

# 从编程评测到 CODIA

2021-12-04

# 目录

- 从传统评测到编程评测
- 编程评测的基本范式
- CODIA 编程评测
- 编程评测研究

# 传统评测

- 各类考试大家都很熟悉了
- 简单总结：
  - 目标：考察知识掌握
  - 形式：客观题、主观题

编辑

围建一个面积为  $360m^2$  的矩形场地，要求矩形场地的一面利用旧墙（利用的旧墙需维修），其它三面围墙要新建，在旧墙的对面的新墙上要留一个宽度为  $2m$  的进出口，如图所示。已知旧墙的维修费用为  $45$  元 /  $m$ ，新墙的造价为  $m$ ，设利用的旧墙的长度为  $x$ （单位： $m$ ），修建此矩形场地围墙的总费用为  $y$ （单位：元）。

（ I ）将  $y$  表示为  $x$  的函数；（ II ）试确定  $x$ ，使修建此矩形场地围墙的总费用最小，并求出最小总费用。



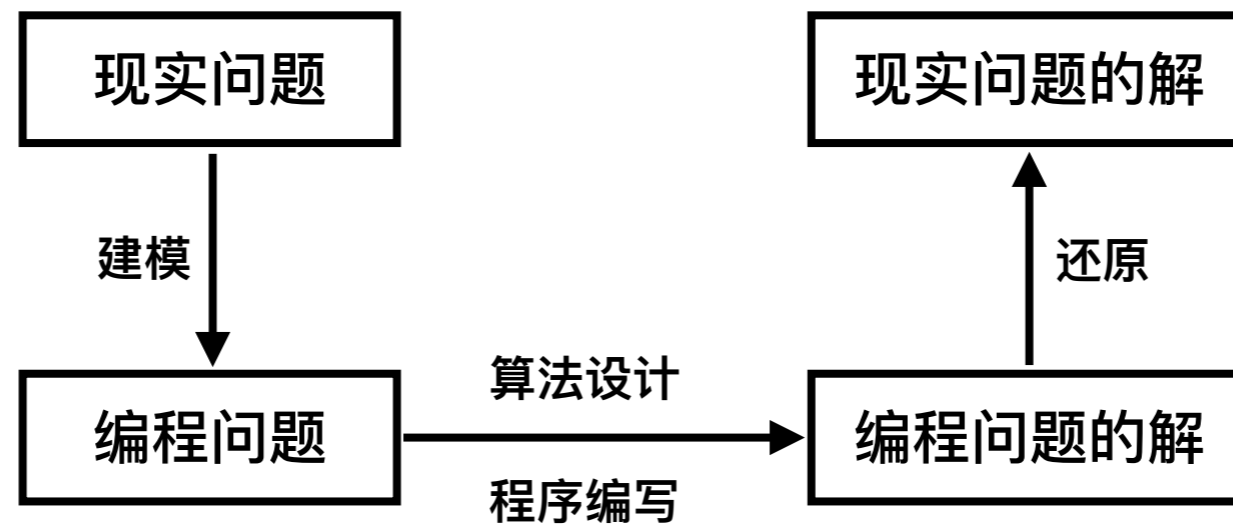
# 从传统评测到编程评测

- 思考问题：
  - 目标是否有区别？
    - 编程并非知识 (knowledge) 而是技能 (ability)
    - 需要考察技能掌握情况
  - 形式是否有变化？
    - 选择题能考察技能吗？
    - 写代码和写作文一样吗？

# 编程评测的基本范式

## 核心目标

- 编程评测的核心目标：考察编程解决问题的能力
  - 解决问题是算法设计的目的



# 编程评测的基本范式

## 核心目标

- 思考：能否考察开发网站？机器人对抗？网络攻防？
  - 编程能力只占所考察能力中的一部分
  - 考察有偏向性
  - 很多时候可简化为编程问题

# 编程评测的基本范式

## 基本形式

形式	提交答案	提交代码
类比	选择题（答案唯一）	主观题（答案不唯一）
提交内容	问题的解	问题的解法
考察目标	解决一个问题	解决一类问题
正确判定	答案正确	通过（AC）

# 编程评测的基本范式

## 举例

- 求  $2+3$
- 提交答案：5
- 输入整数  $a, b$ , 求  $a + b$
- 提交代码：

```
1  a, b = input().split()
2  a, b = int(a), int(b)
3  print(a + b)
```

- 算法竞赛更常采用



# 编程评测的基本范式

## 提交答案的应用

- 问题本身很难求解时，提交答案也能验证解决问题能力
  - 应用：机器学习比赛

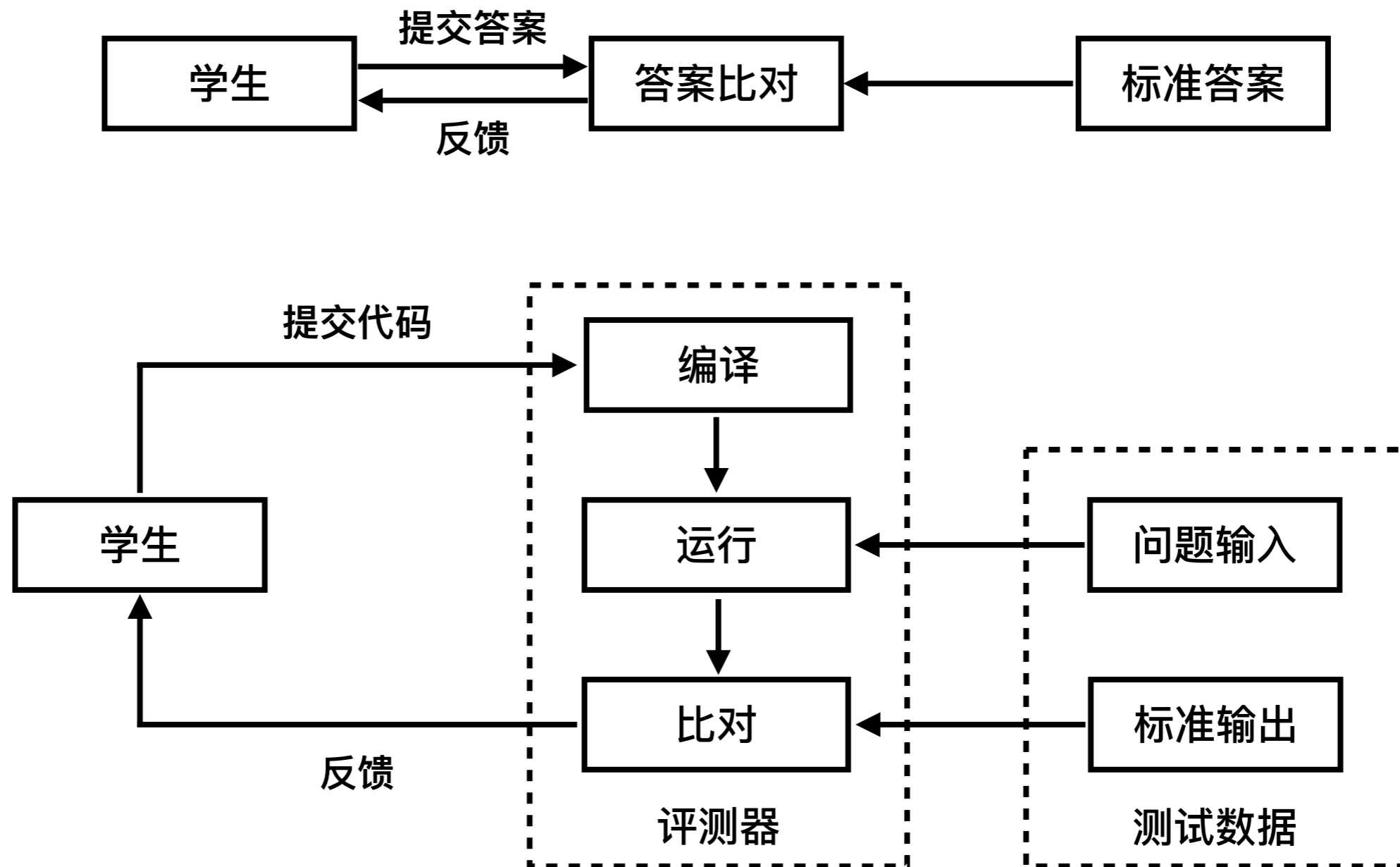
# 编程评测的基本范式

## 评价问题

- 评价能否解决一个问题：
  - 直接比较结果
  - 或通过自定义 metrics
- 如何评价是否能解决一类问题？
  - 验证是否能解决问题的一些实例（比如 10 个）
  - **测试数据**：包含输入（问题实例）和对应输出（问题的解）
  - 注意：输入不可见，能解决未知的问题实例才有意义

# 编程评测的基本范式

## 评价问题



# 编程评测的基本范式

## 编程测试

- 出题：题面，输入输出格式，测试数据设计， .....
- 约束：语言限制，时空限制，提交次数限制， .....
- 赛制：题目数量，测试时长，是否即时反馈， .....

# 目录

- 从传统评测到编程评测
- 编程评测的基本范式
- **CODIA 编程评测**
- 编程评测研究

# CODIA 编程评测

## 评测机制

- 题目组成
  - 问题描述，输入输出描述，样例数据，测试数据

爬楼梯 提交情况  
30 / 73

贡献者 [官方题库](#)

### 题目描述

假设你正在爬楼梯。需要  $n$  阶你才能到达楼顶。

每次你可以爬 1 或 2 个台阶。你有多少种不同的方法可以爬到楼顶呢？

**注意：**给定  $n$  是一个正整数。

**示例 1：**

```
输入： 2
输出： 2
解释： 有两种方法可以爬到楼顶。
1. 1 阶 + 1 阶
2. 2 阶
```

### 测试数据

1	2	3	4	5
---	---	---	---	---

#### 测试数据 2 🗑️

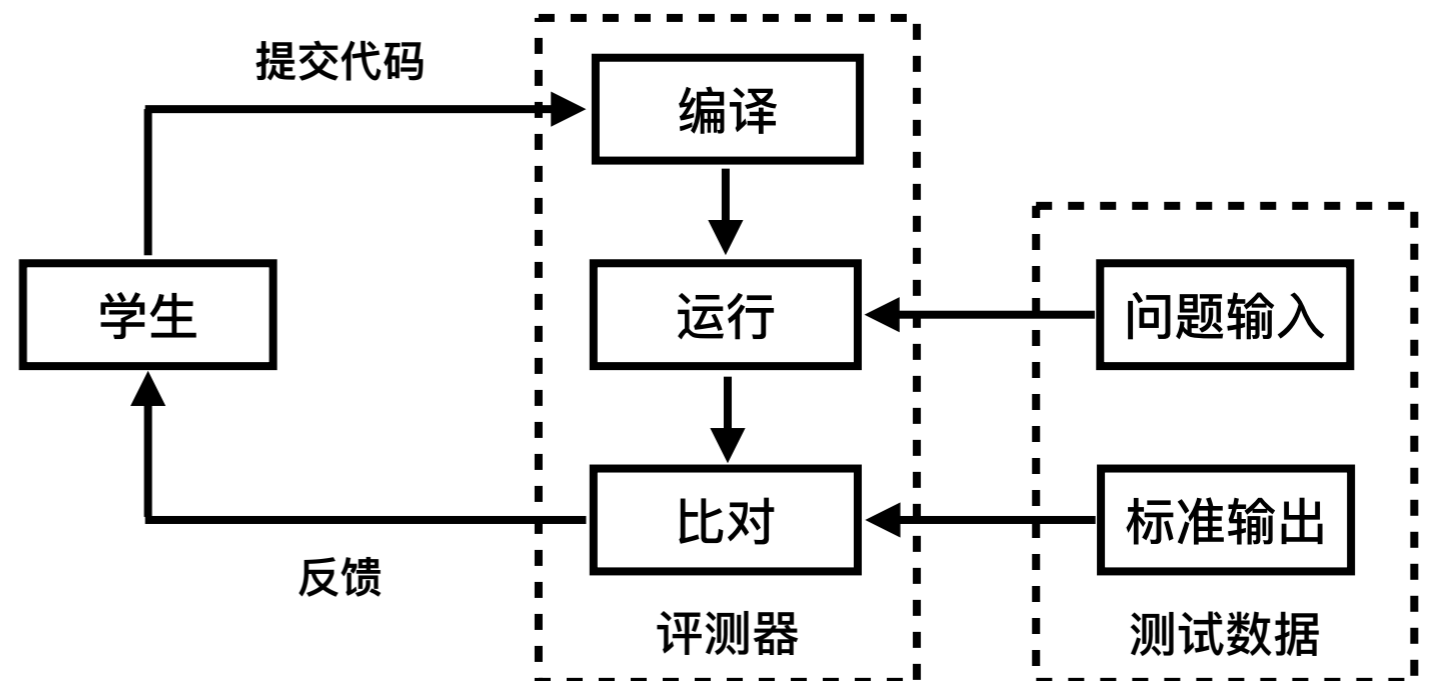
```
5 4
1 2 1
2 3 2
3 4 1
3 5 3

3
```

# CODIA 编程评测

## 评测机制

- 评测器
  - 运行代码，判定结果
  - 规定了不同语言的编译运行方式
  - 可以设定时空限制等
  - 提供反馈：
    - 编译、运行信息
    - 错误类型



# CODIA 编程评测

## 测试组织

- 题包：规定做哪些题目
- 赛制：题包规定完成时间、可参与人员等
- 结果反馈：
  - 即时反馈（类似 ACM）
  - 赛后反馈（类似高中信息学竞赛）





# 目录

- 从传统评测到编程评测
- 编程评测的基本范式
- CODIA 编程评测
- 编程评测研究

# 编程评测研究

## 围绕评价

- 测试数据质量评估和优化
- 测试数据生成
- 对错之外的更多评价标准：
  - 针对代码本身的评价（如代码风格）
  - 编程过程评价（解决问题的 efficiency, systematicity 等）

# 编程评测研究

## 围绕测试

- 自动组卷
- 自适应测试

# 编程评测研究

## 从评测推广

- 评测到诊断/追踪
  - 传统场景下行为和反馈都很少：需要通过序列
  - 从代码出发、从反馈出发：序列是必要的吗？
- 评测到学习：路径推荐、资源交叉推荐

