最优二叉检索树的算法

关键问题

子问题边界界定

如何将该问题归结为更小的子问题

优化函数的递推方程及初值

计算顺序

是否需要标记函数

时间复杂度分析

子问题划分

子问题边界为(i, j)

数据集： $S[i,j] = < x_i,\ x_{i+1}, \dots, x_j >$

存取概率分布：

$P[i,j]=<a_{i-1}, b_i, a_i, b_{i+1}, \dots, b_j, a_j>$

输入实例： $S = < A, \underline{B, C, D}, E >$

$P = < 0.04, 0.1, \underline{0.02, 0.3, 0.02, 0.1},$

$\underline{0.05, 0.2,\ 0.06}, 0.1, 0.01>$

子问题： $S[2,4] = < B, C, D >$

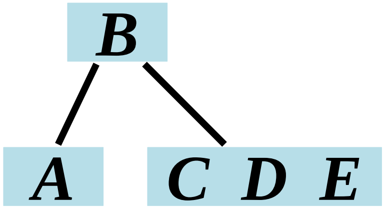
$P[2,4]=<0.02,0.3,0.02,0.1,0.05, 0.2,0.06>$

子问题归约

以 x_k 作为根归结为子问题：

$S[\,i, k-1], P[\,i, k-1]$

$S[\,k+1, j\,], P[\,k+1, j\,]$



$S[1,5] = <A, B, C, D, E>$

$P[1,5] = <$ 0.04, 0.1, 0.02, 0.3, 0.02, 0.1, 0.05, 0.2, 0.06, 0.1, 0.01$>$

$S[1,1] = <A>$,

$P[1,1] = <$ 0.04, 0.1, 0.02 $>$

$S[3,5] = <C,D,E>$,

$P[3,5] = <$ 0.02, 0.1, 0.05, 0.2, 0.06,0.1,0.01$>$

子问题的概率之和

子问题界定 $S[i,j]$ 和 $P[i,j]$，令

$$w[i,j]=\sum_{p=i-1}^{j}a_p+\sum_{q=i}^{j}b_q$$

是$P[i,j]$中所有概率(数据与空隙)之和

实例：$S[2,4]=<B,C,D>$

$P[2,4]=<0.02,0.3,0.02,0.1,0.05,0.2,0.06>$

$$\begin{aligned}w[2,4]&=(0.3+0.1+0.2)\\&\quad+(0.02+0.02+0.05+0.06)\\&=0.75\end{aligned}$$

优化函数的递推方程

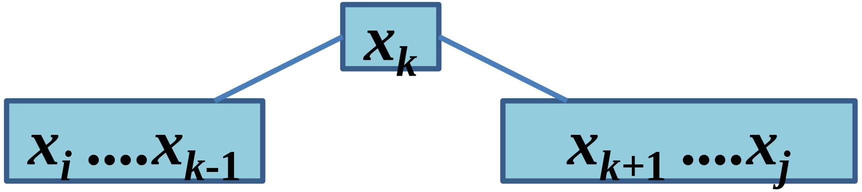
设 $m[i,j]$ 是相对于输入 $S[i,j]$ 和 $P[i,j]$ 的最优二叉搜索树的平均比较次数

递推方程：

$$m[i,j]=\min_{i\le k\le j}\{\underline{m[i,k-1]}+\underline{m[k+1,j]}+w[i,j]\},$$

$$1\le i\le j\le n$$

$$m[i,i-1]=0,\quad i=1,2,\ldots,\ n$$



$m[\,i, j\,]_k$公式的证明

$m[i,j]_k$：根为 x_k 时平均比较次数的最小值

作为子树增加次数

$$
\begin{aligned}
& m[i,j]_k \\
&= (m[i,k-1] + \underline{w[i,k-1]}) + (m[k+1,j] + \underline{w[k+1,j]}) + 1\times b_k \\
&= (m[i,k-1] + m[k+1,j]) + (w[i,k-1] + b_k + w[k+1,j]) \\
&= (m[i,k-1] + m[k+1,j]) + \underline{(\sum_{p=i-1}^{k-1} a_p + \sum_{q=i}^{k-1} b_q)} + b_k + \underline{(\sum_{p=k}^{j} a_p + \sum_{q=k+1}^{j} b_q)} \\
&= (m[i,k-1] + m[k+1,j]) + \sum_{p=i-1}^{j} a_p + \sum_{q=i}^{j} b_q \\
&= \boxed{m[i,k-1] + m[k+1,j] + w[i,j]}
\end{aligned}
$$

代入概率公式

化简

递推方程

$$m[i,j]=\min_{i\le k\le j}\{m[i,k-1]+m[k+1,j]+w[i,j]\},$$

平均比较次数：在所有 k 的情况下

$m[i,j]_k$ 的最小值

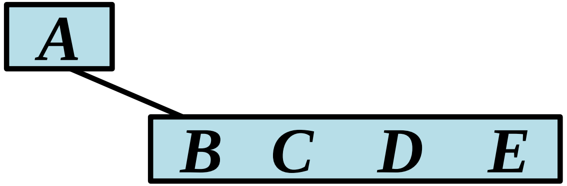
$$\boxed{m[i,j] = \min\{ m[i,j]_k \mid i \le k \le j \}}$$

初值 $m[i, i-1]=0$对应于空的子问题，

例如 S = <A,B,C,D,E>，取 A 作根，

$i = 1$，$k=1$, 左边子问题为空树，

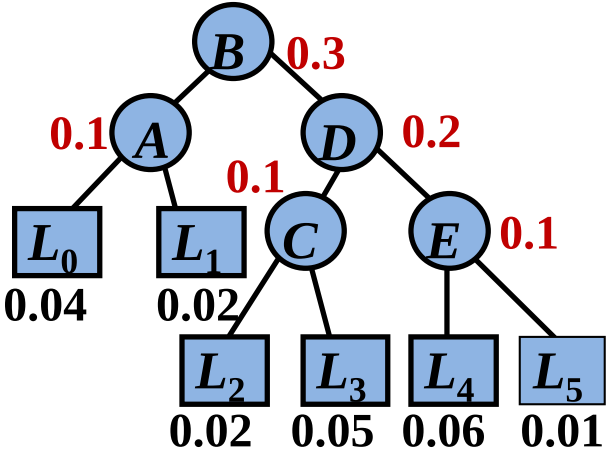
对应于：$S[1,0]$，$m[1,0]=0$的情况.



实例

$$m[i,j]=\min_{i\le k\le j}\{m[i,k-1]+m[k+1,j]+w[i,j]\}$$

$$m[i,i-1]=0$$



以B为根

k=2

m[1,1]=0.16

m[3,5]=0.88

$$m[1,5]=1+\min_{k=2,3,4}\{m[1,k-1]+m[k+1,5]\}$$

$$=1+\{m[1,1]+m[3,5]\}=1+\{0.16+0.88\}=2.04$$

计算复杂性估计

$$\boxed{\begin{array}{l} m[i,j] = \min\limits_{i \le k \le j} \{m[i,k-1] + m[k+1,j] + w[i,j]\} \\ 1 \le i \le j \le n \end{array}}$$

$m[i,i\text{-}1] = 0, \quad i = 1,2,\ldots,n$

i, j 的所有组合$O(n^2)$种

每种要对不同的 k 进行计算，$k=O(n)$

每次计算为常数时间

时间复杂性：　　　$T(n) = O(n^3)$

空间复杂度：　　　$S(n) = O(n^2)$

小结

- 划分子问题,以数据结点作为树根
- 定义优化函数, 列出递推方程与边界条件
- 自底向上计算，设计备忘录 (表格)
- 设立标记函数记录构成最优二叉搜索树或子树时根的位置.
- 时间复杂度估计