



浅谈数据的合理组织

四川省绵阳南山中学 何森



引子

题目越来越难——数据关系越来越复杂！



对组织数据的要求越来越高！



合理组织在解题中越来越重要！

吉祥如意





金明的预算方案

【题意描述】

给出 N 个物品，每个物品都有一个权值 (<50000) 和一个价格 (<10000)。我们称可以直接被购买的物品为主件，称不能被直接购买的物品为附件，附件只有当其主件被购买了才能被购买，一个主件最多有两个附件，附件没有下一级附件。

【任务】

用不超过 M 元钱，购买一些物品，使得被购买的物品的总权值最大。

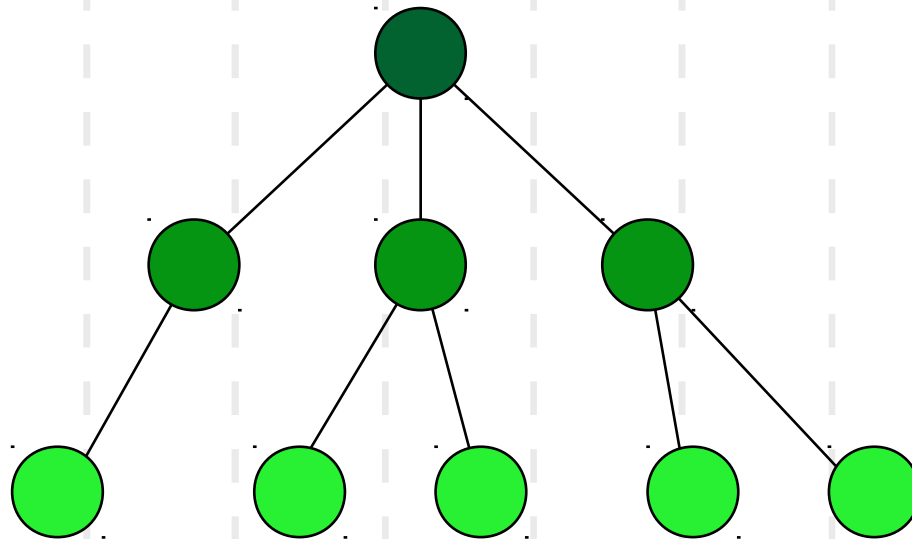
【数据规模】

$N < 60$ $M < 3200$



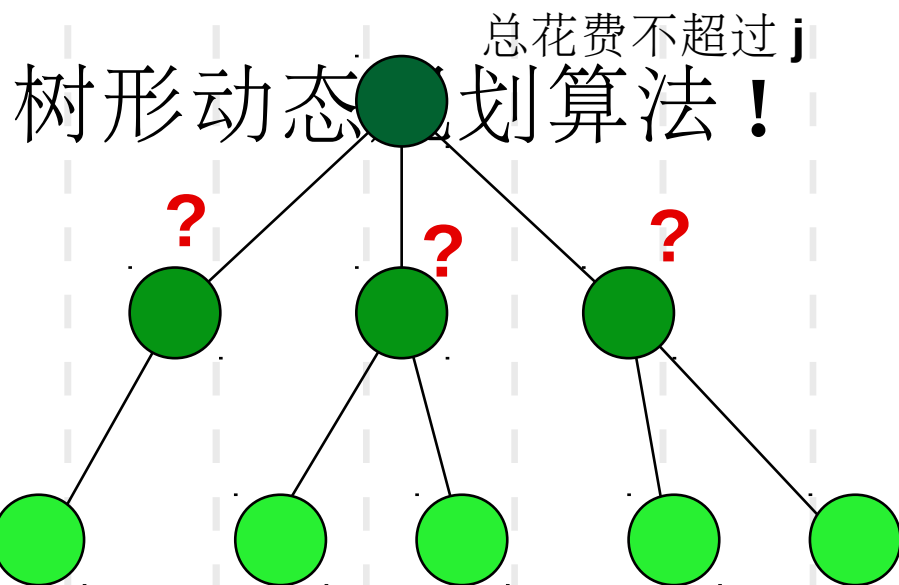
数据的初步组织

题目中给出的主件与附件间形成树形结构，而所有的物品间形成森林结构。为了方便起见，我们给所有的主件都加上一个“上级主件”，这样，所有的物品形成了一棵树。



算法 1

状态 $F[i][j]$ 表示给以 i 为根的子树，总共花费不超过 j 元，所能取得的最大权值和。



枚举量太大，效率不高！



用左儿子右兄弟表示法来表示这一棵树！

状态 **$F[i][j]$** 表示以 **i** 为根的子树总共花费 **j** 元能获得的最大权值和。
我们只需要枚举给左子树分配多少钱，剩下的钱都分给右子树。

状态总数 **$O(MN)$**

状态转移代价 **$O(M)$**

时间复杂度为 **$O(NM^2)$**

$N * M * M \leq 6 * 10^8$ ，不太理想。



我们换一种数据组织方式

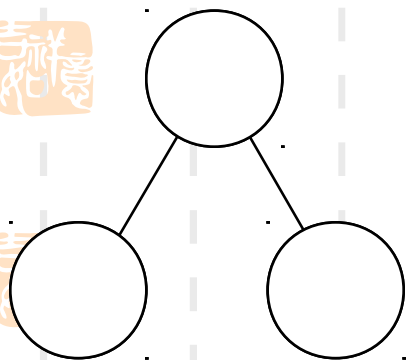


考虑本题特殊条件：

1. 附件没有附件。
2. 每个主件最多只有两个附件。

我们把配套的主件和附件看成一组。

这样，显然对于每一组，可能的购买方案最多只有如下五种：



1. 什么都不买
2. 只购买主件
3. 购买主件和附件 **1**
4. 购买主件和附件 **2**
5. 全购买





类似经典的 0 - 1 背包问题！

- 状态总数 $O(MN)$
- 状态转移代价 $O(1)$

组织数据后，我们可以得到复杂度为 $O(NM)$ 的优秀算法



郁闷的金明

【题意描述】

给出 N 个物品，每个物品都有一个权值 (<50000) 和一个价格 (<10000)。我们称可以直接被购买的物品为主件，称不能被直接购买的物品为附件，附件只有当其主件被购买了才能被购买，**主件可以有任意多附件**，附件没有下一级附件。

【任务】

用不超过 M 元钱，购买一些物品，使得被购买的物品的总权值最大。

【数据规模】

$N < 60$ $M < 3200$

问题分析

题目放宽了“一个主件最多可以有二个附件”这个限制。

我们回想上题的数据组织方式。

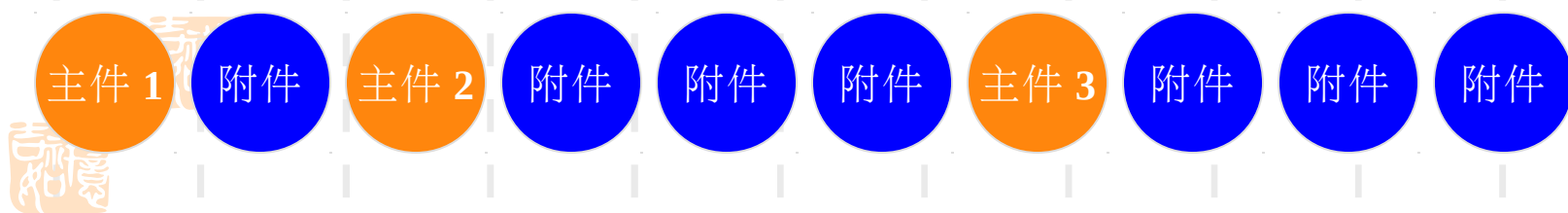
数据组织方式	依然适用	效率
以物品为节点的树形结构	✓	×
以组为元素的序列	×	-

我们需要重新组织数据！

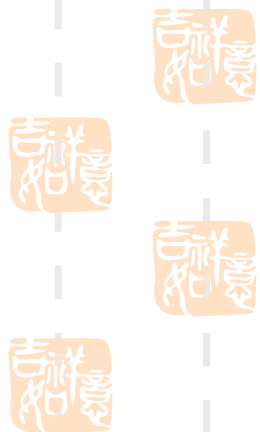


数据组织方案二

重新安排这些物品的顺序，使得每个附件都紧跟其主件，保证其前面的最近的主件就是它附属的主件。如下图：



树 → 序列

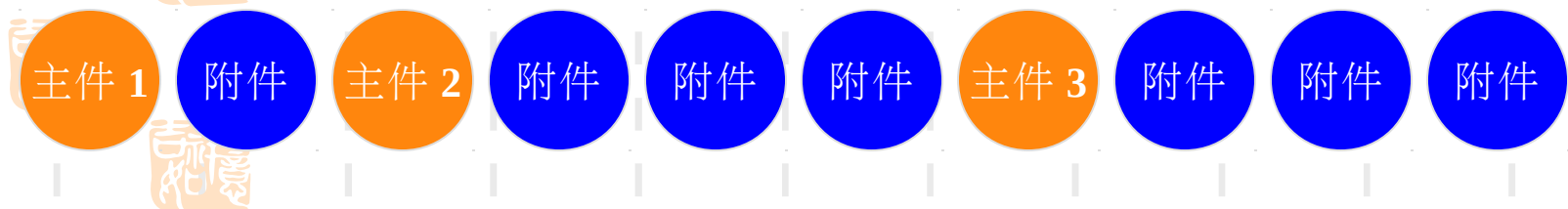


算法 3

这样组织数据以后，一个附件能被购买的必要条件是“其前面的最近的主件被购买了”。

状态 $F[i][j][k]$ 表示从第 i 个物品到第 n 个物品，最多花费 j 元， k 表示 i 物品前面的主件有没有被购买，的最大价值和

$K=0$ 主件 2 没有被购买



算法 3

- 状态总数 $O(NM*2)$
- 状态转移代价 $O(1)$
- 时间复杂度 $O(NM)$

重新组织数据后，我们再次成功地设计出了 $O(NM)$ 的算法。



很郁闷的金明

【题意描述】

给出 N 个物品，每个物品都有一个权值 (<50000) 和一个价格 (<10000)。我们称可以直接被购买的物品为主件，称不能被直接购买的物品为附件，附件只有当其主件被购买了才能被购买，主件可以有任意多附件，可以有多个附件。

【任务】

用不超过 M 元钱，购买一些物品，使得被购买的物品的总权值最大。

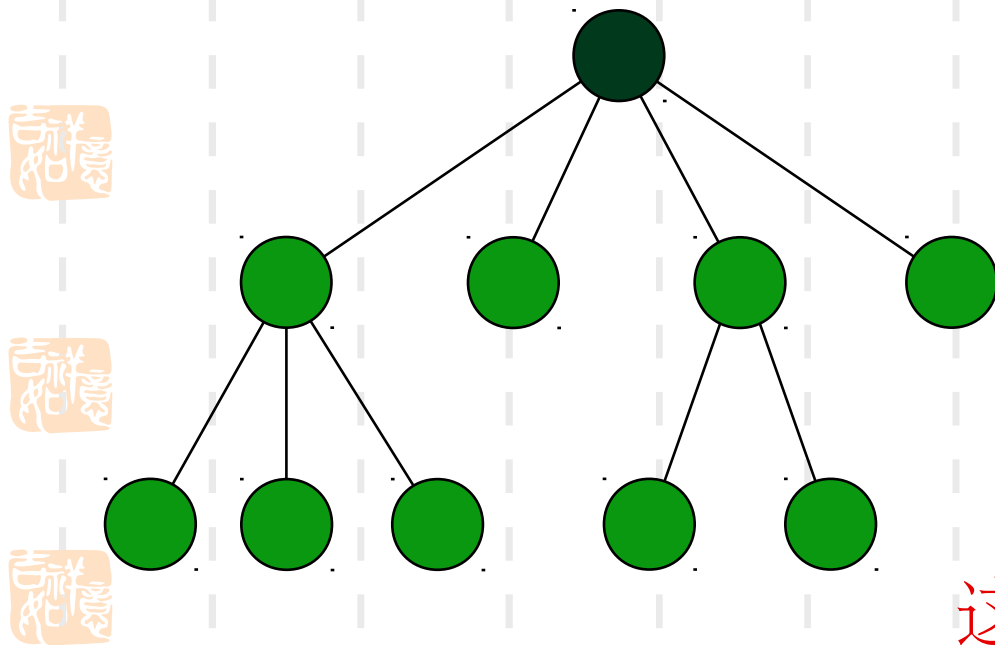
【数据规模】

$N < 60$ $M < 3200$



问题分析

现在的题目在原题的基础条件上不仅增加附件的个数，还出现了多级附件。



这是很一般的树！

一般的树形结构，我们还能不能用前面的数据组织方式呢？

数据组织方式	依然适用	效率
以物品为节点的树形结构	✓	✗
以组为元素的序列	✗	—
附件紧跟其主件的序列	✗	—

说明这些数据组织方式都不合理，需要再次重新组织数据！

组织数据方案三

现在我们再回过头来研究一下前面一种数据组织方式：

把同在一个组的主件放在附件的前面，将树形问题转化到序列上来。

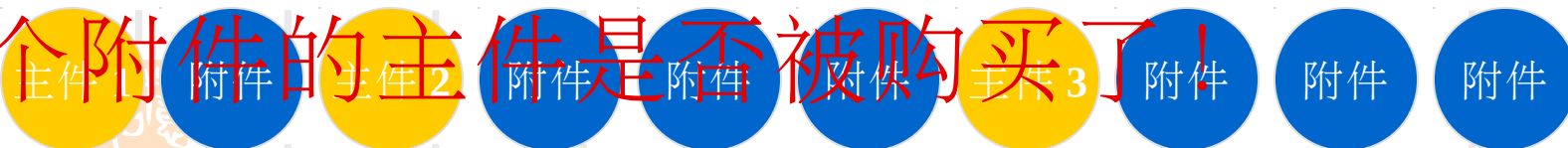
而现在的问题是：树的高度增加了。

考虑：树的先根遍历序。



仔细思考算法 3 的状态转移：

但是我们不能用 $k=0$ 加一维的方法来记录
每个附件的主件是否被购买了！



迁移到本题中，对于一棵子树，如果我们不购买其根结点，那么其子孙都不必讨论了（因为其子孙节点都不能被购买）





这一结论似乎很显然，但是我们并不是要在树形结构中运用这一结论。

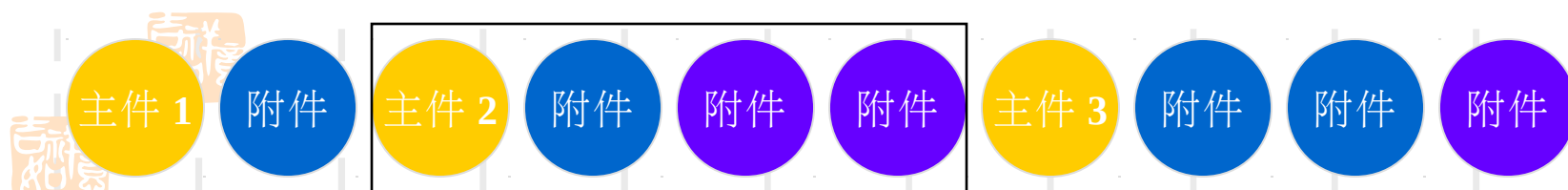
正如上面提到的，我们要在树的先根遍序上进行动态规划，而这一结论正是我们成功的关键。



算法 4

状态 $F[i][j]$ 表示在遍历序列中从第 i 个物品到第 n 个物品，最多花费 j 元，能得到的最大权值和。

没有购买根结点！



直接“跳”过去！

算法 4

■ 状态总数 $O(NM)$

这样，实际上我们避开了“记录主件状态”的问题！成功地实现了状态的合法转移

■ 状态转移代价 $O(1)$

■ 时间复杂度 $O(NM)$

重新组织数据后，我们再一次优解此题。

小结

树形结构  树形动态规划 $O(NM^2)$

合理地组织数据

 线形结构  线形动态规划 $O(NM)$

树的果实

【题意描述】

给出一棵有 N 个节点的有根树（根为 1 号节点），每个节点有权值。

要求对于每一个节点，求：

1. 其子树中权值比该节点大的节点总数
2. 树上所有比该节点大的节点总数
3. 从根节点到该节点路径中比该节点大的节点总数

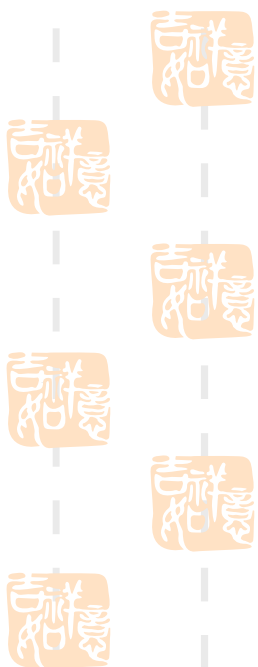
其中 $(1 \leq N \leq 10^5)$



问题分析

吉祥如意

树形上的统计问题！
序列上的统计问题。





对数据的初步组织

我们进行新的组织数据的尝试：利用先根遍历序将树转化为序列，因为这样，同一棵子树构成一个连续的区间，这是一个非常好的性质。

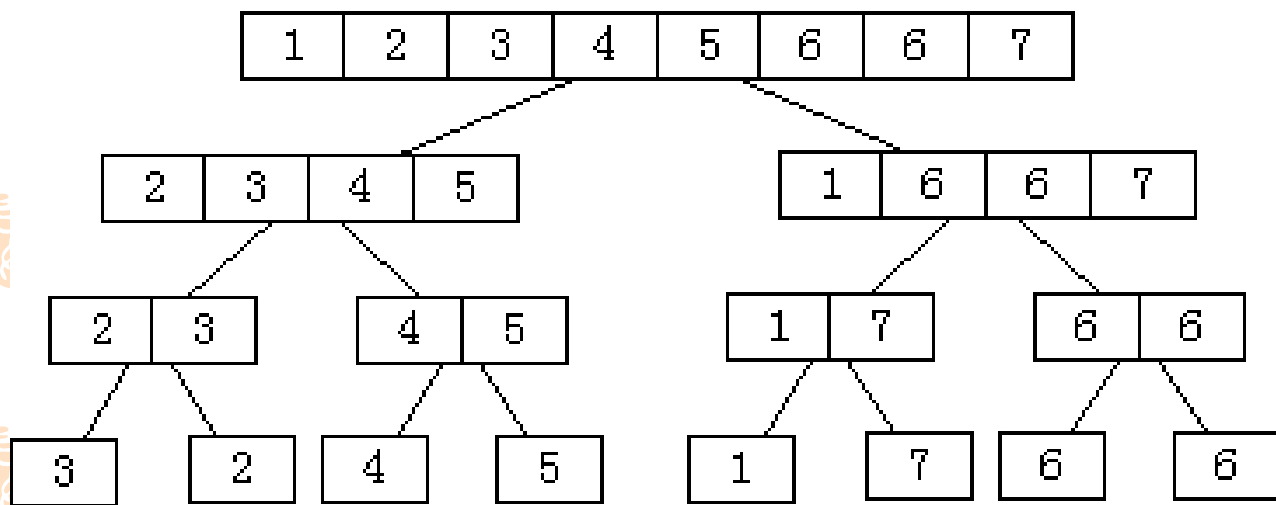


问题转化为：在一个由整数构成的序列上，进行 **N** 次区间询问。询问一个区间中有多少元素的权值比给定的值大



在组织后的数据中尝试求解

我们直接在组织成序列的数据中进行统计。可以利用以有序表为元素的线段树！



预处理——归并排序 $O(N\log_2(N))$

每次统计的时间复杂度为 $O(\log_2^2(N))$

总的时间复杂度为 $O(N\log_2^2(N))$



进一步组织数据

我们重新考虑转化后的问题。

考虑一种特殊情况：

当前的序列中所有元素的权值均大于给定的值

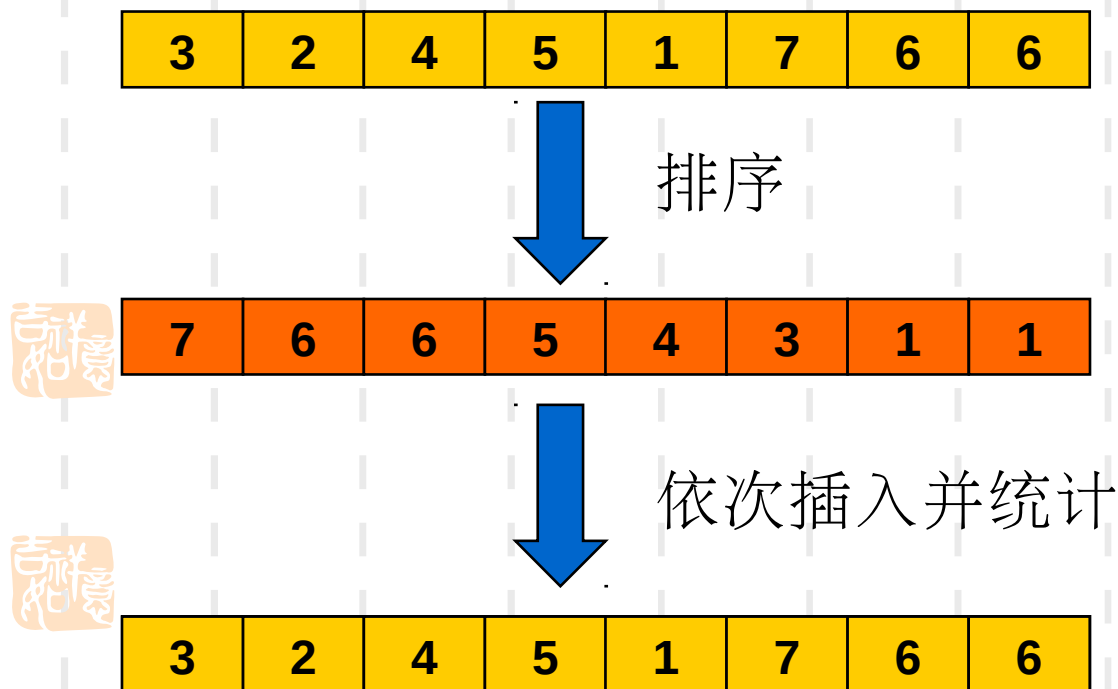
。

答案即是区间中的元素个数！

这是线段树和树状数组的看家本领！



一个很巧妙的方法：从大到小地向线段树里面依次加入元素，边加边统计。



这样，我们的总时间复杂度为 $O(N \log_2(N))$

小结

吉祥如意

树



组织“形态”

序列



嵌套数据结构



组织“顺序”

“从大到小”



一般数据结构





总结

以上我们从几个例题介绍了“数据的合理组织”的重要性及其在解题中的一些应用，然而例题只是一部分题目的代表，具体的题目应该具体分析。

没有最合理的数据组织方式！多思考，多总结，多实践，才是万能的解题之道。



吉祥如意

谢谢

吉祥如意

吉祥如意

吉祥如意

吉祥如意

吉祥如意

吉祥如意