型，可以将分析一中已获取的评价指标，作为其权重参与运算；
- 选取层次分析中相似的统计量，采用层次分析中的指标，限制程度大大降低，不需要目标函数亦不需要通过检验。同时，层次分析中的结论也可参与使用。

### 2. TOPSIS算法过程

1. 将原始数据正向化

我们选用的某些指标，指标值越高、显示被试者能力越强；而亦存在一些指标，其值越高、显示被试者能力越弱。有必要将所有指标进行统一的正向化，方便运算。

指标名称	指标类型
最终得分	极大型指标
题目提交均分	极大型指标
作答题目比	极大型指标
提交次数比	极小型指标
面向用例比	极小型指标

此次分析中，仅存在极小型指标需要正向化，使用极小型转极大型的简单公式：

$$x^* = x_{\max} - x$$

即可完成转换

2. 正向化矩阵标准化

为消除不同量纲的影响，我们将已正向化的矩阵  $X$  进行标准化处理：

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}$$

将标准化矩阵记为  $Z$ ，有

$$z_{ij} = x_{ij} / \sqrt{\sum_{i=1}^n x_{ij}^2}$$

3. 计算得分

若有  $n$  个要评价的对象， $m$  个评价指标的标准化矩阵

$$Z = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1m} \\ z_{21} & z_{22} & \dots & z_{2m} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nm} \end{pmatrix}$$

定义：

$$\begin{aligned} Z_k^+ &= \max\{z_{1k}, z_{2k}, \dots, z_{nk}\} \\ Z_k^- &= \min\{z_{1k}, z_{2k}, \dots, z_{nk}\} \\ Z^+ &= (Z_1^+, Z_2^+, \dots, Z_m^+) \\ Z^- &= (Z_1^-, Z_2^-, \dots, Z_m^-) \end{aligned}$$

定义：

$$\omega_j$$

为第 $j$ 个评价指标相对于总体的权重，其值应由层次分析法得到  
定义第 $i$ 个评价对象与最大值和最小值的距离：

$$D_i^+ = \sqrt{\sum_{j=1}^m \omega_j (Z_j^+ - z_{ij})^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^m \omega_j (Z_j^- - z_{ij})^2}$$

最终，可以算出第 $i$ 个评价对象未归一化的得分：

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

#### 4. 对 $S_i$ 的解释：

使用层次分析法结果加权后， $S_i$ 即可看做被试者的平均得分，越高则能力越强

### 3. 数据处理

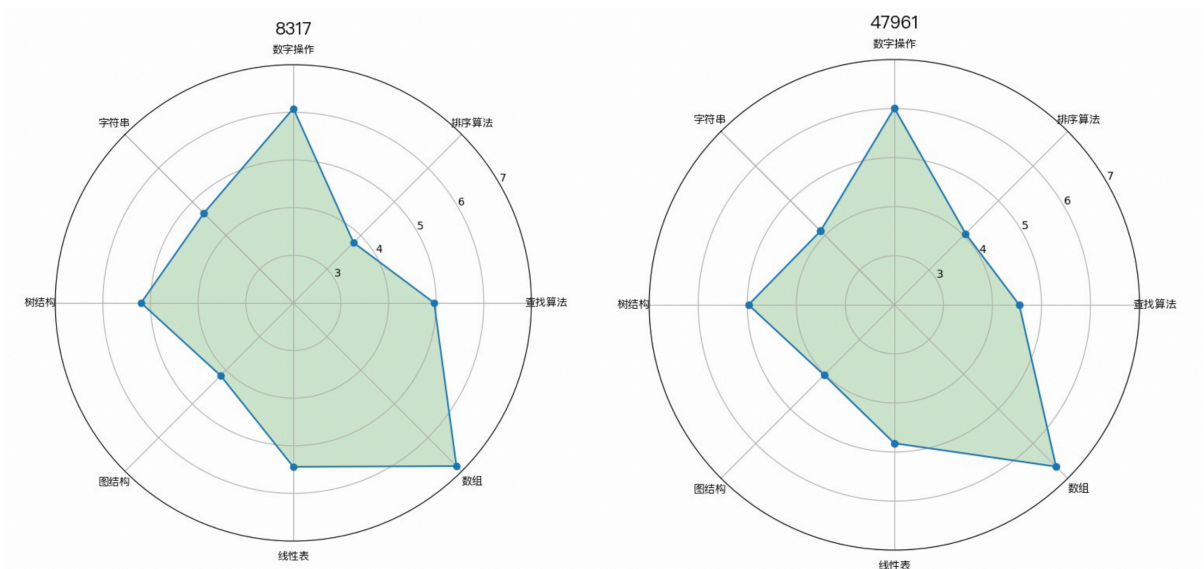
- 依据评价体系与2.2中算法，被试者最终得分越高则其能力越强
- 为了能够统一被试者的能力标准，依据1.0中 *组间题目不同但难度相似* 的前提，我们在分组得到基本统计量后，将不同组别合并处理，依据2.2中算法，得到最终分数。下图为部分同学在 *查找算法* 项目中的得分情况：

$S$  score 项为TOPSIS中所得  $S_i$  同倍数放大后的结果

## TODO Example Fronted Ratio

User id	Mean score of committed	Mean score of submitted	Commit ratio	Example fronted ratio	Submit times commit ratio	S score
49823	96	41.7605634	1.0	0	3.55	4.25346405
60580	100	95.2380952	1.0	0	1.05	4.92420883
60591	100	100	1.0	0	1.15	4.95856851
60592	100	89.1666667	1.0	0	1.2	4.87161747
60601	100	96.1904762	1.0	0	1.05	4.93161604
60605	95.75	63.3904348	1.0	0	2.3	4.55220421

- 依据上述数据制图，得到每位同学的能力分布雷达图，下面为id为8317、47961两位同学的能力分布雷达图：



- 分析雷达图后，我们发现完成率高的同学们，能力分布大体形状一致，但存在少许差别；
  - 即采用TOPSIS法分析同学们的能力分布基本合理。
- 依据上述数据制图，我们还可以得到各组同学之间的能力分布差距（基于各组题目难度大致相同的前提）

