



结果提交类问题

重庆外语学校 雷环中



结果提交类问题

- ✦ 概念：提供输入数据，提交结果
- ✦ 重要技巧：数据分析
- ✦ 随机化
- ✦ 深刻变革
- ✦ 综合策略
- ✦ 原则：最短时间内得到正确结果

数据分析

[例]Crack the code(Baltic OI 2001)

文件 **cra.out** 与 **cra.txt** 出自同一篇文章。

定义函数 $\text{succ}(C,1)$ 为大写字母 C 的后继

有 $\text{succ}('A',1) = 'B'$, $\text{succ}('B',1) = 'C'$,, $\text{succ}('Z',1) = 'A'$

$\text{succ}(C,0) = C$, $\text{succ}(C,n) = \text{succ}(\text{succ}(C,n-1),1)$, $n \in N$

另有密码整数 $a_1 a_2 \dots a_{10}$, 对 **cra.out** 第 i 个字母 c_i 作转换:

1. 如 c_i 为小写, 则变为大写, 转 2

2. 如 c_i 为大写, 则作变换 $c_i = \text{succ}(c_i, a_{(i-1) \bmod 10 + 1})$

cra.out 转换后的结果即 **cra.in**

你的任务是根据提供的 **cra.in** 和 **cra.txt**, 得出 **cra.out**



数据分析法考虑

我们先看一个数据：

信息文本：

cra1.txt :

ALFRED TARSKI WAS ONE OF THE GREATEST MATHEMATICIANS,
LOGICIANS AND PHILOSOPHERS OF THE PASSING CENTURY, AND
ONE OF THE GREATEST LOGICIANS OF ALL TIME. 后略。

密文：


cra1.in

JK XHLMGXC PP ZEU OWL BM 1939 VP ZEU CRGBSHVOLD IJ
UHGU XUK DYYXTM HPJ IYPIO MGLTK BLYV DBMJG,
GCVCPTTSLF LFHMX LM SOG NXHPECW TUKBBHMMER
ZUF ZEUH,EQMDY 1943, CZ QXY YYBULTYFJS SQ VZSKLLHHML
TS IGXHUFIJ. 后略。

结论：文章是英文

数据分析法考虑

- ✚ 联想思考：许多高频词汇都很短（如 **a**，**I**，**the**，**of** 等）
- ✚ $a_1 \dots a_{10}$ 意义：原文和密文单词对应字母差异
- ✚ 建立对应：

原单词 $A_1 A_2 \dots A_k$  加密后单词 $B_1 B_2 \dots B_k$

则有：

$$B_j = \text{succ}(A_j, a_{(m+j-2) \bmod 10 + 1})$$

A_1 是文章第 m 个字母， $j=1, 2, \dots, k$

- ✚ 方法：猜想 + 试探单词对应关系

关键：充分发挥手工和程序各自的优势

举例

信息文本:

cra1.txt :

ALFRED TARSKI WAS ONE OF THE GREATEST MATHEMATICIANS,
LOGICIANS AND PHILOSOPHERS OF THE PASSING CENTURY, AND
ONE OF THE GREATEST LOGICIANS OF ALL TIME. 后略。

密文:

cra1.in

JK XHLMGXC PP ZEU OWL BM 1939 VP ZEU CRGBSHVOLD IJ
UHGU XUK DYYXTM HPJ IYPIO MGLTK BLYV DBMJG,
GCVCP TTS LFHMX LM SOG NXHPECW TUKBBHMMER
ZUF ZEUEH, EQMDY 1943, CZ QXY YYBULTYFJS SQ VZSKLLHHML
TS IGXHUFIJ. 后略。 **AFTER !**

信息: 文字在描述某人生平

EQMDY 的 'E' 是文章第 116 个字母

AFTER $\xrightarrow{(4,11,19,25,7)}$ **EQMDY**

结果: $a_6=4$, $a_7=11$, $a_8=19$, $a_9=25$, $a_{10}=7$

举例

用 $a_6 \dots a_{10}$ 还原文本得：

JK XHLIVED IP ZEU OSA IN 1939 OP ZEU
CNVITAVOLD IF JOHN XUK DYUMAN APJ
IYPED THETK BLYR SINCG. GCVCLIAF
LFHMT AT TI THE HARVARD UNIVERSITY
ZEUH, AFTER 1943, CZ
THE UNIVERSITY OF CALIFORNIA
AT BGXHUFY. 后略。

于是可得 a_1 , a_2 , a_3 , a_4 , a_5

全文可破

再举一例

cra8.txt

My name is Mary.

cra8.in

I AM A HRLDNB HKUP. : I AM NOT CCG.

1 2 3 4

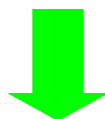


$a_1 a_2 a_3 a_4$

15 16 17 18 19 20



$a_5 a_6 a_7 a_8 a_9 a_{10}$



I AM A CLEVER GIRL. I AM NOT BAD.

经典方法考虑

算法思路：每种字母出现频率不同→频率比较

用 $\mathbf{P}=(P_A \dots P_Z)$ 表示未加密文本中每个字母出现的相对频率

用 $\mathbf{Q}=(Q_A \dots Q_Z)$ 表示密文中第 $i+10*(N-1)$ 位字母组成集合中

每个字母出现的相对频率（用以确定 a_i ）

比较 \mathbf{P} 和 \mathbf{Q} 的相似程度 ---- 估价函数的选择

$$\text{方法一: } f(P, H) = \sum_{ch='A'}^{Z'} |P_{ch} - H_{ch}|$$

$$\text{方法二: } f(P, H) = \sum_{ch='A'}^{Z'} (P_{ch} - H_{ch})^2$$

$$\text{方法三: } f(P, H) = \sum_{ch='A'}^{Z'} \frac{(P_{ch} - H_{ch})^2}{P_{ch}} \quad (\text{分母为 } 0 \text{ 的修正})$$

方法比较

表中数字为该数据算对的密码位数，前五组为官方数据

	一	二	三	四	五	六	七	八	九	总分
方法一	10	10	10	10	8	8	5	1	1	63
方法二	10	10	10	10	8	8	3	1	1	61
方法三	10	10	10	10	10	7	5	2	1	65
官方程序	10	10	10	10	10	6	4	0	1	61
数据分析	10	10	10	10	10	10	10	10	0	80

结论: 1. 以上算法均不理想
2. 程序方法差异不大
3. 结果提交的原因

随着样本容量的减小，正确率显著降低

3. 对于本题结果提交特点，数据分析法相对最好！



数据分析法小结

- 长处：具体问题具体分析，易于实现
短处：不利于处理规模数据
- 关键：观察数据特点
- 精髓：手工与程序运算相协调
- 目的：在最短的时间内得到正确结果



随机化

不同的随机化：

经典随机化

需在规定时间内出解
评测时只运行一次

结果提交类问题的随机化

没有任何限制

目标只有一个：得到正确结果！



得到算法最有利因素，避开不利因素



随机化

例： **Tetris(NOI 2002)**

题目大意：

给出一个初始俄罗斯方块的棋盘状态，
每次可从标准的 **19** 种形状中任选一种，
放到任意位置上，要求在 **100000** 步以内
将棋盘消空。输入数据保证有解。

限制： 棋盘永远不能悬空

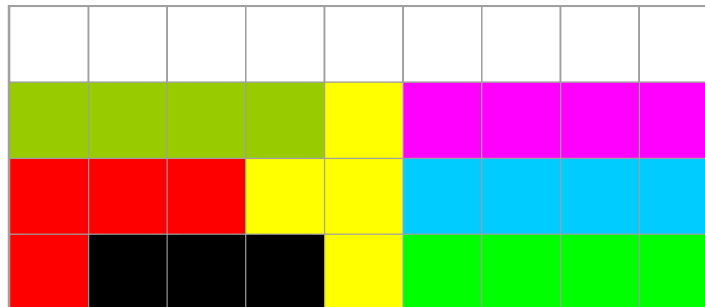
放置不能超出边界

分析测试数据

Tetris1.in

9

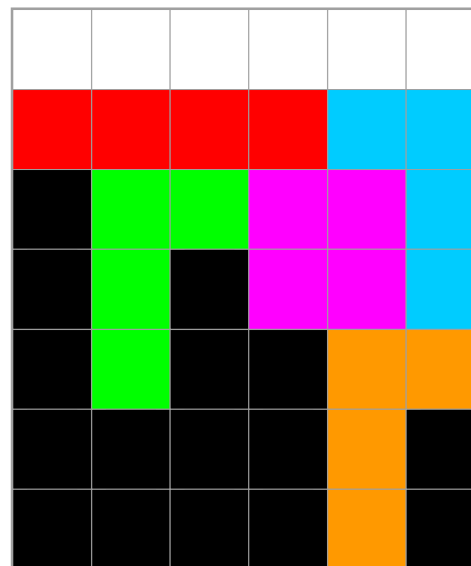
0 1 1 1 0 0 0 0 0



Tetris2.in

6

5 2 4 3 0 2



分析测试数据

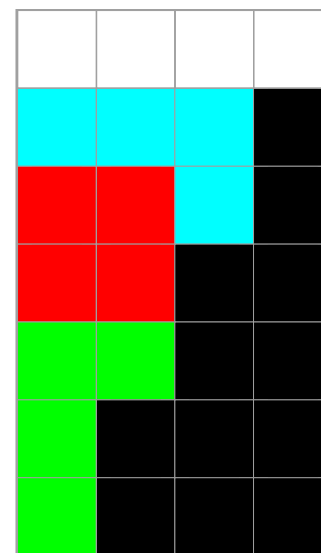
Tetris3.in

200

0 2 4 6 6 8 10 12 12 14 16 18 18 20

22 24 24 26 28 30 30 32 34 36 后略。

数据以四列为单位规律出现。
最基本的 **(0,2,4,6)** 容易处理：



本数据只需一个小程序即可解决



分析测试数据

下面是后面几组数据的规模：（数据略）

Tetris4.in N=16

Tetris5.in N=47

Tetris6.in N=97

Tetris7.in N=100

Tetris8.in N=246

Tetris9.in N=574

Tetris10.in N=1202

综合分析

数据一、二：规模很小，手工能迅速出解

$$N \leq 9$$

数据三：规模较大，规律明显，只需

$$N = 200 \quad \text{一个短程序}$$

数据四、五：规模不小，可手算，但需一

$$N < 50 \quad \text{定时间}$$

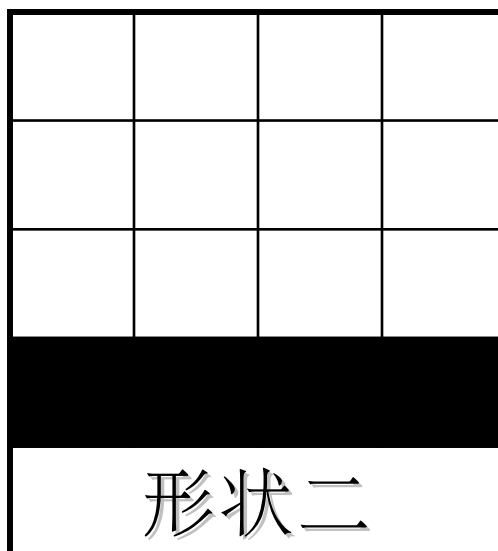
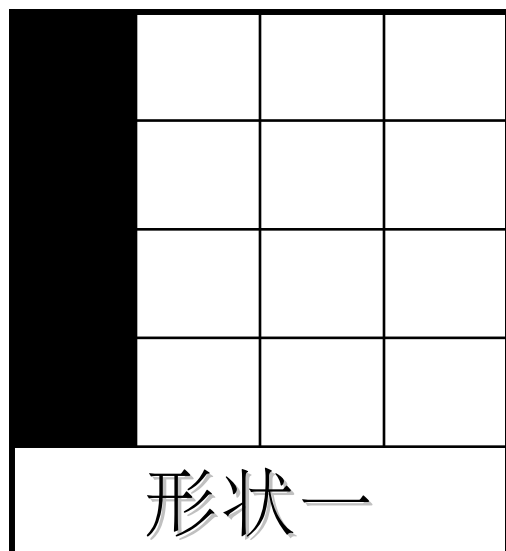
数据六...十：规模较大，规律很不明显或

$$N \geq 97 \quad \text{无规律，手工无法胜任}$$

这其实也是一种数据分析

随机化

两种形状:



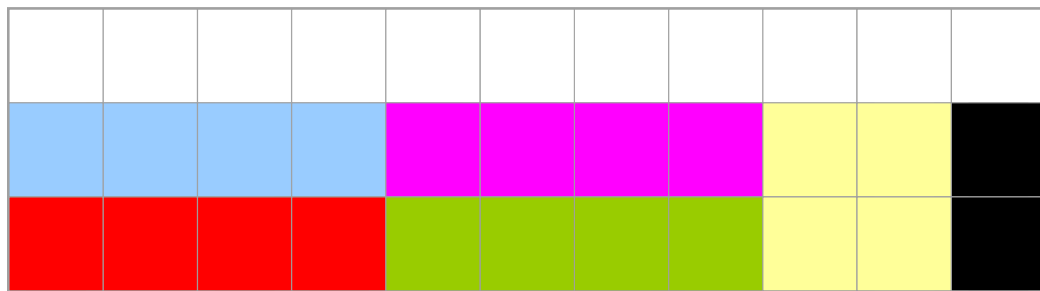
对于任一初状态，可仅用这两种形状使棋盘高度不超过 **3**
然后从左到右找“空”，随机用一个合法的形状填上此空，
直到将棋盘完全弄平

此算法很粗劣，显得过于随机

应在随机的基础上加入一些人的思想

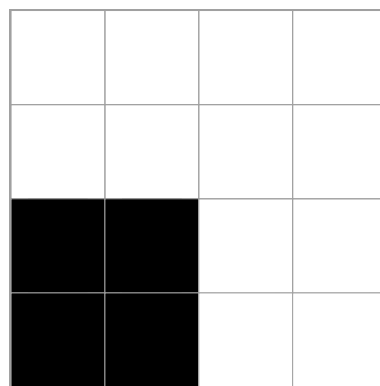
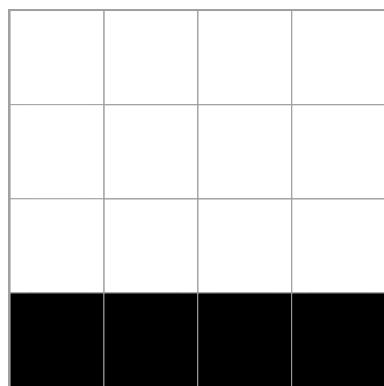
优化一

这样虽合法，但对大数据而言步数太多。其实只需：



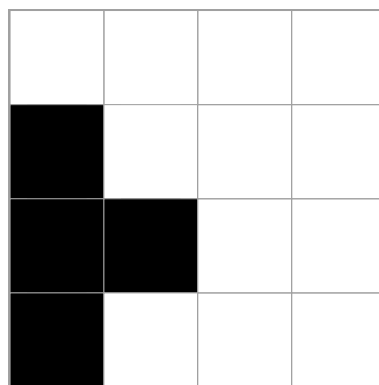
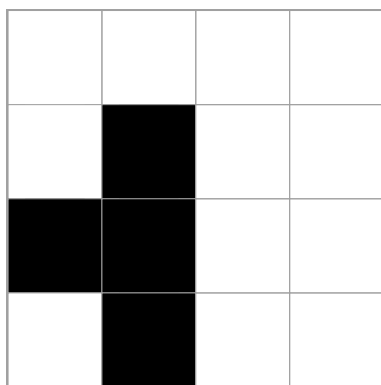
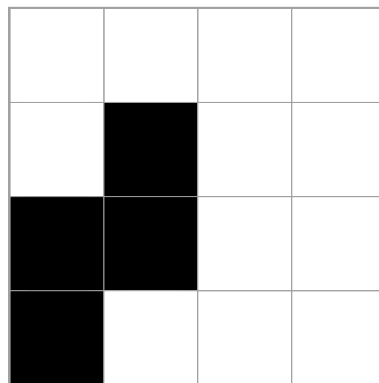
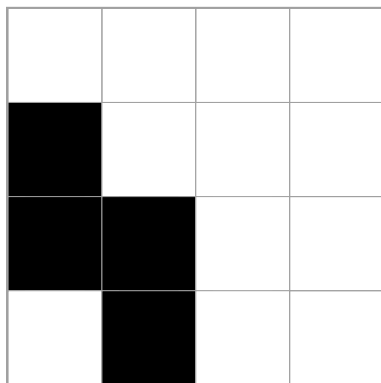
结论：

应尽量使用下面两种形状，以使棋盘尽量显得平整



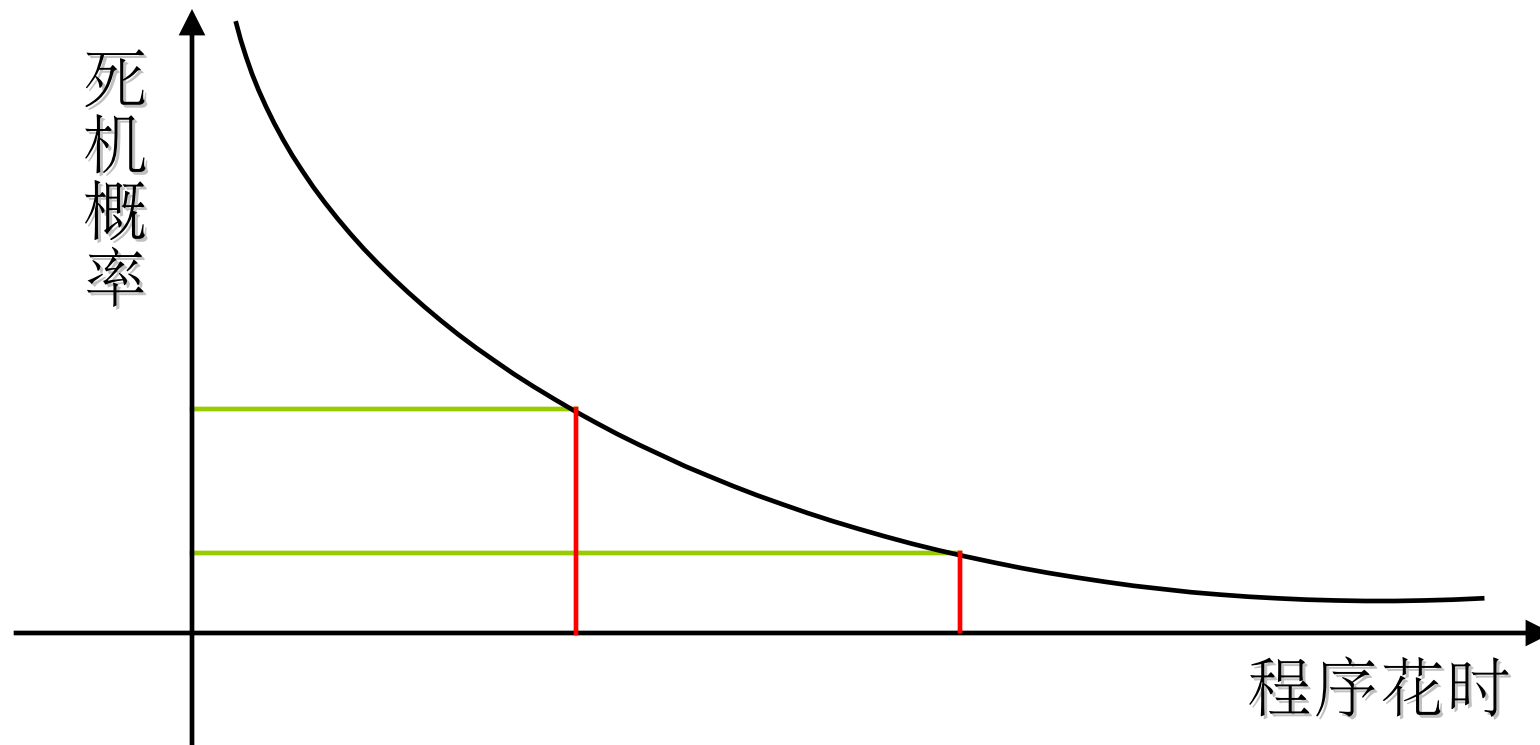
优化二

应尽量避免下面四种形状，以减少突兀



类似的优化还有很多.....

随机化



得到算法最有利因素
避开算法不利因素



经典方法

算法一：贪心

每一步都使棋盘变得更平
直到最后完全平整

算法二：构造

以 **4** 列为单位对棋盘分组

如最后不足 **4** 列，则独为一组

先填平 **4** 列组

再用形状 **2** （横四）结合贪心填平整整个棋盘

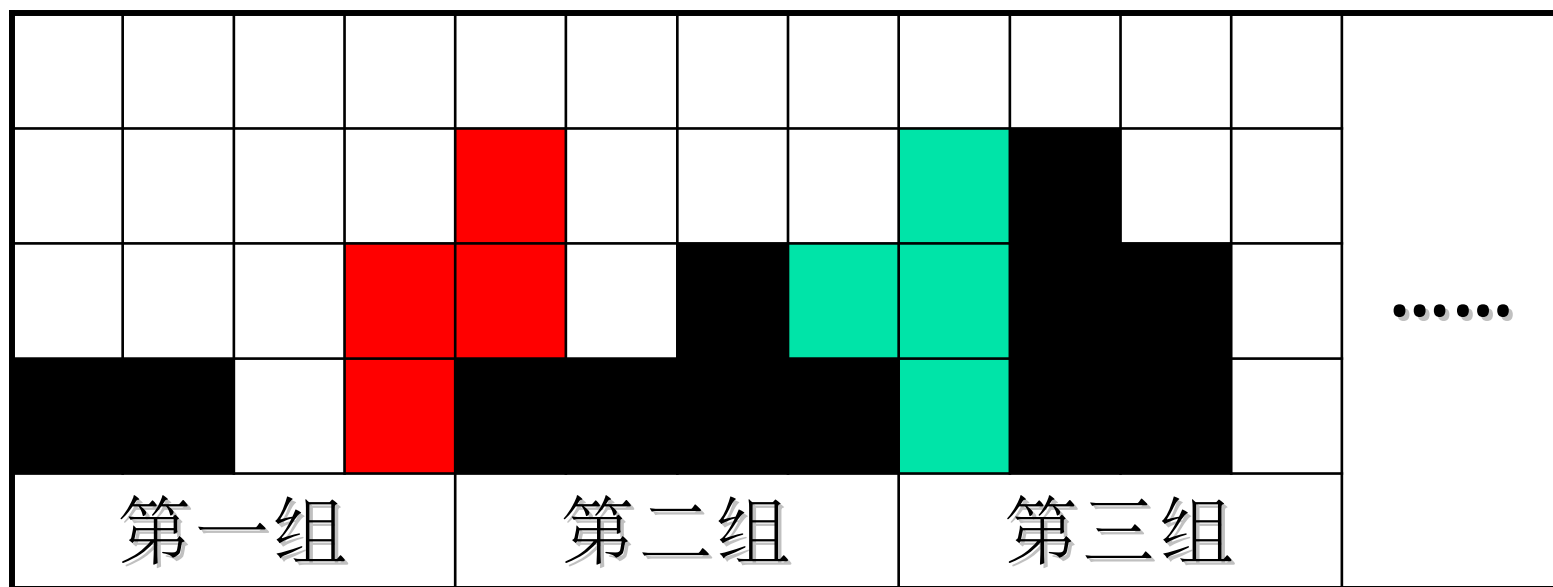
注意组间协调，保证可行性





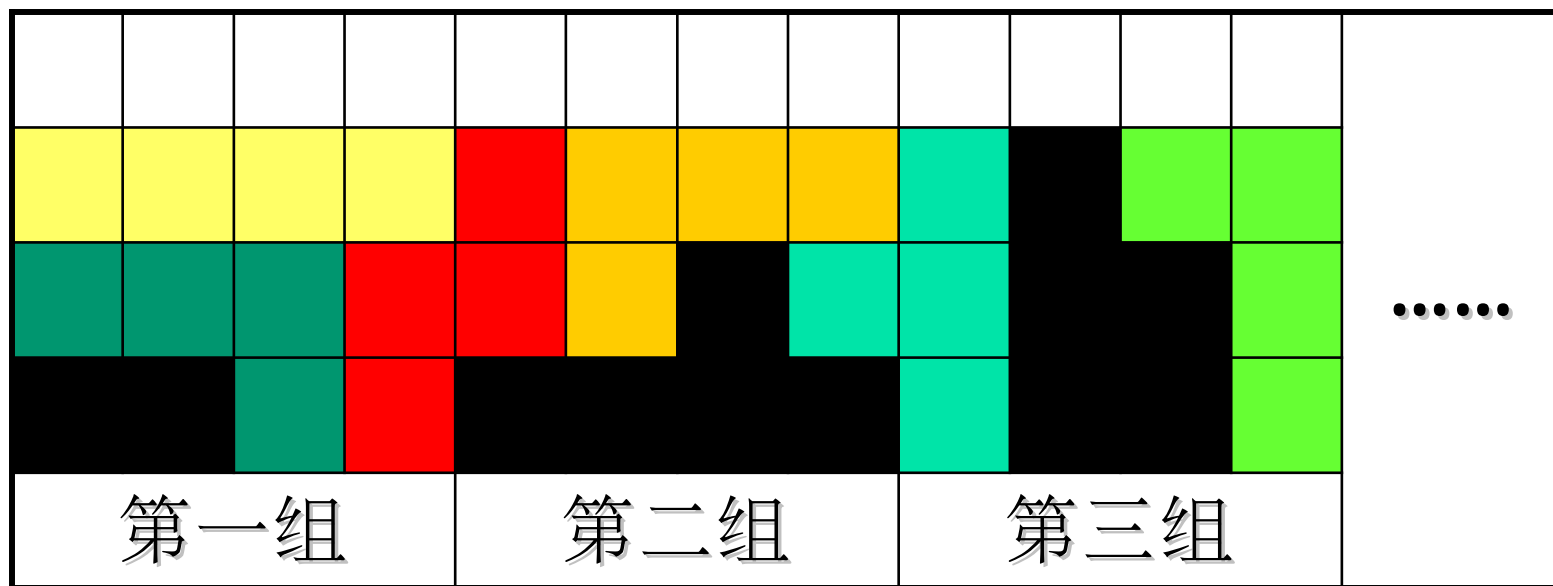
算法二

通过下图的组间调节使它们都变成 **4** 的倍数
调节后三组的块数分为 **4** 、 **8** 、 **8**



算法二

之后每组便很易填平
如下图所示：



算法比较

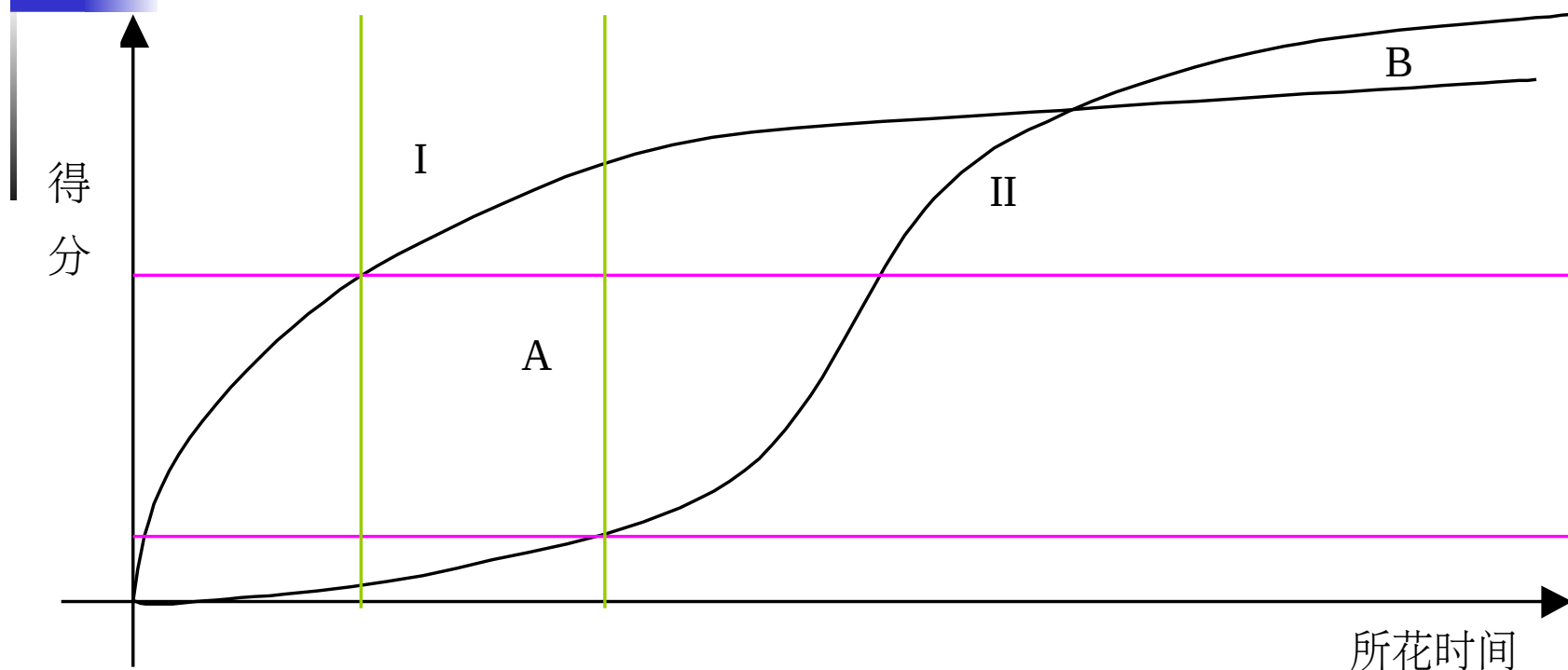
	时间复杂度	空间复杂度	运行速度	算法实现复杂度	解的质量
随机化算法	$O(M)$	$O(N)$	$< 1s$	低	较差
算法一	$O(M \times N)$	$O(N)$	$< 10s$	低	较好
结论：对结果提交，随机化算法和算法一最合适 对非结果提交，算法二最好					较好
不同要求下我们的算法可能不同 因为还要综合考虑算法实现的复杂度					较好

深刻变革

- 大大扩展了程序的可用空间
以至没有严格限制
- 时限更自由，甚至可达几个小时
- 多线程方法崭露头角
- 测试数据作为题目的组成部分出现

.....

综合策略



结果提交类问题

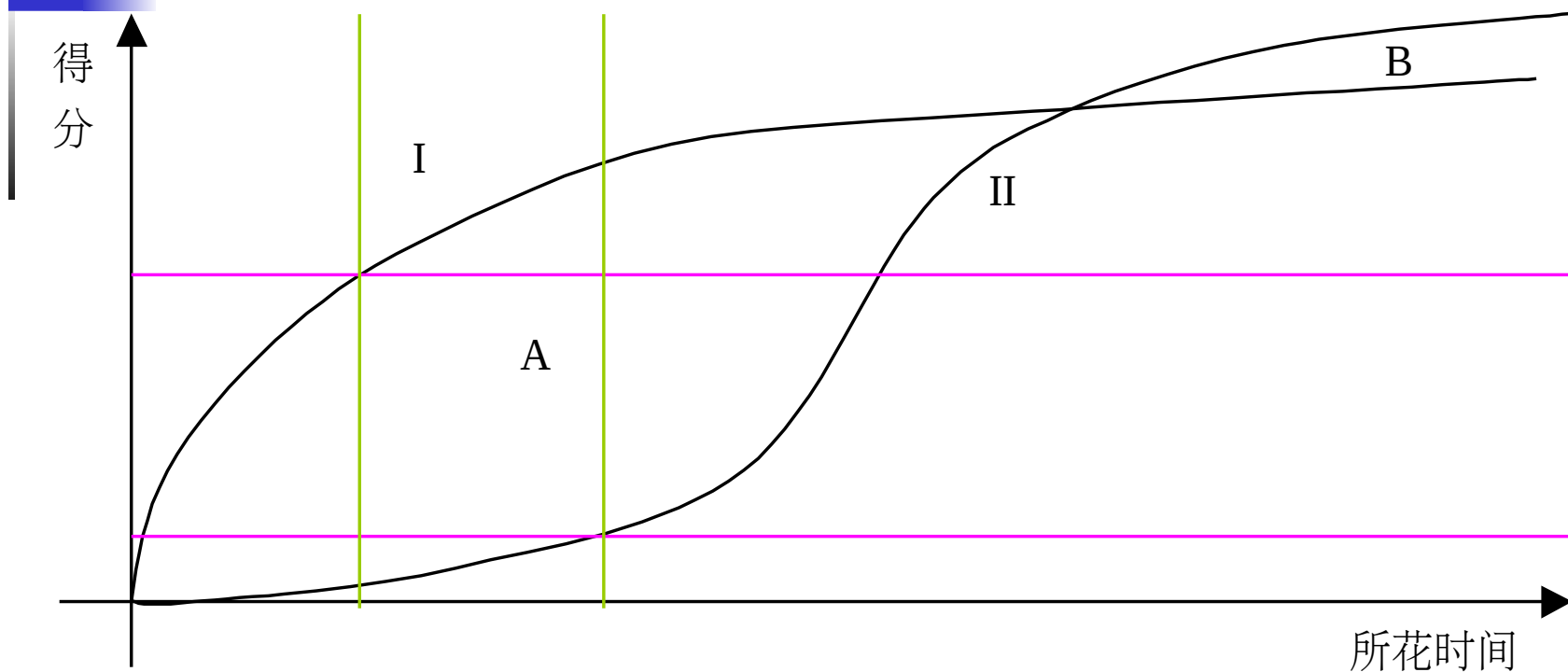
飞流直下 ---- 曲线开始斜率较大，后逐渐减小

经典问题

厚积薄发 ---- 曲线始末斜率较小，中间阶段斜率较大

结果提交 (I) 与经典 (II) 问题完成题目时间与得分关系图

综合策略



区别一：

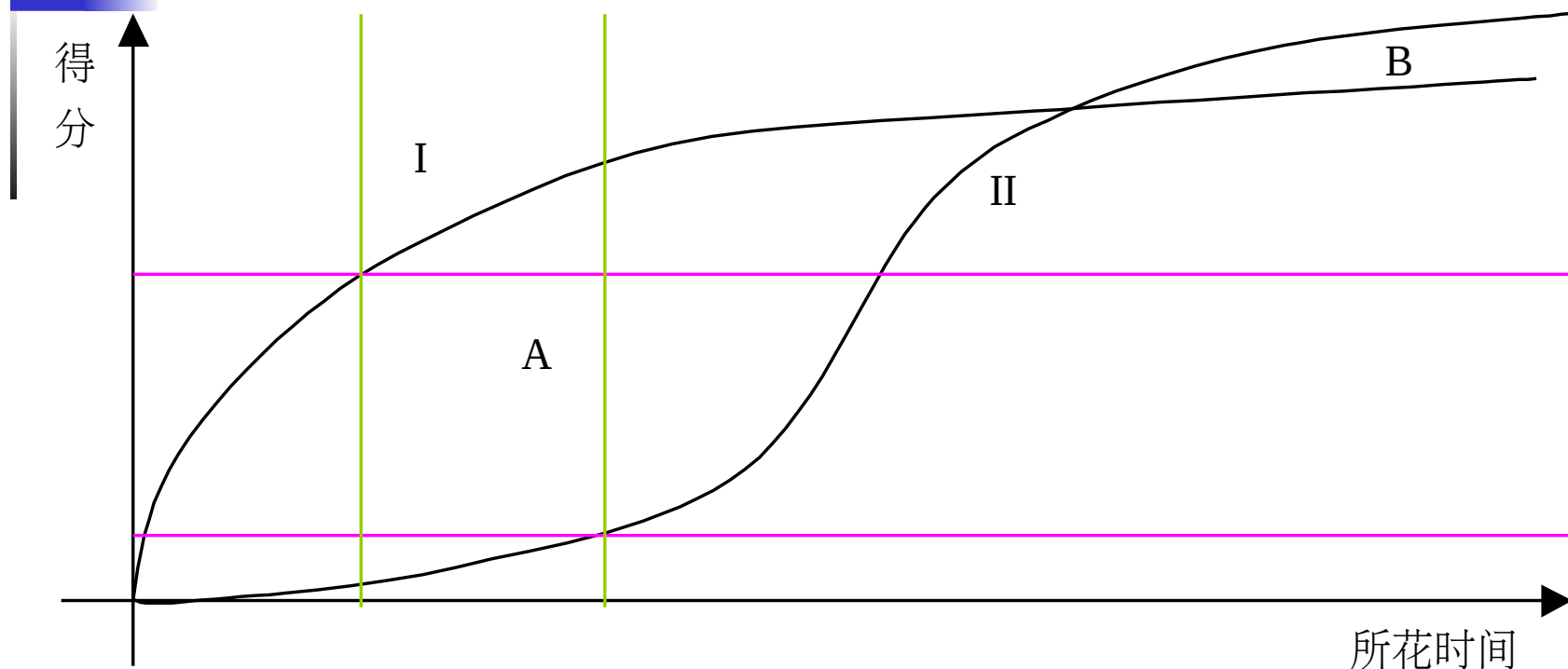
经典问题中常不知何时“适可而止”

算法上得到提高，策略上并不明智

结果提交问题中对能得分数有较准估计，不会多花无谓时间

结果提交 (I) 与经典 (II) 问题完成题目时间与得分关系图

综合策略

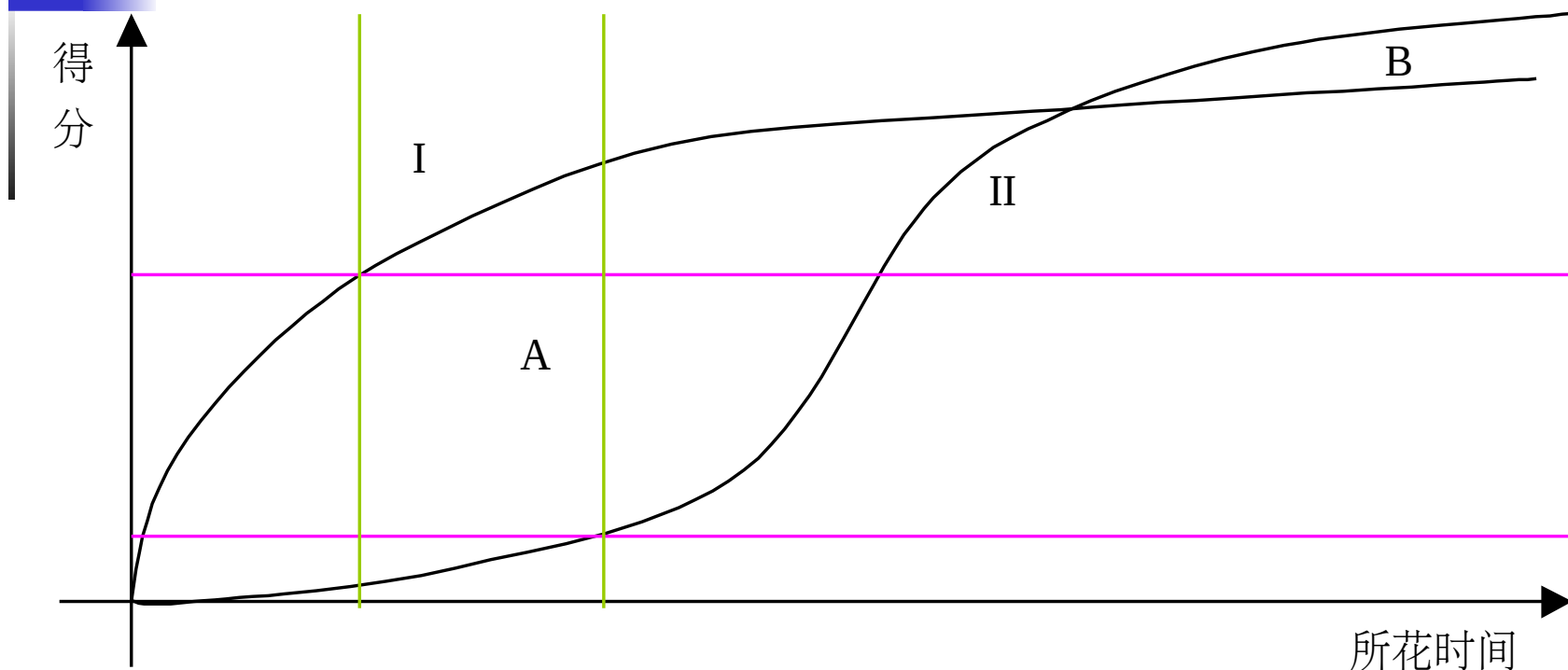


区别二：

两种方法程序效率（得分）都在某段较短时间内迅速上升而这段时间在结果提交问题中显然要相对提前很多

结果提交 (I) 与经典 (II) 问题完成题目时间与得分关系图

综合策略



区域 **A**：曲线 **I** 速度快，效率（得分）高，应把握

的关键
区域 **B**：结果提交类问题不再占优 ---- 优势的相对性

对两种问题应舍其短就其长

关键：获得 得分 / 时间 的最佳值



总结

永远提醒自己：

你所需要的

只是在最短时间内得到正确结果
而不是过程



结果提交类问题

谢谢大家！