D. 形式语言与自动机

命题人：清华大学 陈鸿基
验题人：清华大学 张艺缤
清华大学 迟凯文

题目大意

- 给定一个左括号和右括号数量相等的字符串 s
- 对于 $0 \le l < r < |s|$，由 (l, r) 唯一确定一种划分 s 的方案：
  - 将 s 划分为 $s_1 \cdots s_l$，$s_{l+1} \cdots s_r$ 和 $s_{r+1} \cdots s_{|s|}$ 三部分
  - 如果 $s_1 \cdots s_l s_{r+1} \cdots s_{|s|} s_{l+1} \cdots s_r$ 是合法的括号序列，则称这个划分方案是合法的
- 求有多少种本质不同的合法的划分方案
- $1 \le |s| \le 10^7$

- ~~还有清华食堂的香锅的确很好吃~~

分析

- 记 $f(s)$ 表示一个串中左括号的数量减去右括号的数量，那么一个串 t 是合法的括号序列当且仅当其每个前缀对应的 f 都非负，且整个串 $f(t) = 0$
  - ~~如果想严格证明，可以考虑选修形式语言与自动机~~
  - ~~相信能来到决赛的选手应该都知道这个事实~~
- 把题目给出的限制分解给每一段
  - 对于 $s_1 \cdots s_l$：它必须是一个合法的“括号序列前缀”
  - 对于 $s_{l+1} \cdots s_r$：反向考虑，它必须是一个合法的“括号序列后缀”
  - 对于 $s_{r+1} \cdots s_n$：$s_1 \cdots s_l s_{r+1} \cdots s_n$ 是一个合法的“括号序列前缀”
- 可以证明，划分方式是合法的当且仅当其满足上述要求

分析

- 把题目给出的限制分解给每一段
  - 对于 $s_1 \cdots s_l$： 它必须是一个合法的“括号序列前缀” $\Rightarrow l$ 有限制 $l \le L_0$
  - 对于 $s_{l+1} \cdots s_r$： 反向考虑，它必须是一个合法的“括号序列后缀”
  - 对于 $s_{r+1} \cdots s_n$： $s_1 \cdots s_l s_{r+1} \cdots s_n$ 是一个合法的“括号序列前缀”
- 在什么条件下 $s_{l+1} \cdots s_r$是一个合法的“括号序列后缀”？
- 对任意 $i = l + 1, \cdots, r$， $f(s_i \cdots s_r) \le 0$
- 也就是说 $f(s_1 \cdots s_r) \le f(s_1 \cdots s_{i-1})$
- 记 $S(i) = f(s_1 \cdots s_i)$， l 有限制 $l \ge \min\{i | \forall i \le j \le r, S(i) \ge S(r)\}$
- 这一限制与 r 有关，不妨记做 $L(r)$

分析

- 把题目给出的限制分解给每一段
  - 对于 $s_1 \cdots s_l$：它必须是一个合法的“括号序列前缀” $\Rightarrow l \le L_0$
  - 对于 $s_{l+1} \cdots s_r$：它必须是一个合法的“括号序列后缀” $\Rightarrow l \ge L(r)$
  - 对于 $s_{r+1} \cdots s_n$：$s_1 \cdots s_l s_{r+1} \cdots s_n$ 是一个合法的“括号序列前缀”
- 在什么条件下 $s_1 \cdots s_l s_{r+1} \cdots s_n$ 是一个合法的“括号序列前缀”？
- 假设已知 $s_1 \cdots s_l$ 是合法前缀，则只需满足对 $i = r+1, \cdots, n$，$f(s_1 \cdots s_l s_{r+1} \cdots s_i) = f(s_1 \cdots s_l) + f(s_{r+1} \cdots s_i) \ge 0$
- 可以对每个 r 求出 $d(r) = \min_i f(s_{r+1} \cdots s_i)$，则只需使 $f(s_1 \cdots s_l) + d(r) = S(l) + d(r) \ge 0$

命题人放过的做法

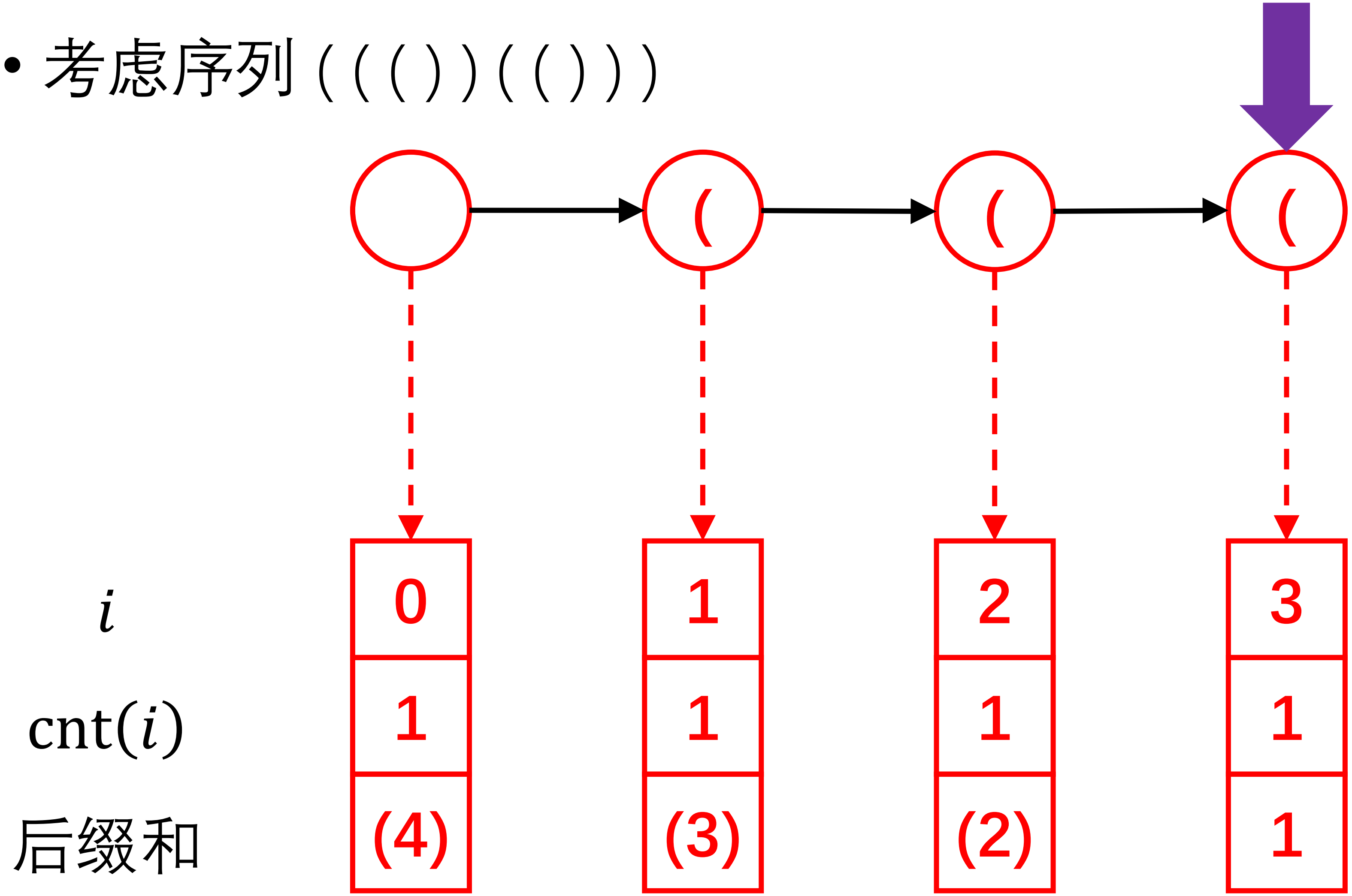
- 把题目给出的限制分解给每一段
  - 对于 $s_1 \cdots s_l$：它必须是一个合法的“括号序列前缀” $\Rightarrow l \le L_0$
  - 对于 $s_{l+1} \cdots s_r$：它必须是一个合法的“括号序列后缀” $\Rightarrow l \ge L(r)$
  - 对于 $s_{r+1} \cdots s_n$：$s_1 \cdots s_l s_{r+1} \cdots s_n$ 是一个合法的“括号序列前缀” $\Rightarrow S(l) + d(r) \ge 0$
- 相当于平面上有一堆点 $\big(l, S(l)\big)$，对每个 r 统计在 $[L(r), \min\{L_0, r-1\}] \times [-d(r), +\infty)$ 中的点的个数
- 二维数点！
- 只要你会用树状数组来统计，并且你的常数和验题人一样小（其中包括不少根据题目性质的优化），那你就能通过本题

复杂度更优秀的做法

- 虽然拿树状数组来维护可以通过本题，但毕竟复杂度是 $O(|s|\log|s|)$
- 这里为喜欢线性做法的选手提供其中一种将复杂度优化到线性的算法

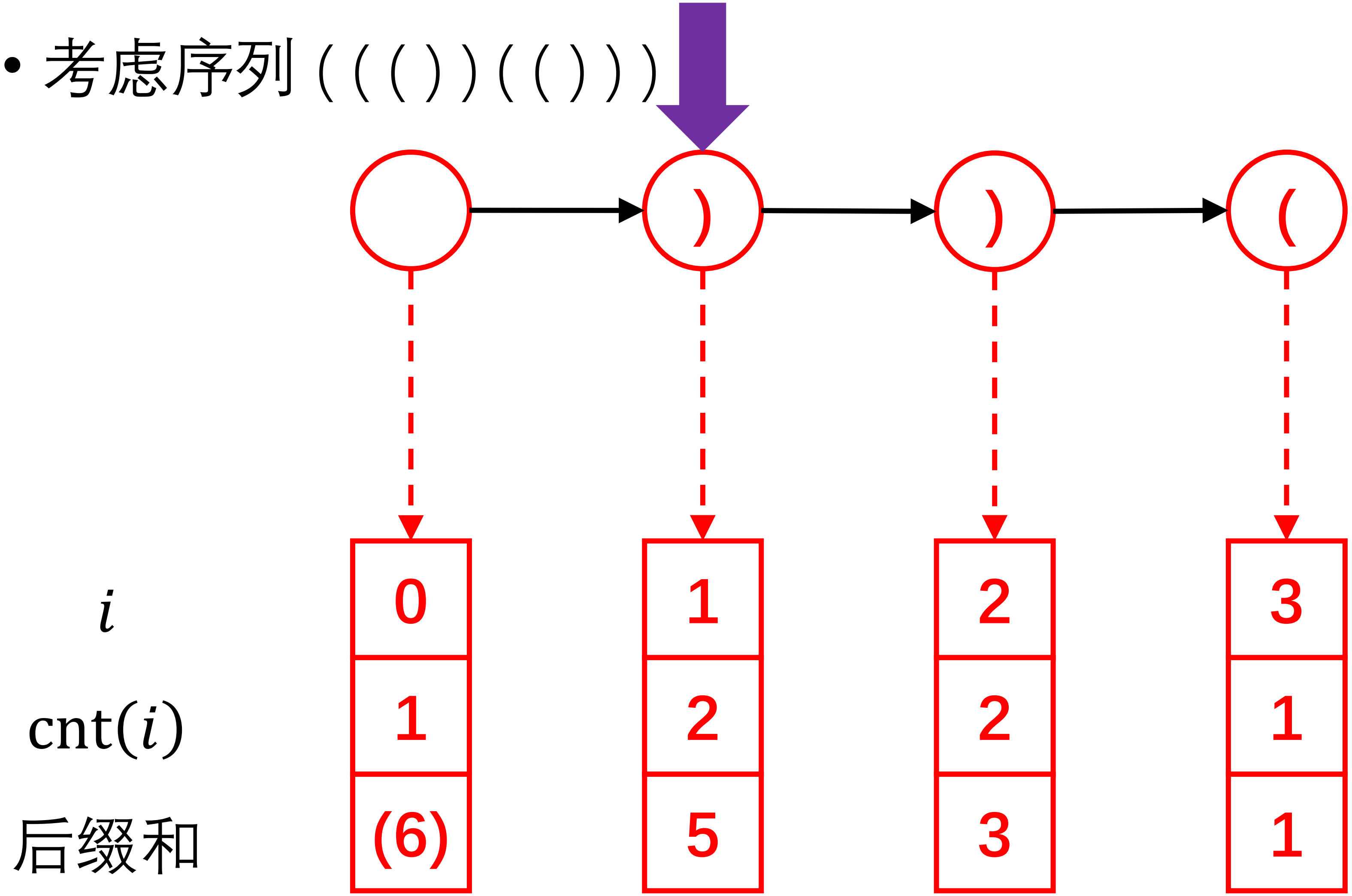
复杂度更优秀的做法

- 考虑序列 ((( )) (( )))



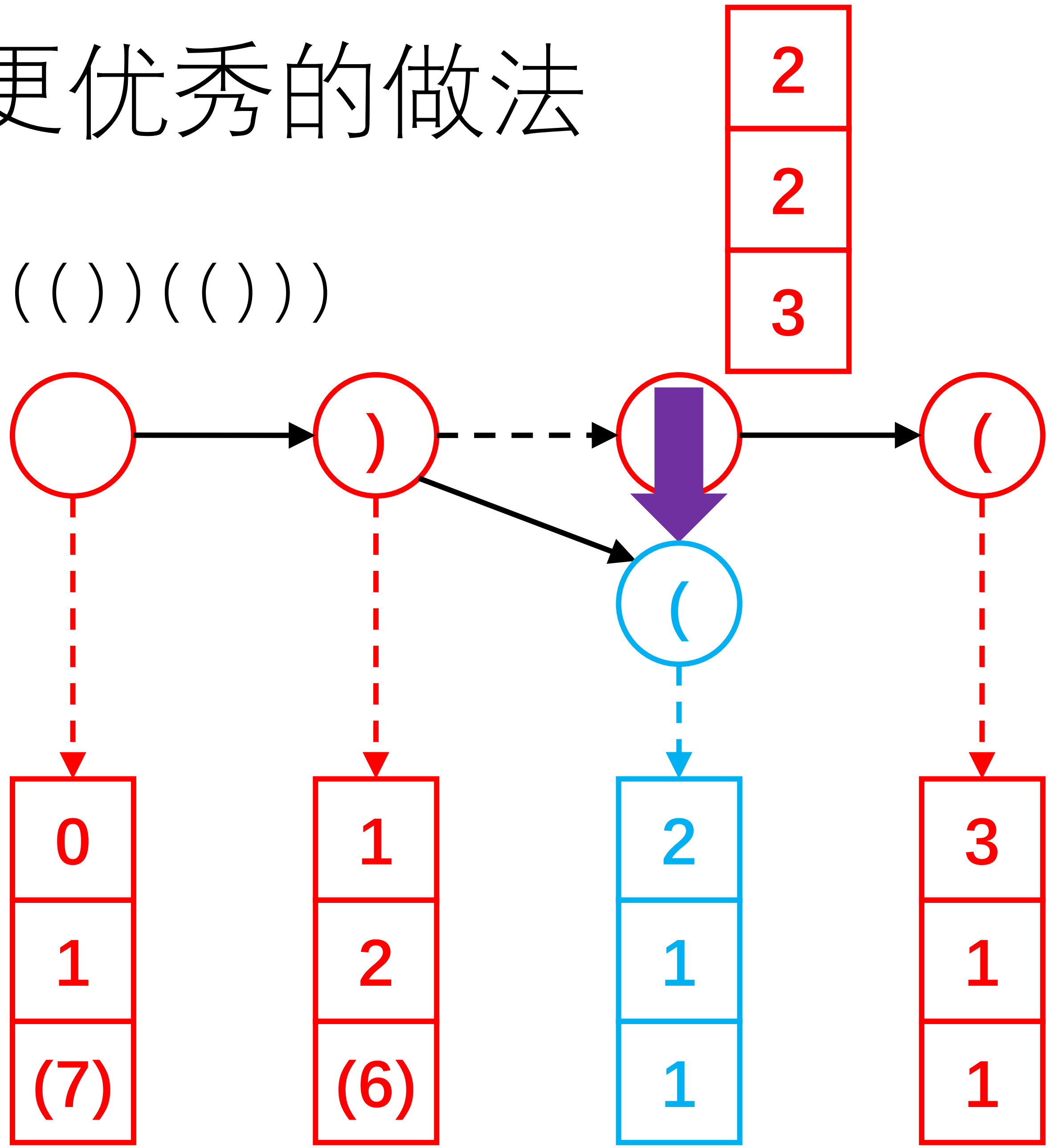
复杂度更优秀的做法

- 考虑序列 ((( )) (( )))



复杂度更优秀的做法

- 考虑序列 ((( )) (( )))



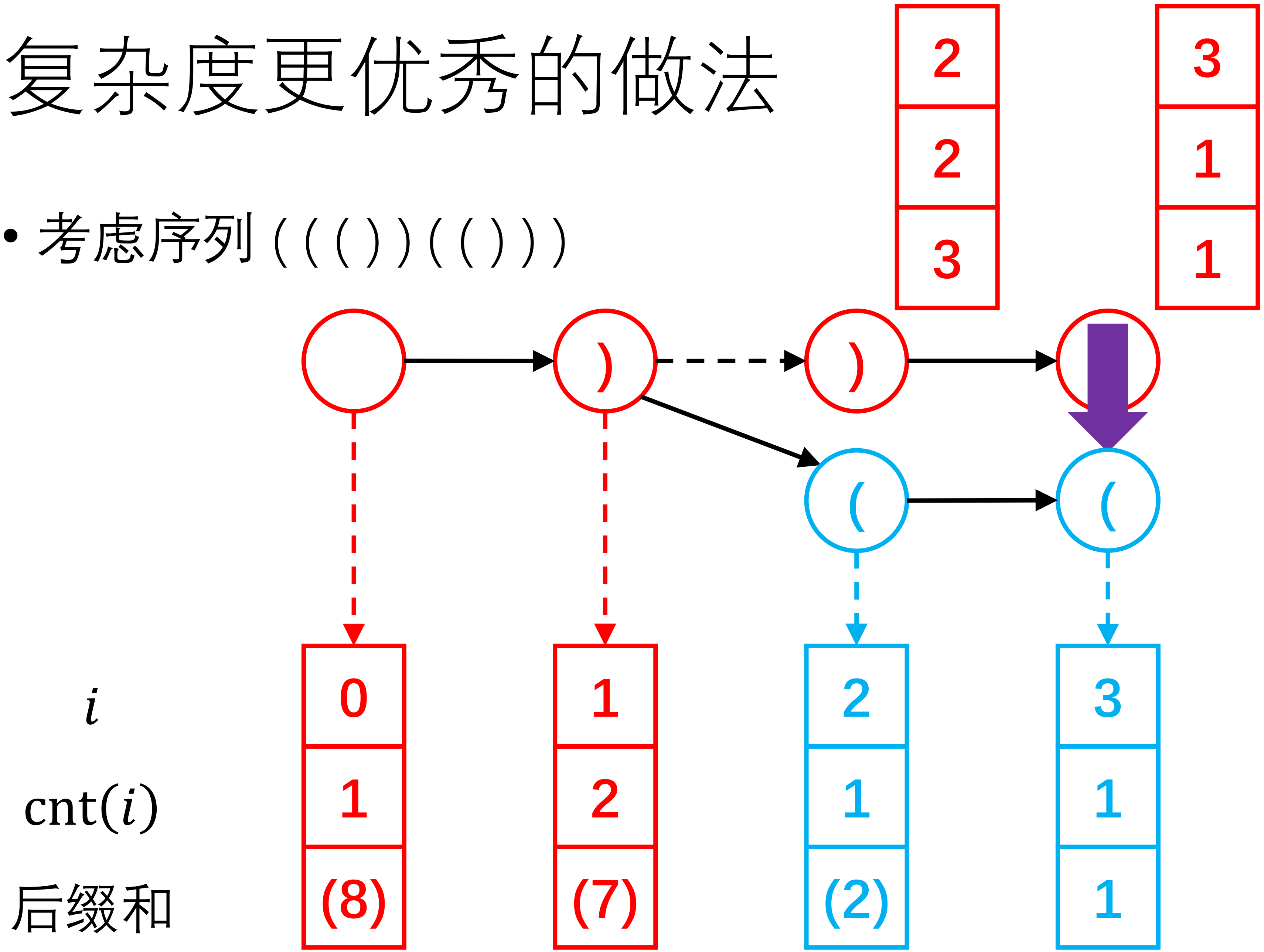
i

$\mathrm{cnt}(i)$

后缀和

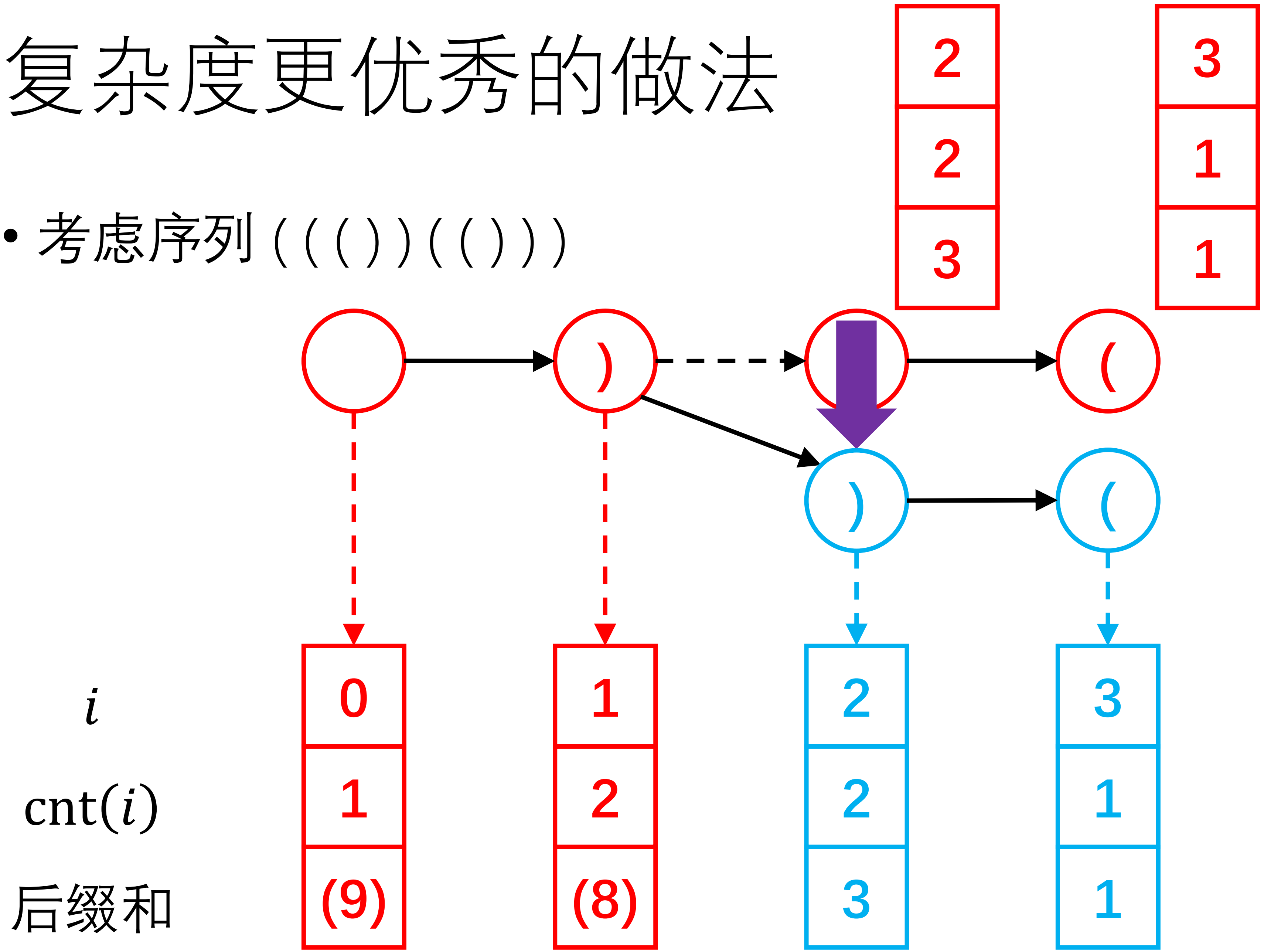
复杂度更优秀的做法

- 考虑序列 ((( )) (( )))



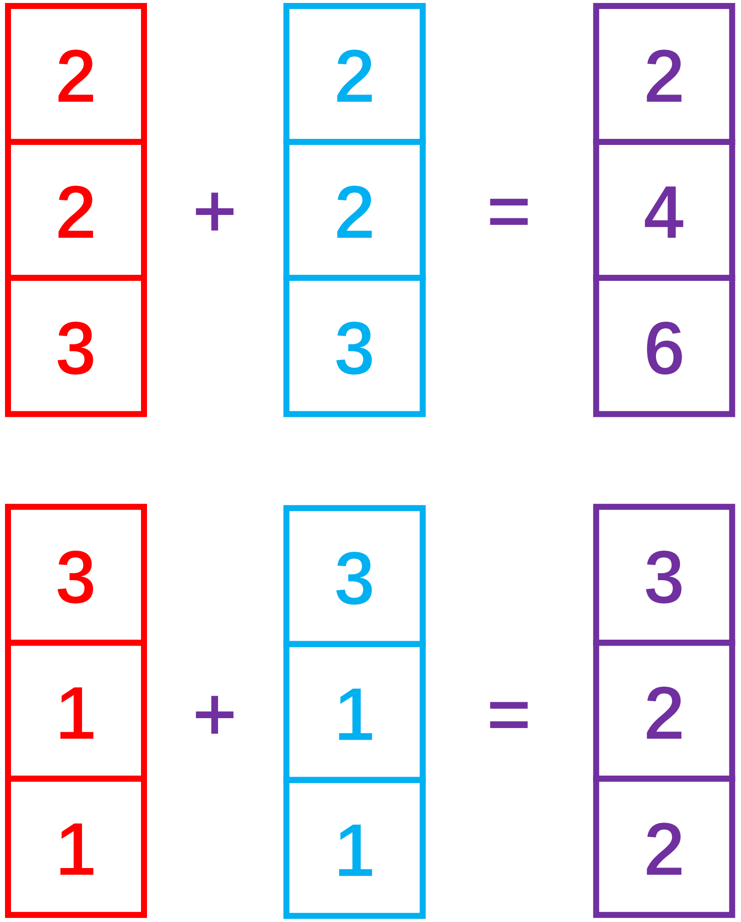
复杂度更优秀的做法

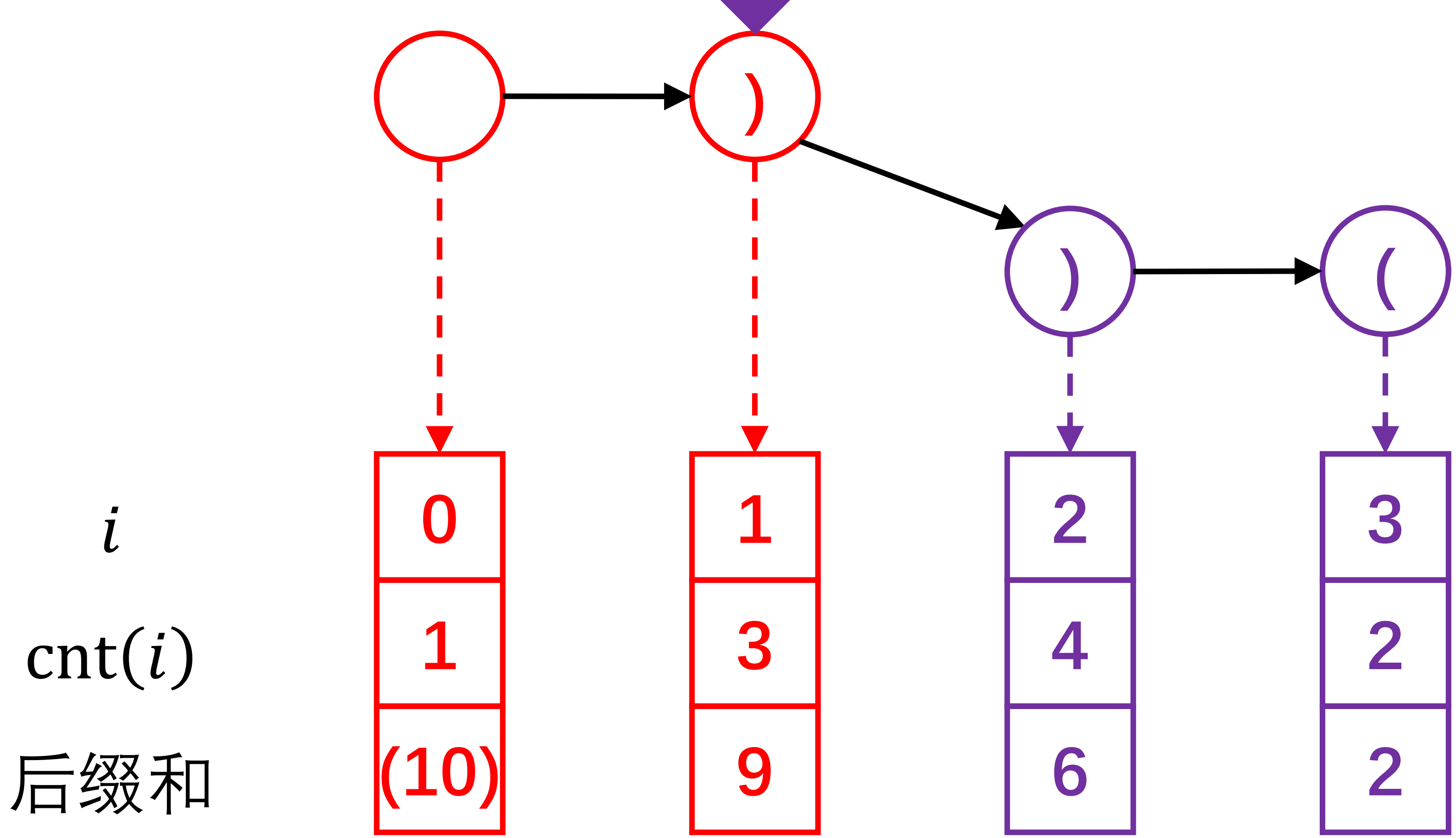
- 考虑序列 ((( )) (( )))



复杂度更优秀的做法

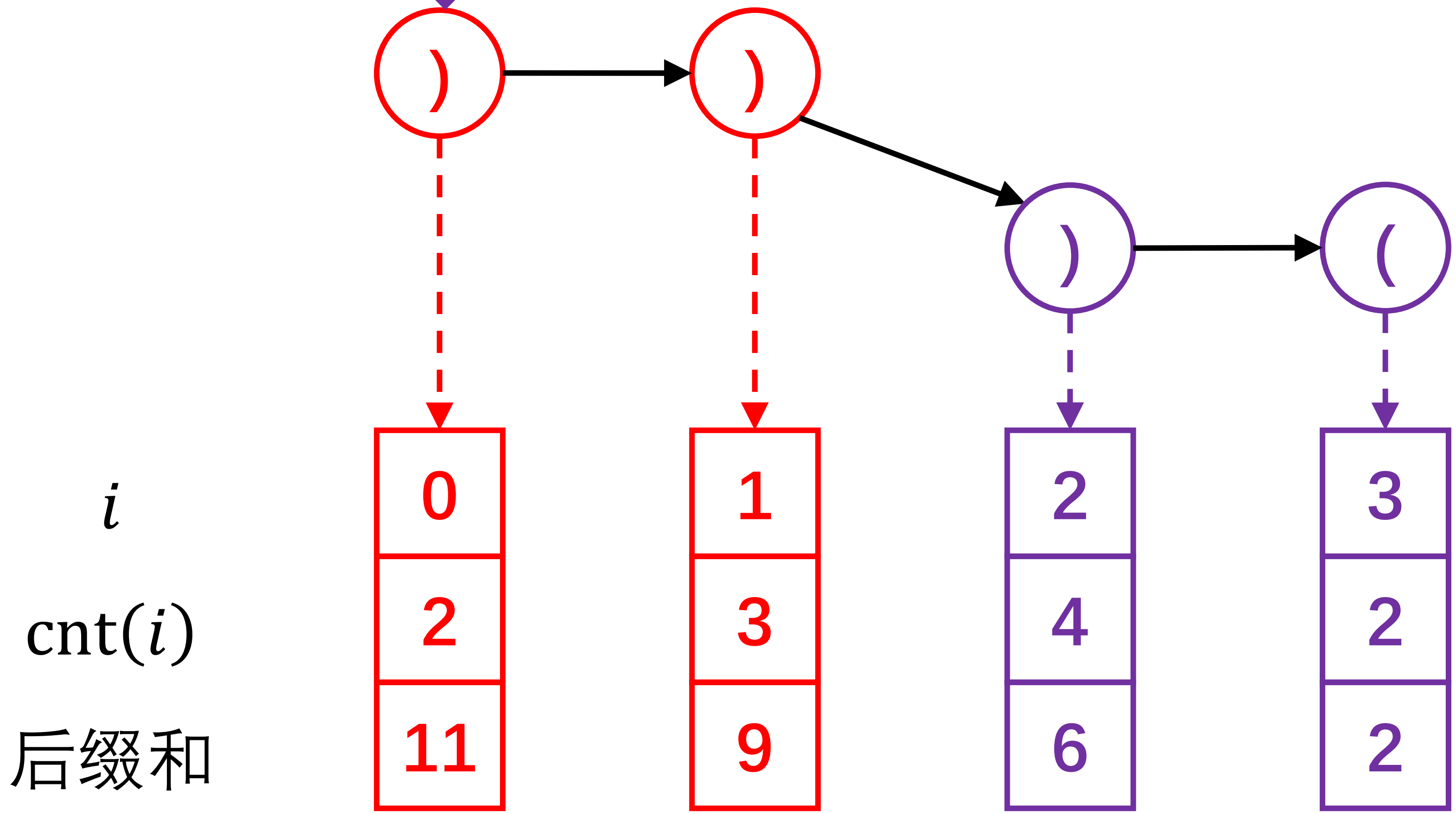
- 考虑序列 ((( )) (( )))





复杂度更优秀的做法

- 考虑序列 ((( ) (( )))



复杂度更优秀的做法

- 虽然拿树状数组来维护可以通过本题，但毕竟复杂度是 $O(|s|\log|s|)$
- 这里为喜欢线性做法的选手提供其中一种将复杂度优化到线性的算法
- 参考刚才的演示，我们可以通过维护一些分支链得到一个线性的做法
- 由于树状数组实在是太快了，同时考虑到本场比赛总体难度比较高，就放实现优秀的树状数组过了