



上海海事大学
SHANGHAI MARITIME UNIVERSITY

自然语言处理

2024–2025 学年第 2 学期



信息工程学院 谢雨波

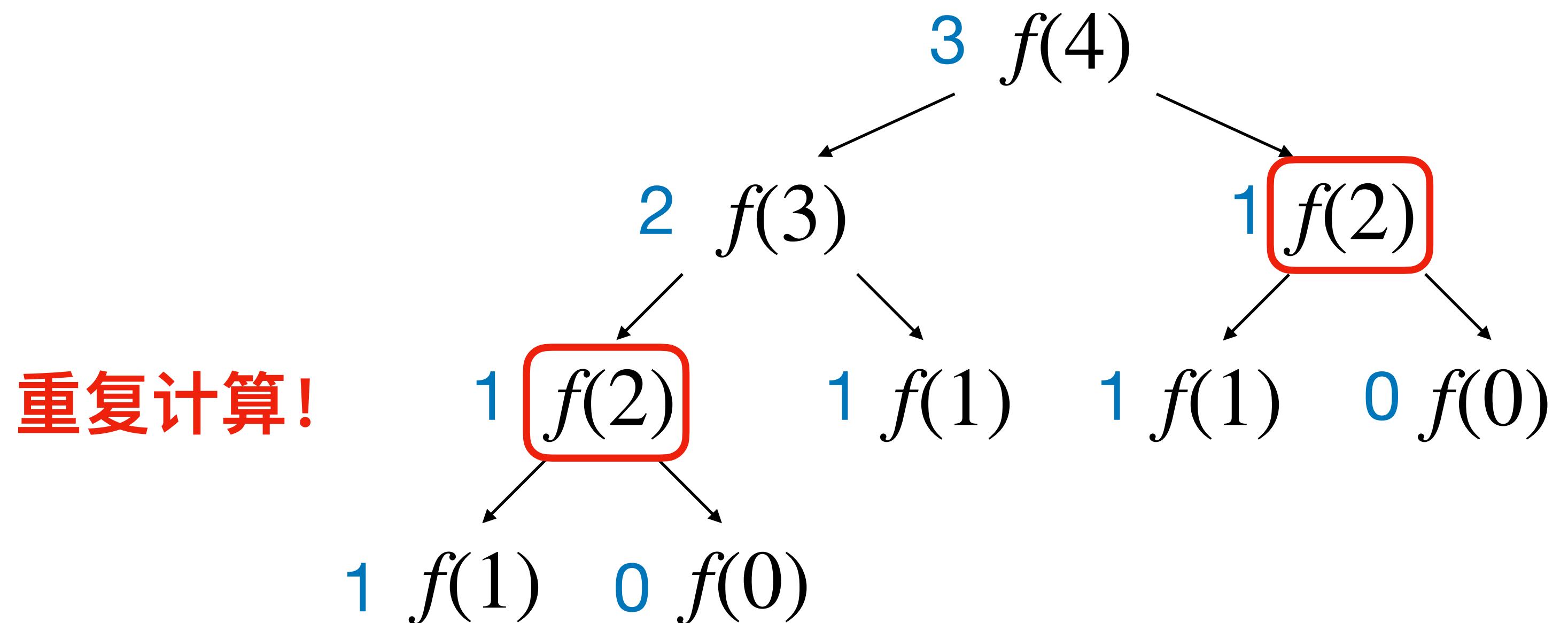
动态规划

动态规划

- 动态规划的核心思想是**将复杂问题分解为更小的子问题，并存储这些子问题的解决方案以避免重复计算**
- 它具有两个关键特性：
 - **重叠子问题**: 问题可以分解为多个子问题，且这些子问题在求解过程中会被重复调用
 - **最优子结构**: 问题的最优解可以通过子问题的最优解组合得到

以斐波那契数列为例

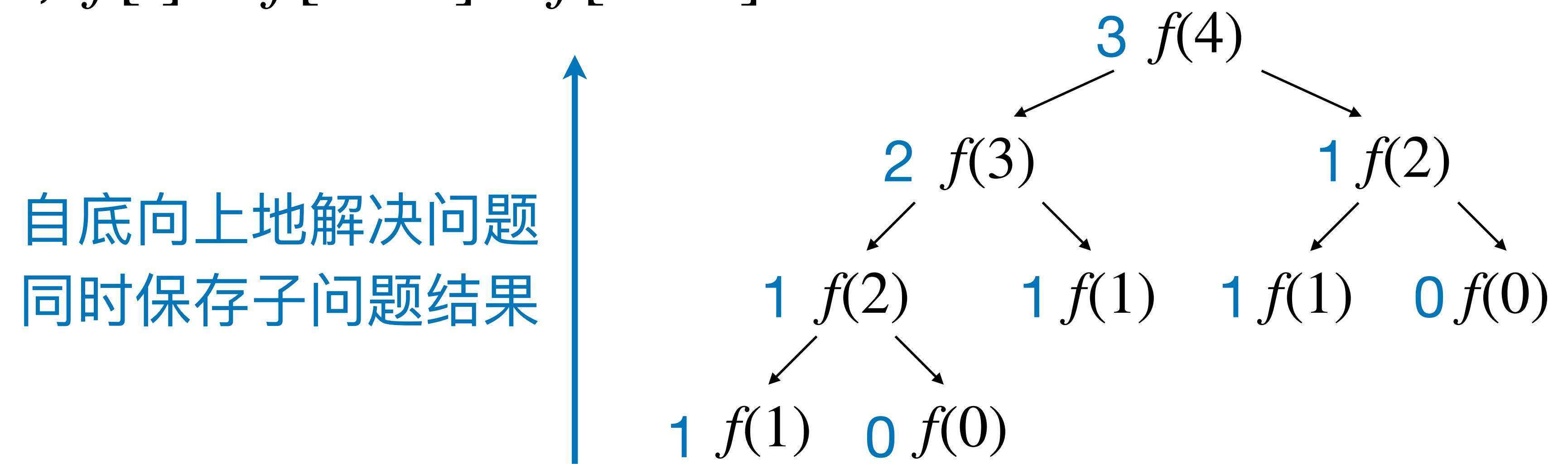
- 斐波那契数列: 0, 1, 1, 2, 3, 5, 8, 13, 21, ...
- 已知 $f(n) = f(n - 1) + f(n - 2)$, 求 $f(4)$



满足：
(1) 重叠子问题
(2) 最优子结构

以斐波那契数列为例

- 对于斐波那契数列问题，我们可以将已计算的子问题结果存储下来，之后再碰到相同的子问题，只需要读取结果即可
- 初始化： $f[0] = 0, f[1] = 1$
- 迭代：对于 $i \geq 2$, $f[i] = f[i - 1] + f[i - 2]$



隐马尔可夫模型与词性标注

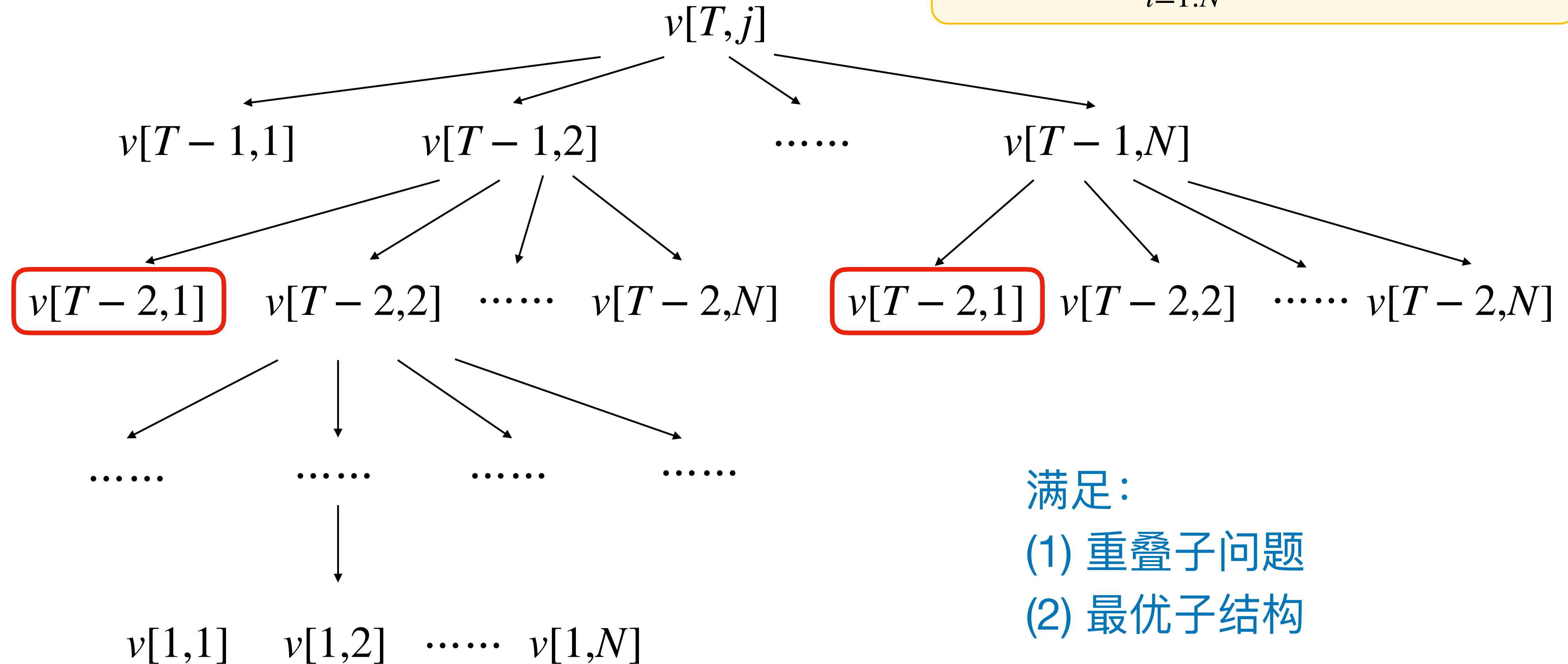
- 目标：找到概率最大的词性标注序列
- 等价于： \max_j 概率最大的词性标注序列，且最后一个词性为 j

$$\max_j \overbrace{\quad\quad\quad}^{\downarrow} v[T, j]$$

目标： $\max_j v[T, j]$

隐马尔可夫模型与词性标注

$$v[t,j] = \max_{i=1:N} v[t-1,i] a_{ij} b_j(o_t)$$



基于动态规划思想的 Viterbi 算法

```
function Viterbi( $o_1, \dots, o_T, A, B, \pi$ ):
```

```
    创建数组  $v[T, N]$ 
```

```
    for  $i$  from 1 to  $N$  do      // 初始化
```

```
         $v[1, i] \leftarrow \pi_i \cdot b_i(o_1)$ 
```

```
        backtrack[1, i]  $\leftarrow 0$ 
```

```
    for  $t$  from 2 to  $T$  do      // 迭代
```

```
        for  $j$  from 1 to  $N$  do
```

```
             $v[t, j] \leftarrow \max_{i=1:N} v[t - 1, i] \cdot a_{ij} \cdot b_j(o_t)$ 
```

```
            backtrack[t, j]  $\leftarrow \arg \max_{i=1:N} v[t - 1, i] \cdot a_{ij} \cdot b_j(o_t)$ 
```

```
    bestprob  $\leftarrow \max_{i=1:N} v[T, i]$ 
```

```
    bestpointer  $\leftarrow \arg \max_{i=1:N} v[T, i]$ 
```

```
    bestpath  $\leftarrow$  从 bestpointer 开始，根据 backtrack 往前回溯
```

```
    返回 bestpath, bestprob
```