

## 第五章 概率分析和随机算法

这个章节的内容说难不难，说简单也不简单，要是有一点概率基础的话看懂还是挺容易的，但是深究下去的话还是有很大难度的。考虑到自己的当前进度与计划，因此对本章不做重点讨论（但是在空闲之余我还是会加强、完善的），所以对习题和 5.4（打\*号的）都没做过多停留。如果有问题可以与我交流。

第一部分《基础知识》到此结束了，下面就开始了真正的算法导论之旅。第六章的内容是堆排序，相信一定会很有趣的，大家一起进步！

雇佣策略伪代码：

HIRE-ASSISTANT( $n$ )

```
1   $best \leftarrow 0$   $\triangleright$  candidate 0 is a least-qualified dummy candidate
2  for  $i \leftarrow 1$  to  $n$ 
3      do interview candidate  $i$ 
4      if candidate  $i$  is better than candidate  $best$ 
5          then  $best \leftarrow i$ 
6      hire candidate  $i$ 
```

**概率分析**是在问题的分析中应用概率技术。

**随机算法**：如果一个算法的行为不只是由输入决定，同时也由随机数生成器所产生的数值决定，则称为这个算法是随机的。

为了分析包括雇用问题在内的许多算法，我们将利用指示器随机变量（indicator random variable）。它为概率和期望之间的转换提供了一个便利的方法。给定一个样本空间 $S$ 和事件 $A$ ，那么事件 $A$ 对应的**指示器随机变量** $I\{A\}$ 定义为：

$$I\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

定理：

给定样本空间 $S$ 和 $S$ 中的事件 $A$ ，令 $X_A = I\{A\}$ ，则 $E[X_A] = Pr\{A\}$

利用指示器随机变量对雇用问题进行分析

假设应聘者以随机的顺序出现，令  $X$  作为一个随机变量，其值等于雇用一个新的办公助理的次数。如果运用期望值的定义

$$E[X] = \sum_{x=1}^n x \Pr\{X = x\}$$

但是，这个计算将变得很复杂。利用随机变量指示器可以简化计算。

为了利用指示器随机变量，我们定义  $n$  个和每个应聘者是否被雇用对应的变量。特别地，令  $X_i$  对应于第  $i$  个应聘者被雇用这个事件的指示器随机变量。

$$X_i = I\{i(st) \text{ candidate was hired}\} = \begin{cases} 1 & \text{if } i(st) \text{ candidate was hired} \\ 0 & \text{if } i(st) \text{ candidate was not hired} \end{cases}$$

且有

$$X = X_1 + X_2 + \cdots + X_n$$

由前面的定理有

$$E[X_i] = \Pr\{i(st) \text{ candidate was hired}\}$$

如果应聘者  $i$  胜过从 1 到  $i-1$  的每一个应聘者，则应聘者  $i$  被雇用。应聘者  $i$  比从 1 到  $i-1$  更有资格的概率是  $1/i$  ( $i$  人中一个人是最优秀的概率为  $1/i$ )

$$\begin{aligned} \therefore E[X_i] &= 1/i \\ \implies E[X] &= E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] \\ &= \sum_{i=1}^n 1/i = \ln n + O(1) \end{aligned}$$

即使面试了  $n$  人，也大概只雇用其中的  $\ln n$  人。

假设应聘者以随机次序出现，算法 HIRE-ASSISTANT 总的雇用费用为  $O(c_h \ln n)$ 。

在雇用策略上加一个操作用以随机的排列应聘者(随机算法)

RANDOMIZED-HIRE-ASSISTANT( $n$ )

```
1  randomly permute the list of candidate
2   $best \leftarrow 0$   $\triangleright$  candidate 0 is a least-qualified dummy candidate
3  for  $i \leftarrow 1$  to  $n$ 
4      do interview candidate  $i$ 
5          if candidate  $i$  is better than candidate  $best$ 
6              then  $best \leftarrow i$ 
7              hire candidate  $i$ 
```

这样做的好处是避免出现最坏情况。这个过程的期望费用是 $O(c_h \ln n)$

### 随机排列数组

许多随机算法通过排列给定的输入数组来使输入随机化。一个比较常用的方法是为数组的每个元素 $A[i]$ 赋一个随机的优先级 $P[i]$ ，然后依据优先级对数组 $A$ 中的元素进行排序。

PERMUTE-BY-SORTING( $A$ )

```
1   $n \leftarrow \text{length}[A]$ 
2  for  $i \leftarrow 1$  to  $n$ 
3      do  $P[i] = \text{RANDOM}(1, n^3)$ 
4  sort  $A$ , using  $P$  as sort keys
5  return  $A$ 
```