

第1章 数学建模

数值方法在现代科学中扮演着重要的角色。科研工作往往在计算机上进行，而不是依赖于种种实验设备。当然，科学实验室的角色极少能被计算机模拟完全取代，后者通常作为实验手段的补充而存在。

例如，在图 1-1a 中描绘的是对两辆纳斯卡汽车进行的空气动力学模拟。为实现这一模拟，我们需要求出相关偏微分方程组 (Partial Differential Equation, PDE) 的数值解，这组方程对流过车体的气流建立数学模型。由于汽车的车身必须光滑平顺，所以我们对它的建模常常用上三次(或更高次)的样条函数。类似的建模方法也在飞机设计当中使用。第 8 章和第 14 章将探讨使用样条函数求解偏微分方程组的数值问题。

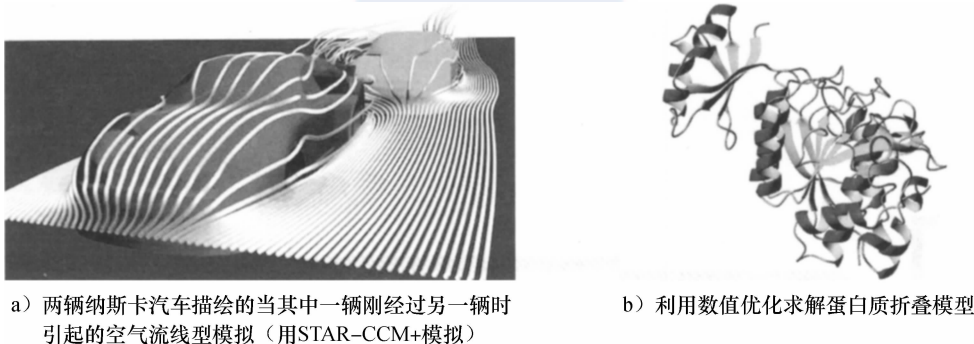


图 1-1

生物数学这一在工业、学术和政界中都火热的研究领域，为我们提供了另外的例子，数值算法在其中扮演了关键性的角色。比如，蛋白质折叠模型通常作为一个大规模的优化 (optimization) 问题得以解决，而蛋白质本身是用“最小能量原则”来确定自身形状的——对大自然来说，这样的设定是小事一桩，不足挂齿，但对人类的计算来说，它就很难了。数值最优化本身是一个完整的主题，本书不便多涉；然而，大多数优化过程的核心内容仍由本书所述的数值方法所组成。

在学习如何分析和应用诸多有效而精确的数值方法之前，我们不妨先一窥数学建模这个主题，它能将现实世界中的问题转化成一系列方程，使得数值分析师有其用武之地。数学上的方程通常不过是表示真实的物理状态的一个模型 (model)，因此，对数值分析家或计算机科学家而言，对这一模型的来源能有所认识，是至关重要的。实际上，数值分析师有时与设计出这些模型的科学家、工程师一起工作。这样的互动之所以重要有着以下的原因：第一，许多算法无法求出真实解，而只能获得近似解，为判断什么样的近似解精确到可以接受，人们必须明了原问题的来源。若是在战场上定位敌方坦克，数厘米的误差足可容忍；但是在激光外科手术中确定肿瘤位置时，这种误差将是无法容忍的。第二，即使算法在理论上能有精确解，但在一台精度有限的计算机上应用时，求得的结果也大多是不精

确的；弄明白有限精度运算对结果准确性的影响，也是数值分析的任务之一。有关有限精度计算的问题，将在第 5 章研究。本章将介绍数值计算的大量应用，它们来自各行各业的数学建模实践。

1.1 计算机动画中的建模

计算机生成的许多主导银幕的图像是由动态模拟技术产生的，即，运用物理定律建立模型，然后用数值方法计算该模型的结果。本节将展示在 2002 年电影《星球大战前传 II：克隆人的进攻》的动画中数字的作用。特别是，我们要仔细考察用于创造绝地大师尤达这个角色的一些数值特效。尤达大师在 1980 年的《星球大战 2：帝国反击战》首次出现，彼时他是由一个线控木偶饰演的。而在 2002 年的电影里，尤达已经鸟枪换炮，由采用大量数值算法的数码技术来打造了。

创建一个数字化的尤达大师，关键就是要让这个角色的举止看上去逼真。尤达身体的动作采用关键帧动画技术来描绘。在关键帧技术里，一个姿势是由某些特定点确定的，之后计算机将采用插值法来自动决定中间帧的相应姿势。（第 8 章将讨论一些插值技术。）动画师对运动的对象可施以多重控制，比如指定它们的速度和方向。当绘制者们确定尤达大师身体如何运动时，计算机必须把其长袍的拂动给刻画出来。

长袍模型

参见图 1-2，我们看到长袍是用三角形来表示的，而其每个顶点的运动必须要确定。将每个顶点看作是一个具有质量、位置、速度，对外力有反应，但没有大小的点状粒子。

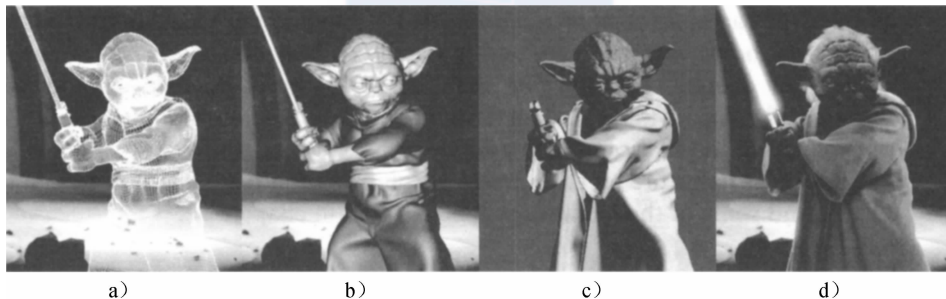


图 1-2

图 1-2 是《星球大战前传 II》中尤达大师和杜库伯爵战斗场景的各个模拟阶段的示意。在 b 和 c 里能看到尤达的两层衣服，它们是分开计算的。一个称作“碰撞检测”的方法保证不会透过外衣看见他的内衣。简化的衣物图饰模型构建一个真正衣服的外观，最终渲染出 d 的效果^[22]。（图片由 Lucasfilm 有限公司提供。《星球大战前传 II：克隆人的进攻》中© 2002 Lucasfilm 公司。版权所有。授权使用。非授权复制违反相关法律。数字化的工作由工业光魔公司完成。）粒子的运动要遵守牛顿第二定律，它由以下方程表示

$$\mathbf{F} = m\mathbf{a} = m(d^2\mathbf{y}/dt^2) \quad (1.1)$$

其中， \mathbf{y} 是时间 t 的距离函数。请注意，因为我们在三维空间中计算，所以方程涉及向量

值函数。由于粒子具有质量($m \neq 0$)，式(1.1)可以改写为二阶常微分方程(ODE)

$$\frac{d^2 y}{dt^2} = \frac{F}{m} \quad (1.2)$$

因为在某个初始时刻粒子的状态已经给定了，所以这个常微分方程是一个初值问题的一部分。在电影的情况下，这个初始时刻指的是在一个可能持续几秒或几分之一秒的场景的开始。

为了保持长袍的形状，相邻粒子会用弹力相互连接到一起。胡克定律规定弹簧对两端施加的力 F_s 是与弹簧当前长度及其不受力时的长度 x_0 的差成正比的。数学表达式为

$$F_s = -k(x - x_0)$$

其中， x 表示弹簧的当前长度， k 是弹性系数。为简单起见，我们给出的是胡克定律在一维情况下的公式，但模拟尤达大师的长袍需要一个三维版本。

在制造尤达大师长袍的动画时，许多其他的力也要考虑到，包括重力、风力、碰撞力、摩擦力，甚至其他凭空造出来的力，它们的存在仅仅是为了实现满足导演要求的动作。在最简单的情况下，这个长袍模型的解析解可能存在。然而，在计算机动画设计中，作用于粒子的力在不断地改变，因此要在每一帧找到一个分析解，即便可能，也是不切实际的。相反，数值方法通过模拟粒子在离散时间下每一步的状况，用来求近似解。常微分方程组初值问题的数值解有很多文献，其中有一些方法会在第11章讲述。

3

对物理定律的扭曲

在模拟《星球大战 II》中尤达大师的长袍的运动时，动画师发现，如果实际执行这个特效，那么长袍将会被强大的加速度给撕开。为解决这个问题，动画师特意在模拟中加入了为这款袍子量身定做的“保护”特效，使得加速度能降低。最终，通过超人的运动，穿在数字演员身上的衣服变形较小，在左侧的图中，我们看到欧比旺·克诺比(由伊万·麦格雷戈饰演)在《星球大战 II》中表演一个对大活人来说极其危险的特技。注意，他的头发就像衣服一样，用一束束被弹簧串联在一起的粒子来模拟^[22]。(图片由 Lucasfilm 有限公司提供。《星球大战前传 II：克隆人的进攻》中 ©2002 Lucasfilm 公司。版权所有。授权使用。非授权复制违反相关法律。数字化的工作由工业光魔公司完成。)



创造令观众信服的模拟效果，是动画电影的一个目标。作为科学模拟的首要目标，准确性可能与这个目标产生了冲突。虽然数值模拟被科学家和娱乐业用于不同目的，但它们有很多共同的数学工具和策略。现在，让我们把注意力转向科学中的模拟。

1.2 物理建模：辐射的传播

辐射传输可以用随机过程描述，用蒙特卡罗模拟来建模。蒙特卡罗方法将在第3章讲

述。辐射传输的模型，也可以由微分-积分方程——玻尔兹曼迁移方程建立。这是一个带有积分式的偏微分方程。第 14 章会讨论偏微分方程数值解，而第 10 章将讨论数值积分方法。结合这些想法以及用于求解大型线性系统的迭代方法(12.2 节)，可用来求玻尔兹曼迁移方程的近似解。

你知道你的身体接受一次牙科 X 射线检查所受的辐射量有多大吗？你大概还记得在检查的时候穿的防护衣吧？还记得在 X 光机打开的时候牙医助理匆忙离开的背影吗？穿上防护衣是为了屏蔽辐射。为了有助于设计这样的保护材料和验证其有效性，我们应该怎样对 X 射线光子的传播建模呢？在这里，光子是随着电子束的开关而产生的。单个光子的能量和方向无法预测，但是 X 射线的整体能量分布和方向能够被近似估算出来。

能量、位置、方向、制造光子的时间是系统模型中的自变量。可以认为每个变量都是随机的，但遵循着一定的分布。例如，光子能量的分布如图 1-3 所示。

假定入射光子在进入被测物之前是沿直线运动的，然后它们以一定的概率(具体概率依赖于射入材料的性质)和该物体的原子碰撞，如图 1-4 所示。这种碰撞通常会把光子的能量转移到原子内部的电子上去，之后光子能量减弱，方向也发生了变化。在最极端的情况下，光子会把自己所有的能量转移给电子，这样它就整个被原子吸收了。

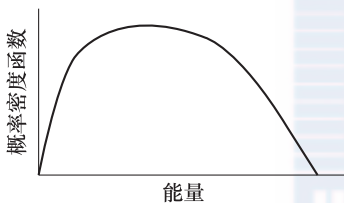


图 1-3 光子能量分布的一个例子

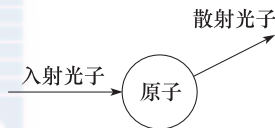
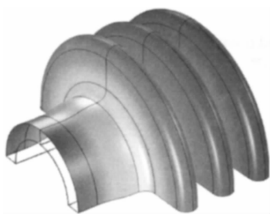


图 1-4 光子与原子的碰撞

为了做好模拟工作，以上事件发生的概率是需要知晓的。这些概率可以从 X 射线以及可能用来屏蔽的材料的理论和实测性质中推算出来。人们可以运行通常包含了百万数量级光子的计算机模拟程序，其中的每个光子都遵守由概率分布决定的随机路径。每个光子都会被追踪，直到它们被吸收，或飞出系统一去不复返为止。模拟的平均结果可以用来确定在特定位置接受到的辐射量。



研究半导体

研究半导体材料中电子的行为需要求解玻尔兹曼迁移方程，该方程包含复杂的积分。确定性方法和蒙特卡罗方法有时可用于这种情形。左边的图是用在辐射迁移问题的确定性模型中的有限元离散化方法(图像源自伦敦帝国学院和 EDF 能源办的应用模型及计算小组，保留版权归拥有者的所有其他版权)。

1.3 运动建模

在世界杯足球赛中,球员们总是尝试利用足球在空中飘忽迷离的曲线和回旋的轨迹来迷惑守门员,把球送入球门。世界级的球员(如巴西的罗伯特·卡洛斯,德国的巴拉克和英格兰的贝克汉姆)尤其擅长从任意球中踢出这样的曲线来。

根据谢菲尔德大学体育工程研究组和福鲁特欧洲公司(Fluent Europe)的计算流体力学(CFD)的研究,足球的形状和表面,以及它的初始方向,对球在空中的轨迹起重要的作用。特别地,这项研究增加了人们对“旋转球”效果的了解,有时这种效果让众多站在最后一道防线的守门员困惑。为了得到这种结果,人们把足球分块数字化了,再如图1-5所示拼接起来。注意,在接缝附近的分块格外细,这是边界层恰当建模所需要的。

有时候任意球的初速会高达每小时70英里^①。风洞实验表明根据足球表面结构和纹理,足球的运动以每小时20~30英里的速度由层流转向湍流,如图1-6所示。

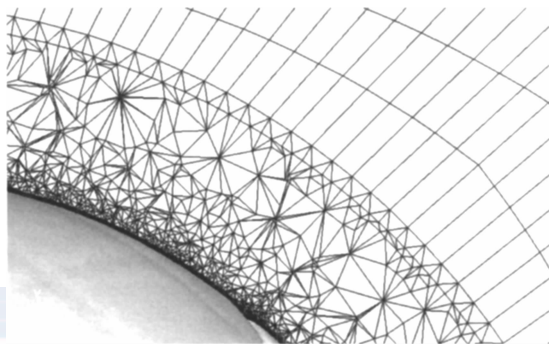
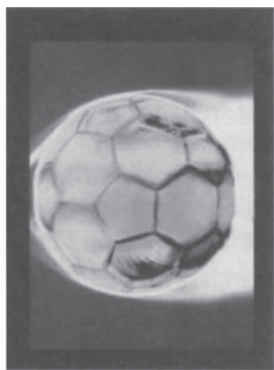
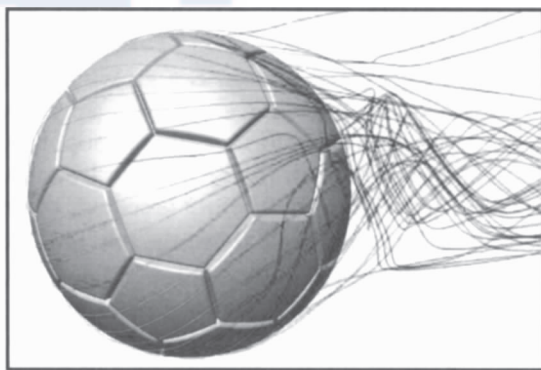


图 1-5 在谢菲尔德大学进行的计算流体力学模拟中,重要的一步是用一台三维接触式激光扫描仪捕捉足球的几何构型。本图显示了由约900万个单元所构成的足球网格的一部分(图片由谢菲尔德和安塞思大学提供)



a) 接受风洞测试的无自旋足球



b) 计算流体力学模拟出的无自旋足球的尾迹线,气流速度是27英里/小时(图片由谢菲尔德和安塞思大学提供)

图 1-6

贝克汉姆在2001年世界杯预选赛英格兰对阵希腊的比赛中有一个令人难忘的进球,谢菲尔德大学开发的技术促进了对这个进球的详细分析。从某种意义上说,贝克汉姆的临门一脚运用了复杂的物理学。虽然谢菲尔德的CFD研究仅仅在超过平均时间时能够准确地模

^① 1英里/小时=0.44704米/秒。

6

拟湍流，因此还不能给出所有情况下的轨迹，但这样的研究已经足以影响从初学到专业的各个层次的运动员。例如，足球制造商们可以利用现有的研究成果来制造轨迹更一致的或更有趣的足球，以满足不同层次球友的需要。这一研究工作也能影响到足球运动员的训练。参考文献[6]和[7]能提供更多的信息。

为此，谢菲尔德大学开发了一个名叫“足球仿真”的模拟程序，它能够预测足球的飞行轨迹。该程序所需的输入，可以通过计算流体力学、风洞测试，以及对球员临门一脚的高速录像来获取。然后，应用该软件可以比较足球在不同初始方向，或被踢飞时不同的旋转状况下的运动轨迹。此外，不同足球的运动轨迹也可以放到一起来比较。

请注意，数值方法在本节的应用和在 1.1 节中的应用一样，涉及微分方程的求解。我们讨论的下一个应用涉及离散现象。

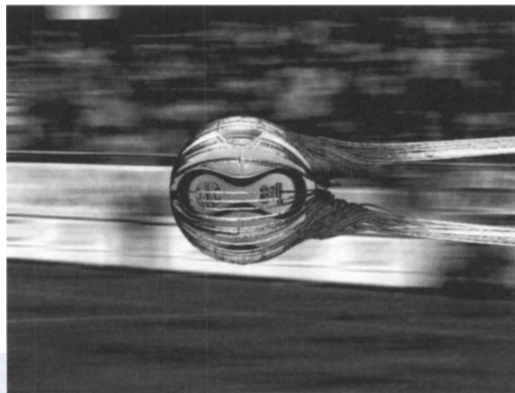


图 1-7 2006 年“团队之星”足球的局部速度着色的高速气流迹线(由谢菲尔德和安塞恩大学授权使用)

7

1.4 生态模型

数值方法在计算生物学领域的应用日益增长。本节将要探索生态学中一个简化的例子。

假设我们要研究某种鸟的种群数目，这些鸟在每年春天出生，至多有三年寿命。我们将跟踪每年这种鸟在繁殖季节之前的总数。这时鸟群可以分成 3 个年龄段：零岁(在上一个春季出生)，一岁，以及两岁。用 $v_0^{(n)}$, $v_1^{(n)}$ 和 $v_2^{(n)}$ 来表示对第 n 年的种群进行观察时，得到第 n 年每年龄段中雌鸟的数目。为了建立一个种群数量模型，我们需要知道：

- 存活率。假设 20% 的新生幼鸟能活到第二年春天，50% 的一龄鸟将会存活到下一年。
- 繁殖率。假设每只一龄的雌鸟生下一窝数量确定的蛋，平均有三只小雌鸟能活到下一个繁殖季；二龄雌鸟能下更多的蛋，假设对于每只这样的雌鸟，能有六只成功活到第二年春天的小雌鸟。

那么我们可以对每年每个年龄段的雌鸟的数目建立以下模型：

$$v_0^{(n+1)} = 3v_1^{(n)} + 6v_2^{(n)}$$

$$v_1^{(n+1)} = 0.2v_0^{(n)}$$

$$v_2^{(n+1)} = 0.5v_1^{(n)}$$

用矩阵-向量的形式来表示，则为

$$\mathbf{v}^{(n+1)} = \mathbf{A}\mathbf{v}^{(n)} = \begin{bmatrix} 0 & 3 & 6 \\ 0.2 & 0 & 0 \\ 0 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} v_0^{(n)} \\ v_1^{(n)} \\ v_2^{(n)} \end{bmatrix} \quad (1.3)$$

8

式(1.3)中的这个矩阵 \mathbf{A} 可以用来预测种群未来的数目以及长期的变化规律。这种可以反映存活率和繁殖率的矩阵，叫作**莱斯利矩阵**(Leslie matrix)。

假设在一开始，种群里有 3000 只雌鸟，每个年龄段各 1000 只。代入式(1.3)，得到

$$\begin{aligned} \mathbf{v}^{(0)} &= \begin{pmatrix} 1000 \\ 1000 \\ 1000 \end{pmatrix} & \mathbf{v}^{(1)} = \mathbf{A}\mathbf{v}^{(0)} &= \begin{pmatrix} 0 & 3 & 6 \\ 0.2 & 0 & 0 \\ 0 & 0.5 & 0 \end{pmatrix} \mathbf{v}^{(0)} = \begin{pmatrix} 9000 \\ 200 \\ 500 \end{pmatrix} \\ \mathbf{v}^{(2)} = \mathbf{A}\mathbf{v}^{(1)} &= \begin{pmatrix} 3600 \\ 1800 \\ 100 \end{pmatrix} & \mathbf{v}^{(3)} = \mathbf{A}\mathbf{v}^{(2)} &= \begin{pmatrix} 6000 \\ 720 \\ 900 \end{pmatrix} \end{aligned}$$

以年份为自变量, 将每个年龄组的鸟的数目作为年份的函数, 可以画出如 1-8a 所示的种群数目示意图. 显然, 种群里鸟的数量呈指数增长. 图 1-8b 显示的是每个年龄段鸟的数目占种群总数的比例, 同样以年份为自变量. 注意, 随着时间的推移, 每年龄段鸟的数目占总数的比例会稳定下来.

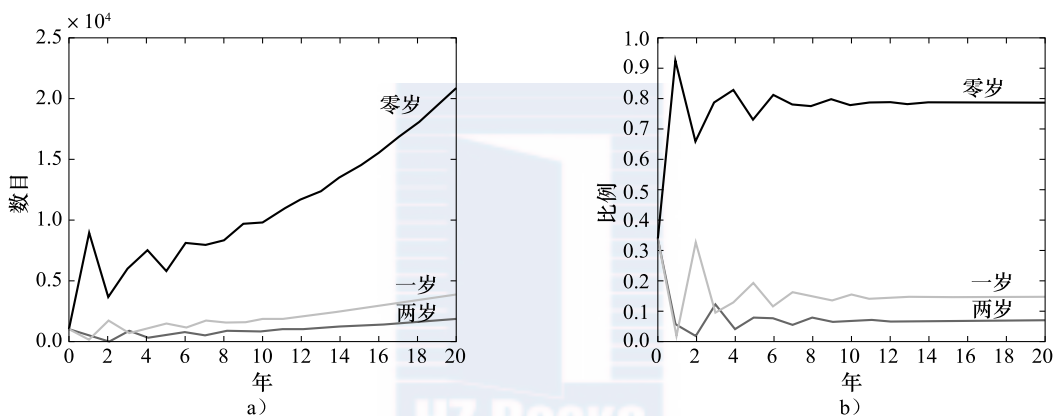


图 1-8 $a_{32} = 0.5$ 及初始数据 $\mathbf{v}^{(0)} = (1000, 1000, 1000)^T$ 时的鸟群总数

为了做得更定量一些, 我们可以对比 $\mathbf{v}^{(20)}$ 和 $\mathbf{v}^{(19)}$ 这两个种群数向量. 可以算出 $\mathbf{v}^{(20)} = (20\ 833, 3873, 1797)^T$, 这表示在第 20 个年头, 每个年龄段的鸟的数目占总数的比例分别为 $(0.7861, 0.1461, 0.0068)^T$. 比较 $\mathbf{v}^{(20)}$ 和 $\mathbf{v}^{(19)}$, 我们发现种群总数在这两年间增长了 1.0760 倍.

注意, $\mathbf{v}^{(n)} = \mathbf{A}\mathbf{v}^{(n-1)} = \mathbf{A}^2\mathbf{v}^{(n-2)} = \cdots = \mathbf{A}^n\mathbf{v}^{(0)}$. 在第 12 章里, 我们将看到上面的这个式子是如何表明这个线性系统的渐近性态(经过长时间后的性态)可以通过矩阵 \mathbf{A} 的主特征值(绝对值最大的特征值)和与之联系的特征向量而被预测出来. \mathbf{A} 的特征值可以被算出是 $1.0759, -0.5380 + 0.5179i, -0.5380 - 0.5179i (i = \sqrt{-1})$. 其中模最大的特征值是 1.0759, 在各分量的和为 1 这一条件的约束下, 相应的特征向量是 $\hat{\mathbf{v}} = (0.7860, 0.1461, 0.0679)^T$. 模最大的特征值决定了经过足够长时间后种群数目的增长率, 特征向量 $\hat{\mathbf{v}}$ 里的元素给出了各年龄段的鸟的数目占种群总数的比例. 像 \mathbf{A} 一样 3×3 矩阵的特征值和特征向量可以通过分析计算得到, 在第 12 章, 我们将学习一些方法来以数值形式模拟更大矩阵的特征值和特征向量.

捕捉策略 (Harvesting Strategies)

我们采用这个模型去研究以下问题: 假设我们为了控制生物数目的指数式增长, 或者

为了利用它们作为食物来源(或者兼有以上两种目的),需要捕捉整个种群中一定比例的成员,那么我们需要捕捉多少,才能既有效控制它们的增长,又不至于让这种生物灭绝呢?

捕捉将会降低生存率.我们希望把生存率降低到一个值上,使得相关问题里矩阵的主特征值的绝对值非常接近 1,这意味着长期来看,这种生物的总数会保持稳定.如果我们把生存率降得太低,使得所有特征值的绝对值都小于 1,那么该生物种群的数量将会指数型减少,以至灭绝.

例如,假设我们捕猎了 20% 的一龄鸟,使得存活率从 0.5 降到 0.3. 这么一来,矩阵 A 的元素 a_{32} 将会从 0.5 降到 0.3,使得最大特征值降到 0.9830. 图 1-9 是这些参数下种群数目变化的模拟图.

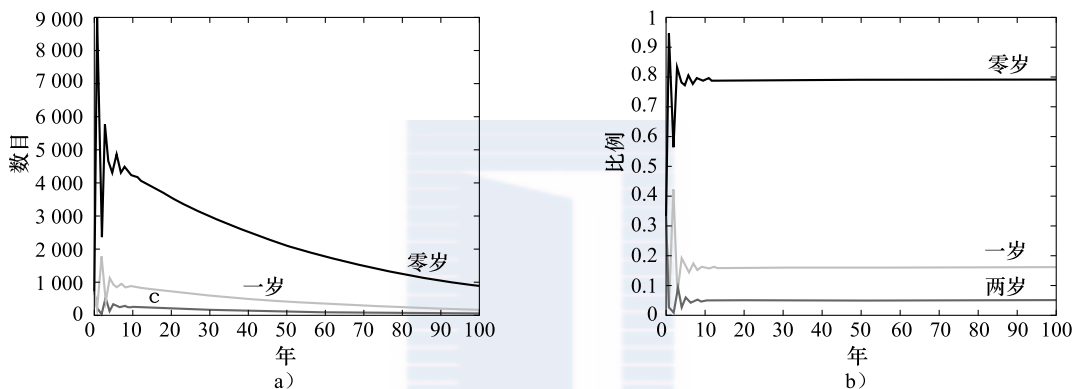


图 1-9 $a_{32} = 0.3$ 及初始数据 $\mathbf{v}^{(0)} = (1000, 1000, 1000)^T$ 时的鸟群总数

注意,此时这种鸟的数量面临指数式的衰减,最终面临灭绝.数目的减少是逐渐的,在 100 年之内,鸟的数目将会从 3000 降到 1148.渐渐地(长期来看),每年鸟的总数会降到上一年的 0.9830. 所以,如果我们花了 100 年的时间,才把鸟的数量降到 1148 只,那么还需要多少年才能把它们数目降到小于 1(也就是灭绝)呢? 这个年数 k 将满足方程 $1148(0.9830)^k = 1$, 经计算得出 $k = -\log(1148)/\log(0.9830) \approx 411$ 年.

人们可以尝试求出 a_{32} 的值使得该问题 A 中模最大特征值大小恰好为 1,从种群数量控制的角度来看,这是一个有趣的数学问题,但它可能会成为徒劳无功之举. 毕竟,上面的这个模型是高度简化的,即使是存活率和繁殖率,也会随着时间变化,而不是像假设那样保持不变. 为了对特别的捕捉策略进行调整以适应环境的变化,必须频繁地监测种群的数目.

1.5 对网络冲浪者和谷歌的建模

向搜索引擎提交查询是信息检索的常用方法. 众多公司为了占据搜索引擎网页上靠前的位置而竞争. 事实上,一些公司的业务就是帮助付费客户使其网页能在被搜索的时候更靠前一些. 这是通过利用搜索引擎算法相关的知识而得以实现的. 虽然有不少关于搜索算法的知识是公开的,但仍有一些仅为公司所有,不为外人所知的部分. 本节将讨论如何基于网页的内容对其排序,本质上是介绍性的. 欢迎感兴趣的读者进一步研究搜索引擎分析的文献,这是一个不断发展的领域. 在这里,我们会考虑一个简单的向量空间模型,用于

执行搜索任务. 这种方法将不考虑万维网的超链接结构, 所以从这个模型得出来的结果应该与考虑网站的超链接结构的算法的结果相结合. 谷歌的 PageRank 算法就是这样的方法, 本节将对其简要讨论, 更详细的叙述见 3.4 节和 12.1.5 节.

1.5.1 向量空间模型

向量空间模型包含了两个主要部分: 文件表和术语的词典. 文件表包含了那些要在上面实施检索的文档, 而词典本身是一个包含各关键字的数据库. 尽管词典的内容可以涵盖每个文档里的所有术语, 但这么做在计算上不现实, 也不为人乐见. 不在词典内的词在被搜索时会返回空值.

当文件表内含 n 个文件, 关键字有 m 个的时候, 向量空间模型构造出一个 $m \times n$ 的文件矩阵 \mathbf{A} ,

$$a_{ij} = \begin{cases} 1 & \text{若文件 } j \text{ 与关键字 } i \text{ 相关} \\ 0 & \text{其他} \end{cases} \quad (1.4)$$

发出查询时, 查询向量 $\mathbf{q} = (q_1, \dots, q_m)^T$ 生成, 如果向量含有第 i 个关键字, 那么 $q_i = 1$, 否则 $q_i = 0$. 这个查询和每个文档的接近度可以由查询向量 \mathbf{q} 与 \mathbf{A} 中代表该文档的列向量的夹角的余弦值来表示. 令 \mathbf{a}_j 表示 \mathbf{A} 的第 j 列 (若 \mathbf{a}_j 中元素不全为零), 则 \mathbf{q} 与 \mathbf{a}_j 之间的夹角 θ_j 满足

$$\cos(\theta_j) = \frac{\mathbf{a}_j^T \mathbf{q}}{\|\mathbf{a}_j\|_2 \|\mathbf{q}\|_2} \quad (1.5)$$

11

由于两向量的夹角越小, 这个夹角的余弦值就越大, 所以文件表要按照与查询内容的相关性来排序, 也就是按这个余弦值降序排列.

表 1-1 应用在三维例子中的文件和词典

词典		文件
电子化	I	战争的艺术
击剑	II	剑术大师
花剑	III	花剑、重剑和佩剑的击剑技术
	IV	构建电子化击剑的热点提示

构造文件矩阵 \mathbf{A} 的主要困难在于如何去判断一个文档与给定的关键字是否相关. 不过完成这个工作的方法有不少. 例如, 在每一个文档中搜索每一个关键字. 如果这个文档包含了某个关键字, 那么我们就认为它和该字是相关的. 当文档和关键字的数目很大

时, 这种做法在计算量上将会付出很大的代价. 一些搜索引擎采用了替代的方法, 就是只读入和分析每个文档的一部分内容. 比如谷歌和雅虎只从每个网页文本里读入 100~500KB 来分析^[12]. 其他要做的选择包括是寻求精确匹配还是允许同义词出现在搜索结果里, 以及是否考虑到词序. 比如我们认为 boat show 和 show boat 这两个词组, 究竟是相同, 还是不同呢? 所有的这些选择对拣选出最相关的文档都有很大的影响^[36]. 为了说明这个方法, 我们举一个简单的例子, 只要文件标题包含关键字, 它就被视为相关.

在小空间内检索

假设我们的词典只含有三个关键字: “electric”, “fencing” 和 “foil”. 文档也只有四个, 分别为 “the art of war”, “the fencing master”, “fencing techniques of foil, epee, and saber” 和 “hot tips on building electric fencing”. 这些文档的标题和词典里的各个单词都用小写字母来表示, 以免产生大小写匹配上的问题. 为便于参考, 这些信息都记录

在表 1-1 里.

为构造矩阵 \mathbf{A} ——第 1、2、3 行分别对应了 “electic”, “fencing” 和 “foil” 3 个词, 而第 1、2、3、4 列对应着文档 I、II、III、IV, 采用式(1.4)所述的对应法得出矩阵

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

图 1-10 以三维空间的向量形式描绘了 \mathbf{A} 的各列(除了第一列之外, 其他列都被归一化成长度为 1).

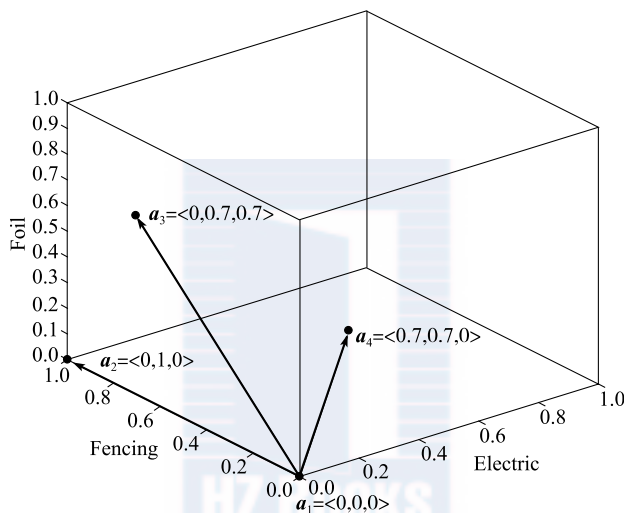


图 1-10 三维向量空间搜索例子的直观形象化

假设我们要检索 fencing 这个词, 那么, 检索向量 $\mathbf{q} = (0, 1, 0)^T$, 我们发现它正好是 \mathbf{A} 的第 2 列. 所以文档 II 可以被判定是与检索内容相关度最大的. 为确定剩下文档的相关次序, 我们利用式(1.5)计算 \mathbf{q} 和 \mathbf{A} 的每个非零列之间夹角的余弦, 得出

$$\cos(\theta_3) = \frac{\mathbf{a}_3^T \mathbf{q}}{\|\mathbf{a}_3\|_2 \|\mathbf{q}\|_2} = \frac{1}{\sqrt{2}}$$

$$\cos(\theta_4) = \frac{\mathbf{a}_4^T \mathbf{q}}{\|\mathbf{a}_4\|_2 \|\mathbf{q}\|_2} = \frac{1}{\sqrt{2}}$$

所以文档 III 和 IV 的相关度是一样的, 排于文档 II 之后.

1.5.2 谷歌的 PageRank 算法

现实世界中的搜索引擎要处理在计算领域中一些最大的数学和计算机科学问题. 有趣的是, “Google” 是单词 “googol” 的变形, 后者表示 10^{100} , 这反映了该公司的目标是把万维网上的所有信息组织到一起.

当你向 Google 提交一个检索(比如查询 numerical analysis)的时候, 搜索引擎将如何在排第一的网页和其他(比如排第 100)的网页中做出区分呢? 有许多因素影响选择, 一个因

素是网页和检索内容的关联,前面已讨论过,另一个重要因素则是网页的“质量”和“重要程度”。“重要”的网页,通常被许多其他网页指向,从而很可能抓住你的目光.一个叫PageRank的算法用来衡量网页的重要程度,它基于一个网络冲浪的模型。

同任何其他关于现实的模型一样,谷歌也是用近似的方法对网络冲浪建模的。PageRank的一个重要功能就是它对冲浪者通过链接,从当前页面跳到另一个页面的百分比的假设。具体的假设是公司的内部数据,但通常认为谷歌估计出,冲浪者通过点击超链接而打开页面的概率是85%,这种做法所打开的链接没有倾向性;而其他15%的情况是冲浪者通过键入URL地址来打开另一网页,很可能是重新载入当前浏览的页面。

这些关于冲浪行为的假设不仅影响了模型的准确性,还影响了用来对这个模型求解的数值技术的效率。注意,谷歌为数以十亿计的网页建立了索引,使得与此相关的计算成了有史以来解决过的最大规模的计算问题!第12章将要更多地讨论解决这类问题的数值方法。

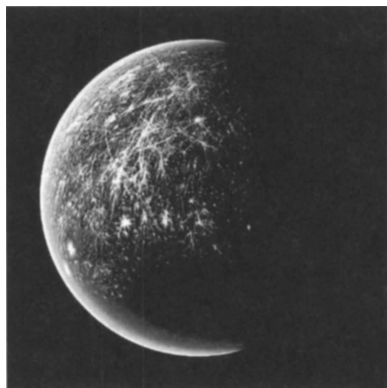


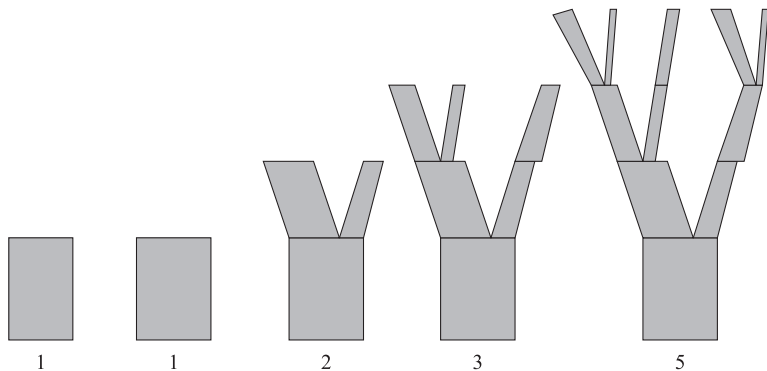
图1-11 由David F. Gleich基于OPTE项目提供的图像和数据绘制的因特网示意图

1.6 第1章习题

1. 斐波那契数及其性质. 定义斐波那契数列 $F_1=1, F_2=1, F_3=F_2+F_1=2, F_4=F_3+F_2=3, \dots$ 一般地, $F_{j+1}=F_j+F_{j-1} (j=2, 3, \dots)$. 写下 F_5, F_6 和 F_7 .

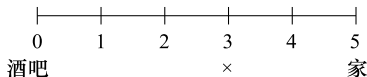
人们经常观察到,一朵花的花瓣数或者一株树的枝条数是一个斐波那契数。例如,大多数雏菊有34, 55或89个花瓣,这些数字分别是第9、第10和第11个斐波那契数。因为4不是斐波那契数,所以四片叶子的三叶草相当稀少。

为什么呢?考虑下面这个树枝生长的模型。我们从主干开始计算($F_1=1$),经历了一个季度的生长($F_2=1$),在第三个季度长出了两个分支——一根较粗,另一根较细。在接下来的一季中粗的那根枝条又长出了一粗一细两个分支,而细的枝条长粗了,准备在下一季才开始分支。所以,此时树上有 $F_4=3$ 根枝条——两粗和一细。当下一季结束时,这两根粗枝分叉了,变成了两粗两细一共四条枝条,而细的那根枝条长粗了,所以此时有 $F_5=5$ 根枝条。而有 $F_4=3$ 根粗枝在下一季将分叉开来。试解释由这个树枝生长模型产生斐波那契数的原因。



2. 醉鬼行走问题. 一个醉鬼在下图的点 x 处开始行进。他每走一步,往右的概率是0.5,往左的概率也

是 0.5. 如果他走到了酒吧, 他就会待在那里, 喝醉为止. 如果他能走到家, 就会上床睡觉. 你想知道他在到达酒吧之前就回到了家的概率 $p(x)$.

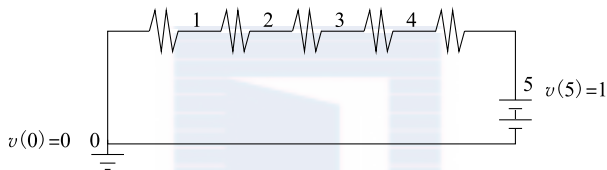


这是一个典型的马尔可夫链问题, 后面将会讨论它. 注意, $p(0)=0$, $p(5)=1$, 因为在这两种情况下该醉鬼都没有必要再移动了. 对于 $x=1, 2, 3, 4$, $p(x)=0.5p(x-1)+0.5p(x+1)$, 因为他下一步往左走或向右走都有一半概率.

- (a) 让该醉鬼在 $x=3$ 处开始, 那么一步之后他又可能站在哪个点? 他一步之后走到这些点的概率分别是多少? 如果他走了两步, 又会怎样?
- (b) 对于每个内点 x , 你认为在这个点上开始步行的醉鬼更有可能到达酒吧还是回到家? 为什么?

这是一个典型的随机步行问题, 虽然它听起来挺傻的, 但它有着实际的物理运用.

考虑一个把相等的电阻串联起来的电路, 所有电阻两端的电压是 1V.



电压函数 $v(x)$ 在 $x=0, 1, 2, 3, 4, 5$ 这几个点上有定义. 把 $x=0$ 这点接地, 使得 $v(0)=0$, 而在 $x=5$ 处和电源之间并没有电阻, 所以 $v(5)=1$. 由基尔霍夫定律, 流入 x 的电流必须和流出 x 的电流相等. 由欧姆定律, 如果点 x 和 y 被值为 R 的电阻隔开, 则通过这两点的电流

$$i_{xy} = \frac{v(x) - v(y)}{R}$$

所以对 $x=1, 2, 3, 4$, 有

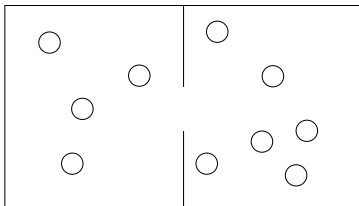
$$\frac{v(x-1) - v(x)}{R} = \frac{v(x) - v(x+1)}{R}$$

两边乘以 R , 合并同类项, 得到 $v(x)=0.5v(x-1)+0.5v(x+1)$. 这和以 $p(x)$ 为未知函数的醉鬼走路问题的方程是完全一致的, 连边值条件也一样.

你能想到其他也用这个方法建模的问题吗? 或许股票价格变动的行为可以? 如果你推广了醉鬼走路的问题, 让他往左走和往右走的概率不同, 比如有 0.6 的概率向右, 有 0.4 的概率向左, 那么怎样推广电阻问题使之与这个醉鬼走路的问题相关呢? [提示: 考虑阻值不一样的电阻。]

3. 埃伦费斯特的瓮. 考虑两个瓮, 里面一共分布着 N 个小球. 每个单位时间里, 你从某个瓮里取出一个球, 将它移动到另外一只瓮里, 选择每个球的概率是相等的, 都是 $1/N$. 所以被挑中的球在给定的瓮中的概率和这个瓮里球的个数成正比. 用 $X(t)$ 表示时间为 t 时左边瓮中的球数, 设 $X(0)=0$, 这样 $X(1)=1$, 因为一开始只能在右边选择一个小球移到左边去.

- (a) 令 $N=100$, 那么 $X(2)$ 、 $X(3)$ 、 $X(4)$ 的可能值是多少? 它们取得这些可能值的概率分别各是多少?
- (b) 如果搬运持续了非常长的时间, 那么你认为 $X(t)$ 最有可能取到的值是多少?



这个模型是 20 世纪早期由 Paul Ehrenfest 和 Tatiana Ehrenfest-Afanassjewa 引入的, 描述了气体分子穿过半透膜的扩散过程. 例如, 我们可以查看位于 http://en.wikipedia.org/wiki/Ehrenfest_model 上关于该模型和热力学第二定律之间关系的讨论. 第 3 章将要讨论物理过程中的蒙特卡罗模拟, 其采用的模型与这里相似. 你能想到其他或许能用这样的过程来建模的状况吗? 除了物理和化学过程之外, 你也许应该考虑譬如经济决策和社会制约之类的过程.

4. 每年都有本科生参加数学建模竞赛(MCM). 请浏览 www.mcm.org. 下面是竞赛参加者需要解决的数学建模问题的一个范例:

在一个周围有众多建筑物环绕的大露天广场上有一个观赏喷泉, 它把水喷到空中. 在刮大风的日子, 风会把喷泉的水吹到路人身上. 喷泉的流量由和测速仪(测量风速和风向)相连的机关控制, 而测速仪放置在相邻建筑物的顶部. 进行这种控制的目的是让路人在欣赏诱人景观和被淋透之间达成一个平衡. 风越大, 喷出的水量应该就越少, 水柱的高度就越低, 使得溅出喷池的水量也越少.

你的任务是设计一个算法, 它能采用测速器的各项数据, 当风的速度和方向改变时, 对喷泉的水流进行调节.

想想看你应该怎样构造这个问题的数学模型, 比较一下你的解决方案和其他学生的解决方案, 他们给的解法可以在 *UMAP: Journal of Undergraduate Mathematics and its Applications*, 2002, 23(3): 187-271 中找到.

5. 考察由以下关键字构成的词典:

巧克力、冰淇淋、碎末

和如下文档表:

D1: 我只吃了蛋糕外的巧克力

D2: 我喜欢巧克力和香草冰淇淋

D3: 孩子们喜欢撒有碎末的巧克力蛋糕

D4: 如果你既有撒在上面的碎末又有巧克力汁, 我可以再来一勺冰淇淋吗?

构造文档矩阵 A , 采用向量空间模型按照它们与查询词“巧克力”, “冰淇淋”的相关性, 对 D1、D2、D3 和 D4 进行排序.

6. 当你向搜索引擎提交一个请求时, 一个有序的网页列表会回复给你.

网页按照与你查询内容的相关性和质量进行排序. 考虑如图 1-12 所示的由网页构成的小型有向图. 假设这是你的搜索引擎产生过索引的网页的集合. 每个顶点表示一个网页. 如果网页 i 有指向网页 j 的链接, 那么表示 i 的点和表示 j 的点之间将由一条有向线段连接. 例如, 我们看到网页 1 就有一个指向网页 4 的链接. 由 Larry Page 和 Sergey Brin 提出的 PageRank 算法^[18]假设网络冲浪者在 85% 的情况下通过点击链接打开另一个网页, 每个网页被打开的概率是一样的; 剩下 15% 的情况里, 冲浪者会键入 URL 地址, 很可能是当前正在访问的网页, 其他的网页被打开的概率也是相同的. 令 $X_i(t)$ 表示冲浪者在 t 步之后来到第 i 个页面上的概率. 假设他从第 1 个页面开始浏览, 那么 $X_1(0) = 1$, $X_i(0) = 0$, $i = 2, 3, 4, 5$.

(a) 当 $1 \leq i \leq 5$ 时, 算出 $X_i(1)$.

(b) 当 $1 \leq i \leq 5$ 时, 算出 $X_i(2)$.

作为对网页质量的衡量, 对所有 i , PageRank 会近似给出 $\lim_{t \rightarrow \infty} X_i(t)$ 的值.

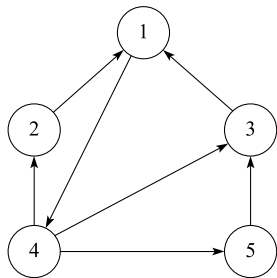


图 1-12 网页的一个小型网络