理性派

身边的数学译丛

谁排第一？

关于评价和排序的科学

[美] Amy N. Langville
Carl D. Meyer 著　　郭斯羽 译



机械工业出版社
CHINA MACHINE PRESS

Who's #1?: The Science of Rating and Ranking

“兰维尔和梅耶带领我们，在评分方法这片风景秀美的土地上进行了一次缜密而又轻松的旅程。这是一次通过体育的透镜来了解评分方法的愉悦的阅读体验，而且也说明了评分是如何通过影片、网络搜索、在线购物和其他应用来影响我们的日常生活的。”

——克里斯·沃林斯基（Chris Volinsky），Netflix奖获胜队成员

“《谁排第一？》是对评分体系背后的基本数学思想的一篇出色综述。排名曾是属于体育爱好者的一个领域，而现在它正越来越成为许多信息时代的应用领域中的不可或缺的工具。兰维尔和梅耶比较了多个模型，并凸显了它们之间的差别，说明了各个模型的数学基础和产生的诱因。本书将会激发读者去对这个激动人心的领域进行进一步的探索。”

——肯尼斯·梅西，梅西评分法的发明者

“《谁排第一？》是一篇正当其时的综述，介绍了对诸如体育队伍、影片、政治家等等事物进行排名和评分中所用的方法。人们对于这一话题有着极为浓厚的兴趣，而读者们现在终于有了一个去处，可以获得不同方法的综合性了解。”

——韦恩·L·温斯顿，《数学和体育：赌徒、经理和体育迷们是如何在棒球、篮球和橄榄球中使用数学的》一书的作者

“这本非常易于理解的著作，为读者提供了跨越许多不同学科的不同数学排名方法的综合介绍，对于从研究者到体育统计数据瘾君子在内的每个人而言，它都会是引人入胜的。”

——塞普·卡姆瓦尔（Sep Kamvar），《自组织信息网络中个性化搜索的数值法》一书的作者

关注 理性派_数学和科学 新浪微博

ISBN 978-7-111-45932-3 定价：49.00元

身边的数学译丛

谁排第一？
关于评价和排序的科学

（美）Amy N. Langville
Carl D. Meyer 著
郭斯羽 译

机械工业出版社

本书是首个关于评分和排名科学的著作。它是搜索排序姊妹篇的第二本。本书主要内容有：排名概述、梅西法、科利法、基纳法、埃洛体系、马尔可夫法、攻防评分法、基于重新排序的排名方法、分差、用户偏好评分、处理平局、加入权重、“假如……会怎样”的问题与敏感性、排名聚合、比较排名的方法、数据等。

本书可作为数学、计算机、网络技术、管理学和数据科学等专业的参考书，也可作为教材使用。

Who's#1? The Science of Rating and Ranking
Amy N. Langville & Carl D. Meyer
Copyright © 2012 by Princeton University Press
All Rights Reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the Publisher.

北京市版权局著作权合同登记号：01-2012-6410

图书在版编目（CIP）数据

谁排第一？关于评价和排序的科学/（美）兰维尔（Langville，A. N.），（美）梅耶（Meyer，C. D.）著；郭斯羽译．—北京：机械工业出版社，2014.5
（身边的数学译丛）
ISBN 978-7-111-45932-3

Ⅰ.①谁… Ⅱ.①兰… ②梅… ③郭… Ⅲ.①排序 Ⅳ.①0223

中国版本图书馆 CIP 数据核字（2014）第 031781 号

机械工业出版社（北京市百万庄大街 22 号 邮政编码 100037）
策划编辑：韩效杰 责任编辑：韩效杰
版式设计：常天培 责任校对：赵 蕊
封面设计：路恩中 责任印制：乔 宇
保定市中画美凯印刷有限公司印刷
2014 年 6 月第 1 版第 1 次印刷
184mm × 260mm · 16.5 印张 · 1 插页 · 392 千字
标准书号：ISBN 978-7-111-45932-3
定价：49.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务	网络服务
社服务中心：(010)88361066	教材网：http://www.cmpedu.com
销售一部：(010)68326294	机工官网：http://www.cmpbook.com
销售二部：(010)88379649	机工官博：http://weibo.com/cmp1952
读者购书热线：(010)88379203	**封面无防伪标均为盗版**

致约翰，

那个我无需任何算法就知道的、过去是、现在是、
而且将来也一直都会是我的第一的人，

以及

致贝萨妮，
得分 = 10，排名 = 1

前　言

“在得到一个漂亮的答案之前，总会有一个更漂亮的问题。”——E. E. 卡明斯（E. E. Cummings）

目的

在更早些时候撰写《网络排名 PR 值及其他——搜索引擎排名的科学》（Google's PageRank and Beyond: The Science of Search Engine Rankings）[49]时，评分和排名方法的威力激起了我们的好奇心。在写作那本“谷歌之书”的时候，我们开始领略到评分和排名的这一领域是多么丰富（和古老），而它又是如何远远地超出了万维网搜索这一应用领域。但要了解这一主题的不同侧面则是枯燥乏味的，因为我们无法找到一个方便的资料源，能够将大量方法和应用干净利索地打成一包。由此便诞生了现在的这本书。它将分散于众多单独成篇的论文、网站与其他来源的相关信息集中于一处，而且除此之外，我们还给出了数个有关评分和排名的新想法，这些想法之前还从未被发表过。我们的目标是通过充足的背景知识以及内容的多样性，来使读者能够逐步去欣赏评分和排名的通用技巧，并准备好去探索超出本书内容之外的更多方法和应用。

读者

正如在**第 6 页**上的第 1 章结尾处开列的清单所指出的那样，排名的应用极多且极为不同。因此，有许多不同类型的读者可能对这一主题产生兴趣，并进而对本书产生兴趣。比方说，这些读者可以是体育迷，或是社会选择方面的理论家，或是数学家、计算机科学家、工程师，也可以是大学与中学中讲授诸如线性代数、最优化、数学建模、图论或其他专门主题的课程的教师，此外更不用提那些对任何能够想象得到的东西都乐意下赌注的人了。

阅读本书的前提要求

在撰写本书的大部分内容时，我们都假定读者曾经读了基础线性代数的课程，另有几章还假定读者具备最优化方面的若干知识。如果您没有上过这类课程，那我们有几点建议。

1. 无论如何，您应该通读全书。书中的每个基于线性代数的方法都被归结为这门课程的一两个最基本的要点：求解一个线性方程组系统，或计算特征向量。即使您不了解这些基本要点，您仍然还是可以实现这些方法。例如，现在已经有免费或付费的

软件来求解线性方程组系统或计算特征向量，而您只需要了解需要输入哪些数据以及如何解读输出就可以了——至于您所用的软件内部是如何做到这些的，则并不需要您去了解。

2. 当然，如果能真正理解这些课程，将能够提高您为特定目的去实现、修改和调整这些方法的能力，因此您也许会愿意考虑一下通过在线教程或经典著作来进行自学。我们最为喜爱的线性代数著作是参考文献［54，76，40］，但如果您全然是一个新手，那就请针对您所疑问的内容，去搜索一下相关的在线基础教程（这样的教程有许多）。几本我们最为喜爱的最优化方面的著作是参考文献［82，60，10，84，85］，在此按内容由浅至深列出。

3. 我们所推荐的最后也是最为理想的方式，则是修读一门基础线性代数的课程，并且在时间和金钱许可的条件下，最好能再修读一门最优化的课程。

利用本书进行教学

我们估计可能会有大学或中学教师使用本书作为专题或数学建模课程中的一个短期的、1～2 节课的单元模块，或者是用于整门课程的教学。在前一种情况下，教师们应该能够发现，他们可以很容易地从本书中抽取出一个短期模块所需的素材。实际上，我们已经考虑到了这一点，并将素材进行了“模块化处理”——即书中各章的内容很少依赖于之前所用的素材。书中各章几乎可以按任意顺序排列。例如，一位线性代数的教师也许希望增加一节有关线性系统应用的课，那么有关科利（Colley）法的第 3 章就是一个很好的选择，该方法是本书中最为简单也最为直接的线性系统方法。同理，在特征值和特征向量理论的课后，教师可以选择第 4 章中的材料来上一堂有关特征系统应用的课。第 6 章中包含了有关马尔可夫链的素材，这一内容常常出现在线性代数课程将要结束的时候。

致谢

- 查尔斯顿学院（College of Charleston）。卢克·英格拉姆（Luke Ingram）和约翰·麦康奈尔（John McConnell）这两位前理科硕士生在 2006 年春完成了一个课程设计项目，对“三月疯狂（March Madness）”篮球巡回赛中的队伍进行了排名。卢克·英格拉姆将这个课程设计项目扩充成为一篇优秀的论文[41]，其中包含了若干有趣的排名思路。之后在 2008 年春，两名数学专业的本科生尼尔·古德森（Neil Goodson）和科林·斯蒂芬森（Colin Stephenson）也着手进行了同样的课程设计，对当年的“三月疯狂”巡回赛的比赛进行了预测。尼尔和科林使用了本书的一个非常初步的草稿，并测试了其中数个排名模型。在 2008 年的巡回赛中，他们在比赛预测方面做出了非常了不起的成绩，以至于赢得了数量惊人的全国性报导。请见**第 151 页**和**第 212 页**上的杂谈。拥有数学学士学位的研究生凯思琳·佩丁斯（Kathryn Pedins）在硕士生阶段，仍然继续专注于线性排序这一学位论文课题的研究。她收集了在书中数个杂谈里所使用的数据，参与了多次集体研讨和讨论会，表现出了作为助手的非凡之处。凯思琳的大部分工作出现在第 8 章和第 15 章。最后，查尔斯顿学院排名和聚类研究组的所有学生都参与了研究讨论周

会，这些周会对本书的成形颇有帮助，因此，本书的第一作者在此对埃米·道格拉斯（Emmie Douglas）、伊拜·巴沙贝（Ibai Basabe）、芭芭拉·波尔（Barbara Ball）、克莱尔·罗杰斯（Clare Rodgers）、赖恩·帕克（Ryan Parker）和帕特里克·莫兰（Patrick Moran）表示感谢。

- 北卡罗来纳州立大学（North Carolina State University）。作为其论文《排名理论及其在大众体育中的应用》（Ranking Theory with Applications to Popular Sports）[34]的一部分，安吉拉·戈万（Anjela Govan）为攻防评分（Offense-Defense rating）理论[35]的发展做出了贡献，她建立了 OD 法和矩阵平衡（matrix balancing）的辛克霍恩-诺普（Sinkhorn-Knopp）理论之间的联系，并对其进行了正式化。此外，她在万维网检索和数据采集方面所做的工作，也成为了许多评分和排名方法的实验的基础，而其中的一些就包括在了本书之中。评价表现不佳的方法，与发现性能不错的方法具有同样的重要性。我们将本书局限于那些具有突出优点的方法之上，因此读者可能看不到安吉拉的许多重要贡献，但她的工作的确极大地提高了本书的质量。查尔斯 D（查克）. 韦塞尔［Charles D.（Chuck）Wessell］在撰写本书时是北卡罗来纳州立大学的一名研究生，现在则是宾夕法尼亚州葛底斯堡的盖茨堡学院（Gettysburg College）的一名数学教员。查克给了我们许多有用的建议，而我们也从他那鹰一般锐利的眼睛中受益良多。他仔细阅读了草稿，并认真地检查了数据［特别是美国橄榄球联盟（NFL）的数据］，改正了多个错误。此外，查克还在一门本科生课程中使用了本书中的材料，而这一课堂的经验帮助我们对文中的阐述进行了进一步的打磨。
- 同事。当大卫·格莱克（David Gleich）在查尔斯顿学院访问期间，我们和他所进行的讨论对第 14 章和第 16 章中的某些内容产生了良好影响。戴维逊学院（Davidson College）的蒂莫西·夏提埃（Timothy Chartier）与第一作者定期通信，而他与他的学生埃里希·克罗伊策（Erich Kreutzer）开展的项目，以及他们对本书早期草稿的反馈，都包含着深刻的见解，特别是由这一合作所带来的第 11 章中关于平局的讨论。卡森-纽曼学院（Carson-Newman College）的肯尼斯·梅西（Kenneth Massey）创建了一个用于体育排名的巨型数据库和网站。本书中几乎所有体育方面的示例都应归功于梅西博士和他的数据。我们感谢他的慷慨、友善、知识以及他在计算机方面的专业技能。
- 资助。第一作者的部分工作由国家科学基金 CAREER 奖励 CCF-0546622 所资助，它资助了合作旅行以及一个短期的学术休假以完成和校订书稿。此外，第一作者还要感谢查尔斯顿学院所提供的支持与培养氛围。从第一天起，大学、学院和系对她的事业和研究都极为热情与支持。她特别要感谢前任与现任系主任迪安娜·卡夫尼（Deanna Caveny）和鲍勃·米尼奥内（Bob Mignone），以及前任院长努南（Noonan），感谢他们富有创造性的领导以及热情的支持。
- 照片。感谢美国科学技术奖章委员会与赖恩 K. 莫里斯（Ryan K. Morris）允许重印**第 4 页**上的 K. 阿罗（K. Arrow）的照片。感谢肯尼斯·梅西允许重印**第 9 页**上他的照片，并感谢詹姆斯·基纳（James Keener）允许重印**第 51 页**上他的照片。**第 92 页**上 J. 克莱恩伯格（J. Kleinberg）的照片由迈克尔·奥科涅夫斯基

（Michael Okoniewski）提供。感谢安吉拉·戈万允许重印**第 94 页**上她的照片。感谢查尔斯顿学院提供**第 152 页**上尼尔·古德森和科林·斯蒂芬森的照片。感谢凯思琳·佩丁斯和山本洋右[⊖]（Yoshitsugu Yamamoto）提供了**第 153 页**上他们的照片。

⊖ 按英文音译。——译者注

目　录

前言 Ⅳ

第 1 章　排名概述 **1**

社会选择与阿罗的不可能性定理 3

阿罗的不可能性定理 4

一个小型范例 4

第 2 章　梅西法 **9**

原始的梅西评分方法 9

梅西的主要思路 9

在范例上应用梅西评分法 11

梅西评分法的高级特性 11

范例：使用高级梅西评分法 12

梅西评分法的总结 13

第 3 章　科利法 **21**

范例 23

科利评分法总结 24

梅西法和科利法之间的联系 24

第 4 章　基纳法 **29**

实力和评分的规则 29

选择体现实力的属性 29

拉普拉斯的承续法则 30

有所偏好，还是不要偏好 31

规范化 32

先有鸡，还是先有蛋 33

评分 33

实力 33

基本方程 34

约束 35

佩隆-弗罗贝尼乌斯 36

重要的性质 37
计算评分向量 37
强制形成不可约性和素性 39
总结 40
NFL 2009—2010 赛季 42
吉姆·基纳对比尔·詹姆斯 45
回到未来 48
基纳能让你致富吗 49
结论 50

第 5 章　埃洛体系 **53**
优美的智慧 55
K 因子 55
逻辑斯谛参数 ξ 56
定常的和值 56
NFL 中的埃洛评分 57
事后预测的准确率 58
事前预测的准确率 59
结合比赛得分 59
$\xi = 1000$，$K = 32$，$H = 15$ 时的事后预测和事前预测 60
对 NFL 比分使用可变的 K 因子 60
使用比分和可变 K 因子的事后和事前预测 62
逐场比赛分析 62
结论 64

第 6 章　马尔可夫法 **67**
马尔可夫法 67
利用胜负进行投票 68
利用负者的分差进行投票 69
胜负双方都按失分进行投票 70
超越比赛得分 71
处理全胜的队伍 73
马尔可夫评分方法总结 75
马尔可夫法和梅西法之间的联系 76

第 7 章　攻防评分法 **79**
攻防评分法的目的 79
OD 法的前提假设 79
但先评定哪一个呢 80
交替求精过程 81

分离 81
结合两个 OD 评分 82
再次回到范例 82
对比得分和码数 83
NFL 2009—2010 赛季的 OD 评分 84
OD 法的数学分析 87
对角线 88
辛克霍恩-诺普 89
OD 矩阵 89
OD 评分和辛克霍恩-诺普 90
作一点弊 91

第 8 章　基于重新排序的排名方法 97
排名差距 98
范例 99
求解最优化问题 101
松弛问题 103
进化方法 103
高级排名差距模型 105
排名差距法的总结 106
排名差距法的性质 106
评分差距 107
范例 109
求解重排问题 110
评分差距法总结 111

第 9 章　分差 113
它是什么（又不是什么） 113
抽头 114
为什么不简单地赌胜负 114
分差是如何操作的 114
赌中分差 115
超/欠赌 115
为什么使用评分难以预测分差 116
利用分差来构建评分（并预测分差） 117
NFL 2009—2010 赛季的分差评分 120
几场决斗 121
其他成对比较 124
结论 125

第 10 章　用户偏好评分 **127**
直接比较 129
直接比较、偏好图和马尔可夫链 130
重心法与马尔可夫链方法对比 132
结论 133

第 11 章　处理平局 **135**
输入平局与输出平局 136
加入平局 136
科利法 136
梅西法 137
马尔可夫法 137
OD 法、基纳法和埃洛法 138
扰动分析给出的理论结果 139
真实数据集上的结果 140
影片排名 140
NHL 冰球队排名 141
诱导平局 142
总结 144

第 12 章　加入权重 **147**
四种基本的加权方案 147
加权梅西法 149
加权科利法 150
加权基纳法 150
加权埃洛法 150
加权马尔可夫法 150
加权 OD 法 151
加权的差距法 151

第 13 章　“假如……会怎样”的问题与敏感性 **155**
秩一更新的影响 155
敏感性 156

第 14 章　排名聚合——第 1 部分 **159**
重温阿罗准则 160
排名聚合方法 163
波达计数法 165
平均排名 166
模拟比赛数据 167

排名聚合的图论方法 172
排名聚合后的一个精化步骤 175
评分聚合 176
由评分聚合矩阵得到评分向量 178
聚合方法总结 181

第 15 章　排名聚合——第 2 部分 183
范例 185
求解 BILP 186
BILP 的多最优解 187
BILP 的 LP 松弛 188
约束松弛 190
敏感性分析 191
限界 191
最优化排名聚合方法总结 193
重温评分差距法 194
评分差距与排名聚合的对比 194
范例 196

第 16 章　比较排名的方法 201
两个排名列表间的定性差异 201
肯德尔测度 τ 203
完全列表上的肯德尔测度 τ 204
部分列表上的肯德尔测度 τ 205
完全列表上的斯皮尔曼加权简捷测度 206
部分列表上的斯皮尔曼加权简捷测度 207
不等长的部分列表 210
评判标准：与已知的标准相比较 211
评判标准：与聚合列表相比较 211
回溯打分 212
事前预测 212
学习曲线 214
与坡形之间的距离 214

第 17 章　数据 217
梅西的体育数据服务器 217
波默罗伊的大学篮球数据 218
搜集你自己的数据 218
生成成对比较矩阵 220

第 18 章　后记 **223**
层次分析法（AHP） 223
雷德蒙法 223
朴-纽曼法 224
逻辑斯谛回归/马尔可夫链法（LRMC） 224
霍赫鲍姆法 224
蒙特卡洛仿真 224
纯粹统计分析 225
还有更多更多 225

词汇表 **231**

参考文献 **235**

索引 **241**

第 1 章　排名概述

我们首次接触到排名这一领域时，正好是在新的千年即将到来之际。如今看来影响深远的 1998 年的论文《搜索引擎分析》（Anatomy of a Search Engine）[⊖]使用了马尔可夫链这一我们最为喜爱的数学工具来对网页进行排名。两个名不见经传的研究生利用马尔可夫链改进了搜索引擎的排名方法，而他们的方法是如此成功，以至于这一方法成为了他们新公司的基础，而这家公司很快就成为了尽人皆知的搜索引擎谷歌。我们阅读的有关排名的资料越多，我们就越来越被这一领域所吸引。从那以来，我们已经就排名这一主题撰写了一本著作《网络排名 PR 值及其他——搜索引擎排名的科学》，此外还有数十篇论文。2009 年，我们甚至主办了首届一年一度的美国东南地区的排名和聚类会议。

作为应用数学家，我们被排名科学所吸引是很自然的，但我们相信它对于一般人而言也同样具有吸引力。我们之所以相信这一点，是出于以下原因。

- 排名问题的简单性十分优美——不过就是将一组事物根据其重要性加以排列——但这个问题的一些解答却是复杂的，而且充满了悖论和谜题。我们将在本章的后续部分以及本书通篇中，为读者介绍若干这类能够激发人们好奇心的难题。
- 排名问题有着悠久而辉煌的历史，它最早可以追溯到 13 世纪。散布于全书的杂谈介绍了排名领域中的某些主要且更为多姿多彩的参与者；
- 现在看来，就研究的活跃程度和研究者的兴趣而言，排名问题的发展已经达到了一个顶峰。不断增长的兴趣的一个起因，来自于现在我们所具有的数据收集能力。从房地产和股票市场到体育比赛与政治再到教育和心理学，在几乎每个可以想象到的领域中，都不缺少有趣而实际的数据集。人们对排名的兴趣不断增长的另一个起因则与许多工业化国家所培养出的文化趋势有关，特别是在美国，我们迷恋于评估，而由于排名和评估之间的密切关系，我们便也因此迷恋上了排名。每天你都可能与成打的排名发生直接接触。例如，当早上开车时，你可以从广播中听到排名前十的歌曲名单。之后你收到了由你的老板或老师给出的量化季评（并由于得分落在了那 95% 之中而欢欢喜喜地离开）。在午餐时间，你瞟了一眼报纸上的体育栏目，发现了你最喜爱的队伍的排名。再晚一点，在下午的休息时

⊖ 似应为《大规模超文本网页搜索引擎分析》（The Anatomy of a Large-scale Hypertextual Web Search Engine）的简称。——译者注

间，你上网去查看自己的虚拟联赛球队的排名。与此同时，在家中，你的信箱里已经放着你所请求的下一部Netflix电影——感谢邮局以及该公司复杂的排名算法。最后一件但却可能是做得最多的一件事，便是你每天都要通过搜索引擎帮助你寻找信息、购买商品或下载文件。只要你使用了这些虚拟机器中的任何一个，你便依靠了先进的排名技术。排名无处不在，以至于它甚至出现在了最近的 xkcd 漫画[⊖]中。

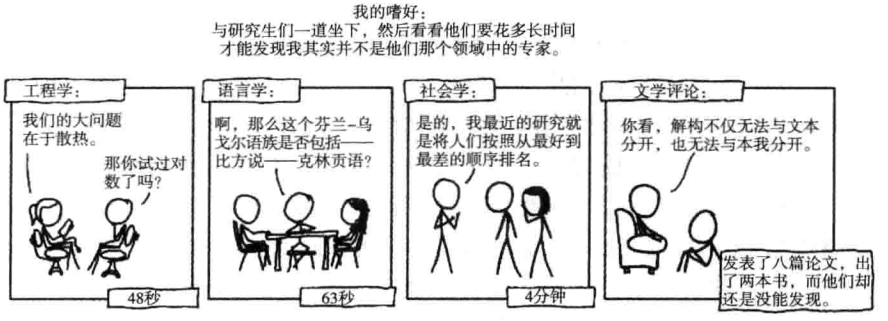


图 1.1　关于排名的 xkcd 漫画

本书中有几章用于介绍构成了若干现在常用技术基础的十来个最为流行的排名方法。但是，我们注意到在此所描述的排名方法，在很大程度上是基于矩阵分析和最优化这一我们自身专业之上的。当然，从统计学、博弈论、经济学等其他专业也产生了大量的排名方法。

- 前面所提到的在线电影租赁公司 Netflix，对于为用户精确地进行电影排名的能力十分看重，以至于在 2007 年时他们主办了一次竞赛，竞赛的奖金是我们这个资本主义时代有史以来最大的赏额之一。他们提供了一百万美元的 Netflix 奖，以奖励能够将影片推荐的准确度提高 10% 的个人或小组。这个比赛完全把我们给迷住了，正如它还迷住了许多其他人一样。实际上，我们应该为这场竞赛所表现出的内在精神而鼓掌欢呼，因为这一做法源自于古时候的帝国和教皇委托任务的行事方法。意大利的某些最伟大的艺术作品，都是在某位艺术家用自己的作品在这类悬赏比赛中胜出之后，才被委托由他来完成的。例如，著名的佛罗伦萨大教堂洗礼堂的大门，便是洛伦佐·吉贝尔蒂（Lorenzo Ghiberti）在 1401 年赢得了羊毛商工会提供赏金的比赛之后而完成的。我们很乐于看到排名的数学问题能够促进竞赛精神的回归，即在竞赛中，任何一个地位不高但志向远大的科学家，都能够有机会参与并获胜。最终 Netflix 奖被 BelKor（**第 134 页**）这个由顶级科学家组成的小组所获得。有关 Netflix 的更多内容，我们将在本书的杂谈中讲述。
- 即使对于不是科学家的人来说，他们与排名之间也有着天然的联系，因为人类似乎从心理上就被相互比较所操纵，而这样的比较就能够产生出排名。一个众所周知的事实是人类难以对任何一组多于 5 个的事物进行排名。不过在另一方面，我们特别擅长成对比较。在其畅销书《眨眼之间：不假思索的决断力》（Blink）中，马尔科姆·格

⊖ 一个漫画网站。——译者注

拉德威尔（Malcolm Gladwell）指出，在眨眼之间做出的仓促决定，常常就能决定生死。物种的进化也认可了这一模式，并奖赏了那些能快速进行比较的生物。实际上，这样的比较每天都会发生好几十次，因为我们会将自己和他人（或之前的自己）加以对比。我看起来比她年轻，她比我高，吉姆比比尔更快，我觉得自己今天更苗条些了，等等。这类成对比较是本书的核心，因为每个排名方法都始于成对比较数据。

- 最后，尽管可能只有少数人会对这个在末尾出场的事实感兴趣，但现如今，研究体育比赛排名是有可能获得博士学位的。实际上，我们的学生之一安吉拉·戈万最近就做了这件事，她靠着博士学位论文《排名理论及其在大众体育中的应用》从北卡罗来纳州立大学毕业了。唔……实际上，她是靠着对美式橄榄球队伍排名的多个算法所进行的数学分析而通过了最终答辩的，但对那些既热爱体育又同样热爱数据和科学的体育迷而言，这就为他们多打开了一扇门。

社会选择与阿罗的不可能性定理

本书是关于对对象进行排名的，而这些对象的类型可以各式各样，如体育队伍、网页和政治候选人等都可以进行排名。最后这一类对象则引导我们去探索社会选择这一领域，它主要研究各种政治投票体系。在整个人类历史中，人们都痴迷于投票体系。如古希腊这样的早期民主政体使用了标准的简单多数投票（plurality voting system）体系，其中每位投票人都只能为他的首选投出一票。自那时起，人们就一直在争论这一简单标准的利弊。实际上，大声反对简单多数投票的法国数学家让-查理·德·波达（Jean-Charles de Borda）和孔多塞侯爵马利·让·安托万·尼古拉·卡里塔（Marie Jean Antoine Nicolas Caritat）都各自提出了新的方法，我们将在后文中讨论这些方法——特别请见**第 17 页**上关于 BCS 排名的杂谈和**第 165 页**上的讨论。

简单多数投票与偏好列表投票（preference list voting）形成了对比，在后者中，每位投票人给出一份*偏好列表*（preference list），将候选人按排名的先后顺序加以排列。这些经过排名的清单通过某种方式被归总起来（更多有关排名聚合的内容请见第 14 章和第 15 章），以决定总的获胜者。某些国家正在践行偏好投票，比如澳大利亚。

民主政体长久以来都在寻找着完美的投票体系。但是，直到 20 世纪，数学家兼经济学家肯尼斯·阿罗（Kenneth Arrow）才考虑到改变研究的焦点。相比于问“完美的投票体系是什么样子的?”，阿罗问的是“是否存在完美的投票体系?”

阿罗的不可能性定理


K. Arrow

本节的标题已经预示了答案。阿罗发现他那个关于完美投票体系存在性问题的答案是否定的。1951 年，作为其博士论文中的一部分，肯尼斯·阿罗证明了所谓的不可能性定理，该定理描述了所有投票体系的内在局限性。这个惊人的定理指出，对于有着三个或更多候选人的情况，没有任何投票体系能同时满足以下 4 条常识性的准则[5]。

1. 阿罗对投票体系的第一个要求，是每个投票者都能按他或她所愿意的任何方式来对候选人进行排名。例如，如果要求某位在职的官员自动地被排在前五，那这种做法将是不公平的。这一要求常常被称为*无限制域准则*（unrestricted domain criterion）。
2. 阿罗的第二个要求则涉及候选人的子集。假设在这一子集中，投票者总是将候选人 A 排在候选人 B 之前，则当候选人集合被扩展为包括所有候选人时，应当仍然保持排名顺序，即该子集之外的候选人的排名顺序改变，不应该影响到 A 和 B 之间的相对排名。这一性质称为*无关选择的独立性*（independence of irrelevant alternatives）。
3. 阿罗的第三个要求称为*帕累托法则*（Pareto Principle），它指出，如果所有投票者在 A 和 B 中都选择了前者，那么一个恰当的投票体系应该总是将 A 排名在 B 之前。
4. 第四个也是最后一个要求称为*非独裁性*（non-dictatorship），它指出，没有任何一个投票者能够对选举施加不成比例的控制力。更精确而言，没有任何投票者有权独断排名。

阿罗在 1951 年提出的这一定理以及包括该定理的论文被普遍认为具有极高价值，1972 年，肯尼斯·阿罗被授予了诺贝尔经济学奖。尽管阿罗的 4 条准则看来是显然的或自明的，但他由此得到的结论却肯定不是这样。他证明，对任何一个投票体系而言，都不可能同时满足所有这 4 条常识性的准则。当然，这其中包括了所有现有的投票体系，也包括了将来可能出现的任何巧妙的新体系。因此不可能性定理迫使我们对包括本书所给出的排名体系在内的投票体系持有一个更为现实的期待。在后文中，我们将会说明，在某些特定的排名问题中，阿罗的某些要求并不那么合适，因此在这样的问题设置中违背阿罗的准则，只会带来些轻微的后果，或者根本就没有任何不良后果。

一个小型范例

本书给出了多个对集合中的对象进行排名的方法。通常而言，贯穿本书的这些对象是各类球队。不过书中的每个思路都可以被扩展到任一需要进行排名的对象集之上。在杂谈以及每一章的示例节中（如许多有趣的非体育领域应用中），都体现出了这一点。为了说明这些排名方法，我们将重复地使用一个小型的示例。由于体育比赛的数据量大，而且也易于获取，

因此我们的范例使用了 NCAA[⊖]美式橄榄球联赛 2005 赛季的数据。具体而言，它包括了大西洋沿岸联盟队伍中一个单独的组，其中每支队伍都与其他所有队伍进行比赛，由此我们便能根据胜负记录和分差来给出简单的传统排名。表 1.1 给出了组中 5 支队伍的这些数据，其中与那些没有出现在表中的队伍的比赛结果被省略了。

表 1.1　小型的 5 支队伍示例的比赛得分数据

	杜克大学（Duke）	迈阿密大学（Miami）	北卡罗来纳大学（UNC）	弗吉尼亚大学（UVA）	弗吉尼亚理工大学（VT）	胜负场次	分差
Duke		7-52	21-24	7-38	0-45	0-4	-124
Miami	52-7		34-16	25-17	27-7	4-0	91
UNC	24-21	16-34		7-5	3-30	2-2	-40
UVA	38-7	17-25	5-7		14-52	1-3	-17
VT	45-0	7-27	30-3	52-14		3-1	90

这个小例子还可以帮助我们从一开始便强调指出语言上的一个不同寻常的地方。我们将对排名（ranking）和评分（rating）这两个词加以区分。尽管这两个词常常可以替换使用，但在本书中我们谨慎并有区别地使用它们。*排名*指的是按名次先后而排列的队伍的列表，而*评分*则指的是数值分数的列表，每支队伍都对应有一个分数。例如，由表 1.1 可见，使用胜负场次的简单排名方法，我们可以得到如下的排名列表。

$$\begin{matrix}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}5\\1\\3\\4\\2\end{pmatrix},$$

这意味着杜克大学排名第五，而迈阿密大学排名第一，等等。另一方面，我们可以利用表 1.1 中的分差来给出这 5 支队伍的如下得分列表。

$$\begin{matrix}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}-124\\91\\-40\\-17\\90\end{pmatrix}。$$

对得分列表进行排序便能得到一个排名列表，因此，由上述分差得到的排名列表为

$$\begin{matrix}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}5\\1\\4\\3\\2\end{pmatrix},$$

这与胜负场次所给出的排名列表之间仅稍有出入。不过要弄清这两者之间的差别到底有多大，却是一个棘手的问题。我们将把这一问题推迟到本书靠后的部分来介绍。实际上，两个排名列表之间的差别，正是我们用第 16 章这一整章来讨论的主题。

⊖ NCAA，National Collegiate Athletic Association，即美国大学体育总会的缩写。——编辑注

最后我们可以注意到，每个评分列表都能给出一个排名列表，但反之则不然。此外，长度为 n 的排名列表是由 1 到 n 的这些整数的一个排列，而评分列表则包含实数（或可能为复数）。

排名与评分

对象的排名是这些对象名次排序所得的一个列表，因此，一个排名向量是整数 1 到 n 的一个排列。

对象的评分为每个对象赋予了一个数值分数。经过排序的评分列表便给出了一个排名列表。

杂谈：著名排名示例一个不计排名的列表

本书中散布着来自不同背景和应用的示例。在这些例子中，我们收集了各种评分以及由此得到的排名，并将这些排名和某些著名的标准排名列表进行了对比。我们将从一个未排名的列表来开始这一本有关排名的书，这多少有些讽刺意味。无论如何，下述未按任何特定顺序所给出的清单，对本书中经常遇到的若干标准列表进行了简要介绍。这一未排名的列表同时还展示了排名得以发挥用处的范围。

- PageRank：这个排名因谷歌而十分出名，谷歌使用 PageRank 算法对网页进行排名，并大获成功[15,49]。在第 6 章中还有关于这一著名排名的更多内容。
- HITS：这个网页排名算法[45, 49]面世的时间和 PageRank 差不多。它目前是搜索引擎 Ask. com 的一部分。我们在第 7 章中描述的攻防（也称 OD）方法便受到了 HITS 的某些核心概念的启发。
- BCS：大学美式橄榄球队的排名是根据他们的 BCS 评分来确定的。BCS（碗系列锦标赛）评分可以帮助确定哪支队伍将被邀请参加哪个碗系列的比赛。第 2 章和第 3 章涵盖了 BCS 所使用的两种评分方法。
- RPI：大学篮球队根据其 RPI（评分百分比指数）来排名，它用于确定哪些队伍将被邀请参加三月疯狂巡回赛。
- Netflix 和 IMDb：电影迷们可以参考历来最佳影片的排名列表，以确定租看哪一部影片。新的影片在线租赁邮递系统 Netflix 和因特网电影数据库（IMDb）都提供了得分最高的影片的排名列表。**第 25 页**和**第 95 页**上的杂谈便使用了 Netflix 奖的数据集来对电影进行排名。
- 人类发展指数：联合国根据人类发展指数（HDI）来对国家进行排名，这一指数是预期寿命、识字率、教育和生活标准的一个综合性指标。HDI 可用于决定援助哪些欠发达国家。
- 大学排名：《美国新闻》（US News）每年都会发布美国境内的学院和大学的排名。**第 77 页**上的杂谈中包括了有关这一排名及其计算的更多内容。
- 国际象棋棋手：世界国际象棋联合会 FIDE 使用第 5 章中的埃洛（Elo）法来对全世界数十万的国际象棋棋手进行排名。

- 食物链：某类物种以其他物种为食，同时又是别的物种的食物。因此，可以构建一幅有向图，它说明了不同物种间的支配关系。分析该图便能回答诸如谁处于食物链顶端或哪些物种是整条链中最为关键的部分之类的问题。
- 考古学：**第 192 页**上的杂谈说明了排名方法是如何将文物按照其在多个地点的挖掘深度来进行相对排序的。
- 法律案件：由一个案子到另一个案子再到又一个案子之间的引用提供了有价值的信息，使我们能对法律案件进行排名。实际上，目前的司法搜索引擎就使用了这一技术，来针对律师助理就某个特定主题所提出的查询，返回相关案件的一个排名列表。
- 社交网络：如脸谱（Facebook）这样的社交网络中的链接结构，可以根据受欢迎程度对网络成员进行排名。在这种情况下，第 6 章和第 7 章中的马尔可夫法或 OD 法是特别好的选择。但是，脸谱的评分可能源自第 5 章中所述的埃洛评分体系——请见**第 64 页**上的杂谈。当然，某些如协调恐怖分子进行 911 攻击的这种网络将具有更为密集的通信往来。使用与找到脸谱中最受欢迎成员相同的方法，还可以识别出这些网络中的关键人物或中央枢纽。

数字——

50 000 000 = 2010 年美国大学美式橄榄球比赛的观赛人数。

——The Line Makers

第 2 章　梅西法

碗系列锦标赛（BCS）是 NCAA 大学美式橄榄球联赛的一个评分体系，它被设计用来确定哪些队伍被邀请参加哪个系列的比赛。BCS 已经因为它给 NCAA 的每支队伍所打的分数而名声在外，而且可能不是个好名声。这些评分有两个来源——人以及计算机。人这一方面的输入数据来自于教练和媒体；计算机方面的输入则来自于 6 个计算机数学模型——它们的细节将在**第 17 页**的杂谈中给出。2001 赛季和 2003 赛季的 BCS 评分被体育迷和分析家认为是最具争议性的，而且大多数人都已经熟知 BCS 选拔体系与某次季后赛结果存在冲突的瑕疵，这些人中甚至还包括了美国总统——请见**第 19 页**上的杂谈。

原始的梅西评分方法


K. Massey

在 1997 年，当时还是布鲁菲尔德学院（Bluefield College）的一名本科生的肯尼斯·梅西，提出了一种对大学美式橄榄球队进行排名的方法。他利用了数学理论中的最小二乘法写成了一篇获奖论文[52]。在本书中，我们称这种方法为梅西法，尽管还有其他几种方法也应归功于肯尼斯·梅西。梅西此后成为了卡森-纽曼学院的一名数学教授，而且现在仍然在改进他的体育排名模型。梅西教授已经提出了多个评分方法，其中之一被碗系列锦标赛（BCS）体系用于确定 NCAA 美式橄榄球季后赛的分组——请见**第 17 页**上的杂谈。在下一章中讲述的科利法同样也是 BCS 所使用的 6 种计算机评分体系之一。由于我们并不清楚 BCS 所使用的梅西法的细节，因而此处将描述我们已经了解了全部细节的那个梅西法，即他在论文中所描述的那一个[52]。对梅西更新的、用于 BCS 的模型感兴趣的读者，可以仔细阅读他的体育排名网站（masseyratings. com）来获得更多信息。

梅西的主要思路

梅西最小二乘法最根本的思路可以总结为一个理想化的方程

$$r_i - r_j = y_k,$$

式中，y_k 是比赛 k 中获胜方所取得的优势；而 r_i 和 r_j 分别为队伍 i 和 j 的评分。换而言之，两支队伍的评分 r_i 和 r_j 之间的差值完美地预测了这两队进行比赛时获胜方的优势。

任何一个评分体系所要达到的目的，是为一个具有 n 支队伍且至今已总共进行了 m 场联赛的联盟中的每支队伍赋予一个评分。当然，我们并不知道这些队伍的评分 r_i，但是我们确实知道谁和谁进行了比赛，而获胜方的优势又是多少。因此，对每一场比赛 k，都对应有一个上述形式的方程，由此产生了一个包含 n 个未知数的 m 个线性方程的系统，可写为

$$\boldsymbol{Xr} = \boldsymbol{y}$$

系数矩阵 $\boldsymbol{X}$ 中的每一行都几乎全部为 0，只是在位置 i 上有一个 1，而在位置 j 上有一个 −1，这意味着队伍 i 在这场比赛中击败了队伍 j。因此，$\boldsymbol{X}_{m\times n}$ 是一个非常稀疏的矩阵。向量 $\boldsymbol{y}_{m\times 1}$ 是表示获胜方优势的右向量，而 $r_{n\times 1}$ 是未知的评分向量。通常来说，$m \gg n$，因此该线性系统是一个高度超定且矛盾的系统。不过这并不是说就无法对它进行求解，因为由正规方程组（normal equations）$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Xr} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$ 可以得到一个最小二乘解。这个最小二乘向量——即正规方程组的解——（从方差的最小化的意义上而言）是原方程组 $\boldsymbol{Xr} = \boldsymbol{y}$ 中评分向量 r 的最佳线性无偏估计[54]。

梅西发现，由于 $\boldsymbol{X}$ 本身的结构上的特点，使用系数矩阵 $\boldsymbol{M} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$ 更为有利。实际上，$\boldsymbol{M}$ 根本不用计算就能得到，因为我们有如下事实，即该矩阵的对角元素 $\boldsymbol{M}_{ii}$ 就是队伍 i 所完成的比赛场次的总数，而非对角元素 $\boldsymbol{M}_{ij}$（$i \neq j$）就是队伍 i 与队伍 j 对阵场数的相反数。通过类似的便捷方法，还可以根据累计分差来获得正规方程组的右侧项 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$。右向量 $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$ 的第 i 个元素是队伍 i 在赛季中进行的所有比赛中的分差之和，我们由此定义 $\boldsymbol{p} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$。因此，梅西的最小二乘系统就变为

$$\boldsymbol{Mr} = \boldsymbol{p},$$

式中，$\boldsymbol{M}_{n\times n}$ 是上述的梅西矩阵；$\boldsymbol{r}_{n\times 1}$ 是未知评分构成的向量；而 $\boldsymbol{p}_{n\times 1}$ 是表示累积分差的右向量。

梅西矩阵 $\boldsymbol{M}$ 具有几点值得注意的性质。第一，$\boldsymbol{M}$ 的大小比 $\boldsymbol{X}$ 要小得多。实际上，它是一个 n 阶对称方阵；第二，它是一个对角优势的 M-矩阵；第三，它每行的元素和为 0，因此 $\boldsymbol{M}$ 的各列是线性相关的。这带来了一个小问题，即由于 $\mathrm{rank}(\boldsymbol{M}) < n$，线性系统 $\boldsymbol{Mr} = \boldsymbol{p}$ 不具有唯一解。梅西在该问题中用于避开这一麻烦的解决方法，是在 $\boldsymbol{M}$ 中任选一行，以全 1 行来代替它，并将 $\boldsymbol{p}$ 中的对应元素设为 0。这一做法在该线性系统中增加了一个约束，它要求评分的总和必须为 0，而由此也得到了一个满秩的系数矩阵。在马尔可夫链稳态向量的直接求解法中也使用了类似的技巧，请见参考文献［75］的第 3 章。经过行调整后的新系统记为 $\overline{\boldsymbol{M}}\boldsymbol{r} = \overline{\boldsymbol{p}}$。

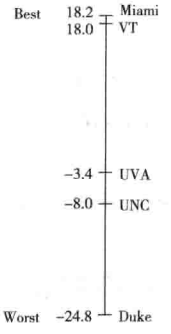
在范例上应用梅西评分法

现在我们为第 5 页的范例中那 5 支队伍来产生梅西评分。每支队伍都正好与其余各队比赛了 1 次，因此 $\boldsymbol{M}$ 的非对角元素都是 -1，而对角元素都是 4。但为了保证 $\boldsymbol{M}$ 满秩（并因此获得唯一的最小二乘解），我们使用梅西在参考文献［52］中给出的技巧，用所有评分之和等于 0 的这一约束来代替最后一行。因此，梅西最小二乘系统 $\overline{\boldsymbol{M}}\boldsymbol{r} = \overline{\boldsymbol{p}}$ 为

$$\begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} -124 \\ 91 \\ -40 \\ -17 \\ 0 \end{pmatrix},$$

它给出了如下表所示的梅西评分和排名列表。

队伍	评分 $\boldsymbol{r}$	排名
Duke	-24.8	第 5
Miami	18.2	第 1
UNC	-8.0	第 4
UVA	-3.4	第 3
VT	18.0	第 2



用表格的形式表示评分和排名值十分有用，但有时如果以上面右图所示的数线图表示则会更好一些。数线表示可以让我们更为容易地观察队伍之间的相对差距，而这一附加信息正是我们在**第 176 页**上评分聚合小节中所要用到的东西。

梅西评分法的高级特性

对于这个微型的数据集而言，以上的结果是合理的，但对于梅西法，则还有更多东西可说[52]。实际上，还有两个向量应当提及。梅西从总的评分向量 $\boldsymbol{r}$ 中创造了两个新的向量，即一个攻击评分向量 $\boldsymbol{o}$ 和一个防守评分向量 $\boldsymbol{d}$。梅西假设队伍 i 的总评分为该队的进攻和防守评分之和，即 $r_i = o_i + d_i$。梅西通过一些巧妙的代数运算将 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 从 $\boldsymbol{r}$ 中抽取出来。但是，要了解他所使用的代数方法，我们还需要一些额外的记法。右向量 $\boldsymbol{p}$ 的第 i 个元素 p_i 中保存了队伍 i 在该赛季所进行的所有比赛的累计分差，现在将其分解为 $\boldsymbol{p} = \boldsymbol{f} - \boldsymbol{a}$。向量 $\boldsymbol{f}$ 为“得分”向量，记录了每支队伍在整个赛季中所获得的总得分，而向量 $\boldsymbol{a}$ 称为“失分”向量，它记录了每支队伍在整个赛季中所丢掉的分数之和。要理解梅西求取 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 的方法，我们还需要进行另一个分解。梅西系数矩阵 $\boldsymbol{M}$ 可以被分解为 $\boldsymbol{M} = \boldsymbol{T} - \boldsymbol{P}$，其中，$\boldsymbol{T}$ 是一个对角阵，记录了每支队伍所进行的比赛总数，而 $\boldsymbol{P}$ 是一个非对角阵，记录了赛季中所有队伍之间

捉对厮杀的场数。我们从原始的梅西最小二乘系统 $\boldsymbol{Mr} = \boldsymbol{p}$ 开始，并通过一系列的替换来说明这两个新向量 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 是如何推得的。

$$\begin{aligned}\boldsymbol{Mr} &= \boldsymbol{p}\\ (\boldsymbol{T}-\boldsymbol{p})\boldsymbol{r} &= \boldsymbol{p}\\ (\boldsymbol{T}-\boldsymbol{p})(\boldsymbol{o}+\boldsymbol{d}) &= \boldsymbol{p}\\ \boldsymbol{To}-\boldsymbol{Po}+\boldsymbol{Td}-\boldsymbol{Pd} &= \boldsymbol{p}\\ \boldsymbol{To}-\boldsymbol{Po}+\boldsymbol{Td}-\boldsymbol{Pd} &= \boldsymbol{f}-\boldsymbol{a}。\end{aligned}$$

现在，上述的最后一个方程可以被分为两个独立的方程。

$$\boldsymbol{To}-\boldsymbol{Pd}=\boldsymbol{f} 和 \boldsymbol{Po}-\boldsymbol{Td}=\boldsymbol{a}。$$

左边的方程 $\boldsymbol{To}-\boldsymbol{Pd}=\boldsymbol{f}$ 说明，一支队伍在整个赛季中所得的总分数，可以通过将该队的进攻评分与比赛场数相乘，再减去对手的防守评分的总和而得到。继续对左边的方程进行替换，我们就可以更加接近于解出这两个新的评分向量 $\boldsymbol{o}$ 和 $\boldsymbol{d}$。

$$\begin{aligned}\boldsymbol{To}-\boldsymbol{Pd} &= \boldsymbol{f}\\ \boldsymbol{T}(\boldsymbol{r}-\boldsymbol{d})-\boldsymbol{Pd} &= \boldsymbol{f}\\ (\boldsymbol{T}+\boldsymbol{P})\boldsymbol{d} &= \boldsymbol{Tr}-\boldsymbol{f}。\end{aligned}$$

注意上述最后一行中的右侧项 $\boldsymbol{Tr}-\boldsymbol{f}$ 是一个由常值构成的向量，因为 $\boldsymbol{r}$ 已经得到了。因此，（给定 $\boldsymbol{r}$ 时）求解 $\boldsymbol{d}$ 的梅西线性系统为

$$(\boldsymbol{T}+\boldsymbol{P})\boldsymbol{d}=\boldsymbol{Tr}-\boldsymbol{f}。$$

最后，一旦 $\boldsymbol{r}$ 和 $\boldsymbol{d}$ 都已得到，$\boldsymbol{o}$ 可以通过 $\boldsymbol{r}=\boldsymbol{o}+\boldsymbol{d}$ 来求得。

范例：使用高级梅西评分法

梅西对 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 的创造和使用，为体育排名模型研究开启了一条有趣的道路，具体来说，就是对比赛的最后比分进行预测，或至少是分差的预测。因此，我们将梅西法应用到这个小例子上。特别地，梅西使用这两个新的评分向量 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 来预测未来比赛的得分。他认为 o_i 给出了“队伍 i 在对阵具有平均防守水平的对手时的期望得分”。将它与已知的防守评分相结合，梅西便给出了最终得分的预测。具体来说，如果队伍 i 和 j 进行比赛，则队伍 i 从队伍 j 上获取的预计得分由 $o_i - d_j$ 给出。表 2.1 给出了完整的梅西法中所有这三个评分向量。请注意，该方法预测 VT 对阵杜克大学的最终比分为 20.7 -（-26.8）= 47.5 比 2.0 -（-2.7）= 4.7，即约为 48 比 5。这与 2005 年的实际比赛比分相当接近，当年 VT 以 45 比 0 击败了杜克大学。不过，如果你兴奋于高级梅西模型在博彩方面所具有的潜力的话，那在此之前请注意并非所有的预测都能如此之好。例如，该模型预测迈阿密大学将以 25 比 24 击败 VT，而这与 2005 年数据中 25 比 7 的实际比分相差甚远。这个预测的不稳定性部分应归咎于所使用的数据量

太小（仅仅只有 5 支队伍）。许多体育迷都对分差的问题感兴趣，特别是那些喜欢博彩的体育迷们，因此我们在**第 113 页**的第 9 章的内容中给出了有关分差和最优“分差评分”的完整讨论。

表 2.1　梅西评分和排名向量的完整列表

队伍	r	根据 r 所得的排名	o	进攻排名	d	防守排名
Duke	-24.8	第 5	2.0	第 4	-26.8	第 5
Miami	18.2	第 1	22.0	第 1	-3.8	第 2
UNC	-8.0	第 4	1.4	第 5	-9.4	第 3
UVA	-3.4	第 3	7.8	第 3	-11.2	第 4
VT	18.0	第 2	20.7	第 2	-2.7	第 1

梅西评分法的总结

以下是描述体育队伍排名的梅西法时所使用的符号。

n　联盟中的队伍数 = 矩阵 $\boldsymbol{T}$，$\boldsymbol{P}$ 和 $\boldsymbol{M}$ 的阶次

m　进行的比赛数

$\boldsymbol{X}_{m\times n}$　按队伍给出的比赛情况矩阵

若队伍 i 在比赛 k 中获胜，则 $\boldsymbol{X}_{ki}=1$；若队伍 i 在比赛 k 中告负，则 $\boldsymbol{X}_{ki}=-1$；其他情况为 0

$\boldsymbol{T}_{n\times n}$　包含了所进行的总比赛数信息的对角阵

$\boldsymbol{T}_{ii}$ = 队伍 i 在赛季中所进行的比赛总场数

$\boldsymbol{P}_{n\times n}$　包含了逐对比赛数据的非对角阵

$\boldsymbol{P}_{ij}$ = 队伍 i 和 j 之间的对阵次数

$\boldsymbol{M}_{n\times n}$　对称 M-方阵，称为梅西矩阵（Massey matrix）

$\boldsymbol{M}=\boldsymbol{T}-\boldsymbol{P}$

$\overline{\boldsymbol{M}}_{n\times n}$　修正的梅西矩阵（adjusted Massey matrix），将 $\boldsymbol{M}$ 中的一行替换为 $\boldsymbol{e}^{\mathrm{T}}$（全 1 向量）而得

$\boldsymbol{f}_{n\times 1}$　累积得分向量

f_i = 队伍 i 在赛季中的总得分

$\boldsymbol{a}_{n\times 1}$　累积失分向量

a_i = 队伍 i 在赛季中的总失分

$\boldsymbol{p}_{n\times 1}$　累积分差向量

$p_i=f_i-a_i$ 为队伍 i 在赛季中的总分差

$\overline{\boldsymbol{p}}_{n\times 1}$　修正的分差向量，通过将 $\boldsymbol{p}$ 中的一个元素以 0 代替而得

$\boldsymbol{r}_{n\times 1}$　梅西最小二乘系统给出的总评分向量

$\boldsymbol{o}_{n\times 1}$　梅西最小二乘系统给出的攻击评分向量

$\boldsymbol{d}_{n\times 1}$　梅西最小二乘系统给出的防守评分向量

梅西最小二乘算法

1. 求解系统 $\overline{\boldsymbol{M}}\boldsymbol{r} = \bar{\boldsymbol{p}}$ 以获得梅西总评分向量 $\boldsymbol{r}$；
2. 求解系统 $(\boldsymbol{T} + \boldsymbol{P})\boldsymbol{d} = \boldsymbol{T}\boldsymbol{r} - \boldsymbol{f}$ 以获得防守评分向量 $\boldsymbol{d}$；
3. 计算攻击评分向量 $\boldsymbol{o} = \boldsymbol{r} - \boldsymbol{d}$。

以下所列即为梅西法的若干性质。

- 梅西法使用得分数据来对队伍进行评分，不过其他的比赛统计数据也同样可用。而在其他领域中，任何对象的成对比较数据也均属可用。例如，在对国家的贸易能力进行评分时，可以使用出口总量这一数据。出口量（可能按人口数量进行了平均处理）更大的国家，将赢得这一虚拟的贸易竞赛。通过这种方式，可以支持并强化**第 6 页**上介绍的联合国人类发展指数 HDI；
- 梅西法产生了三个向量，即总评分向量 $\boldsymbol{r}$、攻击评分向量 $\boldsymbol{o}$ 和防守评分向量 $\boldsymbol{d}$。通过巧妙地组合 $\boldsymbol{o}$ 和 $\boldsymbol{d}$，我们可以对分差进行估计。（请见**第 113 页**上第 9 章的内容。）

杂谈：梅西法网页排名

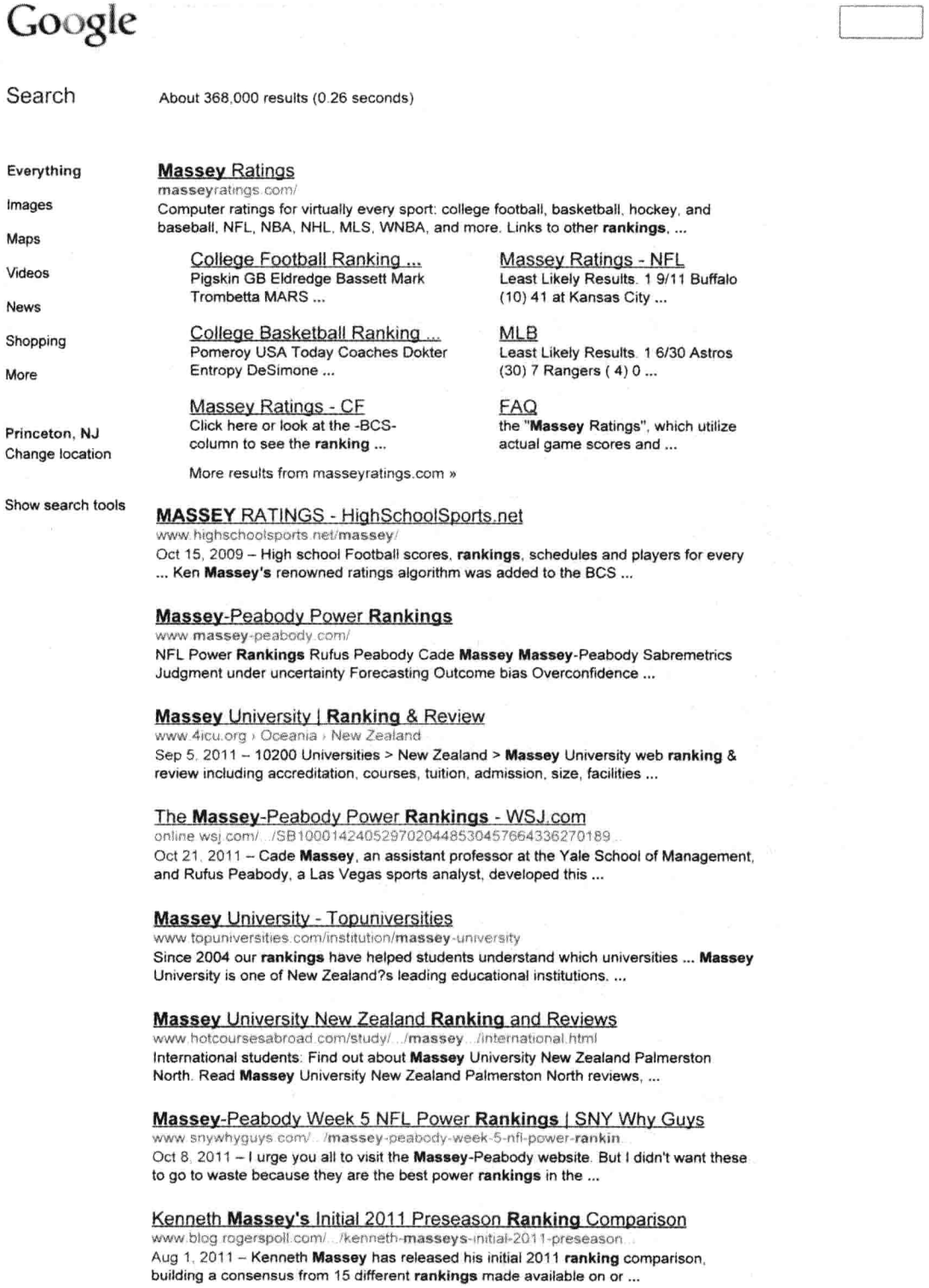
在这个杂谈中，我们将把梅西法应用于网页，以强调说明该方法可以用于任何对象集合的排名。（实际上，只要稍微想一下，就会发现对于本书中的所有方法而言，这一点都成立。）由于呈现给用户的网页，是依据用户查询所做出的响应，并根据其排名高低进行了排列，因此网页排名是每个主要搜索引擎的基础组成部分。例如，图 2.1 便给出了在谷歌中输入查询项“Massey ranking”时返回的排名第 1 到第 10 的网页。

将梅西法应用于万维网的诀窍，在于如何定义网页之间的比赛或成对较量，而要做到这一点，有多种可能的方式可供选择。例如，如果像 Alexa 搜索引擎[78, 49]那样有网页的流量统计数据可用，则如果 $t_i > t_j$ 成立的话，我们称网页 i 击败了网页 j，其中 t_i 和 t_j 分别为这两个网页的流量大小的度量。此时，$t_i - t_j$ 表示“得分”之差。网页 i 的著名的 PageRank[15] 度量 π_i 是另一个可以用来引出对比数据的指标，此时当 $\pi_i > \pi_j$ 时，网页 i 以 $\pi_i - \pi_j$ 的分差击败网页 j。梅西网页排名方法同样也是由同一个理想化基本方程

$$r_i - r_j = y_k$$

开始进行，只不过此时要根据所选用的数据，来对比赛 k 中的获胜方优势 y_k 进行修改。例如，在确定假想比赛中的胜者时，如果使用流量度量，则 $y_k = t_i - t_j$；而如果使用 PageRank 度量，则 $y_k = \pi_i - \pi_j$。

之后，梅西法网页排名便按普通的方式进行。由于梅西法可以产生进攻和防守这两个评



Google

Search

About 368,000 results (0.26 seconds)

Everything
Images
Maps
Videos
News
Shopping
More

Princeton, NJ
Change location

Show search tools

Massey Ratings
masseyratings.com/
Computer ratings for virtually every sport: college football, basketball, hockey, and baseball, NFL, NBA, NHL, MLS, WNBA, and more. Links to other **rankings**, ...

College Football Ranking ...
Pigskin GB Eldredge Bassett Mark Trombetta MARS ...

Massey Ratings - NFL
Least Likely Results. 1 9/11 Buffalo (10) 41 at Kansas City ...

College Basketball Ranking ...
Pomeroy USA Today Coaches Dokter Entropy DeSimone ...

MLB
Least Likely Results. 1 6/30 Astros (30) 7 Rangers (4) 0 ...

Massey Ratings - CF
Click here or look at the -BCS- column to see the **ranking** ...

FAQ
the "**Massey** Ratings", which utilize actual game scores and ...

More results from masseyratings.com »

MASSEY RATINGS - HighSchoolSports.net
www.highschoolsports.net/**massey**/
Oct 15, 2009 – High school Football scores, **rankings**, schedules and players for every ... Ken **Massey's** renowned ratings algorithm was added to the BCS ...

Massey-Peabody Power **Rankings**
www.**massey**-peabody.com/
NFL Power **Rankings** Rufus Peabody Cade **Massey** **Massey**-Peabody Sabremetrics Judgment under uncertainty Forecasting Outcome bias Overconfidence ...

Massey University | **Ranking** & Review
www.4icu.org › Oceania › New Zealand
Sep 5, 2011 – 10200 Universities > New Zealand > **Massey** University web **ranking** & review including accreditation, courses, tuition, admission, size, facilities ...

The **Massey**-Peabody Power **Rankings** - WSJ.com
online.wsj.com/.../SB10001424052970204485304576643362701 89...
Oct 21, 2011 – Cade **Massey**, an assistant professor at the Yale School of Management, and Rufus Peabody, a Las Vegas sports analyst, developed this ...

Massey University - Topuniversities
www.topuniversities.com/institution/**massey**-university
Since 2004 our **rankings** have helped students understand which universities ... **Massey** University is one of New Zealand?s leading educational institutions. ...

Massey University New Zealand **Ranking** and Reviews
www.hotcoursesabroad.com/study/.../**massey**.../international.html
International students: Find out about **Massey** University New Zealand Palmerston North. Read **Massey** University New Zealand Palmerston North reviews, ...

Massey-Peabody Week 5 NFL Power **Rankings** | SNY Why Guys
www.snywhyguys.com/.../**massey**-peabody-week-5-nfl-power-**rankin**...
Oct 8, 2011 – I urge you all to visit the **Massey**-Peabody website. But I didn't want these to go to waste because they are the best power **rankings** in the ...

Kenneth **Massey's** Initial 2011 Preseason **Ranking** Comparison
www.blog.rogerspoll.com/.../kenneth-**masseys**-initial-2011-preseason...
Aug 1, 2011 – Kenneth **Massey** has released his initial 2011 **ranking** comparison, building a consensus from 15 different **rankings** made available on or ...

图 2.1　在谷歌搜索引擎中“Massey ranking”的查询结果

分向量，因此每个网页都将获得两个评分。这两个评分分别对应了枢纽和权威评分，即在网页排名行业中广为人知的双排名[45, 49]。第 7 章中将讨论更多有关体育排名领域中的枢纽和权威的内容。

第 12 章给出了梅西模型的一个时间加权的改进，对于网页的梅西法而言，这一改进同样具有振奋人心的应用潜力。这个改进用一种自然的方式，来将时间信息整合到排名之中。对于万维网而言，这意味着诸如标记了网页最近更新的时间戳之类的可用时间信息将能够被加以利用。

垃圾信息是当今许多搜索引擎越来越关注的问题——不过相比于现在的网页排名方法，诸如梅西法这样的排名方法则不大容易受到垃圾信息的影响，因为它可以构建在诸如站点流量这类网页所有者更难控制的度量之上，这一点正与目前几乎所有主要搜索引擎排名所使用的基于链接的度量形成了鲜明对比。网页所有者肯定可以控制他们自己网页的出链，并已经知道了如何略施小计来影响他们的入链[49]。我们相信，最好的、对垃圾信息最具抵抗力的排名，可以通过网页的多个度量的组合来构建得到，正如在第 6 章和第 14 章的聚合方法中所详细说明的那样。

至此，对梅西方法在网页排名方面的适用性，我们已经在杂谈中描绘出了一幅前景光明的图景。现在我们转到最后一个问题，即问题规模这个非常现实的问题，一个可以决定方法成败的问题。如果一个方法极为巧妙，但却不能很好地扩展到具有数百万乃至数十亿结点的网络规模的图的话，那它对于网页排名而言就毫无用处。幸运的是，梅西法具有一个可资利用的结构，使其能够被用于万维网。我们将描述一种可能的做法，来根据万维网的超链接结构直接构建梅西线性系统。这一超链接数据是现成可用的，因为它是目前几乎所有搜索引擎所使用的基本数据。更为复杂的模型可以使用更为复杂的数据，如上述的流量和 PageRank 度量等。就目前而言，我们使用基本的超链接数据来说明梅西法在网络规模问题上的可扩展性。

令 $\boldsymbol{W}$ 为万维网的一个加权邻接矩阵，即 w_{ij} 为网页 i 和 j 之间有向超链接的权值。因此，如果在网页 i 和 j 之间没有超链接，则 $w_{ij}=0$。在梅西法的这个万维网实现中，$\boldsymbol{W}$ 记录了“得分”差信息，即 $w_{ij}>0$ 意味着网页 i 通过出链链接到网页 j。梅西矩阵 $\boldsymbol{M}$ 的大小为 $n\times n$，其中 n 为网页数量。由于 m_{ij} 是网页 i 和 j 交手次数的相反数，所以只要 $w_{ij}>0$，便有 $m_{ij}=m_{ji}=-1$。因此，如果已知加权邻接矩阵，则只需增加几行代码，就可以构造出万维网的梅西系统。我们用 MATLAB 语言的方式给出这些代码。

```
A = W' > 0; % A 是一个二值邻接矩阵，由加权邻接矩阵 W 的转置所导出
M = -A - A' + diag (Ae) + diag (A'e); % 万维网梅西矩阵 M
```

梅西线性系统 $\boldsymbol{Mr}=\boldsymbol{P}$ 的最后一部分是右向量 $\boldsymbol{p}$。如同 $\boldsymbol{M}$ 一样，$\boldsymbol{p}$ 可以很容易求得。$\boldsymbol{p}$ 中第 i 个元素，记为 p_i，等于指向页面 i 的所有入链的总数（或所有入链的权值之和）减去发自页面 i 的所有出链的总数。因此，对于万维网梅西系统，有

```
p = W'e - We
```

尽管需要求解的线性系统 $\overline{\boldsymbol{M}}\boldsymbol{r}=\overline{\boldsymbol{p}}$ 非常庞大，但它的规模并未超过如 PageRank 或 HITS[49] 这样的现有排名系统。实际上，它可以通过如雅可比法、SOR 法或 GMRES 法[32, 75, 74, 66] 这样的迭代方法来高效存储和求解。

杂谈：BCS 评分

BCS 评分决定了哪些队伍有资格进入大学美式橄榄球季后赛，并为高额奖金而战。那么这些重要的评分是如何确定的呢？它们由三个同等重要的部分所组成，其中两个是人工调查结果［哈里斯大学美式橄榄球互动调查（Harris Interactive College Football Poll）和《今日美国》教练员调查（USA Today Coaches Poll）］，而第三个则是由杰夫·安德森（Jeff Anderson）和克里斯·赫斯特（Chris Hester）、理查德·比林斯利（Richard Billingsley）、威斯利·科利（Wesley Colley）、肯尼斯·梅西（Kenneth Massey）、杰夫·萨加林（Jeff Sagarin）以及彼得·沃尔夫（Peter Wolfe）所提供的 6 个计算机排名所给出的归总结果。

这些排名通过一个称为波达计数（Borda Count）的投票方法加以整合——在**第 165 页**上给出了波达计数法的完整描述。对于 BCS 而言，它大致是这样运作的。在人工调查（从 10 月份到 12 月份，每周都进行调查）中，投票者要在评分票上填入他们认为的前 25 位的队伍。每个队伍被赋予了一个 $1 \leqslant r \leqslant 25$ 的评分，该评分与其排名正好相反——即在每张票上，排名最前的队伍的评分为 $r = 25$，第二名的队伍的评分为 $r = 24$，等等，直到最后一名，其评分为 $r = 1$。对计算机排名也使用同样的评分体系。在每次人工调查中，队伍是根据所得的总评分——即它们在所有选票中所得评分之和——来评价的。对于可能有所改变的投票者人数，其处理方法是求取每支队伍的“BCS 商”，其定义为队伍总分与满分之比。通过下面的例子，可以理解得更为清晰。

- 如果在给定的赛季中，哈里斯大学美式橄榄球互动调查拥有 113 名投票者，则一支队伍所能获得的最高总评分为 $113 \times 25 = 2\,825$，当且仅当所有这 113 名投票者都为该队打出 $r = 25$ 的评分时，或说当投票者们都将该队排在第一位时，才能达到这一总评分。类似地，一支队伍所能获得的最低总评分为 113，只有当所有投票者都给该队打了 $r = 1$ 的评分（即所有投票者都将该队排在了第 25 位）时，才会出现这一总评分。因此，如果 r_i 是第 i 名投票者对某队（比方说北卡罗来纳州立大学，以下简称 NC）所给出的哈里斯评分，则

$$\text{NC 的哈里斯 BCS 商} = \frac{1}{2825}\sum_{i=1}^{113} r_i。$$

- 如果在同样的赛季，《今日美国》教练员调查拥有 59 名投票者，则可能的最高总评分为 $59 \times 25 = 1\,475$。因此，如果 r_i 是第 i 名投票者为 NC 所打的《今日美国》教练员评分，则

$$\text{NC 的《今日美国》教练员 BCS 商} = \frac{1}{1475}\sum_{i=1}^{59} r_i$$

- 正如人类投票者所做的那样，6 个计算机模型中的每一个都给出了各自的前 25 名队伍的数值评分——即排名最先的队伍得分为 $r = 25$，排名第二的得分为 $r = 24$，等等。但是，在计算每支队伍的综合评分之前，每支队伍的最高评分和最低评分将被舍弃。因此，每支队伍只能得到 4 个计算机评分。这 4 个评分之和便给出了每支队伍的总评分。因此，可能获得的最高的计算机总评分为 $4 \times 25 = 100$。例如，当 NC 由 6 个计算机模型所获得的评分排序为 $r_1 \geqslant r_2 \geqslant r_3 \geqslant r_4 \geqslant r_5 \geqslant r_6$ 时，有

$$\text{NC 的计算机 BCS 商} = \frac{1}{100}(r_2 + r_3 + r_4 + r_5)$$

归总。在每个调查期结束时的 BCS 评分和排名，由上述 BCS 商简单求和而得。尽管对排名不会产生影响，但有时公布的是这三个 BCS 商的均值而非总和，因为这一均值使最终的评

分落在 (0,1] 区间内。换而言之，

$$\text{最终的 BCS 评分} = \frac{\text{哈里斯 BCS 商} + \text{教练员 BCS 商} + \text{计算机 BCS 商}}{3}。$$

BCS 最近的计算机模型。由 BCS 所使用的最近的计算机模型包括下面各位所提出的方法。不过，有关他们方法的详细说明一般无法得知。许多开发者并不希望他们秘密的细节被完全披露，因此他们只给出了一些提示。

1. 杰夫·安德森和克里斯·赫斯特（www. andersonsports. com）几乎没有给出什么细节，只是说他们为击败高手的队伍提供了额外的奖励，而且他们没有考虑获胜方优势，此外，他们的评分不仅依赖于胜负记录，也依赖于所在联盟的强弱。
2. 理查德·比林斯利（www. cfrc. com）也什么都没有公开。别人推测他的公式的主要部分是胜负记录，包括根据对手以往记录、评分和排名所定的对手强弱（更强调最近的表现），以及比赛地点和防守表现等次要考虑。
3. 威斯利·科利（www. colleyrankings. com）的方法将在**第 21 页**上第 3 章的内容中详细给出。
4. 肯尼斯·梅西的原始方法在**第 9 页**上已进行了介绍，不过他暗示现在使用的是更为精细（但并未彻底说明）的方法——请见 masseyratings. com。
5. 杰夫·萨加林（www. usatoday. com/sports/sagarin. htm）称他的模型主要使用了埃洛系统，细节在**第 53 页**第 5 章的内容中给出。
6. 彼得·沃尔夫（prwolfe. bol. ucla. edu/cfootball/ratings. htm）称他的方法是由参考文献［13］中的布拉德利（Bradley）-特里（Terry）模型改进而来。他表示该方法是一个最大似然方法，其中队伍 i 被赋予了一个评分值 π_i，用于预测当队伍 i 与队伍 j 相遇时的结果，其中 i 击败 j 的似然度为 $\pi_i/(\pi_i+\pi_j)$。所有结果都与之前的实际结果相符的概率 P 的求取方法，则是将根据每场比赛所得的单个概率全部乘起来。而评分值的选取应使 P 的取值最大化。詹姆斯·基纳在参考文献［42］中也讨论了类似的方法。

杂谈：不留破绽

BCS 评分从 NCAA 甲组联盟美式橄榄球冠军杯组（FBS，前身为甲-A 组）中选出两支队伍，来进行“BCS 全国锦标赛”的对决。根据合同要求，来自美国橄榄球教练协会的投票成员应投票将这场比赛的胜者选为“BCS 全国冠军”。此外，由各联盟签署的一份合同要求他们承认这场比赛的胜者为“官方且唯一的冠军”，从而消除了由于相互竞争的评分单位所造成的出现多个冠军的可能性。

杂谈：圣母大学规则

BCS 评分决定了哪些学校在季后赛中占有一席之地，但这个评分中却夹杂了一套复杂的规则，当初设计它们的目的，是为了保护不同的体育联盟以及历史上的传统强队。这些规

则中最为古怪的一个例子恐怕要数“圣母大学规则”，只要圣母大学取得前八名，该规则就令其自动获得一个季后赛的席位。还有更奇怪的。圣母大学可以保证获得一份单独且特有的报酬，据报道，它占到了BCS季后赛净收入的1/66——而这无关于他们的胜负记录，也不管他们是否获得了季后赛的参赛资格！据报道，在2009年，战斗的爱尔兰人队被支付了大约170万美元，而在该赛季中，他们仅获得了6胜6负的战绩，还把教练查理·韦斯（Charlie Weis）给炒掉了。据估计，如果他们能够参加一场季后赛，则将得到约600万美元的奖励。

对圣母大学规则的争论仍在继续。在FBS中有近半数的队伍得不到赢取或争夺BCS冠军的真正机会，而与此同时，圣母大学却被赋予了独一无二的优势和机遇，这一点让众多粉丝有如骨鲠在喉。

杂谈：奥巴马和司法部

在2008年的总统竞选期间，当选总统奥巴马在星期一美式足球之夜中亮相。ESPN体育比赛主持人克里斯·伯曼（Chris Berman）请他说一说，在与体育有关的事情中，有哪一件是他将要加以改变的。奥巴马提到了他对使用BCS评分系统来决定季后赛参赛队的厌恶。如同许多粉丝一样，他更希望通过有8支队伍参加的季后赛来决出全国冠军的头衔。许多粉丝都希望总统能简单地将这一建议签署为法律。

在2009年，犹他州参议员奥林·哈奇（Orrin Hatch）请求美国司法部重新检视BCS的合法性，以确定它是否有违背反垄断法的地方。促使哈奇做出这一举动的是如下事实，即在2008年，尽管犹他大学一场未败，但由于BCS系统的古怪，它却未能获得争夺冠军的资格。哈奇和其他人引述了居于这一“巨大不公”之核心的反垄断问题，称BCS对于某些联盟有所偏袒。比方说，在BCS的规则下，6个联盟的冠军将自动参加最顶级的季后赛。

大西洋沿岸联盟　→橙碗
大12联盟　→节日碗
大东部联盟　→无特定的BCS碗赛
大10联盟　→玫瑰碗
太平洋-12联盟　→玫瑰碗
东南联盟　→糖碗

此外，这6个联盟因为在这些比赛中出场而得到了比其他联盟更多的金钱。但是，联盟的组成目前处于一种流动性的状态，因此当你在读到这一部分时，可能已经发生了若干变化。

截至2011年5月，BCS已接到通知，说明他们正在受到美国司法部的彻查。据报道，司法部反垄断部门的领导、助理首席检察官克里斯汀·瓦尔尼（Christine Varney）给马克·埃默特（Mark Emmert，当时的NCAA主席）发去了一封两页纸的信，信中指出：“不断出现的严重问题表明，目前的碗系列锦标赛（BCS）体系，其行为可能并不符合联邦反垄断法中所体现的竞争原则。”

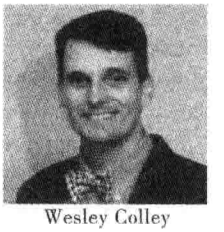
数字——

$ 21 200 000 = 2010 年每个 BCS 联盟所获得的收入。

$ 6 000 000 = 支付给 SEC 和大 10 联盟的额外收入（这两个联盟都有两支队伍参加了 BCS 比赛）。

——illegalshift. com

第 3 章　科利法


Wesley Colley

2001 年，接受了天体物理学家教育的威斯利·科利博士，撰写了一篇关于他对体育队伍进行排名的新方法的论文[22]。他的这一副产品大获成功，以至于如同梅西的模型一样，它现在也被纳入了 BCS 对 NCAA 大学美式橄榄球队进行排名的方法之中。这个我们称为科利评分法的方法，是对一个最简单而且最古老的评分体系的改进，这个评分体系所用的便是获胜率。根据获胜率为队伍 i 打出的评分值 r_i 为

$$r_i = \frac{w_i}{t_i},$$

式中，w_i 和 t_i 分别为队伍 i 所赢得的比赛场数和所进行的比赛总场数。在世界范围内的各种娱乐性联赛和巡回赛中，胜率法都是主要的评分方法。实际上，它也为大多数职业联赛所使用。尽管简单而且易于使用，但这一评分体系确实有着几个明显的缺陷。首先，在类似橄榄球这样的体育项目中，大多数队伍都会与同一批对手交锋相同数量的场次，因此常常会发生评分持平的情况；其次，分析中并未以任何方式对对手的强弱加以考虑，因此相比于击败联盟中最强的队伍，击败联盟中最弱的队伍也能够在评分中带来同样的好处，而这可以说是不公平的；再次，有时胜率评分会给出不太正常的结果，例如，在赛季开始时，所有队伍的赛季前评分都是 $\frac{0}{0}$，此外，随着赛季的推进，一场未胜的队伍所得的评分将为 0。威斯利·科利提出了一个我们称为科利法的体育队伍评分方法，以求能够弥补某些缺陷。

科利法的主要思想

科利法源自对传统胜率公式的一个小小的修改

$$r_i = \frac{1 + w_i}{2 + t_i} \tag{3.1}$$

在本节中我们将说明，这一修改的主要优点在于它考虑了赛程强度，即一支队伍的对手强弱。

科利对胜率公式的调整，其思路源于确定骰桌下注的拉普拉斯的“承续法则”（rule of succession）[29]。尽管对于传统胜率公式进行的这一修改看来变动不大，但相比于传统的公式而言，它却具有了若干优点。相比于无意义的 $\frac{0}{0}$ 的赛季前评分，现在每支队伍在赛季开始时都以相同的评分 $\frac{1}{2}$ 开始。此外，如果队伍 i 在赛季的第一周输给了对手，则科利评分将给出 $r_i = \frac{1}{3}$，他认为这一数值比 $r_i = 0$ 更为合理。

使用拉普拉斯法则，即式（3.1）而非标准胜率的另一个优点，与所谓赛程强度的概念有关。我们的想法是，相比于击败较弱的对手而言，一支队伍应能从战胜强队中获得更多的奖励。实际上，队伍 i 的评分应该与其对手的评分相关联。科利认为，拉普拉斯的法则——尽管是隐含地——但却包含了队伍 i 的赛程强度。注意到在式（3.1）中，所有队伍都从 $r_i = 1/2$ 开始，而随着赛季的推进，各队的评分将或高或低地偏离这一起始评分。实际上，一支队伍（以比赛获胜的形式）所得的好处，就是另一支队伍的损失，因此这些评分之间是相互依赖的。这样的相互依赖性在式（3.1）中并不明显，而只能通过仔细地剖析才能加以揭示。首先，我们将一支队伍获胜的场数加以分解。

$$\begin{aligned} w_i &= \frac{w_i - l_i}{2} + \frac{w_i + l_i}{2} \\ &= \frac{w_i - l_i}{2} + \frac{t_i}{2} \\ &= \frac{w_i - l_i}{2} + \sum_{j=1}^{t_i} \frac{1}{2}。 \end{aligned}$$

由于所有的队伍在赛季初都从 $r_j = 1/2$ 这个评分开始，因此总和 $\sum_{j=1}^{t_i} \frac{1}{2}$ 在初始时等于 $\sum_{j \in O_i} r_j$，其中，O_i 为队伍 i 的对手集合。随着赛季的展开，和式 $\sum_{j=1}^{t_i} \frac{1}{2}$ 便有可能不等于 $\sum_{j \in O_i} r_j$，不过它可以通过对手的累积评分来很好地加以近似（记住评分总是在 1/2 附近摇摆）。因此，

$$w_i \approx \frac{w_i - l_i}{2} + \sum_{j \in O_i} r_j。$$

假设上式为等式，并代入式（3.1），则可得

$$r_i = \frac{1 + (w_i - l_i)/2 + \sum_{j \in O_i} r_j}{2 + t_i}。 \tag{3.2}$$

当然，正如之前的评分体系一样，我们的目标是找出这些未知的 r_i。式（3.2）表明，在这种情况下，未知的 r_i 依赖于其他未知数，即 r_j。这也说明了科利法是如何将对手的强弱结合到队伍的评分之中的。通过更多一些代数运算，并利用矩阵的表示方法，我们发现 r_i 对其余 r_j 的依赖性并不是个问题。r_i 可以被容易地求出。实际上，式（3.2）可以被很紧凑地写为一个线性系统 $\boldsymbol{Cr} = \boldsymbol{b}$，其中，$\boldsymbol{r}_{n\times 1}$ 是未知的科利评分向量，$\boldsymbol{b}_{n\times 1}$ 是右向量，定义为 $b_i = 1 + \frac{1}{2}(w_i - l_i)$，而 $\boldsymbol{C}_{n\times n}$ 是科利系数矩阵，定义为

$$\boldsymbol{C}_{ij} = \begin{cases} 2 + t_i, i = j, \\ -n_{ij}, i \neq j, \end{cases}$$

式中，n_{ij} 为队伍 i 和 j 的交手次数。可以证明，科利系统 $\boldsymbol{Cr} = \boldsymbol{b}$ 总是具有唯一解，因为 $\boldsymbol{C}_{n\times n}$ 是可逆的。

范例

现在是时候在那个使用表 1.1 数据的小型示例上来测试一下科利法了。此时的科利线性系统 $\boldsymbol{Cr} = \boldsymbol{b}$ 如下。

$$\begin{pmatrix} 6 & -1 & -1 & -1 & -1 \\ -1 & 6 & -1 & -1 & -1 \\ -1 & -1 & 6 & -1 & -1 \\ -1 & -1 & -1 & 6 & -1 \\ -1 & -1 & -1 & -1 & 6 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ 1 \\ 0 \\ 2 \end{pmatrix}。$$

注意 $\boldsymbol{C}$ 是一个实对称正定阵。这些性质意味着 $\boldsymbol{C}$ 具有形如 $\boldsymbol{C} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}$ 的乔列斯基分解（又称平方根法，Cholesky decomposition）[54]，其中，$\boldsymbol{U}$ 是一个上三角阵。因此，如果得到了乔列斯基分解，就能非常高效地求解科利系统 $\boldsymbol{Cr} = \boldsymbol{b}$。但是，对于许多体育领域的应用而言，这一系统的规模都较小，从而足以使用诸如 MATLAB 这样的软件包，按标准的数值计算方法如高斯消元法、克雷洛夫法（Krylov methods）等，可以迅速计算评分向量 $\boldsymbol{r}$。将科利评分法用于我们的范例，可得如表 3.1 所示的评分以及排名。

表 3.1　5 支球队示例中的科利评分结果

球队	$\boldsymbol{r}$	排名
Duke	0.21	第 5
Miami	0.79	第 1
UNC	0.50	第 3
UVA	0.36	第 4
VT	0.65	第 2

科利法最引人注目的一点是它没有以任何方式将比分考虑在内。根据各方观点的不同，这既可以视为该模型的一个优点，也可以看做是一个缺点。注意，相比于**第 13 页**上梅西法的排名，UNC 和 UVA 的排名发生了互换。这就是不使用比分所带来的一个有趣的结果。梅西法根据 UVA 所得的更高的总分差，将它排在了 UNC 之前，尽管 UNC 的胜负记录更好些。科利认为，通过忽略比分，他的方法实现了无偏性——他用这个词来指代方法的如下能力，即能够避免因强队在弱队身上获取大的分差而带来的潜在问题。像梅西法这类使用了比分的方法，则会受到这一偏差的影响。

科利评分法总结

$\boldsymbol{C}_{n\times n}$　称为科利矩阵的实对称正定阵；

$$\boldsymbol{C}_{ij} = \begin{cases} 2 + t_i, i = j, \\ -n_{ij}, i \neq j \end{cases}$$

t_i　队伍 i 所进行的比赛总场数

n_{ij}　队伍 i 和 j 对阵的场数

$\boldsymbol{b}_{n\times 1}$　右向量；$b_i = 1 + \frac{1}{2}(w_i - l_i)$

w_i　队伍 i 累积获胜的场数

l_i　队伍 i 累积失利的场数

$\boldsymbol{r}_{n\times 1}$　科利系统给出的广义评分向量

n　联盟中队伍的总数 $= \boldsymbol{C}$ 的阶次

科利算法

求解系统 $\boldsymbol{Cr} = \boldsymbol{b}$ 以获取科利评分向量 $\boldsymbol{r}$。

以下列出的是科利法的若干性质。

- 科利法给出的结果是无偏的，这意味着这些结果仅根据胜负信息来得到，而没有使用得分的数据。因此，科利评分不会受到下面这种情况的影响，即某些队伍故意在较弱的对手身上获取高分。此外，某些体育项目或联赛比其他的项目或联赛更容易受到高分差的影响。例如，在一场典型的 NBA 比赛中，得分差通常要小于一场 NCAA 篮球赛中的分差。因此，科利法的适用性可能依赖于所考虑的联赛、项目或应用；
- 科利评分具有守恒性。每支队伍都以 1/2 这个初始评分开始一个新的赛季，而随着赛季的推进，一支队伍的评分将根据比赛结果在这个中心点上下来回波动。但总体而言，所有评分 r_i 的平均值 $\bar{\boldsymbol{r}}$（即 $\bar{\boldsymbol{r}} = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{r}/n$）恒为 1/2。因此，存在所有评分的总体守恒性。当一支队伍的评分增加时，另一支队伍的评分就要相应降低；
- 科利法同样十分适用于非体育的应用，即使在这些应用中，与分差数据对应的等价物可能无法获取或者并不适用。

梅西法和科利法之间的联系

尽管从原理上来看似乎十分不同，但在梅西法和科利法之间，却存在着一个惊人的联系。这两个方法通过公式 $\boldsymbol{C} = 2\boldsymbol{I} + \boldsymbol{M}$ 联系在了一起。因此，可以很容易地将梅西法“科利化”或

反之。例如，原始的梅西法 $\boldsymbol{Mr} = \boldsymbol{p}$ 可以科利化为

$$(2\boldsymbol{I} + \boldsymbol{M})\boldsymbol{r} = \boldsymbol{p}(\text{同样可写为 } \boldsymbol{Cr} = \boldsymbol{p})$$

科利化后的梅西法使用了包含得分信息的右向量 $\boldsymbol{p}$，而不是仅仅使用胜负信息的右向量 $\boldsymbol{b}$。此外，在系数矩阵上加上 $2\boldsymbol{I}$，则相当于将拉普拉斯法则的技巧加入了进来，并使系统变为非奇异的，从而使梅西法中通过“替换一个公式”来消除奇异性的方法变得没有必要了。科利化后的梅西法 $\boldsymbol{Cr} = \boldsymbol{p}$ 不再是无偏的，因为对 $\boldsymbol{p}$ 的运用使得比分信息被加入了进来。将科利化后的梅西法应用于范例所给出的评分向量如表3.2所示，它给出的排名与标准梅西法的结果相同，但评分的值则与**第11页**的表中所给出的结果稍有出入。

表3.2　5支球队示例中科利化后的梅西评分结果

球队	r	排名
Duke	-17.7	第5
Miami	13.0	第1
UNC	-5.7	第4
UVA	-2.4	第3
VT	12.9	第2

通过类似的方式也可以将科利法进行梅西化，梅西化后的科利法需要求解线性系统 $\boldsymbol{Mr} = \boldsymbol{b}$。

杂谈：影视排名——科利和梅西与 Netflix 相遇

Netflix 是一个影片租赁公司，通过网址 www. netflix. com 在线运营。Netflix 会员在按月支付了会员费后，便可以每月租借一定数量的影片。这些影片将直接邮递给用户，看完后再邮递回 Netflix。Netflix 定期请求用户对他们所租看的影片进行评分。这一用户-影片信息被加以收集和分析，以帮助公司今后为用户提供影片推荐。2007 年，Netflix 提供了 1 百万美元的奖金，用来奖励能够将现有推荐系统的准确性提高 10% 的个人或小组。为此次竞赛所发布的数据中的一个小样本如下所示。

$$\boldsymbol{U} = \begin{array}{c} \\ \text{用户 1} \\ \text{用户 2} \\ \text{用户 3} \\ \text{用户 4} \\ \text{用户 5} \\ \text{用户 6} \end{array} \begin{array}{c} \begin{array}{cccc} \text{影片 1} & \text{影片 2} & \text{影片 3} & \text{影片 4} \end{array} \\ \begin{pmatrix} 5 & 4 & 3 & 0 \\ 5 & 5 & 3 & 1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 2 & 0 \\ 4 & 0 & 0 & 3 \\ 1 & 0 & 0 & 4 \end{pmatrix} \end{array}。$$

$\boldsymbol{U}$ 是一个记录了评分的用户×影片的矩阵。合法的评分值为 1 到 5 的整数值，5 为最高分。0 表示用户没有对该影片进行评分。为了能够在这样的矩阵上应用体育比赛排名的思路，我们必须考虑影片间的成对比较。比方说，为了使用科利法来对影片进行排名，我们需要从评分矩阵 $\boldsymbol{U}$ 中产生出如图3.1所示的影片-影片比较图。注意在影片1和2之间有两条链接，其中

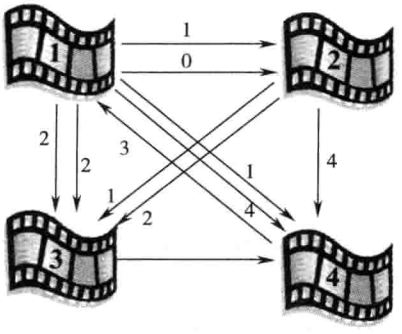


图 3.1　与梅西法相关的影片-影片比较图

一条的权值为 1，因为用户 1 对影片 1 的评分比对影片 2 的评分要多 1 分；另一条链接的权值为 0，因为用户 2 也对这两部影片进行了评分，但给出的分值相同。在图 3.1 中，每当用户同时对两部影片都进行了评分，那么这两部影片之间便存在一条以它们为首尾的链接。

如**第 24 页**上所述，科利系统 $\boldsymbol{Cr} = \boldsymbol{b}$ 可以通过恒等式 $\boldsymbol{C} = 2\boldsymbol{I} + \boldsymbol{M}$ 由梅西矩阵构建得到。图 3.1 对应的梅西矩阵为

$$\boldsymbol{M} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 7 & -2 & -2 & -3 \\ -2 & 5 & -2 & -1 \\ -2 & -2 & 5 & -1 \\ -3 & -1 & -1 & 5 \end{pmatrix} \end{array} \Rightarrow \boldsymbol{C} = 2\boldsymbol{I} + \boldsymbol{M} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 9 & -2 & -2 & -3 \\ -2 & 7 & -2 & -1 \\ -2 & -2 & 7 & -1 \\ -3 & -1 & -1 & 7 \end{pmatrix} \end{array}。$$

科利右向量 $\boldsymbol{b}$ 和梅西右向量 $\boldsymbol{p}$ 分别为

$$\boldsymbol{b} = \begin{pmatrix} 3 \\ 2 \\ -0.5 \\ -0.5 \end{pmatrix} 和 \boldsymbol{p} = \begin{pmatrix} 7 \\ 6 \\ -5 \\ -8 \end{pmatrix}。$$

表 3.3 给出了对这 4 部影片的科利评分和梅西评分及排名。

表 3.3　4 部示例影片的科利评分和梅西评分及排名

影片	科利法 r	科利法排名	梅西法 r	梅西法排名
1	0.67	1	0.65	2
2	0.63	2	1.01	1
3	0.34	4	-0.55	3
4	0.35	3	-1.11	4

对于这个 Netflix 数据集而言，平局是一个需要加以考虑的问题。由于正确的影片评分是 1 到 5 之间的整数，因此经常会发生两名用户对两部影片给出相同评分的情况。如果我们认为这种比较没有反映出哪部影片更好的信息的话，就可以简单地将这样的平分从数据集中去除。此时，“无平分”的 $\boldsymbol{M}$ 和 $\boldsymbol{C}$ 矩阵如下。

$$\boldsymbol{M}=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c}\begin{array}{cccc}1 & 2 & 3 & 4\end{array}\\ \begin{pmatrix}6 & -1 & -2 & -3\\ -1 & 4 & -2 & -1\\ -2 & -2 & 5 & -1\\ -3 & -1 & -1 & 5\end{pmatrix}\end{array},\text{从而}\ \boldsymbol{C}=\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array}\begin{array}{c}\begin{array}{cccc}1 & 2 & 3 & 4\end{array}\\ \begin{pmatrix}8 & -1 & -2 & -3\\ -1 & 6 & -2 & -1\\ -2 & -2 & 7 & -1\\ -3 & -1 & -1 & 7\end{pmatrix}\end{array}。$$

由此给出了表3.4中的“无平分”评分和排名。注意，由于对平分情况的处理方式有所不同，因此科利法对这4部影片的排名发生了变化。这说明对于平分的处理必须非常谨慎，我们将在第11章中再来讨论这个问题。

表3.4　4部影片示例中的科利法和梅西法的“无平分”评分和排名向量

影片	科利法 r	科利法排名	梅西法 r	梅西法排名
1	0.28	2	0.60	2
2	0.33	1	1.09	1
3	0.05	3	-0.55	3
4	-0.39	4	-1.13	4

我们在Netflix竞赛的一个数据样本集上运行了科利法和梅西法。该数据样本集包含了Netflix数据库（在2007年时）的所有17 770部影片，但只选取了用户集的一个子集，即13 141名“超级用户”——为1000部或更多影片进行了评分的用户[⊖]。各方法给出了评分最高的25部影片列举于表3.5中。在Netflix数据中，还包括了时间戳信息，记录了用户对影片进行评分的时间，因此，通过结合时间加权的思路来改进结果也是可行的，这一思路将在第12章中介绍。

⊖　感谢大卫·格莱克为我们提供这一数据集。

表 3.5　由科利法和梅西法给出的排名前 25 位的影片

排名	科利法的结果	梅西法的结果
1	《指环王Ⅲ：王者归来》	《指环王Ⅲ：王者归来》
2	《指环王Ⅱ：双塔奇兵》	《迷失：第一季》
3	《迷失：第一季》	《指环王Ⅱ：双塔奇兵》
4	《星球大战Ⅴ：帝国反击战》	《太空堡垒卡拉狄加：第一季》
5	《太空堡垒卡拉狄加：第一季》	《星球大战Ⅴ：帝国反击战》
6	《夺宝奇兵》	《夺宝奇兵》
7	《星球大战Ⅳ：新希望》	《星球大战Ⅳ：新希望》
8	《肖申克的救赎》	《肖申克的救赎》
9	《教父》	《教父》
10	《星球大战Ⅵ：杰迪归来》	《星球大战Ⅵ：杰迪归来》
11	《黑道家族：第五季》	《黑道家族：第五季》
12	《辛德勒的名单》	《好家伙》
13	《指环王Ⅰ：魔戒现身》	《兄弟连》
14	《兄弟连》	《辛普森一家：第五季》
15	《黑道家族：第一季》	《指环王Ⅰ：魔戒现身》
16	《辛普森一家：第五季》	《辛德勒的名单》
17	《角斗士》	《黑道家族：第一季》
18	《犯罪现场调查：第四季》	《海底总动员》
19	《黑道家族：第二季》	《辛普森一家：第六季》
20	《黑道家族：第四季》	《角斗士》
21	《辛普森一家：第六季》	《辛普森一家：第四季》
22	《辛普森一家：第四季》	《黑道家族：第二季》
23	《海底总动员》	《教父》㊀
24	《玩具总动员》	《黑道家族：第四季》
25	《沉默的羔羊》	《犯罪现场调查：第四季》

数字——

$ 2 782 275 172 = 单部影片所赚得的最高总收入（截至 2011 年）。——《阿凡达》（2009 年发行）。

$ 1 843 201 268 = 第二高的总收入（截至 2011 年）。——《泰坦尼克》（1997 年发行）。

——en. wikipedia. org

㊀ 原文如此。——译者注

第 4 章　基纳法

詹姆斯-P. 基纳在 1993 年发表于《SIAM 评论》的一篇论文中，提出了他的评分方法[42]。类似于其他许多方法，基纳法利用竞争者（队伍）之间竞争（比赛）的非负统计量，来为每支队伍产生一个数值评分。在某些行业里，该评分被称为能力评分。当然，一旦为每支队伍给出了一个数值评分，就自然能根据评分得到队伍排名。

基纳法将给定队伍的评分与该队的绝对实力联系了起来，而绝对实力又依赖于队伍的相对实力——即相对于其对手实力而言的该队伍的强弱。随着内容的展开，这一点将变得更为清楚。基纳将一切都植根于两条规则，也正是它们决定了一支队伍的实力和评分之间的关系。

实力和评分的规则

1. 一支队伍的实力应根据它与对手之间的相互作用以及对手的实力来加以评判。
2. 在给定的联赛中，每支队伍的评分应与其实力成正比例关系。更具体来说，如果 s_i 和 r_i 分别为队伍 i 的实力值和评分值，则应该存在一个比例常数 λ 使得 $s_i = \lambda r_i$，且对于联赛中的每支队伍而言，λ 的值都相等。

选择体现实力的属性

为了将上述两条规则转变为可以用来计算每支队伍的数值评分的机制，让我们对问题中的某些变量赋予不同的量。首先，让我们确定所考虑的竞争或体育项目中的某些属性或统计数据，我们认为在这些属性之上，可以合理地进行队伍实力的相对比较，并令

$$a_{ij} = \text{队伍 } i \text{ 与队伍 } j \text{ 竞争时给出的统计数据值}$$

举例来说，我们可以考虑使用的最简单的统计数据之一，就是队伍 i 在本赛季中击败队伍 j 的次数。对于每个平局，我们给两支参赛队都赋予 1/2。换而言之，若 W_{ij} 是队伍 i 最近击败队伍 j 的次数，且若 T_{ij} 是队伍 i 和 j 战平的次数，则可以设

$$a_{ij} = W_{ij} + \frac{T_{ij}}{2}。$$

如果你觉得队伍 i 从队伍 j 上所赢取的分数 S_{ij} 是体现 i 相对于 j 的实力的一个更为相关的属性，则可以定义

$$a_{ij} = S_{ij}。$$

若队伍 i 和 j 交手不止一次，则 S_{ij} 可以取 i 从 j 所赢得的总的分数。

除了比赛得分和获胜次数外，你也可以选用其他属性。例如，如果你是一名老式的大学橄榄球球迷，并相信更强的队伍是那些在奔跑中胜过他人的队伍，那么你可能愿意取

$$a_{ij} = \text{队伍 } i \text{ 对队伍 } j \text{ 所完成的累积冲球码数。}$$

类似地，一名 NFL 的现代行家则可能认为传球才是最重要的，此时便有

$$a_{ij} = \text{队伍 } i \text{ 对队伍 } j \text{ 所完成的传球码数。}$$

不管你选择哪个属性来定义 a_{ij}，你都必须在整个赛季中不断地更新属性数据（当然，如果你是在所有比赛都已经结束之后才来对联赛进行评价的话，那就可以不进行更新）。但是在更新 a_{ij} 时，你还必须确定一件事，即某个属性在赛季初的值是否应当与赛季结束时的值同等重要——即你是否希望按时间对 a_{ij} 进行加权？

所有这些灵活性使我们可以对排名进行许多变动和微调，而构建评分和排名模型之所以如此有趣，也正是由于这一点。正如你已经开始了解的那样，基纳法特别适于进行修修补补，而随着我们介绍的深入，我们将有更多的机会接触这些方面。

拉普拉斯的承续法则

无论选择了何种属性，该属性的统计数据都很少能够以原始的形式来产生一种成功的评分方法。例如，考虑得分 S_{ij}。如果队伍 i 和 j 都是防守薄弱而进攻出色，那它们在相互对阵时就都能获得很高的分数。另一方面，如果队伍 p 和 q 都是防守牢固而进攻乏力，则它们之间的比赛就很可能只有很低的分数。如此一来，在任何后续的评分系统中，相比于小的 S_{pq} 和 S_{qp} 值，大的 S_{ij} 和 S_{ji} 值将对排名产生不适当的影响。当比较队伍 i 和 j 时，更好的做法是考虑两队的总得分，并设

$$a_{ij} = \frac{S_{ij}}{S_{ij} + S_{ji}} \tag{4.1}$$

类似的注意事项对于其他属性也同样成立——例如，冲码码数、传球码数、失误次数等。

利用式（4.1）这样产生的比例 a_{ij} 将比原始值更好，但正如基纳所指出的，我们实际上应该使用类似于

$$a_{ij} = \frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2} \tag{4.2}$$

这样的做法。这样做的缘由在于拉普拉斯的承续法则[29]，其直观思路是，如果

$$p_{ij} = \frac{S_{ij}}{(S_{ij} + S_{ji})}, \text{则 } 0 \leqslant p_{ij} \leqslant 1。$$

且 p_{ij} 可以被解释为队伍 i 在今后击败队伍 j 的概率。如果队伍 i 在过去的比赛中以累积总分 X 比 0 击败了队伍 j（其中 $X > 0$），则可以合理地认定队伍 i 在某种程度上强于队伍 j（如果 X 的值较小，则队伍 i 稍强于 j，而如果 X 的值很大，则 i 远强于 j）。但是，无论 $X > 0$ 的值是多少，我们都有

$$p_{ij} = 1 \text{ 以及 } p_{ji} = 0。 \tag{4.3}$$

这意味着队伍 j 在今后根本不可能击败队伍 i，而这显然是不成立的。另一方面，如果式（4.2）被解读为队伍 i 在今后击败队伍 j 的概率 p_{ij}，则由式（4.2）可见 $0 < p_{ij} < 1$，而如果在过去队伍 i 以累计得分 X 比 0 击败了队伍 j，则

$$p_{ij} = \frac{X + 1}{X + 2} \to \begin{cases} 1/2, \text{当 } X \to 0 \text{ 时}, \\ 1, \text{当 } X \to \infty \text{ 时}, \end{cases}$$

这比式（4.3）更为合理。此外，如果 $S_{ij} \approx S_{ji}$，且都为大值，则 $p_{ij} \approx 1/2$，但随着分差 $S_{ij} - S_{ji} > 0$ 增加，p_{ij} 变得更接近于 1，这也是合理的。因此，相比于式（4.1），式（4.2）更适于使用。

有所偏好，还是不要偏好

偏好考虑的是如何补偿“我们就是行”的情况，此时强队将毫不手软地在较弱的对手身上捞取大把的分数，这么做或者是为了提高自身的评分，或者就是想“落井下石”。如果使用式（4.2），则基纳建议对每个 a_{ij} 采用一个非线性的*偏好函数*，如

$$h(x) = \frac{1}{2} + \frac{\operatorname{sgn}\{x - (1/2)\} \sqrt{|2x - 1|}}{2}。 \tag{4.4}$$

以帮助减小强弱之间的差距。换而言之，就是以 $h(a_{ij})$ 来代替 a_{ij}。$h(x)$ 的曲线如图 4.1 所示，可以发现该函数具有以下性质，即 $h(0) = 0$，$h(1) = 1$，且 $h(1/2) = 1/2$。使用 $h(a_{ij})$ 来代替 a_{ij}，可以带来我们所希望的效果，即最强者和最弱者之间差距不那么极端。利用 $h(x)$ 实现的偏好，在值 1/2 附近还造成了对两个原始值之间差距的人为拉大，在区分实力接近的队伍时，这可能将有所帮助。

一旦你了解了式（4.4）中基纳的偏好函数并用过一遍以后，就应该能够构造出许多你自己的偏好函数。例如，你可能希望定制一个偏好函数，使它的偏好能力较 $h(x)$ 更强或更弱一些，或者你希望函数对 a_{ij} 值较大的一端的影响与对 a_{ij} 值较小的一端的影响有所不同。基纳提出的偏好的思路，其价值在于你能够针对特定赛式的建模需要来“调整”你

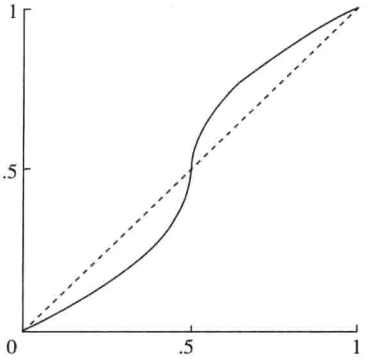


图 4.1　偏好函数 $h(x)$ 的图形

的系统。偏好只不过是对基纳法进行调整的不计其数的方法中的一种而已。

偏好并非总是必要的。例如，偏好对参考文献［34，35，1］中所研究的 NFL 排名并没有显著影响。这可能是由于在所考虑的时间段内，各支 NFL 球队的实力相当均衡，因此也就不那么需要进行精细调整。当然，对于 NCAA 的体育项目或其他竞赛而言，情况往往大不相同，因此不要忽视基纳的这个偏好的思路。

规范化

在确定了如何定义非负 a_{ij} 值的属性，并确定了是否希望通过重定义 $a_{ij} \leftarrow h(a_{ij})$ 来使这些值有所偏好之后，如果并不是所有队伍都进行了同样多场比赛的话，那么还需要对 a_{ij} 进行最后一点改动（或者说规范化）。此时，我们进行如下替换

$$a_{ij} \leftarrow \frac{a_{ij}}{n_i} \quad （其中，n_i = 队伍\ i\ 所进行的比赛场数） \tag{4.5}$$

为了理解这一做法的必要性，假设我们用 S_{ij} 来按照式（4.1）或式（4.2）定义 a_{ij}。如果某些队伍比其他队伍比赛了更多场次，那么它们就有可能获得更多的得分，从而带来更大的 a_{ij} 值，而这将影响到最终由 a_{ij} 所导出的任何“实力”度量。对于得分之外的其他统计数据来说，这一点同样成立——比方说，如果使用的统计数据是冲球码数或传球码数的话，那么比赛场次更多的队伍就可能累积获得更高的统计数据，并由此影响由这些统计数据给出的任何评分。

注意！偏好本身就具有规范化的效果，因此，如果使用了偏好函数，而各队的比赛场数之间的差异又不是很大时，式（4.5）可能就并非必需——而且这时还会面临着数据“过度规范化”的风险。此外，针对不同情况，需要不同的规范化策略。利用式（4.5）是最为简单的做法，也是一个不错的出发点，但你可能还需要创造其他的策略并加以实验，以获得最优的结果——这又是一个可以进行调整的地方。

先有鸡，还是先有蛋

一旦定义了 a_{ij} 并对其进行了偏好和规范化后，就可以将数据组织为一个方阵的形式

$$\boldsymbol{A} = [a_{ij}]_{m\times m} \quad (\text{其中}, m = \text{联赛中队伍的数量})。$$

将数据以这种方式加以表达，可以使我们运用矩阵论中某些极其有力的思路来量化“实力”并产生数值评分。尽管现在看来还不明显，但有必要指出，每支队伍的“实力值”和它的“评分值”之间有着微妙的区别。在后文中，这一区别将变得越发明显。尽管“实力值”和“评分值”并非等同，但它们显然是有所关联的。这就将我们带入了一个先有鸡还是先有蛋的境地，因为我们希望用“评分值”来衡量“实力值”，但每支队伍的“实力”又会影响队伍的“评分”。

评分

让我们不管那么多，干脆闭着眼睛随便选一个，就从评分开始吧。我们希望在当前赛季中的某个时刻 t，根据这 m 支队伍中每支队伍的、我们所认为的“实力”（我们现在仍然还不知道它到底是什么）来构建出队伍的评分值。令

$$r_j(t) = \text{队伍 } j \text{ 在时刻 } t \text{ 的量化评分},$$

并令 $\boldsymbol{r}(t)$ 表示包含了所有 m 个评分的列向量。可以理解，评分向量 $\boldsymbol{r}(t)$ 中的每个值 $r_j(t)$ 都会随时间改变，因此记号中明确表示出的时间其实可以省略——即我们简单地用 r_j 来代替 $r_j(t)$，并设

$$\boldsymbol{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = \text{评分向量}。 \tag{4.6}$$

尽管 $\boldsymbol{r}$ 中的值现在对我们而言是未知的，但我们仍假设它们存在且可能通过某种方式加以确定。

实力

现在我们把式（4.6）中（目前仅在理论上存在）的评分与“实力”的概念联系起来。回忆一下，基纳的第一条规则指出，一支队伍的实力应该由它与对手比赛的表现加以衡量，同时根据对手的实力加以调整。换而言之，赢得十场轻而易举的较量，应该与赢得十场与强手的比赛要有所区别。

队伍 i 对阵队伍 j 的表现如何，正是我们希望统计数据 a_{ij} 加以衡量的东西，而队伍 j 到底多强（或弱）则是我们希望评分值 r_j 加以衡量的东西。因此，采用如下的定义是合理的。

相对实力

队伍 i 相比于队伍 j 的相对实力定义为

$$s_{ij} = a_{ij} r_j 。$$

我们很自然地把队伍 i 的总实力或绝对实力取为队伍 i 与联盟中所有其他队伍相比的相对实力之和。换而言之，可以合理地定义绝对实力（或简称实力）如下。

绝对实力

队伍 i 的绝对实力（或简称实力）定义为

$$s_i = \sum_{j=1}^{m} s_{ij} = \sum_{j=1}^{m} a_{ij} r_j，且 \boldsymbol{s} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix} 为实力向量。 \tag{4.7}$$

注意，实力向量 $\boldsymbol{s}$ 可以表示为

$$\boldsymbol{s} = \begin{pmatrix} \sum_j a_{1j} r_j \\ \sum_j a_{2j} r_j \\ \vdots \\ \sum_j a_{mj} r_j \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = \boldsymbol{A}\boldsymbol{r}。 \tag{4.8}$$

基本方程

基纳关于实力和评分之间关系的第二条规则，要求每支队伍的实力和评分成正比例，即存在一个比例常数 λ，使得对于每个 i 都有 $s_i = \lambda r_i$。用式（4.6）和式（4.7）中的评分向量 $\boldsymbol{r}$ 和实力向量 $\boldsymbol{s}$ 来表达，则这一规则要求对某个常数 λ，有 $\boldsymbol{s} = \lambda \boldsymbol{r}$。但是，式（4.8）说明 $\boldsymbol{s} = \boldsymbol{A}\boldsymbol{r}$，因此我们可以得出结论，即评分向量 $\boldsymbol{r}$ 与矩阵 $\boldsymbol{A}$ 中的统计数据 a_{ij} 可通过以下方程被联系起来

$$\boldsymbol{A}\boldsymbol{r} = \lambda \boldsymbol{r}。 \tag{4.9}$$

这一方程就是基纳法的根本！

用线性代数的语言来说，方程（4.9）指出评分向量 $\boldsymbol{r}$ 必定是一个特征向量，而比例常数 λ 必定是矩阵 $\boldsymbol{A}$ 的一个特征值。有关特征值和特征向量的知识，在因特网上可以找到许多，

但我们推荐的正式出版的文献来源则是参考文献［54］的第 489 页。

乍一看去，确定评分向量 $\boldsymbol{r}$ 的问题不仅已经解决，而且式（4.9）的求解看来出奇的简单——只要找到某个矩阵 $\boldsymbol{A}$ 的特征值和特征向量就可以了。但是，如果我们不能把确定 $\boldsymbol{r}$ 的问题局限为一般特征值问题 $\boldsymbol{Ar} = \lambda \boldsymbol{r}$ 的一个特殊的情形的话，那么我们将会遇到麻烦。其中的原因有很多，下面就举出若干。

1. 对于一个一般的 $m \times m$ 矩阵 $\boldsymbol{A}$，求解式（4.9）可以得出多达 m 个不同的 λ 值。这将使我们陷入窘境，因为我们将不得不从中挑选出一个 λ 值。而 λ 的选择又会进而影响所得到的评分向量，因为一旦 λ 确定，则评分向量 $\boldsymbol{r}$ 将作为方程 $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{r} = 0$ 的解而随之确定。
2. 尽管统计数据矩阵 $\boldsymbol{A}$ 中只包含实数，但由式（4.9）所得的某些（甚至全部）λ 值则可能是复数，而这又将使得与之对应的特征向量中包含复数。对于评分和排名而言，这样的 $\boldsymbol{r}$ 是毫无用处的。例如，将一支队伍的评分设为 $6 + 5\mathrm{i}$ 而另一支的评分设为 $6 - 5\mathrm{i}$ 是毫无意义的，因为我们不可能对这两个数进行比较，并确定谁更大一些。
3. 即使在最好的情况下，求解式（4.9）得到了实数的 λ 值，但它们还可能是负数。而哪怕得到了一个正特征值 λ，与之相应的特征向量 $\boldsymbol{r}$ 可能（而且也常常会）包含某些负值元素。虽说负的评分还不至于像复数评分那么糟糕，但这仍然不是最佳的结果。
4. 最后，我们不得不面对以下问题，即到底打算如何实际计算出矩阵的特征值和特征向量，以便获得 $\boldsymbol{r}$。对于规模较大的一般方阵 $\boldsymbol{A}$，$\boldsymbol{Ar} = \lambda \boldsymbol{r}$ 的求解所涉及的编程和计算复杂性将迫使人们去购买或得到用于成熟的特征值计算的软件包的使用权——而大多数这样的软件包都如此昂贵，以至于它们的包装盒上都应该镀上金才配得上那高昂的身价。

约束

基纳避开了以上所有绊脚石，他是通过对球队之间相互作用的程度以及由此所得的 $\boldsymbol{A}_{m \times m}$ 中的统计数据 a_{ij} 施加了 3 个并不苛刻的约束来做到这一点的。

Ⅰ. **非负性**。不管用什么属性来确定所用的统计数据 a_{ij}，也不管如何对这些数据进行调整、偏好或规范化，最后都必须保证每个统计数据 a_{ij} 均非负——即 $\boldsymbol{A} = [a_{ij}] \geqslant \boldsymbol{0}$ 是一个非负矩阵。

Ⅱ. **不可约性**。联赛必须有足够多的以往交手记录，以保证能够对任意两支队伍进行比较，即使它们并未直接相互交手。更准确地来说，如果队伍 i 和 j 是联赛中的两支

不同球队，那它们必须能够通过一系列牵涉到其他队伍 $\{k_1, k_2, \cdots, k_p\}$ 的以往的交锋而被“联系”起来，从而通过这些队伍形成一条比赛串

$$i \leftrightarrow k_1 \leftrightarrow k_2 \leftrightarrow \cdots \leftrightarrow k_p \leftrightarrow j \quad （其中，a_{ik_1} > 0, a_{k_1k_2} > 0, \cdots, a_{k_pj} > 0） \tag{4.10}$$

对这一约束的技术性说法，便是称这个联赛（或矩阵 $\boldsymbol{A}$）是不可约的；

Ⅲ. **素性**。这只是Ⅱ中不可约性约束的一个更为严格的版本，即我们现在要求每两支队伍都被均等数量的比赛所联系。换而言之，式（4.10）中队伍 i 和 j 之间的联系，对应了一条由 p 场具有正统计数据的比赛所构成的链。对于另外两支队伍，Ⅱ仅要求它们是联系的，但它们可以通过一条由 q 场具有正统计数据的比赛所构成的链而联系起来，而 $p \neq q$。素性则要求所有队伍都必须通过相同数量的比赛场数被联系起来——即必须存在一个值 p，使得式（4.10）对所有的 i 和 j 都成立。素性约束实际等价于要求对于某个幂次 p，有 $\boldsymbol{A}^p > \boldsymbol{0}$。之所以要求素性成立，是因为它可以使得评分向量 $\boldsymbol{r}$ 的最终计算易于执行。

佩隆-弗罗贝尼乌斯

通过施加上述的约束Ⅰ和Ⅱ，强有力的佩隆-弗罗贝尼乌斯定理便能被用来承担由基纳的基本方程 $\boldsymbol{Ar} = \lambda \boldsymbol{r}$ 获取唯一评分向量的任务。相比于我们这里所能展示（以及我们所需要）的部分，完整的定理中还有着多得多的内容，但其核心本质将在下面给出——完整的描述请见参考文献［54］的第 661 页。

佩隆-弗罗贝尼乌斯定理

若 $\boldsymbol{A}_{m\times m} \geqslant \boldsymbol{0}$ 不可约，则以下陈述均成立。

- 在满足 $\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$ 的所有 λ_i 值以及对应的向量 $\boldsymbol{x}_i \neq \boldsymbol{0}$ 中，存在一个值 λ 和向量 $\boldsymbol{x}$ 满足 $\boldsymbol{Ar} = \lambda \boldsymbol{r}$，且
  - ▶λ 为实数
  - ▶$\lambda > 0$
  - ▶对所有的 i，有 $\lambda \geqslant |\lambda_i|$
  - ▶$x > 0$
- 除了 $\boldsymbol{x}$ 的正数倍数之外，不管特征值 λ_i 为何值，不存在 $\boldsymbol{A}$ 的其他非负特征向量 $\boldsymbol{x}_i$；
- 存在唯一一个向量 $\boldsymbol{r}$（具体而言就是 $r = x / \sum_j x_j$），满足

$$\boldsymbol{Ar} = \lambda \boldsymbol{r}, \boldsymbol{r} > \boldsymbol{0}, \text{且} \sum_{j=1}^{m} r_j = 1。 \tag{4.11}$$

- λ 值和 $\boldsymbol{r}$ 向量分别称为佩隆值和佩隆向量。对我们来说，佩隆值 λ 就是式（4.9）中的比例常数，而唯一的佩隆向量 $\boldsymbol{r}$ 就是我们的评分向量。

重要的性质

请注意，由于是由 $\boldsymbol{A}$ 的佩隆向量所定义得到的，所以评分向量 $\boldsymbol{r}$ 不仅由 $\boldsymbol{A}$ 中的统计数据 a_{ij} 所唯一确定，而且 $\boldsymbol{r}$ 还具有如下性质，即每支队伍的评分都满足

$$0 < r_i < 1 \text{ 且 } \sum_{j=1}^{m} r_i = 1。$$

这一点十分重要，因为它使每支队伍都能得到平等的比较，并且我们能够用百分比来描述实力——例如，我们可以作出一幅漂亮的饼图，来图形化地说明一支队伍相比于联盟中其他队伍而言的强弱。评分之和为 1 则意味着，每当某支队伍的评分有所提高时，就必然会有另一支或多支队伍的评分随之下降，因此在整个给定的赛季或多个年份及不同赛季中，我们都得以维持一种均衡。此外，这种均衡使我们可以对由两个不同的属性所给出的评分进行公平比较。例如，假设我们根据得分 S_{ij} 得到一组基纳评分，而又根据获胜（以及平局）场数 W_{ij} 来得到另一组基纳评分，则这两组评分和所得的排名可以被加以比较、协调甚至聚合起来，以产生有时被称为一致（或总体）的评分和排名。排名聚合的方法将在第 14 章中详细讨论。

计算评分向量

假定已经保证所有统计数据 a_{ij} 均非负（即 $\boldsymbol{A} \geqslant 0$），则接下来首先要检查约束 Ⅱ 所要求的不可约性是否得到满足。假设不可约性得到了满足，如果不是这样，那我们必须通过某些修改来做到这一点，在**第 39 页**上就给出了若干可能的修改方法。一旦确定了 $\boldsymbol{A}$ 不可约，我们就有两种可选的方式来计算 $\boldsymbol{r}$。

强力法。如果能够使用某个镀金包装的、能够计算特征值和特征向量的数值计算软件，那么你就可以简单地将矩阵 $\boldsymbol{A}$ 送入软件，并让它吐出所有的特征值和特征向量。

- 查看返回的特征值列表，并找到其中为正实数且幅值比所有其他特征值都更大的那一个。佩隆-弗罗贝尼乌斯定理确保将只存在一个这样的值，它就是佩隆根——即我们所要寻找的 λ 的值。
- 接下来让软件返回 λ 对应的特征向量 $\boldsymbol{x}$。这个向量并不一定是评分向量 $\boldsymbol{r}$。根据你所使用的软件，这个向量 $\boldsymbol{x} = (x_1, x_2, \cdots, x_m)$ 可能作为一个全为负数或全为正数的向量被返回（但不会有零——如果有，那就是什么地方出错了）。即使 $\boldsymbol{x} > \boldsymbol{0}$，其元素之和也可能不等于 1，因此一般需要通过令

$$\boldsymbol{r} = \frac{\boldsymbol{x}}{\sum_{i=1}^{m} x_i},$$

来迫使和为 1 这一点得以成立。

这就是佩隆向量，而 $\boldsymbol{r}$ 即为所要获取的评分向量。

幂法。如果Ⅲ中的素性条件得到了满足，那就有一条相对简单的途径来计算 $\boldsymbol{r}$，这个计算不需要多少编程技巧或计算能力。该方法称为幂法，因为它依赖于如下事实，即随着 k 的增加，作用于任意正向量 $\boldsymbol{x}_0$ 的 $\boldsymbol{A}$ 的幂 $\boldsymbol{A}^k$ 都将驱使乘积 $\boldsymbol{A}^k\boldsymbol{x}_0$ 趋向于与佩隆向量同向——如果有兴趣知道原因的话，请参见参考文献［54］的第 533 页。因此，通过 $\boldsymbol{A}^k\boldsymbol{x}_0$ 元素之和对其进行比例伸缩，并令 $k\to\infty$，就可以得到佩隆向量（即评分向量）$\boldsymbol{r}$，为

$$\boldsymbol{r}=\lim_{k\to\infty}\frac{\boldsymbol{A}^k\boldsymbol{x}_0}{\sum_{i=1}^{m}\boldsymbol{A}^k\boldsymbol{x}_0}。$$

$\boldsymbol{r}$ 的幂法计算过程如下。

— 选择一个初始正向量 $\boldsymbol{x}_0$ 来开始计算。均匀向量

$$\boldsymbol{x}_0=\begin{pmatrix}1/m\\1/m\\\vdots\\1/m\end{pmatrix}$$

通常都会是初始向量的一个不错选择。

— 相比于直接计算幂 $\boldsymbol{A}^k$ 并与 $\boldsymbol{x}_0$ 相乘，利用如下的迭代计算方法所需要的计算量要少得多。设

$$\boldsymbol{y}_k=\boldsymbol{A}\boldsymbol{x}_k,\boldsymbol{\nu}_k=\sum_{i=1}^{m}(\boldsymbol{y}_k)_i,\boldsymbol{x}_{k+1}=\frac{\boldsymbol{y}_k}{\boldsymbol{\nu}_k}\quad(\text{其中},k=0,1,2,\cdots)。\tag{4.12}$$

直接验证这一迭代可以给出所需要的序列

$$\boldsymbol{x}_1=\frac{\boldsymbol{A}\boldsymbol{x}_0}{\sum_{i=1}^{m}(\boldsymbol{A}\boldsymbol{x}_0)_i},$$

$$\boldsymbol{x}_2=\frac{\boldsymbol{A}\boldsymbol{x}_1}{\sum_{i=1}^{m}(\boldsymbol{A}\boldsymbol{x}_1)_i}=\frac{\boldsymbol{A}^2\boldsymbol{x}_0}{\sum_{i=1}^{m}(\boldsymbol{A}^2\boldsymbol{x}_0)_i},$$

$$\boldsymbol{x}_3=\frac{\boldsymbol{A}\boldsymbol{x}_2}{\sum_{i=1}^{m}(\boldsymbol{A}\boldsymbol{x}_2)_i}=\frac{\boldsymbol{A}^3\boldsymbol{x}_0}{\sum_{i=1}^{m}(\boldsymbol{A}^3\boldsymbol{x}_0)_i},$$

$$\vdots$$

素性条件可以保证

$$\text{当 } k\to\infty \text{ 时},\boldsymbol{x}_k\to\boldsymbol{r}。$$

（数学细节请见参考文献［54］的第 674 页）在实际运算中，当 $\boldsymbol{x}_k$ 的元素已经收敛到了足够多的有效位数，从而能将联盟中的队伍相互区分开来时，迭代式（4.12）就可以终止了。如果考虑的队伍数量越多，则需要的有效位数也越多——请参见**第 42 页**上 NFL 的评分结果，来体会一下所需要的有效位数为多少。只是注意不要太矫枉过正，除非你想要对多得出乎寻常的队伍进行评分和排名，否则你没必要收敛到 16 个有效数位。

强制形成不可约性和素性

正如在**第 35 页**上对约束的讨论中所说明的那样，不可约性和素性条件要求所考虑的队伍之间进行了充分的交手，而这些交手对于良好定义的评分向量 $\boldsymbol{r}$ 及其幂法求取而言是必要的。这就带来了一个自然而然的问题："我们如何才能实实在在地核实我们的联赛（或矩阵 $\boldsymbol{A}$）是否具有这一类连通性呢?"

对不可约性的检查没有捷径可走，实际上就是根据定义来检查——即对联赛中的每两支队伍 (i,j)，验证其是否存在一系列的比赛

$$i \leftrightarrow k_1 \leftrightarrow k_2 \leftrightarrow \cdots \leftrightarrow k_p \leftrightarrow j \quad (\text{其中}, a_{ik_1} > 0, a_{k_1k_2} > 0, \cdots, a_{k_pj} > 0)$$

将它们联系起来。类似地，对于素性，必须验证是否存在一个均等的比赛场数，能够将联赛中任意两支队伍联系起来。这些检查都十分枯燥冗长，但只要队伍数量 m 不是太大，计算机就可以帮助我们完成。

但是，检查连通性的计算方面的问题常常都不是最大的问题，因为对于大多数竞赛而言，这两种连通性很少都能得到满足。如果是从比赛开始时，就以一种不间断的方式（例如逐周地）来构建评分（和排名），并在整个赛季中不断对之进行更新时，就尤其如此。在赛季的初期，是极不可能拥有足够的连通性来保证不可约性和素性成立的，而且通常来说，即使到了一个赛季的尾声，这些连通性条件也不能得到满足。

因此，更重要的问题是："我们如何才能够在竞赛中强制形成不可约性和素性，从而我们根本都无需对其进行检查?"一个特别简单的解决方法，就是对 $\boldsymbol{A}$ 施加一个小的扰动并以之来代替原有的 $\boldsymbol{A}$，这一重定义如下

$$\boldsymbol{A} \leftarrow \boldsymbol{A} + \boldsymbol{E}, \text{其中}, \boldsymbol{E} = \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}} = \begin{pmatrix} \varepsilon & \varepsilon & \cdots & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \cdots & \varepsilon & \varepsilon \\ \vdots & \vdots & & \vdots & \vdots \\ \varepsilon & \varepsilon & \cdots & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & \cdots & \varepsilon & \varepsilon \end{pmatrix}$$

式中，$\boldsymbol{e}$ 为全 1 列；$\varepsilon > 0$ 为某个数，它相比于扰动前的矩阵 $\boldsymbol{A}$ 中最小的非零元素值而言，也仍是一个小数。这一做法的效果相当于在每两支队伍之间都引入了一场虚拟的比赛，这些虚拟比赛的统计数据都足够小，从而不会扭曲真实比赛中所蕴含的信息。注意，在这个经过扰动的竞赛中，每支队伍现在都与任一支其他队伍直接相连（而且具有正的统计数据）[⊖]，因此不可约性和素性就都得到了保证。

如果已知进行了足够数量的比赛来保证不可约性的话，则可以用一种更轻微的扰动来保证素性。如果竞赛（或矩阵 $\boldsymbol{A}$）是不可约的，则在 $\boldsymbol{A}$ 的任一对角位置加上任一值 $\varepsilon > 0$ 便能形成素性。换而言之，就是在某支队伍和其自身之间增加一场统计数据可以忽略的虚拟比赛，或等价地，将 $\boldsymbol{A}$ 重定义为

⊖ 如果每支队伍和其自身之间也进行了虚拟比赛这一点让你不安的话，也可以将 $\boldsymbol{E}$ 的对角元素设为 0。

$$\boldsymbol{A} \leftarrow \boldsymbol{A} + \boldsymbol{E}, \text{其中}, \boldsymbol{E} = \varepsilon \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}, \boldsymbol{e}_i^{\mathrm{T}} = (0,0,\cdots,1,0,\cdots,0)$$

便能造成素性[54，第678页]。

总结

现在所有的部件都已经摆在桌面上了，让我们将它们拼起来。以下是一个总结，说明了如何根据基纳对佩隆-弗罗贝尼乌斯定理的运用，来构造评分和排名体系。

1. 首先选择一个要考虑的竞赛或体育项目的特定属性，它可以成为比较每支队伍相对于其他队伍的强弱的良好基础。这类属性的例子包括队伍 i 击败（或战平）队伍 j 的次数，或队伍 i 从队伍 j 身上取得的分数。

   — 对于给定的竞赛，当考虑更为具体的特性时，可以构建出不止一种方案——例如，一支橄榄球队在对阵另一支队伍时所完成的冲码数（或传球码数），或一支篮球队对阵另一支球队时投入的三分球数或罚球数等。甚至连防守性的属性也是可以使用的——例如，一支橄榄球队在对阵另一支球队时所拦截的传球次数，或在篮球比赛中队伍 i 对阵队伍 j 时的盖帽次数等。
   — 由竞赛的各个更为细致的方面所得到的不同的评分（和排名）可以被聚合形成一致（或总体）评分（或排名）。例如，可以根据攻击方面的属性来构建一个或多个基纳方案，并根据防守方面的属性来构建其他方案，然后将所得的多个评分结果聚合为一个主评分和排名列表。有关这一点是如何做到的细节将在第 14 章中给出。在进发到排名聚合之前，首先要透彻理解一个基于某种特定属性的切实系统。

2. 无论选取了何种属性，都要根据以往的比赛来整理出相应的统计数据，并置

   $$a_{ij} = \text{队伍 } i \text{ 对阵队伍 } j \text{ 时得到的统计数据值。}$$

   保证每个 a_{ij} 均为非负数是完全有必要的！
3. 调整第 2 步中的原始统计数据 a_{ij} 以处理异常情况。比方说，如果使用 $a_{ij} = S_{ij} =$ 队伍 i 在队伍 j 身上获取的分数，则如**第 31 页**上所说明的那样，我们应该对所有的 i 和 j，重新定义 a_{ij} 为

   $$a_{ij} = \frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2}。$$

4. 如果对数据进行调整后还觉得数据中存在着不平衡的现象，即某些 a_{ij} 远大于（或小于）它们应取的值（这可能是由于某支队伍人为地在较弱的对手身上将统计数据拔

得很高所造成的），则应当通过构建一个如**第 31 页**的讨论中所给出的偏好函数

$$h(x) = \frac{1}{2} + \frac{\operatorname{sgn}\{x - (1/2)\}\sqrt{|2x-1|}}{2}$$

来恢复数据的平衡，并进行替换 $a_{ij} \leftarrow h(a_{ij})$。

5. 如果并非所有的队伍都进行了相同场数的比赛，则通过替换

$$a_{ij} \leftarrow \frac{a_{ij}}{n_i} \quad （其中，n_i = 队伍\ i\ 进行的比赛场数）$$

来对第 4 步中的 a_{ij} 进行规范化处理以应对这种情况，并将数据组织为一个非负矩阵 $\boldsymbol{A} = [a_{ij}] \geqslant \boldsymbol{0}$。

6. 或者检查联赛中是否已经进行了足够多的比赛来保证竞赛（或矩阵 $\boldsymbol{A}$）满足第 36 页上所定义的不可约性和素性条件，或者如**第 39 页**上所述的那样，通过增加统计数据可忽略的虚拟比赛来对 $\boldsymbol{A}$ 进行扰动，从而迫使这些条件得以成立。

— 如果不能肯定（或者就是不想去检查）约束 Ⅰ 和 Ⅱ 是否成立，那么通过替换 $\boldsymbol{A} \leftarrow \boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}$ 可以使不可约性和素性被强制成立，其中，$\boldsymbol{e}$ 为全 1 列，$\varepsilon > 0$ 是一个小数，它相比于被扰动前的 $\boldsymbol{A}$ 中最小正元素值而言也为小。

— 如果不可约性已经得到满足——即对每两支队伍 i 和 j，存在一系列比赛，使得对某个 p 满足

$$i \leftrightarrow k_1 \leftrightarrow k_2 \leftrightarrow \cdots \leftrightarrow k_p \leftrightarrow j \quad （其中，a_{ik_1} > 0, a_{k_1k_2} > 0, \cdots, a_{k_pj} > 0）$$

（p 可能随 (i,j) 对的不同而不同）——则联赛（或矩阵 $\boldsymbol{A}$）可通过简单地在 $\boldsymbol{A}$ 的任一对角位置上加上任意 $\varepsilon > 0$ 而使得素性成立（即使得存在一个一致相同的 p，对所有的 (i,j) 对，都存在一系列比赛使 i 和 j 连通）。这一做法相当于进行了一场某支队伍与其自身之间的虚拟比赛，并且这场比赛的统计数据的影响可以忽略不计，该做法也等同于进行了替换 $\boldsymbol{A} \leftarrow \boldsymbol{A} + \varepsilon \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}$，其中，$\varepsilon > 0$，且对某个 i，有 $\boldsymbol{e}_i^{\mathrm{T}} = (0,0,\cdots,1,0,\cdots,0)$。

7. 利用**第 38 页**上所描述的幂法求取评分向量 $\boldsymbol{r}$。如果已经按第 6 步中所总结的方式迫使不可约性和素性得以成立，则可以如下对**第 38 页**上的幂法迭代稍作修改。

— 对于扰动后的矩阵 $\boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}$，如下执行幂法。

- 初始时置 $\boldsymbol{r} \leftarrow (1/m)\boldsymbol{e}$，其中 $\boldsymbol{e}$ 为全 1 列。
- 重复以下步骤，直至 $\boldsymbol{r}$ 中的元素收敛至某个预先给定的有效位数。

BEGIN

1) $\sigma \leftarrow \varepsilon \sum_{j=1}^{m} r_j \ (= \varepsilon \boldsymbol{e}^{\mathrm{T}}\boldsymbol{r})$

2) $\boldsymbol{r} \leftarrow \boldsymbol{A}\boldsymbol{r} + \sigma e \ (= [\boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]\boldsymbol{r})$

3) $v \leftarrow \sum_{j=1}^{m} r_j \ (= \boldsymbol{e}^{\mathrm{T}}[\boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]\boldsymbol{r})$

4) $\boldsymbol{r} \leftarrow \boldsymbol{r}/v \ (= [\boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]\boldsymbol{r}/\boldsymbol{e}^{\mathrm{T}}[\boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]\boldsymbol{r})$

REPEAT

— 如果使用的扰动后矩阵为 $\boldsymbol{A} + \varepsilon \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}$，则幂法稍有不同。

- 初始时置 $\boldsymbol{r} \leftarrow (1/m)\boldsymbol{e}$，其中，$\boldsymbol{e}$ 为全 1 列。
- 重复以下步骤，直至 $\boldsymbol{r}$ 中的元素收敛至某个预先给定的有效位数。

BEGIN

1) $\sigma \leftarrow \varepsilon r_i (= \varepsilon \boldsymbol{e}_i^{\mathrm{T}} \boldsymbol{r})$

2) $\boldsymbol{r} \leftarrow \boldsymbol{A}\boldsymbol{r} + \sigma \boldsymbol{e}_i (= [\boldsymbol{A} + \varepsilon \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}]\boldsymbol{r})$

3) $v \leftarrow \sum_{j=1}^{m} r_j (= \boldsymbol{e}^{\mathrm{T}}[\boldsymbol{A} + \varepsilon \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}]\boldsymbol{r})$

4) $\boldsymbol{r} \leftarrow \boldsymbol{r}/v (= [\boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}]r/\boldsymbol{e}^{\mathrm{T}}[\boldsymbol{A} + \varepsilon \boldsymbol{e}_i \boldsymbol{e}_i^{\mathrm{T}}]\boldsymbol{r})$

REPEAT

— 这些步骤都十分简单，在任何一个不错的电子表格软件中都可以手工完成。

NFL 2009—2010 赛季

为了说明本章的思路，我们使用 NFL 2009—2010 赛季 17 周的常规赛比分来构建基纳评分和排名。这些队伍在下表中按英文字母顺序列出。

序　数	队　名	序　数	队　名
1.	芝加哥熊	17.	纽约喷气机
2.	辛辛那提猛虎	18.	底特律雄狮
3.	布法罗比尔	19.	旧金山 49 人
4.	丹佛野马	20.	绿湾包装工
5.	克里夫兰布朗	21.	卡罗来纳黑豹
6.	坦帕湾海盗	22.	新英格兰爱国者
7.	亚利桑那红雀	23.	奥克兰突袭者
8.	圣迭戈闪电	24.	圣路易斯公羊
9.	堪萨斯城酋长	25.	巴尔的摩乌鸦
10.	印第安纳小马	26.	华盛顿红皮
11.	达拉斯牛仔	27.	新奥尔良圣徒
12.	迈阿密海豚	28.	西雅图海鹰
13.	费城老鹰	29.	匹兹堡钢人
14.	亚特兰大猎鹰	30.	休斯敦德州人
15.	纽约巨人	31.	田纳西泰坦
16.	杰克逊维尔美洲虎	32.	明尼苏达维京人

根据这个序号排列，我们设

S_{ij} = 队伍 i 对队伍 j 在常规赛季的比赛中所取得的累积得分

则原始得分的矩阵如下所示。例如，在下面的矩阵中 $S_{12} = 10$，表明芝加哥熊在常规赛对阵辛辛那提猛虎时得到了 10 分，而 $S_{21} = 45$ 则意味着辛辛那提猛虎从芝加哥熊身上得到了 45 分。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	0	10	0	0	30	0	21	0	0	0	0	0	20	14	0	0	0	85	6	29	0	0	0	17	7	0	0	25	17	0	0	46
2	45	0	0	7	39	0	0	24	17	0	0	0	0	0	0	0	0	23	0	31	0	0	17	0	34	0	0	0	41	17	0	10
3	0	0	0	0	3	33	0	0	16	30	0	41	0	3	0	15	29	0	0	0	20	34	0	0	0	0	7	0	0	10	17	0
4	0	12	0	0	27	0	0	37	68	16	17	0	27	0	26	0	0	0	0	0	0	20	42	0	7	17	0	0	10	0	0	0
5	6	27	6	6	0	0	0	23	41	0	0	0	0	0	0	23	0	37	0	3	0	0	23	0	3	0	0	0	27	0	0	20
6	0	0	20	0	0	0	0	0	0	0	21	23	14	27	0	0	3	0	0	38	27	7	0	0	0	13	27	24	0	0	0	0
7	41	0	0	0	0	0	0	0	0	10	0	0	0	0	24	31	0	31	25	7	21	0	0	52	0	0	0	58	0	28	17	30
8	0	27	0	55	30	0	0	0	80	0	20	23	31	0	21	0	0	0	0	0	0	0	48	0	26	23	0	0	28	0	42	0
9	0	10	10	57	34	0	0	21	0	0	20	0	14	0	16	21	0	0	0	0	0	0	26	0	24	14	0	0	27	0	0	0
10	0	0	7	28	0	0	31	0	0	0	0	27	0	0	0	49	15	0	18	0	0	35	0	42	17	0	0	34	0	55	58	0
11	0	0	0	10	0	34	0	17	26	0	0	0	44	37	55	0	0	0	0	7	21	0	24	0	0	24	24	38	0	0	0	0
12	0	0	52	0	0	25	0	13	0	23	0	0	0	7	0	14	61	0	0	0	24	39	0	0	0	0	34	0	24	20	24	0
13	24	0	0	30	0	33	0	23	34	0	16	0	0	34	85	0	0	0	27	0	38	0	9	0	0	54	22	0	0	0	0	0
14	21	0	31	0	0	40	0	0	0	0	21	19	7	0	31	0	10	0	45	0	47	10	0	0	0	31	50	0	0	0	0	0
15	0	0	0	6	0	24	17	20	27	0	64	0	55	34	0	0	0	0	0	0	9	0	44	0	0	68	27	0	0	0	0	7
16	0	0	18	0	17	0	17	0	24	43	0	10	0	0	0	0	24	0	3	0	0	7	0	23	0	0	0	0	0	54	50	0
17	0	37	32	0	0	26	0	0	0	29	0	52	0	7	0	22	0	0	0	0	17	30	38	0	0	0	10	0	0	24	24	0
18	47	13	0	0	38	0	24	0	0	0	0	0	0	0	0	0	0	0	6	12	0	0	0	10	3	19	27	20	20	0	0	23
19	10	0	0	0	0	0	44	0	0	14	0	0	13	10	0	20	0	20	0	24	0	0	0	63	0	0	0	40	0	21	27	24
20	42	24	0	0	31	28	33	0	0	0	17	0	0	0	0	0	0	60	30	0	0	0	0	36	27	0	0	48	36	0	0	49
21	0	0	9	0	0	44	34	0	0	0	7	17	10	48	41	0	6	0	0	0	0	10	0	0	0	20	43	0	0	0	0	26
22	0	0	42	17	0	35	0	0	0	34	0	48	0	26	0	35	40	0	0	0	20	0	0	0	27	0	17	0	0	27	59	0
23	0	20	0	23	9	0	0	36	23	0	7	0	13	0	7	0	0	0	0	0	0	0	0	0	13	13	0	0	27	6	0	0
24	0	0	0	0	0	0	23	0	0	6	0	0	0	0	0	20	0	17	6	17	0	0	0	0	0	7	23	17	0	13	7	10
25	31	21	0	30	50	0	0	31	38	15	0	0	0	0	0	0	0	48	0	14	0	21	21	0	0	0	0	0	40	0	0	31
26	0	0	0	27	0	16	0	20	6	0	6	0	41	17	29	0	0	14	0	0	17	0	34	9	0	0	30	0	0	0	0	0
27	0	0	27	0	0	55	0	0	0	0	17	46	48	61	48	0	24	45	0	0	40	38	0	28	0	33	0	0	0	0	0	0
28	19	0	0	0	0	7	23	0	0	17	17	0	0	0	0	41	0	32	30	10	0	0	0	55	0	0	0	0	0	7	13	9
29	14	32	0	28	33	0	0	38	24	0	0	30	0	0	0	0	0	28	0	37	0	0	24	0	40	0	0	0	0	0	13	27
30	0	28	31	0	0	0	21	0	0	44	0	27	0	0	0	42	7	0	24	0	0	34	29	16	0	0	0	34	0	0	51	0
31	0	0	41	0	0	0	20	17	0	26	0	27	0	0	0	47	17	0	34	0	0	0	0	47	0	0	0	17	10	51	0	0
32	66	30	0	0	34	0	17	0	0	0	0	0	0	0	44	0	0	54	27	68	7	0	0	38	33	0	0	35	17	0	0	0

NFL 2009—2010 赛季得分

由这些原始得分，我们利用**第 31 页**上的式（4.2）来构造出矩阵

$$\left[\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}\right]_{32\times 32},$$

然后将**第 31 页**上式（4.4）中的基纳偏好函数 $h(x)$ 应用于该矩阵的每个元素上，得到非负矩阵

$$\boldsymbol{A}_{32\times 32}=[a_{ij}]=\left[h\left(\frac{S_{ij}+1}{S_{ij}+S_{ji}+2}\right)\right]。$$

第 32 页上的规范化不必进行，因为所有的队伍都进行了相同场数的比赛（16 场比赛——每支队伍都有一次轮空）。此外，因为 $\boldsymbol{A}^2>\boldsymbol{0}$，所以 $\boldsymbol{A}$ 是素的（因而也是不可约的），因此也无需进行扰动。$\boldsymbol{A}$ 的佩隆值为 $\lambda\approx 15.832$，而评分向量 $\boldsymbol{r}$（精确到 5 个有效数位）如下。

队　伍	评　分	队　伍	评　分
芝加哥熊	0.029410	纽约喷气机	0.034683
辛辛那提猛虎	0.031483	底特律雄狮	0.025595
布法罗比尔	0.029066	旧金山 49 人	0.031876
丹佛野马	0.031789	绿湾包装工	0.035722
克里夫兰布朗	0.027923	卡罗来纳黑豹	0.030785
坦帕湾海盗	0.026194	新英格兰爱国者	0.035051
亚利桑那红雀	0.032346	奥克兰突袭者	0.026222
圣迭戈闪电	0.035026	圣路易斯公羊	0.024881
堪萨斯城酋长	0.028006	巴尔的摩乌鸦	0.033821
印第安纳小马	0.034817	华盛顿红皮	0.029107
达拉斯牛仔	0.034710	新奥尔良圣徒	0.036139
迈阿密海豚	0.029805	西雅图海鹰	0.027262
费城老鹰	0.033883	匹兹堡钢人	0.033529
亚特兰大猎鹰	0.032690	休斯敦德州人	0.033415
纽约巨人	0.030480	田纳西泰坦	0.030538
杰克逊维尔美洲虎	0.028962	明尼苏达维京人	0.034783

NFL 2009—2010 赛季的基纳评分　　(4.13)

按评分由高到低排序后，便得到了如下的2009—2010赛季NFL球队排名。

排名	队　伍	评　分	排名	队　伍	评　分
1.	新奥尔良圣徒	0.036139	17.	辛辛那提猛虎	0.031483
2.	绿湾包装工	0.035722	18.	卡罗来纳黑豹	0.030785
3.	新英格兰爱国者	0.035051	19.	田纳西泰坦	0.030538
4.	圣迭戈闪电	0.035026	20.	纽约巨人	0.030480
5.	印第安纳小马	0.034817	21.	迈阿密海豚	0.029805
6.	明尼苏达维京人	0.034783	22.	芝加哥熊	0.029410
7.	达拉斯牛仔	0.034710	23.	华盛顿红皮	0.029107
8.	纽约喷气机	0.034683	24.	布法罗比尔	0.029066
9.	费城老鹰	0.033883	25.	杰克逊维尔美洲虎	0.028962
10.	巴尔的摩乌鸦	0.033821	26.	堪萨斯城酋长	0.028006
11.	匹兹堡钢人	0.033529	27.	克里夫兰布朗	0.027923
12.	休斯敦德州人	0.033415	28.	西雅图海鹰	0.027262
13.	亚特兰大猎鹰	0.032690	29.	奥克兰突袭者	0.026222
14.	亚利桑那红雀	0.032346	30.	坦帕湾海盗	0.026194
15.	旧金山49人	0.031876	31.	底特律雄狮	0.025595
16.	丹佛野马	0.031789	32.	圣路易斯公羊	0.024881

(4.14)

NFL 2009—2010赛季的基纳排名

大多数熟悉NFL 2009—2010赛季的人可能都会认同一点，即该排名（以及相应的评分）看起来相当不错，因为它准确地反映了在季后赛中所发生的事情。圣徒队排名第1，而实际上，圣徒队赢得了超级碗！他们以31比17的比分击败了小马队。此外，如**第61页**上的图5.2所示，我们给出的排名前10的队伍在2009—2010赛季都进入了季后赛，仅此一点就增加了基纳法的可信度。

在我们的评分中，小马队名列第5，但是如果他们不是自动缴械，有意放弃了赛季的最后两场比赛以避免首发队员受伤的话，他们可能会得到更高的评分。一个可能有趣且有启发性的练习，就是将某些（或全部）队伍的最后两场常规赛的得分去除或进行一定的加权，之后再将所得的评分和排名与以上结果进行比较——我们并没有去做这件事，因此如果你自己决定按此进行的话，请将最后的结果告诉我们。

另一方面，在回顾了第44届超级碗的重播后，我们可以说小马队在对阵圣徒队时看起来确实像第5名，而未受伤病困扰的爱国者队或闪电队都很可能对圣徒队发起更有力的挑战。一个有趣的课题将是对比分进行加权，使得赛季初或某些赛季末的得分的重要性低于赛季中期的关键时期的得分——这可以通过对每一对队伍的得分进行处理来完成。

吉姆·基纳对比尔·詹姆斯

韦恩·温斯顿（Wayne Winston）的那本有意思的书[83]，是以一段讨论开头的，其中所说的是一个常常被称为棒球之毕达哥拉斯定理的东西，该定理由著名的作家、历史学家和统计学家比尔·詹姆斯（Bill James）正式阐述，他曾花了相当长的时间对棒球数据进行分析。詹姆斯发现，一支棒球队在一个赛季中所达到的胜率，可以通过一个毕达哥拉斯式的期望公式很好地加以近似，该公式指出

$$\%\ \text{胜率} \approx \frac{\text{得分跑垒数}^2}{\text{得分跑垒数}^2 + \text{跑垒数}^2} = \frac{1}{1+\rho^2} \quad \left(\text{其中,}\rho = \frac{\text{跑垒数}}{\text{得分跑垒数}}\right)$$

在定量体育分析中，这是一个使用广泛的思路——它已经被用于这个星球上差不多所有的主要体育项目之中。但是，不同项目的公式中的指数会有所不同。换而言之，一旦选定了感兴趣的体育项目后，就必须将公式中的 ρ^2 调整为 ρ^x，其中的 x 对于该项目而言是最优的。为了达到这一目的，温斯顿建议使用绝对平均偏差（或称为 MAD），即若有

$$\omega_i = \text{队伍 } i \text{ 在一个赛季中获胜的场数} \tag{4.15}$$

且

$$\rho_i = \frac{\text{队伍 } i \text{ 的失分}}{\text{队伍 } i \text{ 的得分}} \tag{4.16}$$

而且如果在联赛中有 m 支队伍，则对给定的 x 值的绝对平均偏差为

$$\text{MAD}(x) = \frac{1}{m}\sum_{i=1}^{m}\left|\omega_i - \frac{1}{1+\rho_i^x}\right| \tag{4.17}$$

或等价地用向量的 1-范数[54, 第274页]来表示，

$$\text{MAD}(x)\ \frac{\|\boldsymbol{w}-\boldsymbol{p}(x)\|_1}{m}, \text{其中}\begin{cases}\boldsymbol{w} = (w_1, w_2, \cdots, w_m)^{\mathrm{T}} \\ \boldsymbol{p}(x) = ((1+\rho_1^x)^{-1}, (1+\rho_2^x)^{-1}, \cdots, (1+\rho_m^x)^{-1})^{\mathrm{T}}\end{cases}$$

一旦选定了一个项目，那么就需要找到使 MAD(x) 对该项目为最小的值 x^*。如果真想进行修改的话，则式（4.16）中的“得分”可以用所考虑项目中的其他内容加以替代——例如，一个有趣的练习，便是在橄榄球比赛中使用

$$\rho_i = \frac{\text{队伍 } i \text{ 让对手获得的跑码数}}{\text{队伍 } i \text{ 获得的跑码数}}$$

并看看将会发生些什么。

让我们在 NFL 2009—2010 常规赛季上试一下詹姆斯的毕达哥拉斯定理的点子吧。换而言之，我们估计

$$\%\ \text{常规赛胜率} \approx \frac{1}{1+\rho^x} \quad \left(\text{其中,}\rho = \frac{\text{失分}}{\text{得分}}\right)$$

式中，x 由**第 43 页**上给出的 NFL 2009—2010 赛季得分数据所确定。然后让我们将这些结果与式（4.13）和式（4.14）中基纳评分所估计的胜率加以比较。在 2009—2010 常规赛季中每支 NFL 球队的胜率如下表所示⊖，即我们所希望获得的结果。

⊖ 为方便起见，百分数 * % 都通过将数值乘以 100 进行了转化。

队　伍	胜率（%）	队　伍	胜率（%）
芝加哥熊	43.75	纽约喷气机	56.25
辛辛那提猛虎	62.50	底特律雄狮	12.50
布法罗比尔	37.50	旧金山49人	50.00
丹佛野马	50.00	绿湾包装工	68.75
克里夫兰布朗	31.25	卡罗来纳黑豹	50.00
坦帕湾海盗	18.75	新英格兰爱国者	62.50
亚利桑那红雀	62.50	奥克兰突袭者	31.25
圣迭戈闪电	81.25	圣路易斯公羊	06.25
堪萨斯城酋长	25.00	巴尔的摩乌鸦	56.25
印第安纳小马	87.50	华盛顿红皮	25.00
达拉斯牛仔	68.75	新奥尔良圣徒	81.25
迈阿密海豚	43.75	西雅图海鹰	31.25
费城老鹰	68.75	匹兹堡钢人	56.25
亚特兰大猎鹰	56.25	休斯敦德州人	56.25
纽约巨人	50.00	田纳西泰坦	50.00
杰克逊维尔美洲虎	43.75	明尼苏达维京人	75.00

NFL 2009—2010 赛季的常规赛胜率　　(4.18)

式（4.17）中指数 x 的最优值 x^* 是通过直接计算求得的，即我们对足够多的不同的 x 值来计算 MAD(x)，并通过目测来确定最优值 x^*。使用**第 43 页**上 NFL 2009—2010 赛季的得分数据，我们对 1 到 4 之间的 1500 个等间隔的 x 值求取了 MAD(x)，并估计

$$x^* = 2.27, \text{且 MAD}(x^*) = 0.0621 \tag{4.19}$$

MAD(x) 如图 4.2 所示。这意味着由詹姆斯的广义毕达哥拉斯公式 $1/(1+\rho)^{x^*}$ 所给出的预测与 NFL 2009—2010 赛季各队的实际胜率相比，平均仅偏离了 6.21%。对于这样一个简单的公式来说，这可不算差了！

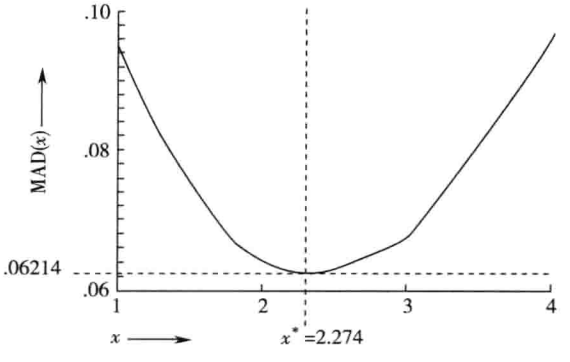


图 4.2　NFL 的最优毕达哥拉斯指数

温斯顿在参考文献［83］中称，休斯敦火箭队的总经理达里尔·莫雷（Daryl Morey）对更早些的 NFL 赛季（不过不清楚是哪些赛季）也进行了类似的计算，莫雷所得到的毕达哥拉斯指数为2.37。直观来看，最优值x^*应该随时间而变化。不过，莫雷和温斯顿的

计算以及我们给出的值表明，NFL 的 x^* 在各个赛季间的变动可能相对较小。此外，温斯顿对 NFL 2006—2007 赛季给出了 MAD(2.37) = 0.061，这与我们得到的 2009—2010 赛季的 MAD(2.27) = 0.062 值相当接近。

现在让我们来看一下，在**第 43 页**的式（4.13）中的 $\boldsymbol{r}$ 是否能给出同样好的结果。并没有任何东西说明，基纳评分是一个能够适用于某类非线性估计公式的毕达哥拉斯对象。但是，式（4.13）中的评分 $\boldsymbol{r}$ 和式（4.18）中的胜率 $\boldsymbol{w}$ 之间的相关系数为

$$R_{rw} = \frac{(\boldsymbol{r}-\mu_r\boldsymbol{e})^{\mathrm{T}}(\boldsymbol{w}-\mu_w\boldsymbol{e})}{\|\boldsymbol{r}-\mu_r\boldsymbol{e}\|_2\,\|\boldsymbol{w}-\mu_w\boldsymbol{e}\|_2} \approx 0.934, \text{其中}, \boldsymbol{e} = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}, \mu_* = \text{均值} \qquad (4.20)$$

这意味着 $\boldsymbol{r}$ 和 $\boldsymbol{w}$ 的数据之间存在着强线性关系[54, 第296页]，因此与其直接将评分 r_i 纳入毕达哥拉斯框架，不如寻找形如 $\alpha+\beta r_i \approx w_i$（或等价地，$\alpha\boldsymbol{e}+\beta\boldsymbol{r}\approx\boldsymbol{w}$）的最优线性估计。高斯-马尔可夫定理[54, 第448页]指出，a 和 β 的“最佳”值将使 $\|\boldsymbol{Ax}-\boldsymbol{w}\|_2^2$ 最小化，其中 $\boldsymbol{A} = \begin{pmatrix}1 & r_1\\1 & r_2\\\vdots & \vdots\\1 & r_m\end{pmatrix}$，$\boldsymbol{x} = \begin{pmatrix}\alpha\\\beta\end{pmatrix}$，$\boldsymbol{w} = \begin{pmatrix}w_1\\w_2\\\vdots\\w_m\end{pmatrix}$。寻找 a 和 β 的工作最终归结为求解正规方程 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$[54, 第226页]，在我们的问题中，这是一个包含两个未知数的且由两个方程构成的方程组。由规范方程求解 x 的计算直截了当，大多数较好的计算器和电子表格软件都提供了这一功能。对于 NFL 2009—2010 赛季的数据，取 5 位有效数字的解为 $\alpha = -1.2983$ 和 $\beta = 57.545$，因此根据基纳（佩隆-弗罗贝尼乌斯）评分所得的常规赛胜率的最小二乘估计（圆整至 5 个有效数字）为

$$\text{队伍 } i \text{ 的估计胜率} = \alpha+\beta r_i = -1.2983 + 57.545 r_i。$$

下表给出了 NFL 球队 2009—2010 常规赛的实际胜率以及毕达哥拉斯和基纳估计（圆整至 3 个有效数字）。

队伍	胜率（%）	毕达哥拉斯估计（%）	基纳估计（%）	队伍	胜率（%）	毕达哥拉斯估计（%）	基纳估计（%）
芝加哥熊	43.8	42.2	39.4	纽约喷气机	56.3	70.9	69.8
辛辛那提猛虎	62.5	52.7	51.3	底特律雄狮	12.5	18.9	17.5
布法罗比尔	37.4	36.9	37.5	旧金山 49 人	50.0	59.1	53.6
丹佛野马	50.0	50.4	53.1	绿湾包装工	68.8	73.3	75.7
克里夫兰布朗	31.3	27.4	30.9	卡罗来纳黑豹	50.0	51.3	47.3
坦帕湾海盗	18.8	24.4	20.9	新英格兰爱国者	62.5	71.6	71.9
亚利桑那红雀	62.5	58.1	56.3	奥克兰突袭者	31.3	18.2	21.1
圣迭戈闪电	81.3	69.0	71.8	圣路易斯公羊	6.3	11.0	13.4
堪萨斯城酋长	25.0	30.2	31.3	巴尔的摩乌鸦	56.3	71.6	64.8
印第安纳小马	87.5	66.7	70.5	华盛顿红皮	25.0	36.9	37.7
达拉斯牛仔	68.8	69.9	69.9	新奥尔良圣徒	81.3	71.6	78.1
迈阿密海豚	43.8	45.4	41.7	西雅图海鹰	31.3	31.9	27.1
费城老鹰	68.8	63.5	65.2	匹兹堡钢人	56.3	57.2	63.1
亚特兰大猎鹰	56.3	56.3	58.3	休斯敦德州人	56.3	58.7	62.5
纽约巨人	50.0	46.5	45.6	田纳西泰坦	50.0	42.8	45.9
杰克逊维尔美洲虎	43.8	35.0	36.8	明尼苏达维京人	75.0	71.9	70.3

现在便是妙处了。基纳估计的 MAD（取 3 位有效数字）为

$$\mathrm{MAD}_r = \frac{1}{32}\sum_{i=1}^{32} |(-1.2981 + 57.547r_i) - w_i| = 0.0591。$$

换而言之，利用基纳评分对 NFL 2009—2010 赛季的胜率进行预测，得到的结果与各队实际胜率之间的平均偏差仅为 5.91%，而最优毕达哥拉斯估计的各队平均偏差为 6.28%。我们使用 MAD 是由于温斯顿在参考文献［83］中用的就是这个指标，不过误差平方和的均值即均方误差（MSE）也是另一个常用的指标。对于我们的例子，有

$$\mathrm{MSE}_{\text{毕达哥拉斯}} = 0.0065 \text{ 和} \mathrm{MSE}_{\text{基纳}} = 0.0050$$

- 结语。基纳（实际上是佩隆-弗罗贝尼乌斯）赢得了 2009—2010 的点球大战！

佩隆-弗罗贝尼乌斯（又名詹姆斯·基纳）应当至少和毕达哥拉斯（又名比尔·詹姆斯）一样好，这一点是合理的，因为后者没有考虑队伍得分或失分的困难程度（或容易程度），而前者却考虑到了这一点。基纳的领先优势是否足够明显呢？这就由你来定了。如果上面的例子激起了你的兴趣，那就请自行开展更多的实验吧，并请让我们知道你的实验结果。

回到未来

假设你能搭上埃米特·布朗（Emmett Brown）博士的钚动力 DeLorean 汽车和马蒂·麦克弗莱（Marty McFly）兜一次风，回到 NFL 2009—2010 赛季之初，而且，正如比夫·坦能（Biff Tannen）带着 2000 期《格雷斯体育年鉴》（Grays Sports Almanac）回到 1955 年去发财一样，在跳上车并坐在通量电容器边之前，你可以将**第 43 页**上基纳的评分，即式（4.13）简要地记在手掌上，那么利用这个评分，你能够正确地选出多少比赛的胜者呢？当然，真正的挑战在于如何预测将来而非过去，不过，选取一个评分方案并与其他方案加以对比，这样的回顾仍然是很有价值的。

后见之明对先见之明

贯穿全书，我们都把利用当前评分来预测以往比赛胜者的做法称为事后预测，而称利用当前评分来预测今后比赛胜者的做法为事前预测。自然地，我们可以预期事后预测的准确度要高于事前预测，但与广为流传的陈词滥调相反，事后回顾也并非就像 1.5 的视力那样看得清楚。对每支队伍而言，没有一个单独的评分值可以完美地反映出整个联赛的所有以往的胜负信息。

既然大多数评分都未考虑主场优势，要进行事后预测或事前预测，通常意味着你必须通过某种方式加入一个“主场因素”。如果能带着**第 44 页**上的基纳评分，即式（4.14）回到 NFL 2009—2010 赛季初来选择 NFL 2009—2010 赛季所有 267 场比赛的胜者的话，就必须在基纳评分中为主队增加一个 0.0008 的主场优势因素，以体现主队对客队的得分差约为 2.4 分的这一

事实[⊖]。这样一来，我们将可以按 73.4% 的准确率正确选出 267 场比赛中 196 场的胜者。这就是事后预测的准确率。

一件有意思的事情，就是对比基纳的事后预测和根据赛季末常可获取的其他评分所得的结果。在因特网上并不缺少这样的评分网站，但大多数都没有公开确定评分的细节。如果你对众多评分者在 NFL 2009—2010 赛季的表现感到好奇，请见托德·贝克（Todd Beck）的预测结果跟踪网站 www. thepredictiontracker. com/nflresults. php，它对众多评分者进行了比较。

基纳能让你致富吗

基纳评分能相当不错地估计给定球队所能达到的胜率，但如果你是个赌徒，那这可能并非你所关注的主要问题。赌徒要赚钱的话，需要比书呆子更为狡猾，归根到底，前者要在预测分差时击败后者才行（**第 113 页**上第 9 章的内容中包含了有关分差以及最优“分差评分”的完整讨论）。正在阅读此段的“玩家”们无疑都已经问过了下面这个问题：“我应该如何利用基纳评分来预测给定交手中的分差?”遗憾的是，如果你试图使用基纳法来估计分差的话，你很可能会输钱。对于入门者而言，队与队之间的分差和相应的基纳评分差之间的相关性很差。实际上，如果考察 NFL 2009—2010 赛季的 267 场比赛的话，由式（4.20）计算得到的比赛分差和对应的基纳评分差之间的相关系数仅约为 0.579，也就是说，两者之间并不存在多少线性关系。

实际上，我们甚至可以说，在橄榄球项目中，通过任何一个单一的评分值列表来产生准确的分差估计，几乎是不可能的。一些一般性的原因将在**第 113 页**上的第 9 章部分详加说明，但如果我们考虑基纳评分的话，我们可以来问一问，为什么这一特定的评分值不能给出准确的分差估计。要理解这个问题的答案，就需要了解那些隐藏在数学演算纸下面的东西。总的来说，基纳法是马尔可夫链的一个变体，这条链是通过在 32 支 NFL 球队进行随机游走所定义得到的，而正因为如此，它的评分仅给出了在长时间内的期望胜率信息。**第 67 页**上的第 6 章对马尔可夫方法进行了详细的讨论。

为了能够直观理解这一点，让我们假设有一个狂热的球迷，他每周都从一个拥有 NFL 球队的城市跑到另一个，而下一周去往哪个城市则是由统计数据矩阵 $\boldsymbol{A} = [a_{ij}]$ 中的元素所决定的。如果将 $\boldsymbol{A}$ 中的各列进行归一化，使得每一列的元素和为 1，则正如在**第 31 页**上所说的那样，a_{ij} 可以被视为队伍 i 击败队伍 j 的概率。但除了用这种方法来加以解释之外，还可以认为 a_{ij} 是

⊖ 这一点并非严格正确，因为尽管总是可以指定一个“主队”，但有三场比赛——具体来说就是 2009 年 10 月 25 日的海盗对爱国者、2009 年 12 月 3 日的比尔对喷气机和 2010 年 2 月 7 日的小马对圣徒——是在中立球场进行的，不过这点出入并未大到足以造成显著的不同。

$$a_{ij} = \text{该球迷现在位于城市 } j \text{ 时，下一周去往城市 } i \text{ 的概率}$$

如果 $a_{ij} > a_{kj}$，那么就队伍 j 而言，队伍 i 比 k 要更强。但是对我们的旅行球迷来说，这就意味着他落在城市 j 以后，便注定要将更多些的时间花在城市 i 而非城市 k 上。将这一逻辑应用于所有城市 j，便意味着长期而言他将把更多时间花在获胜的球队所在的城市中——胜率越高，花的时间越多。换而言之，如果能够跟踪我们这位球迷的移动来确定出他待在每个城市的精确的时间比例，那我们实际上就衡量出了每支队伍在长期内期望能够积累起来的获胜比率。

旅行球迷-基纳之间的关联

根据随机游走的球迷在每个城市中所花费的时间比例所得的“球迷评分”$\boldsymbol{r}_F$，定性而言与基纳评分 $\boldsymbol{r}_K$ 是等同的，即它们都反映了同一个事物——具体而言就是长时期内的胜率。

直观的理由如下。由**第 38 页**上的幂法，即式（4.12）所产生的序列 $\boldsymbol{x}_0$，$\boldsymbol{x}_1$，$\boldsymbol{x}_2$，… 将收敛到基纳评分 $\boldsymbol{r}_K$。在开始时将基纳矩阵归一化，使得每列元素之和为 1，这将使得每次迭代中的归一化（即式（4.12）除以 v_k）操作变得不必要，而幂法也简化为

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\,\boldsymbol{x}_k \quad \text{其中，} k = 0,1,2,\cdots, \boldsymbol{x}_0 = \begin{pmatrix} 1/m \\ 1/m \\ \vdots \\ 1/m \end{pmatrix} \tag{4.21}$$

迭代结果 $\boldsymbol{x}_k$ 给出了我们这位球迷所经路径的一个估计。$\boldsymbol{x}_0$ 使他以等概率由任一给定城市开始旅程。马尔可夫链理论[54，第687页]确保了 $\boldsymbol{x}_k$ 中的元素 i 就是该球迷在第 k 周时处于城市 i 的概率。如果 $\boldsymbol{x}_k \to \boldsymbol{r}_F$，则 $[\boldsymbol{r}_F]_i$ 就是该球迷在无限长的时间内花费在城市 i 上的时间的期望比例。不同的归一化策略可以在 $\boldsymbol{r}_K$ 和 $\boldsymbol{r}_F$ 之间造成小的差别[⊖]，但定性结果仍然相同。因此，基纳评分反映了该球迷在每个城市中花费的期望时间，这也就意味着基纳评分仅仅反映了长期胜率。

结论

基纳评分本质上是一个极限过程的结果，它反映了长期胜率，但没有给出对象短期情况的任何信息，对于预测分差而言，它几乎毫无用处。因此，结论是“否”——基纳的点子不能帮你赌中分差从而发家致富。但这个点子却是驱使两位年轻人跻身世界上最富有者的关键——如果这使你感到好奇，请阅读以下的杂谈。

⊖ 对于 NFL 数据而言，这一直观解释是成立的，因为基纳的 $\boldsymbol{A}$ 矩阵各列元素总和之间的差别较小，因此不同的归一化策略只会带来小的扰动。如果 $\boldsymbol{A}$ 的各列元素总和之间的差别更大的话，$\boldsymbol{r}_K$ 和 $\boldsymbol{r}_F$ 之间的差别也将变得更大。

杂谈：河中落叶

对我们这些应用数学家而言，有一个很有吸引力的比喻，即将数学思想看做知识之树上的树叶。数学家的工作便是摇晃这棵树或者击打这些叶子，来让它们落入时间之河。随着它们在岁月的漩涡中漂来漂去，许多人都会在它们优雅地漂过其身旁时诧异于它们的美丽，而其他人则对此毫无察觉。某些人将会拾起它们，并将它们和由其他科学分枝上落下的叶子与细枝编织在一起，并将如此所得的更大的作品扔回河中，以便为下游的人们带来裨益和欢娱。


O. Perron


F. G. Frobenius

1907 年到 1912 年间，奥斯卡·佩隆（Oscar Perron，1880—1975）和费迪南·基奥里·弗罗贝尼乌斯（Ferdinand Georg Frobenius，1849—1917）就在摇着这棵树。他们将非负矩阵这根枝条上的叶片摇晃松动。他们的想法不过是让这些叶片中的某一些落入河中，以使他人能够欣赏它们的绚丽，除此之外别无他求。排名和评分根本就不存在于他们的脑海之中。事实上，佩隆和弗罗贝尼乌斯根本想不到他们的树叶会被编入多大一摞东西里面！但是就名气而言，他们的非负矩阵理论仅在一个相当专业化的圈子里才为人所知。


J. Keener

詹姆斯·基纳在 1993 年从河中拾起这篇璀璨的叶片，并据此构建他的评分和排名体系，这可能并非出于偶然。基纳是一名数学家，专攻数学在生物学中的应用，而佩隆-弗罗贝尼乌斯理论为这一领域中的研究者所熟知，因为它自然而然地出现于进化系统分析以及生物动态学的其他方面。但是，基纳在参考文献［42］中指出，利用佩隆-弗罗贝尼乌斯的思想来进行评分和排名，这一做法并非是他的原创。更早些时候，T. H. 魏（T. H. Wei）[81] 在 1952 年、莫里斯-G. 肯德尔（Maurice G. Kendall）[44] 在 1955 以及 T. L. 萨蒂（T. L. Saaty）[67] 在 1987 年就把这些树叶放入了河中漂浮着。基纳对评分和排名的兴趣看起来不过就是他在严肃的科学研究之外的一点娱乐消遣。当他瞥到了这片树叶那多彩的闪光时，便从河中拾起了它，赞叹了那么一会儿，旋即又将它放回了河中。


Larry Page　Sergey Brin

1996—1998 年，这片叶子被卷入了流向**谷歌**的一个漩涡之中。斯坦福大学的博士生拉里·佩奇（Larry Page）和谢尔盖·布林（Sergey Brin）出于对万维网共同的兴趣建立了合作关系。他们的目标是改进当时搜索引擎的技术，其核心是对网页重要性进行评分和排名的 **PageRank** 算法。这些叶片被再次从河中捞起，并按类似于基纳的方式被使用。尽管并不清楚布林和佩奇在时间之河的回流中看到的到底是佩隆-弗罗贝尼乌斯那些鲜艳叶片中的哪一片，但从 PageRank 的专利文档和早期的内部技术报告中可以明显看出，他们的思路和基纳以及某些更早的研究者的思路是一脉相承的。有关 PageRank 在数学上的方方面面以及它与佩隆-弗罗贝尼乌斯理论之间的关系可以在参考文献［49］中找到。布林和佩奇利用了佩隆向量所具有的评分和排名能力，将其与网络爬虫软件编织在一起，并将成果转化为了如今的谷歌这一特大企业。靠着从河中捞起的几片艳丽的叶子，他们打造出了一艘巨舰，并将长期停留在更深的主河道上。

这些故事强调了有关数学叶片的普适真理。当你从树上摇落一片叶子，让它落入河中的时候，如果它在下游不远处为人所注意，那你可能会获得一定的声望，但却不会因此而变得无比富有。如果你的叶子顺流而下漂得很远，但却未能吸引他人的注意，那它的颜色将会褪去，并最终为水所浸透，并在没有吸引到多少关注时就消失在水面之下。但是，肯定有别的某些人最终也将从同一根枝干上摇落类似的叶子，并能在河流的不同地方赢得瞩目。最终这些叶子将会被收集起来，并织就为具有极大价值的物品。

上述的内容都是为了强调一点，即这条河流中堆满了树叶，其中许多的颜色都差不多，甚至就是来自于同一根枝干，但名声和财富却常常避开了那些将它们摇晃松动的人们。佩隆-弗罗贝尼乌斯、基纳及其先驱以及谷歌小伙儿的故事说明，认可与财富更多地被给予了那些将想法应用于一个有用的目的的人，而不是那些仅仅提出某个想法的人。

历史方面的备注：正如在之前的段落中所暗示的那样，利用特征值和特征向量来进行评分和排名，这一方法已经被实践了很长时间——至少有 60 年——而这一领域有着悠久的历史。有些人称这些方法为“谱评分”，而塞巴斯提亚诺·维格纳（Sebastiano Vigna）最近撰写了一本有关这一领域的不错的历史介绍。对更多有关谱评分的历史细节感兴趣的读者请见维格纳的论文[79]。

数字——

$ 8 580 000 000 = 谷歌 2011 年第 1 季度的收入。

——现在你对评分和排名的兴趣是不是提高了那么一点点？

——searchenginewatch. com

第 5 章　埃洛体系


Arpad Elo

阿帕德·埃洛（Árpád Élö，1903—1992）是出生于匈牙利的威斯康星密尔沃基马凯特大学物理学教授。此外，他还是一名热忱的（而且是优秀的）国际象棋棋手。正是这一点驱使他创造出一个有效的方法，来对棋手进行评分和排名。他的评分体系于 1960 年被美国国际象棋联合会（USCF）认可，进而在 1970 年得到了世界国际象棋联合会（FIDE）的承认。埃洛的思想最终在国际象棋这个领域之外也流行开来，并被加以改进、扩展和修改，来对其他的体育项目和竞赛进行评分。埃洛所用的前提条件，是每名棋手的表现都是一个服从正态分布的随机变量 X，其均值 μ 只能缓慢地随时间而改变。也就是说，尽管在下一场比赛时，一名棋手可能表现得更好些或更差些，但 μ 实际上在短期内是恒定的，而 μ 的改变需要很长的时间才能发生。

因此，埃洛推断，一旦确定了一名棋手的评分，那唯一能够改变这一评分的因素，便是该名棋手的表现优于或差于其平均表现的程度[⊖]。他提出了一个简单的线性修正方法，该修正与棋手偏离均值的程度成比例。更具体点来说，如果一名棋手最近的表现（或得分[⊜]）为 S，则他的原有评分 $r_{(\text{old})}$ 将通过下式而被更新为新的评分。

$$r_{(\text{new})} = r_{(\text{old})} + K(S - \mu) \tag{5.1}$$

式中，K 为一个常数——埃洛开始时设 $K = 10$。随着可用的国际象棋统计数据变得越来越多，人们发现棋手的表现通常并非是正态分布的，因此 USCF 和 FIDE 都将埃洛的原假设用一条新的假设来代替，这条新假设要求两名棋手得分差的期望值为这两名棋手评分差值的逻辑斯谛函数（logistic function，在**第 54 页**上对此进行了详细说明）。这一改动同时影响了式（5.1）中的 μ 和 K，但这个评分仍然被称为“埃洛评分”。

1997 年，鲍勃·朗尼安（Bob Runyan）修改了埃洛体系来对足球比赛进行评分，同时，自 1985 年以来就一直为《今日美国》提供体育评分的杰夫·萨加林，也开始为橄榄球比赛来修改埃洛体系。

⊖ 棋手的表现是一个服从正态分布的随机变量的假设，以及根据对均值的偏离来定义评分的思想，在关于这个领域的文献中都能找到。阿什伯恩（Ashburn）和柯尔维特（Colvert）的论文[6]以及其中的参考文献为那些有意了解统计知识的读者提供了这方面的信息。

⊜ 在国际象棋中，一场胜利可得 1 分，而一场平局得 1/2 分。可以使用一场比赛的得分或一次巡回赛中的累积得分来衡量棋手的表现。

在埃洛体系的现有形式下，其工作机制如下。我们必须从每名竞争者的某个初始评分集（见第56页）开始——为了能够将国际象棋之外的其他项目也考虑在内，我们将把竞争者设想为球队。当每次球队 i 和球队 j 相互交手时，其各自的评分 $r_{i(\text{old})}$ 和 $r_{j(\text{old})}$ 就按照类似于式（5.1）的公式被更新为 $r_{i(\text{new})}$ 和 $r_{j(\text{new})}$。不过，现在所有的因素都将在相互比较的基础上来考虑，即式（5.1）中的 S 变为

$$S_{ij} = \begin{cases} 1, \text{若 } i \text{ 击败 } j, \\ 0, \text{若 } i \text{ 被 } j \text{ 击败}, \\ 1/2, \text{若 } i \text{ 与 } j \text{ 战平} \end{cases} \tag{5.2}$$

而 μ 变为

$$\mu_{ij} = \text{队伍 } i \text{ 在对阵队伍 } j \text{ 时所得分数的期望值}$$

新假设指出，在队伍 i 和 j 比赛之前，μ_{ij} 是评分差

$$d_{ij} = r_{i(\text{old})} - r_{j(\text{old})}$$

的逻辑斯谛函数。标准的逻辑斯谛函数定义为 $f(x) = 1/(1 + e^{-x})$，但在国际象棋评分中所用的则是以10为底的版本

$$L(x) = \frac{1}{1 + 10^{-x}}。$$

函数 $f(x)$ 和 $L(x)$ 本质上是等同的，因为 $10^{-x} = e^{-x(\ln 10)}$，它们的曲线都具有如下S形的特征形状。

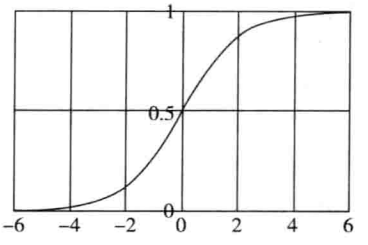


图 5.1　逻辑斯谛函数的曲线图

埃洛国际象棋评分中 μ_{ij} 的精确定义为

$$\mu_{ij} = L(d_{ij}/400) = \frac{1}{1 + 10^{-d_{ij}/400}} \quad (\text{其中}, d_{ij} = r_{i(\text{old})} - r_{j(\text{old})}) \tag{5.3}$$

因此更新球队（或棋手）i 和球队 j 各自评分的公式分别如下。

埃洛评分公式

$$r_{i(\text{new})} = r_{i(\text{old})} + K(S_{ij} - \mu_{ij}) \quad \text{且} \quad r_{j(\text{new})} = r_{j(\text{old})} + K(S_{ji} - \mu_{ji}),$$

式中，S_{ij} 由式（5.2）给出；μ_{ij} 由式（5.3）给出。

但是，在使用这些公式来构建你自己的评分系统之前，还需要对 K 以及式（5.3）中的400这个值有更多些的了解。

优美的智慧

埃洛公式那简单的优雅掩盖了其中蕴含的智慧，即相比于较强的棋手击败较弱的对手而言，埃洛隐含地为弱者击败强手给出了更高的奖励。例如，假设一名平均水平的棋手（记为 *avg*）的评分为1500分，而一名强手（记为 *str*）的评分为1900分，则

$$\mu_{avg,str} = \frac{1}{1 + 10^{-(1500-1900)/400}} = \frac{1}{11} \approx 0.09,$$

而

$$\mu_{str,avg} = \frac{1}{1 + 10^{-(1900-1500)/400}} = \frac{1}{1.1} \approx 0.91$$

因此，平均水平的棋手因击败强手而得到的奖励为

$$r_{avg(\text{new})} - r_{avg(\text{old})} = K(S_{avg,str} - \mu_{avg,str}) = K(1 - 0.09) = 0.91K$$

而强手击败平均水平的棋手所得的奖励仅为

$$r_{str(\text{new})} - r_{str(\text{old})} = K(S_{str,avg} - \mu_{str,avg}) = K(1 - 0.91) = 0.09K$$

K 因子

国际象棋界所称的“K 因子”目前仍然是一个有争议性的话题，不同的国际象棋组织使用了不同的值。这个因子的目的在于适当地根据实际与期望得分之间的差值来平衡以往的评分。

如果 K 过大。则实际得分与期望得分之间的差值便被赋予了过大的权重，这将导致评分的变动过大——例如，一个大的 K 值意味着，哪怕表现仅比期望值好那么一点点，却也能造成评分中的很大改变；另一方面，如果 K 过小，则埃洛的公式将丧失处理更好或更差表现的能力，而评分也将变得过于停滞——例如，一个小的 K 值意味着，即使一名棋手的表现有了明显改进，却也不能对其评分产生多少影响。

国际象棋： 国际象棋的 K 因子一般可以根据比赛的水平高低而改变。例如，FIDE 进行了如下设定：

$K = 25$　对未完成30场受认可的比赛的新棋手

$K = 15$　对进行了 >30 场比赛但评分从未超过2400分的棋手

$K = 10$　对曾经达到过2400分的棋手

在写作本书时，FIDE 正在考虑将以上 K 因子分别改为30，30 和20。

足球： 某些足球的评分体系允许根据赛事的重要性来提高 K 值。例如，在因特网上经常可以看到如下的 K 值：

$K = 60$　对于世界杯决赛圈的比赛
$K = 50$　对于各洲的杯赛和洲际锦标赛
$K = 40$　对于世界杯预选赛和主要的锦标赛
$K = 30$　对于所有其他的锦标赛
$K = 20$　对于友谊赛

K 因子是使得构建评分系统变得有趣的因素之一，因为它使得每名评分者都能自由地定制他或她的系统，以适应进行评分的特定赛事，并加入其自身的特色。

逻辑斯谛参数 ξ

逻辑斯谛函数即式（5.3）中的参数 $\xi = 400$ 是对国际象棋这个项目而设的，它影响了评分之间的差别。我们通过比较两位评分分别为 r_i 和 r_j 的棋手，来看看这个参数有什么影响。进行少许代数计算就可以发现

$$\mu_{ij} = \frac{10^{r_i/400}}{10^{r_i/400} + 10^{r_j/400}} \Rightarrow \frac{\mu_{ij}}{\mu_{ji}} = \frac{10^{r_i/400}}{10^{r_j/400}} \Rightarrow \mu_{ij} = \mu_{ji}\left[10^{(r_i - r_j)/400}\right]。$$

由于 μ_{ij} 是棋手 i 对阵棋手 j 时所期望获得的分数，这就意味着如果 $r_i - r_j = 400$，则我们期望棋手 i 比棋手 j 要强 10 倍。一般而言，棋手 i 对棋手 j 拥有 400 点的评分优势，就意味着棋手 i 击败棋手 j 的概率预期将是棋手 j 击败棋手 i 的概率的 10 倍。

如果我们跳出国际象棋的领域，那么与上述相同的分析表明，将式（5.3）中的 400 用任意的 $\xi > 0$ 值来代替，将得到

$$\begin{aligned} \mu_{ij} &= L(d_{ij}/\xi) = \frac{1}{1 + 10^{-d_{ij}/\xi}} = \frac{10^{r_i/\xi}}{10^{r_i/\xi} + 10^{r_j/\xi}} \\ &\Rightarrow \frac{\mu_{ij}}{\mu_{ji}} = \frac{10^{r_i/\xi}}{10^{r_j/\xi}} \Rightarrow \mu_{ij} = \mu_{ji}\left[10^{(r_i - r_j)/\xi}\right] \end{aligned} \tag{5.4}$$

现在，如果队伍 i 对队伍 j 具有 ξ 个评分的领先优势，则队伍 i 击败队伍 j 的概率预期将是队伍 j 击败队伍 i 的概率的 10 倍。调整 ξ 值也是调整评分体系以使其对于特定需求达到最佳的另一种方式。

定常的和值

如果得分 S_{ij} 只依赖于胜负和平局，即如同为国际象棋所定义的式（5.2）那样的话，则有 $S_{ij} + S_{ji} = 1$，而正如下面将要说明的那样，不管对埃洛评分进行了多少次更新，上式将使埃洛评分的总和保持为一个常数值。这一定常和值的特性对于 S_{ij} 满足 $S_{ij} + S_{ji} = 1$ 的任何竞赛而言都同样成立。下面是对这一论断的正式陈述和证明。

埃洛和保持为常数

只要有 $S_{ij}+S_{ji}=1$，则无论得分 S_{ij} 是如何定义的，在任何时刻 $t>0$ 时，所有埃洛评分 $r_i(t)$ 的和总是等于初始评分 $r_i(0)$ 的和。换而言之，如果有 m 支队伍，则

$$\sum_{k=1}^{m} r_k(0)=\sigma \Rightarrow \sum_{k=1}^{m} r_k(t)=\sigma$$

特别地，如果初始时不对任何竞争者有所偏好的话，则你可能会给每个参与者都赋予一个初始评分 $r_i(0)=0$，而这将会使得之后的埃洛评分的总和保持为 0，从而评分的均值也总是为 0。

显然，将每个初始评分设为 x 将保证任何时刻的平均评分总是等于 x。

证明：注意到无论逻辑斯谛参数 ξ 的值是多少，由式（5.4）都有

$$\mu_{ij}+\mu_{ji}=1。$$

这个等式连同 $S_{ij}+S_{ji}=1$ 一起，便意味着在进行了一场队伍 i 和 j 之间的比赛之后，对旧的评分 $r_{i(\text{old})}$ 和 $r_{j(\text{old})}$ 的改变量分别为（见**第 54 页**）$K(S_{ij}-\mu_{ij})$ 和 $K(S_{ji}-\mu_{ji})$，因此

$$K(S_{ij}-\mu_{ij})+K(S_{ji}-\mu_{ji})=0 \tag{5.5}$$

于是，

$$\begin{aligned}\sum_{k=1}^{m} r_{k(\text{new})} &= \sum_{k\neq i,j}^{m} r_{k(\text{old})}+\left[r_{i(\text{old})}+K(S_{ij}-\mu_{ij})\right]+\left[r_{j(\text{old})}+K(S_{ji}-\mu_{ji})\right]\\ &= \sum_{k=1}^{m} r_{k(\text{old})}。\ \blacksquare\end{aligned}$$

注意！式（5.5）说明队伍 i 的评分的变化量就是队伍 j 的评分变化量的相反数，但这并不意味着 $r_{i(\text{new})}+r_{j(\text{new})}=0$。

NFL 中的埃洛评分

为了了解埃洛评分在国际象棋之外的其他领域表现如何，我们实现了一个基本的埃洛评分方案，它仅依赖于胜负情况来对 NFL 2009—2010 赛季的各支队伍进行评分和排名。我们考虑了所有 267 场常规赛和季后赛，并使用了如**第 42 页**上的表格中所给出的各队的序号。

尽管我们可以使用之前数年的排名或季前赛的排名，但我们还是从一个公平的起点开始，将所有的初始评分都设为 0。因此，在赛季中的任何时刻，所有的评分之和为 0，而最终的评分和也为 0，故而它们的均值同样为 0。

结合**第 54 页**上的式（5.2）给出的胜负得分规则，我们使用了 $K = 32$ 这个 K 因子（该值被数个比赛网站所使用，也与国际象棋界所使用的值接近），但将逻辑斯谛参数由 400 增加到了 $\xi = 1000$，以给出更大的最终评分差。最终的排名对于 ξ 的值并不敏感，而我们对 ξ 的增加也没有明显地改变最终排名，这一排名如下所示。

排名	队伍	评分	排名	队伍	评分
1.	新奥尔良圣徒	173.66	17.	卡罗来纳黑豹	11.474
2.	印第安纳小马	170.33	18.	旧金山 49 人	-1.2844
3.	圣迭戈闪电	127.58	19.	纽约巨人	-5.3217
4.	明尼苏达维京人	103.50	20.	丹佛野马	-11.126
5.	达拉斯牛仔	89.128	21.	迈阿密海豚	-26.717
6.	费城老鹰	69.533	22.	芝加哥熊	-28.142
7.	绿湾包装工	67.829	23.	杰克逊维尔美洲虎	-36.214
8.	亚利桑那红雀	53.227	24.	布法罗比尔	-53.350
9.	纽约喷气机	50.143	25.	克里夫兰布朗	-74.664
10.	新英格兰爱国者	39.633	26.	奥克兰突袭者	-83.319
11.	休斯敦德州人	33.902	27.	西雅图海鹰	-88.845
12.	辛辛那提猛虎	33.012	28.	堪萨斯城酋长	-109.28
13.	巴尔的摩乌鸦	32.083	29.	华盛顿红皮	-110.21
14	亚特兰大猎鹰	28.118	30.	坦帕湾海盗	-130.10
15.	匹兹堡钢人	27.125	31.	底特律雄狮	-170.81
16.	田纳西泰坦	13.222	32.	圣路易斯公羊	-194.12

(5.6)

NFL 2009—2010 赛季仅根据胜负记录以及 $\xi = 1000$ 和 $K = 32$ 得到的埃洛排名

由于这些埃洛评分是严格按照胜负情况来给出的，因此它们如果和**第 46 页**上式（4.18）中给出的胜率之间没有良好的相关关系，那倒是会让人感到奇怪。确实，由式（4.20）可知相关系数为 $R_{rw} = 0.9921$。如我们在**第 47 页**和**第 48 页**上对基纳评分所做的那样进行线性回归，可得

$$\text{队伍 } i \text{ 的估计胜率} = 0.5 + 0.0022268 r_i$$

由此可得

$$\text{MAD} = 0.017958 \text{ 和 } \text{MSE} = 0.0006$$

尽管这些结果相当不错，但我们却不清楚这些结果是否具有显著性，因为本质上它们都归结为利用胜负信息来估计胜率。

一个更有意思的问题则是，埃洛方法在预测 NFL 的胜者方面表现如何？

事后预测的准确率

埃洛方法的长项在于估计胜率，尽管这并不令人吃惊，但我们还不清楚如果按第 48 页所描述的那样将最终的埃洛评分带回过去，并对每场比赛的胜者进行事后预测，那我们能够期望得到怎样的结果。回顾一下 2009—2010 赛季所有 267 场 NFL 比赛的胜者，并将它们与埃洛方法的预测结果——即最终评分较高的队伍被预测为胜者——进行比较，我们可以发现埃洛评分正确地选出了 201 场比赛的胜者，因此

$$\text{埃洛方法的事后预测准确率} = 75.3\%$$

尽管对于预测胜者而言，一个看来合理的做法是为每支主队增加一个主场优势项，但就本例而言，增加主场优势项并不能改进事后预测准确率。

正如在第 48 页上对基纳法的讨论中所提及的，“事后来看的视力是 1.5”这样的开场白对于评分体系而言并不成立——即使是好的评分体系，也无法在事后对所有的比赛都预测正确。埃洛评分那 75.3% 的事后预测准确率已经非常好了——请见**第 123 页**上的对比。

事前预测的准确率

对排名体系更为严酷的考验，则是在第 j 场比赛之后，利用计算得到的评分向量 $\boldsymbol{r}(j)$ 对第 $j+1$ 场比赛中的胜者进行预测，看看这时能做到多好。当然，我们预期这一事前预测的准确率要低于事后预测，因为事后预测使用了已经进行了的所有比赛的全部知识，而事前预测则只能使用当前正在分析的比赛之前的信息。

相比于事后预测，在埃洛事前预测中，主场优势的重要性要高得多，因为如果所有初始评分都相等（正如在我们的例子中那样），则在开始时，当还没有足够多的比赛结果来更新每支队伍的评分时，对于主场的考虑是埃洛方法能够用来对两支队伍加以区别对待的唯一因素。在根据 $\xi = 1000$ 和 $K = 32$ 计算得到的逐场比赛后的评分来进行事前预测时，如果在主队的评分中增加 $H = 15$ 的主场优势项，则埃洛方法能够正确预测 267 场比赛中 166 场的胜者，因此

$$\text{埃洛方法的事前预测准确率} = 62.2\%。$$

结合比赛得分

相比于仅利用胜负来构建评分，让我们利用队伍 i 每次对阵队伍 j 时所得的分数 P_{ij} 来将埃洛方法变得更为有趣一点。但我们不是直接使用原始的得分，而是按照导出**第 31 页**上式（4.2）的思路来对“得分”进行重定义，使得队伍 i 对阵队伍 j 所得的分数为

$$S_{ij} = \frac{P_{ij} + 1}{P_{ij} + P_{ji} + 2} \tag{5.7}$$

除了 $0 < S_{ij} < 1$ 这一点之外，该定义还保证了 $S_{ij} + S_{ji} = 1$，因此 S_{ij} 可以被解释为队伍 i 击败队伍 j 的概率（假设平局都已被排除在外）。此外，如果所有的初始评分均为 0（如同在前面的例子中那样），则**第 56 页**上的讨论意味着埃洛评分在整个赛季期间都将不断变化，但它们的和总为 0，而平均评分也为 0。

当对式（5.7）定义的评分使用（与基本埃洛评分中胜负方案相同）$\xi = 1000$ 和 $K = 32$ 时，埃洛方法给出了如下修改后的评分和排名。

排名	队　伍	评　分	排名	队　伍	评　分
1.	绿湾包装工	58.825	17.	丹佛野马	4.1006
2.	明尼苏达维京人	55.217	18.	辛辛那提猛虎	-.75014
3.	新奥尔良圣徒	49.495	19.	纽约巨人	-3.5097
4.	纽约喷气机	47.215	20.	迈阿密海豚	-9.3122
5.	达拉斯牛仔	43.074	21.	田纳西泰坦	-9.8351
6.	巴尔的摩乌鸦	40.357	22.	芝加哥熊	-16.050
7.	圣迭戈闪电	39.974	23.	布法罗比尔	-23.287
8.	印第安纳小马	39.260	24.	华盛顿红皮	-29.039
9.	新英格兰爱国者	37.860	25.	堪萨斯城酋长	-34.647
10.	旧金山49人	33.189	26.	西雅图海鹰	-35.150
11.	休斯敦德州人	18.447	27.	杰克逊维尔美洲虎	-37.050
12.	亚特兰大猎鹰	18.387	28.	克里夫兰布朗	-47.089
13.	费城老鹰	13.984	29.	坦帕湾海盗	-54.373
14.	匹兹堡钢人	9.1308	30.	奥克兰突袭者	-62.652
15.	亚利桑那红雀	6.1216	31.	底特律雄狮	-72.800
16.	卡罗来纳黑豹	5.2596	32.	圣路易斯公羊	-84.352

使用比分且 $\xi = 1000$ 和 $K = 32$ 时 NFL 2009—2010 赛季的埃洛排名　　(5.8)

$\xi = 1000$，$K = 32$，$H = 15$ 时的事后预测和事前预测

式（5.8）中的评分可以用类似于之前所述的方法来给出事后和事前预测，但不同于仅考虑胜负情况的事后预测，当考虑比分时，主场优势项 H 便显得很重要了。如果对于事后预测和事前预测都使用了 $H = 15$（通过实验确定），则式（5.8）中的评分可以正确地在事后预测 194 场（或 72.7%）比赛的胜者，以及在事前预测 175 场（或 65.5%）比赛的胜者。

对 NFL 比分使用可变的 K 因子

式（5.8）中的评分受到了一个缺陷的影响，即它们是在所有比赛都同样重要的假设下得到的——即对于所有比赛而言，K 因子都相同。但是，我们可以认为在决定最终排名时，NFL 季后赛应该比常规赛更加重要。此外，在 NFL 中存在一种潜规则，即在常规赛季的最后一两场比赛中，强队或者让它们的关键球员休息，或者直截了当地拒绝派他们上场出赛，以避免在比赛中受伤——在 2009—2010 赛季的最后两周内，这种情况就确实发生了。

因此，在这一段时间内，更好的队伍的表现在他们的真实得分能力之下，而在埃洛方法看来，他们的对手则显得比他们的真实实力更强。这表明我们应该使用可变的 K 因子来处理这种情况。

让我们通过如下的（相当随意的）设置来试图纠正这一问题

$K = 32$　对于前 15 周；

$K = 16$　对于最后两周；

$K = 64$　对于季后赛。

进行了如此修改（但保持 $\xi = 1000$）之后，埃洛评分和排名变为了式（5.9）所示的那样。

排名	队　伍	评　分	排名	队　伍	评　分
1.	新奥尔良圣徒	67.672	17.	亚利桑那红雀	1.4959
2.	明尼苏达维京人	63.080	18.	辛辛那提猛虎	1.4707
3.	印第安纳小马	57.297	19.	卡罗来纳黑豹	-3.2548
4.	绿湾包装工	48.227	20.	迈阿密海豚	-7.6586
5.	纽约喷气机	38.781	21.	田纳西泰坦	-7.7187
6.	圣迭戈闪电	35.864	22.	芝加哥熊	-18.565
7.	巴尔的摩乌鸦	35.264	23.	华盛顿红皮	-22.432
8.	新英格兰爱国者	28.496	24.	布法罗比尔	-22.709
9.	旧金山49人	26.047	25.	西雅图海鹰	-29.918
10.	达拉斯牛仔	22.742	26.	杰克逊维尔美洲虎	-31.326
11.	休斯敦德州人	16.289	27.	堪萨斯城酋长	-35.945
12.	费城老鹰	14.492	28.	克里夫兰布朗	-51.611
13.	亚特兰大猎鹰	10.531	29.	坦帕湾海盗	-54.044
14.	匹兹堡钢人	7.5351	30.	奥克兰突袭者	-58.546
15.	丹佛野马	7.0388	31.	底特律雄狮	-68.265
16.	纽约巨人	6.9994	32.	圣路易斯公羊	-77.329

（5.9）

使用比分且 $\xi = 1000$ 和 $K = 32, 16, 64$ 时 NFL 2009—2010 赛季的埃洛排名

当只考虑最终排名时，式（5.9）中的结果看起来比式（5.8）更好，因为式（5.9）更准确地反映了这个赛季尘埃落定时所发生的真实情况。在式（5.9）中，圣徒队最终排名第1，而正如由图5.2中的季后赛结果所示，圣徒队赢得了超级碗。此外，在季后赛中，相比于小马队，维京人队给圣徒队制造了更大的麻烦，而在式（5.9）的最终评分和排名中，这一点也得到了体现。

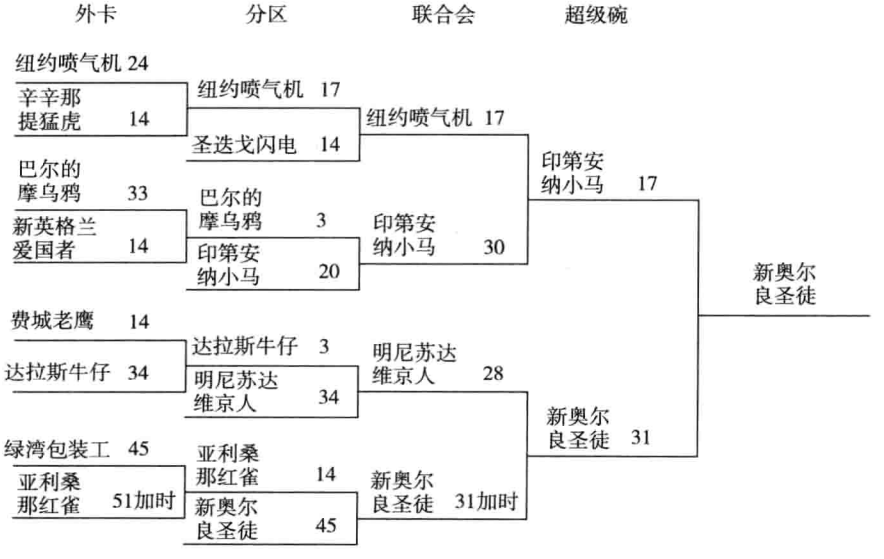


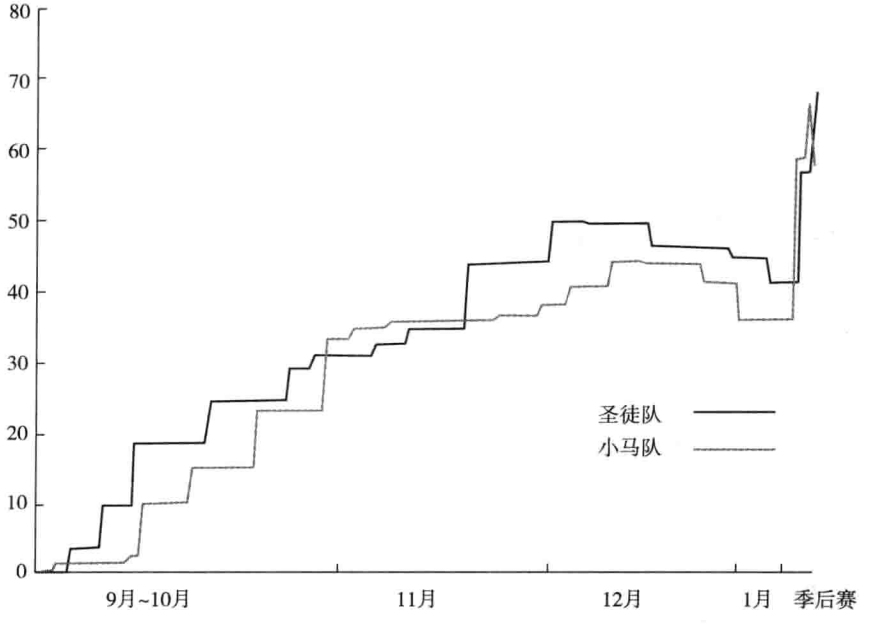
图5.2　NFL 2009—2010 赛季的季后赛结果

使用比分和可变 K 因子的事后和事前预测

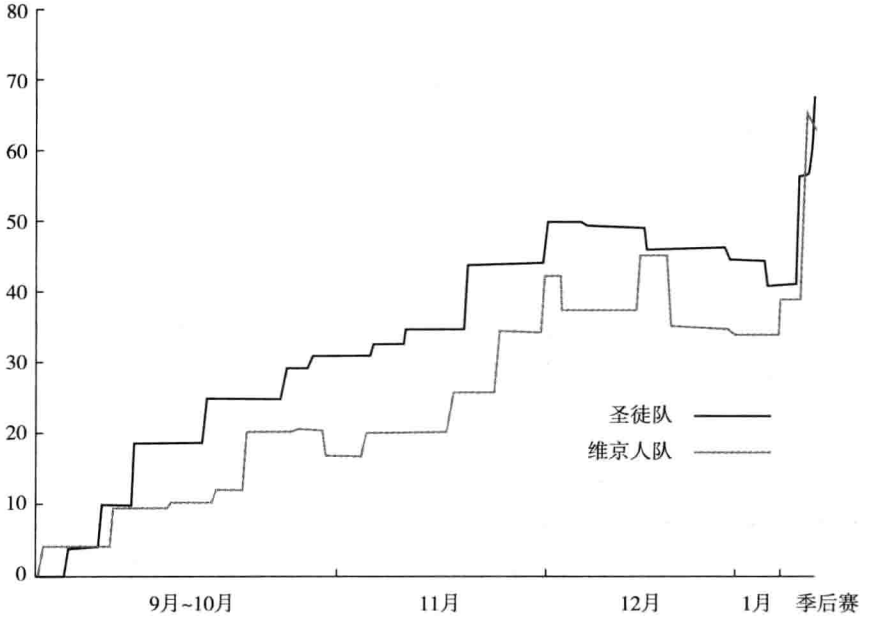
式（5.9）中的评分比式（5.8）更好地反映了季后赛的结果，但式（5.9）预测比赛胜者的能力只比式（5.8）好了一点点。将 K 因子由常数值 $K = 32$ 变为可变量 $K = 32$，16，64 的这一修改，将会改变评分值，因此主场优势项 H 也必须相应改变——对于事后预测而言，最佳的选择是 $H = 0$，而对事前预测则是 $H = 9.5$。在这些参数值下，可变 K 因子的埃洛体系正确地事后预测出了 194 场（或 72.7%）比赛的胜者，以及事前预测出了 176（或 65.9%）比赛的胜者。

逐场比赛分析

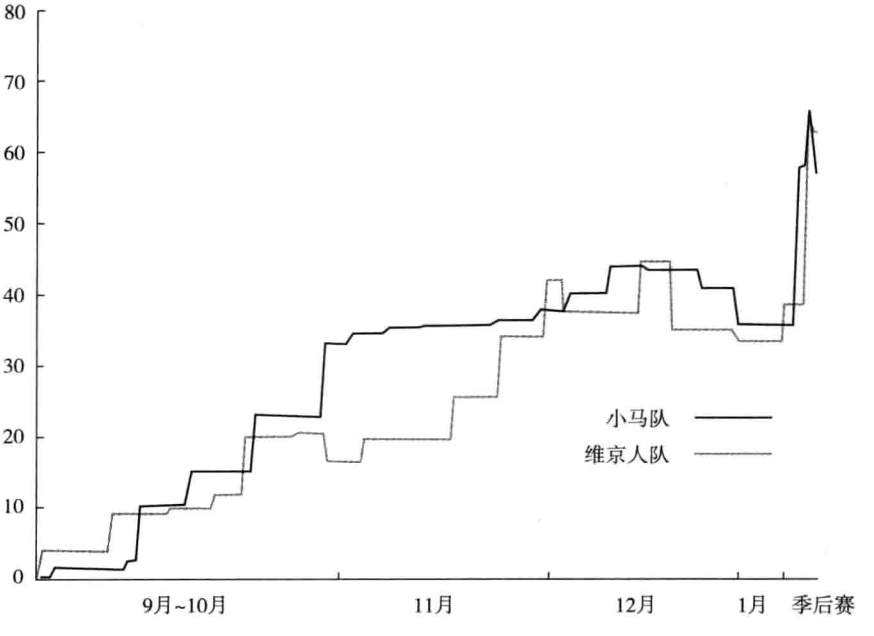
通过逐场比赛的方式跟踪某支队伍在整个赛季中的评分，能够产生出对整体实力的一种反映，而这种反映在最终的评分中可能并不明显。例如，当埃洛体系对季后赛赋予更高的权值并对“赛季末放弃的比赛”打上折扣（如式（5.9）中所做的那样）时，一支队伍有可能仅仅因为季后赛的出色表现而获得高于其应得的最终排名。比如我们来考虑一下式（5.9）中排名前三的队伍逐场比赛后的评分。下面 3 幅图跟踪了圣徒队对小马队、圣徒队对维京人队以及小马队对维京人队的逐场比赛后的评分。



圣徒队对小马队的埃洛评分的逐场跟踪图



圣徒队对维京人队的埃洛评分的逐场跟踪图



小马队对维京人队的埃洛评分的逐场跟踪图

注意，圣徒队几乎一直都在维京人队和小马队之前，因此这清晰地表明圣徒队在整个 2009—2010 赛季中，就总体而言都是实力更强的队。但是小马队和维京人队应该如何比较呢？如果将它们各自的评分在整个赛季期间进行积分（即计算最后一幅图中这两支队伍的评分曲

线下方区域的面积），则结果表明小马队在维京人队之前。换而言之，尽管在最终的评分中维京人队的排名比小马队靠前，但这一逐场比赛分析表明，在整个赛季中，小马队总的来说比维京人队更强。身为一名中立的、观看了小马队和维京人队在2009—2010赛季几乎每场比赛［是的，本人（卡尔·梅耶）拥有假日票和一台数字摄像机］的球迷，我倾向于赞同这一结论。

结论

埃洛评分体系是具有简单优雅美的一个典范。此外，它可以适用于范围很广的应用领域。如果你试图从无到有地构建一个评分体系，那它身后的思路和原理将对你大有帮助。

不管在埃洛评分和脸书之间是否真的存在联系，下面的杂谈都有助于强调埃洛评分体系的适应性。

杂谈：埃洛、辣妹和《社交网络》（The Social Network）

埃洛进电影了。唔，至少他的排名方法进电影了。在影片《社交网络》的一个场景中，爱德华多·萨维林（Eduardo Saverin，脸谱的共同创始人之一）走进了哈佛大学的一间寝室里，而马克·扎克伯格（Mark Zuckerberg，脸谱的共同创始人中的另一位）则正在便携式计算机上敲下最后一个键，这时他们的对话如下。

马克："搞定！时机恰到好处。爱德华多就在这儿，他就要得到那个关键部分了。"

（在谈论了马克女友的困扰之后）

爱德华多："你还好吧？"

马克："我需要你。"

爱德华多："我就在这里等候吩咐。"

马克："不，我需要那个给棋手排名的算法。"

爱德华多："你没事吧？"

马克："我们要给姑娘们也来个排名。"

爱德华多："你是指别的学生吗？"

马克："是的。"

爱德华多："你觉得这真是个好点子吗？"

马克："我需要那个算法——我需要那个算法！"

接下来，爱德华多在他们的哈佛寝室的窗户上写下了**第54页**上的那个埃洛公式，即式（5.3）。呃，差不多是那个公式吧。爱德华多把公式写成了

$$E_a = \frac{1}{1 + 10(R_b - R_a)/400}$$

而非

$$E_a = \frac{1}{1 + 10^{(R_b - R_a)/400}} = \frac{1}{1 + 10^{-(R_a - R_b)/400}}。$$

我们也许应该在这里替制片人说句话，因为不管怎样，爱德华多是用一支蜡笔在窗户上写

字的。影片将埃洛法说成了某种只有天才型的书呆子才能欣赏的惊人复杂的东西，而现在你已经认识到，这并不正确。此外，影片直接暗示了埃洛法是扎克伯格的网站Facemash——脸谱的前身——对哈佛的辣妹们进行打分的基础。认为扎克伯格可能使用了埃洛法，这一猜测是有道理的，因为埃洛法是一个近乎完美的、通过简单的“非此即彼”的成对比较来进行对象评分和排名的方式，但是，可能除了扎克伯格本人之外，没有人确切地知道他是否使用了埃洛法来对哈佛的女生进行过评分和排名。如果你知道的话，请告诉我们。

数字——

2895 = 有史以来最高的棋手积分。——鲍比·菲舍尔（Bobby Fischer，1971 年 10 月）。

2886 = 有史以来第二高的棋手积分。——加里·卡斯帕罗夫（Gary Kasparov，1993 年 3 月）。

——chessbase. com

第 6 章　马尔可夫法



这个新的体育评分方法使用了马尔可夫的一个老技术，因此我们称之为马尔可夫法[⊖]。1906 年，马尔可夫发明了他的那些之后被命名为马尔可夫链的方法，它们用于对随机过程进行描述。尽管马尔可夫最初将他的方法用于对普希金的诗歌《叶甫盖尼·奥涅金》中的原音和辅音序列进行语言分析，但自那时以来，这些链得到了太多太多的应用[8, 80]。就在最近，我们学校的研究生安吉拉·戈万（2008 年于北卡罗来纳大学获博士学位）[34] 和卢克·英格拉姆（2007 年于查尔斯顿学院获理学硕士学位）[41]，各自利用马尔可夫链成功地对 NFL 橄榄球队和 NCAA 篮球队进行了排名。

马尔可夫法

马尔可夫法的主要思想

马尔可夫评分法可以用一个词来总结：投票。两支队伍之间的每次较量，都是弱队给强队投票的机会。

可以有许多方法来衡量对队伍所投的票。最简单的投票方法可能得算是使用胜负情况。失利的队伍为每支击败它的队伍投出一票。更高级的模型则考虑了比赛中的得分。在这种情况下，负者投出的票数等于它在与更强的对手交手时输掉的分差。一个再高级些的模型则是让两支队伍分别向对方投出与各自失分相等数量的选票。最终，获得票数最多的队伍将赢得最高排名。这个思想实际上是谷歌用来进行网页排名的 PageRank 算法的一个修改[49]。正如**第 71 页**的讨论所说明的那样，在这一建模过程中可以进行微调的地方简直数不胜数，不过，在介绍各种增强功能之前，先来直接试一试我们的范例。

⊖ 至少还有另外两种现存的体育排名模型也是基于马尔可夫链的。其中一个是彼得·穆查（Peter Mucha）及其同事的产物，称为随机游走者评分法[17]；另一个则是乔尔·索科尔（Joel Sokol）和保罗·基瓦姆（Paul Kvam）的成果，称为 LRMC[48]。

利用胜负进行投票

仅使用胜负数据的投票矩阵 $\boldsymbol{V}$ 如下所示。

$$\boldsymbol{V} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

由于杜克大学输给了所有的对手，因此，它为其他每支队伍投出了权重相等的票数。为了从这个投票矩阵中梳理出一个排名向量，我们要对 $\boldsymbol{V}$ 中各行进行归一化，以此来产生一个（接近于）随机的矩阵 $\boldsymbol{N}$。

$$\boldsymbol{N} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$\boldsymbol{N}$ 目前是准随机的，因为迈阿密大学未尝败绩。这与网页排名领域中众所周知的悬挂结点问题[49]类似。在网页排名领域中对这个问题的解决方法，是将所有的全 0 行 $\mathbf{0}^{\mathrm{T}}$ 用 $1/n\boldsymbol{e}^{\mathrm{T}}$ 来代替。我们在此借用这一思路，使得 $\boldsymbol{N}$ 变为一个随机矩阵 $\boldsymbol{S}$。

$$\boldsymbol{S} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 1/4 & 1/4 & 1/4 & 1/4 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

尽管悬挂结点（没有出链的页面）在万维网环境中十分常见[51, 26]，但在一个完整的体育赛季期间，它们却极为罕见。还有其他几种处理一场未负的队伍的方法，在**第 73 页**上的讨论中，我们将给出上述利用均匀分布行来替代的方法之外的其他几种可选途径。

再次模仿网页的 PageRank 算法，我们计算出这个随机矩阵的稳态向量。该稳态向量即为 $\boldsymbol{S}$ 的主特征向量，可通过求解特征系统 $\boldsymbol{Sr} = \boldsymbol{r}$ 来得到[54]。下面的小故事简要解释了使用稳态向量来作为队伍排名方法的理由。故事的主角是一名墙头草式的球迷，他或她总是支持当前最好的那支队伍。矩阵 $\boldsymbol{S}$ 可用如图 6.1 所示的图来表示。这名墙头草球迷由这个网络中的任一结点开始，并根据离开当前所支持队伍的链接来选择他要支持的下一支队伍。例如，如果该球迷从 UNC 开始，并问 UNC：“谁是最好的队伍？”UNC 将会回答：“迈阿密大学或 VT，因为他们都击败了我们。”于是这个球迷将接着靠扔硬币来决定要支持的队伍，并假设他选中了 VT，之后他又继续询问相同的问题。这一次，VT 将回答说迈阿密大学是最好的队伍，因此他又跳上了迈阿密大学的彩车。一到了迈阿密大学的营地后，他又提出了他的问题。不过由于迈阿密大学未曾被击败，因此他随机地跳到其他队伍中的任意一支。（这是由于我们增加了 $1/5\boldsymbol{e}^{\mathrm{T}}$ 这一行的缘故，不过还有其他几种巧妙的策略可以处理从未告负的球队，

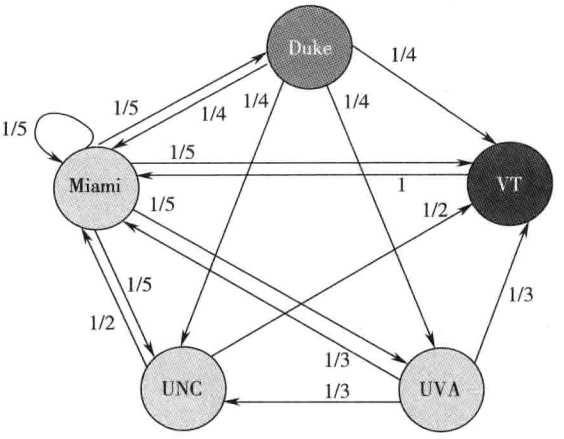


图 6.1　墙头草球迷在这幅马尔可夫图上进行了一次随机游走

同时仍然能够给出具有良好收敛性能的马尔可夫链。请见**第 73 页**。）如果墙头草球迷一直如此行事，而我们则对他在每支队伍上停留的时间比例加以监视，那么我们就能得到各支队伍重要程度的一个衡量。墙头草球迷访问得最为频繁的那支队伍就是排名最高的队伍，因为该球迷总是从其他强队那里被引向这支队伍。从数学上来说，这名墙头草球迷就是在由一条马尔可夫链所定义的图上进行了一次随机游走，而他在相当长的时间内在链的各个状态上消耗的时间比例，就正好是该链的稳态向量或主特征向量。对于这个 5 支队伍的例子而言，S 的稳态评分向量 r 和相应的排名如表 6.1 所示。

表 6.1　马尔可夫法：在 5 支队伍示例中利用胜负信息的评分结果

队　伍	r	排　名
Duke	0.087	第 5
Miami	0.438	第 1
UNC	0.146	第 3
UVA	0.110	第 4
VT	0.219	第 2

墙头草球迷在马尔可夫图上进行随机游走

马尔可夫评分向量给出了墙头草球迷在马尔可夫图上进行长时间随机游走时在每支队伍上所消耗的时间的比例。

利用负者的分差进行投票

当然，球队可能更愿意为对手投出不止一票。继续我们那位墙头草球迷的故事。UNC 可能愿意给 VT 投出更多的票数，而给迈阿密大学的票数则较少，因为 VT 以 27 分的优势

击败了 UNC，而迈阿密大学的优势仅为 18 分。在这个更高级的模型中，投票和随机矩阵为

$$\boldsymbol{V} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 45 & 3 & 31 & 45 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 18 & 0 & 0 & 27 \\ 0 & 8 & 2 & 0 & 38 \\ 0 & 20 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

和

$$\boldsymbol{S} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 45/124 & 3/124 & 31/124 & 45/124 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 0 & 18/45 & 0 & 0 & 27/45 \\ 0 & 8/48 & 2/48 & 0 & 38/48 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}。$$

这一模型给出的评分和排名列于表 6.2 中。相比于基本的投票模型，利用分差的投票给出了稍有不同的排名，但评分则有显著差异。如果我们希望不仅能预测胜者，还能预测分差的话，这些差别将可能十分重要。

表 6.2　马尔可夫法：5 支队伍示例的分差评分结果

队　伍	r	排　名
Duke	0.088	第 5
Miami	0.442	第 1
UNC	0.095	第 4
UVA	0.110	第 3
VT	0.265	第 2

如果两支队伍在一个赛季中交手不止一次，则建模者可以选择构造投票矩阵的不同方式。矩阵中相应的元素既可以是这两支队伍在整个赛季的所有交手中分差的累积值，也可以是分差的平均值。

胜负双方都按失分进行投票

在一个更为高级的模型中，我们让胜负双方都投出等于其所丢分数（以下记为 point）的票数。这一模型使用了完整的得分信息，此时，$\boldsymbol{V}$ 和 $\boldsymbol{S}$ 矩阵（我们记为 $\boldsymbol{V}_{\text{point}}$ 和 $\boldsymbol{S}_{\text{point}}$ 以便和**第 72 页**上的矩阵区分开来）为

$$\boldsymbol{V}_{\text{point}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 52 & 24 & 38 & 45 \\ 7 & 0 & 16 & 17 & 7 \\ 21 & 34 & 0 & 5 & 30 \\ 7 & 25 & 7 & 0 & 52 \\ 0 & 27 & 3 & 14 & 0 \end{pmatrix} \end{array}$$

和

$$S_{\text{point}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 52/159 & 24/159 & 38/159 & 45/159 \\ 7/47 & 0 & 16/47 & 17/47 & 7/47 \\ 21/90 & 34/90 & 0 & 5/90 & 30/90 \\ 7/91 & 25/91 & 7/91 & 0 & 52/91 \\ 0 & 27/44 & 3/44 & 14/44 & 0 \end{pmatrix} \end{array}。$$

注意，$\boldsymbol{S}$ 矩阵的这种表示方式有一个额外的好处，即我们不需要再考虑全胜的问题。这时，只有当一支队伍以不失 1 分的战绩击败所有对手时，才会在矩阵中出现全 0 行，而这几乎是不可能发生的。这一模型给出的评分和排名向量如表 6.3 所示。

表 6.3　马尔可夫法：5 支队伍示例的胜负双方失分投票评分结果

队　伍	r	排　名
Duke	0.095	第 5
Miami	0.296	第 1
UNC	0.149	第 4
UVA	0.216	第 3
VT	0.244	第 2

超越比赛得分

体育迷们都知道比赛得分只是与竞赛有关的众多统计数据之一。例如，在橄榄球中，还有总的前进码数、失误次数、控球时间、冲球码数、传球码数等没完没了的统计数据。尽管比分与比赛结果之间有着直接的关联，但它绝不是我们希望加以预测的唯一东西（想一想分差预测）。同时结合了这些额外统计数据的模型，自然将是一个更加完整的模型。利用马尔可夫模型，可以特别容易地通过为每种统计数据增加一个矩阵来加入更多的统计信息。例如，使用如表 6.4 所示的前进码数信息，我们可以构建一个码数（以下记为 yardage）的投票矩阵，其中各队根据对手所获得的码数来投出选票。

表 6.4　5 支队伍示例中的总前进码数数据

	Duke	Miami	UNC	UVA	VT
Duke		100-557	209-357	207-315	35-362
Miami	557-100		321-188	500-452	304-167
UNC	357-209	188-321		270-199	196-338
UVA	315-207	452-500	199-270		334-552
VT	362-35	167-304	338-196	552-334	

$$\boldsymbol{V}_{\text{yardage}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 577 & 357 & 315 & 362 \\ 100 & 0 & 188 & 452 & 167 \\ 209 & 321 & 0 & 199 & 338 \\ 207 & 500 & 270 & 0 & 552 \\ 35 & 304 & 196 & 334 & 0 \end{pmatrix} \end{array}。$$

元素（Duke，Miami）的值为 577，它表明杜克大学在对阵迈阿密大学时，让对手获得了 577 码的前进码数，因此杜克大学将根据这个相当大的值来给迈阿密大学投票。按行归一化可以给出如下的随机矩阵 $\boldsymbol{S}_{\text{yardage}}$，

$$\boldsymbol{S}_{\text{yardage}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 577/1611 & 357/1611 & 315/1611 & 362/1611 \\ 100/907 & 0 & 188/907 & 452/907 & 167/907 \\ 209/1067 & 321/1067 & 0 & 199/1067 & 338/1067 \\ 207/1529 & 500/1529 & 270/1529 & 0 & 552/1529 \\ 35/869 & 304/869 & 196/869 & 334/869 & 0 \end{pmatrix} \end{array}。$$

由 $\boldsymbol{S}_{\text{yardage}}$ 给出的评分和排名向量见表 6.5。

表 6.5　马尔可夫法：5 支队伍示例的码数评分结果

队　伍	r	排　名
Duke	0.105	第 5
Miami	0.249	第 2
UNC	0.170	第 4
UVA	0.260	第 1
VT	0.216	第 3

当然，得分或码数这些数据并没有什么特别之处——任何统计信息都可以被加入进去。关键在于，我们只需牢记投票这一行为所类比的对象。因此，通过类似的方式，我们可以得到 $\boldsymbol{S}_{\text{turnover}}$ 和 $\boldsymbol{S}_{\text{pass}}$ 为

$$\boldsymbol{S}_{\text{turnover}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 1/9 & 3/9 & 4/9 & 1/9 \\ 3/9 & 0 & 4/9 & 1/9 & 1/9 \\ 2/7 & 3/7 & 0 & 1/7 & 1/7 \\ 1/5 & 0 & 1/5 & 0 & 3/5 \\ 1/10 & 6/10 & 2/10 & 1/10 & 0 \end{pmatrix} \end{array}$$

和

$$\boldsymbol{S}_{\text{pass}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 29.7/118.6 & 30.8/118.6 & 28/118.6 & 30.1/118.6 \\ 30.3/123.6 & 0 & 36.3/123.6 & 31/123.6 & 26/123.6 \\ 29.1/111.6 & 23.7/111.6 & 0 & 27.5/111.6 & 31.3/111.6 \\ 32/131.9 & 29/131.9 & 32.5/131.9 & 0 & 38.4/131.9 \\ 29.9/114.2 & 34/114.2 & 28.7/114.2 & 21.6/114.2 & 0 \end{pmatrix} \end{array}$$

下一步就是将每种单独统计数据的随机矩阵结合为一个单一的矩阵 $\boldsymbol{S}$，使其同时包含了多种统计数据。我们同样有多种途径可以做到这一点。例如，可以将这个单一全局矩阵 $\boldsymbol{S}$ 构建

为各个单独的随机矩阵 $\boldsymbol{S}_{\text{point}}$，$\boldsymbol{S}_{\text{yardage}}$，$\boldsymbol{S}_{\text{turnover}}$ 和 $\boldsymbol{S}_{\text{pass}}$ 的一个凸组合

$$\boldsymbol{S} = \alpha_1 \boldsymbol{S}_{\text{point}} + \alpha_2 \boldsymbol{S}_{\text{yardage}} + \alpha_3 \boldsymbol{S}_{\text{turnover}} + \alpha_4 \boldsymbol{S}_{\text{pass}}$$

式中，$\alpha_1, \alpha_2, \alpha_3, \alpha_4 \geqslant 0$，且 $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$。采用凸组合可以保证 $\boldsymbol{S}$ 仍为随机，对于评分向量的存在性而言，这一点很重要。权值 α_i 可以根据来自体育分析家的信息明智地加以确定。不过，出于目前模型演示的考虑，我们简单地设 $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 1/4$，因此

$$\boldsymbol{S} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0.2617 & 0.2414 & 0.2788 & 0.2182 \\ 0.2094 & 0 & 0.3215 & 0.3055 & 0.1636 \\ 0.2439 & 0.3299 & 0 & 0.1578 & 0.2684 \\ 0.1637 & 0.2054 & 0.1750 & 0 & 0.4559 \\ 0.1005 & 0.4653 & 0.1863 & 0.2479 & 0 \end{pmatrix} \end{array} \text{且} \boldsymbol{r} = \begin{pmatrix} .15 \\ .24 \\ .20 \\ .21 \end{pmatrix}。$$

马尔可夫模型的美妙之处，在于它所具有的灵活性和能力，使我们可以对它进行几乎是无限的调整。差不多任何一种统计信息（主场优势、天气、伤病等）都可以被添加到模型中。当然，精妙之处在于如何调整参数 α_i。

处理全胜的队伍

本章前述的某些产生马尔可夫矩阵元素的方法，很可能会遭遇到全0行的问题，这将使得马尔可夫矩阵成为准随机的。这一准随机性将给马尔可夫模型带来麻烦，因为要产生排名向量，就需要矩阵是按行随机的。在**第68页**的例子中，我们利用了网页排名中的一个标准技巧来处理未尝败绩的迈阿密大学队——将所有的 $\mathbf{0}^{\mathrm{T}}$ 行用 $1/n\boldsymbol{e}^{\mathrm{T}}$ 来代替。尽管这种做法在数学上解决了所面临的问题，并产生了一个随机矩阵，但还有其他处理全胜队伍的方法，这些方法在体育评分的问题中更为合理。

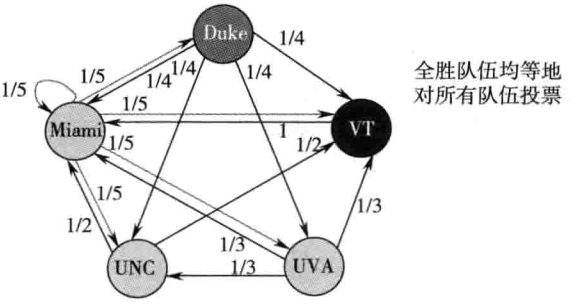
例如，与其迫使最好的队伍给包括自己在内的各队投出相同的票数，为什么不让最好的队伍只给自己投票呢。此时，对每支全胜的队伍，用 $\boldsymbol{e}_i^{\mathrm{T}}$ 即恒等矩阵第 i 行的行向量来取代 $\mathbf{0}^{\mathrm{T}}$ 行。现在的数据便生成了随机矩阵以及马尔可夫链。但是，这一修正带来了另一个问题。实际上，这个问题是一个与矩阵可约性有关的问题，它对所有马尔可夫模型而言都是一个潜在的问题。在我们考虑的情况中，这一问题肯定存在。要使得稳态向量存在且唯一，马尔可夫链必须是不可约的（同时是非周期性的，不过这一点几乎总是能够满足）[54]。一条不可约的链是指其中的任意两支队伍之间都存在着一条通路的链。当全胜的队伍仅对自己投票时，链是可约的，因为这些队伍此时就成为了链的吸收态。换而言之，在图上随机游走的墙头草球迷将最终陷入到一支全胜的队伍中。我们在此再一次使用网页排名中的一个技巧，通过简单地在随机化后的投票矩阵 $\boldsymbol{S}$ 上加上所谓的跳转矩阵 $\boldsymbol{E}$[49]。因此，对于 $0 \leqslant \beta \leqslant 1$，马尔可夫矩阵 $\overline{\boldsymbol{S}}$ 为

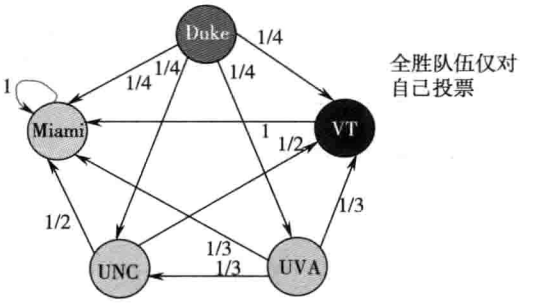
$$\overline{\boldsymbol{S}} = \beta \boldsymbol{S} + (1 - \beta)/n\boldsymbol{E}$$

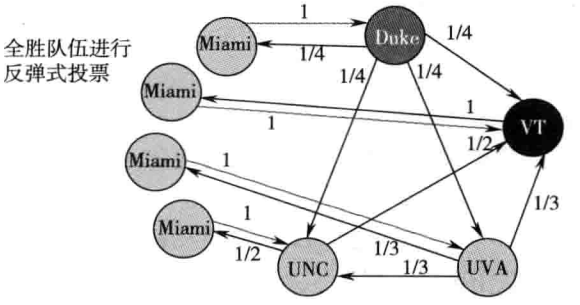
式中，$\boldsymbol{E}$ 为全1矩阵；n 为队伍数量。现在 $\overline{\boldsymbol{S}}$ 是不可约的，因为每支队伍都直接与其他队伍通过至少某个小的概率而联系在一起，而 $\overline{\boldsymbol{S}}$ 的稳态向量便也存在且唯一。这一评分向量依赖于

标量β的选择。通常来说，β越大，模型就越符合原始的数据。对于万维网数据，$\beta = 0.85$是一个常被引述的取值[15]。但对于体育数据，一个远低得多的β值可能更为适宜。实验表明$\beta = 0.6$能很好地用于NFL数据[34]，而$\beta = 0.5$适用于NCAA篮球赛[41]。无论如何，β的选择看来是需要通过特定的应用、数据、体育项目甚至赛季来加以确定的。

处理全胜队伍的另一种方法，则是让墙头草球迷返回他在此步的出发点，然后再开始随机游走。这一反弹的想法已经在网页排名中被成功实现[26, 53, 49, 77]，它漂亮地模型化了对浏览器的“后退”按钮的使用，而这正是上网者的自然反应。尽管在体育排名领域，这一反弹的思路说起来并不那么恰当，但它仍然可能是值得探究的。以下的示意图对处理全胜队伍的三种方法进行了总结。







马尔可夫评分方法总结

下面的阴影框归总了我们在描述体育队伍排名的马尔可夫方法中所使用的记法。

马尔可夫评分方法的记法

k　需要结合到马尔可夫模型中的统计数据种类数

$\boldsymbol{V}_{\text{stat1}}, \boldsymbol{V}_{\text{stat2}}, \cdots, \boldsymbol{V}_{\text{stat}k}$　对每种比赛统计数据的原始投票矩阵

$[\boldsymbol{V}_{\text{stat}}]_{ij}$ = 队伍 i 根据统计数据 stat 对队伍 j 所投的票数

$\boldsymbol{S}_{\text{stat1}}, \boldsymbol{S}_{\text{stat2}}, \cdots, \boldsymbol{S}_{\text{stat}k}$　由相应的投票矩阵 $\boldsymbol{V}_{\text{stat1}}, \boldsymbol{V}_{\text{stat2}}, \cdots, \boldsymbol{V}_{\text{stat}k}$ 所得的随机矩阵

$\boldsymbol{S}$ 由 $\boldsymbol{S}_{\text{stat1}}, \boldsymbol{S}_{\text{stat2}}, \cdots, \boldsymbol{S}_{\text{stat}k}$　构建的最终的随机矩阵

$$\boldsymbol{S} = \alpha_1 \boldsymbol{S}_{\text{stat1}} + \alpha_2 \boldsymbol{S}_{\text{stat2}} + \cdots + \alpha_k \boldsymbol{S}_{\text{stat}k}$$

α_i　与统计数据 i 相对应的权重；$\sum_{i=1}^{k} \alpha_i = 1$ 且 $\alpha_i \geqslant 0$

$\bar{\boldsymbol{S}}$　确保了不可约性的随机马尔可夫矩阵；

$$\bar{\boldsymbol{S}} = \beta \boldsymbol{S} + (1 - \beta)/n\boldsymbol{E} \quad (0 < \beta < 1)$$

$\boldsymbol{r}$　马尔可夫评分向量；$\bar{\boldsymbol{S}}$ 的稳态向量（即主特征向量）

n　联盟中队伍的数量 = $\bar{\boldsymbol{S}}$ 的阶次

由于大多数数学软件如 MATLAB 和 Mathematica 等都具有内建的特征向量计算函数，因此马尔可夫法的步骤十分紧凑。

队伍评分的马尔可夫法

1. 利用感兴趣的 k 种比赛统计数据的投票矩阵来生成 $\boldsymbol{S}$。

$$\boldsymbol{S} = \alpha_1 \boldsymbol{S}_{\text{stat1}} + \alpha_2 \boldsymbol{S}_{\text{stat2}} + \cdots + \alpha_k \boldsymbol{S}_{\text{stat}k}$$

式中，$\alpha_i \geqslant 0$，且 $\sum_{i=1}^{k} \alpha_i = 1$；

2. 计算 $\boldsymbol{S}$ 的稳态向量或主特征向量 $\boldsymbol{r}$。（如果 $\boldsymbol{S}$ 可约，则利用不可约的 $\bar{\boldsymbol{S}} = \beta \boldsymbol{S} + (1 - \beta)/n\boldsymbol{E}$ 来计算，其中，$0 < \beta < 1$）

我们利用马尔可夫评分的一系列性质来结束本节。

- 马尔可夫模型可以将许多不同的比赛统计数据非常自然地融合在一起。
- 评分向量 $\boldsymbol{r} > 0$ 可以被很好地解释为（在投票图上随机游走的）墙头草球迷，在长时间内花费于任一支特定队伍上的时间比例。
- 马尔可夫法与著名的网页排名算法 PageRank 有关[14]。

- 没有遭受失利的队伍将造成马尔可夫矩阵中的全0行，这会使得矩阵变为准随机并带来问题。为了能够得出一个排名向量，每个准随机行都必须通过某种方式被强行随机化。**第73页**上的讨论简要地介绍了几种这样的方法，迫使全胜的队伍也必须进行投票。
- 如在PageRank中那样，马尔可夫法的稳态向量已被证明服从幂律分布。这意味着少数几支队伍将赢得大多数的评分，而大部分队伍则只能共同分享少量的剩余评分。这一情况所带来的一个后果，便是马尔可夫法对于输入数据中小的变动极为敏感，特别是当这些变动牵涉到排在幂律分布后段的低排名队伍时尤其如此。在体育领域中，这意味着少许的扰乱将给排名带来巨大、有时甚至是怪异的影响。请见参考文献［21］中对此现象的详细说明。

马尔可夫法和梅西法之间的联系

尽管马尔可夫法和梅西法看来几乎没什么共通之处，但图形表示方法却解释了两者之间的一个有趣的联系。正如在本章前些时候所指出的，在马尔可夫法中可以有多种方式来进行投票。由于梅西法是建立在分差的基础上的，因此利用分差进行投票的马尔可夫法与梅西法的联系最为紧密。考虑一下图6.2，其中描绘了与**第70页**上的分差投票矩阵相关联的马尔可夫图。这幅图实际上恰好囊括了梅西法所用的相同信息。回忆一下梅西法的基本方程，即对于队伍 i 和 j 之间进行的比赛 k，将产生一个形如 $r_i - r_j = y_k$ 的线性方程，该方程意指队伍 i 在此次交锋中以 y_k 分的优势击败了队伍 j。因此，对于我们的5支队伍的例子，由于VT以27分的优势击败了UNC，梅西法将给出方程 $r_{VT} - r_{UNC} = 27$。如果我们利用这一分差信息来作图，便得到了图6.3，它不过是将马尔可夫图中的箭头调转了朝向。两幅图都使用了分差来对链接进行加权，不过链接的朝向有所不同。在梅西图中，链接的方向体现了优势，而在马尔可夫图中则体现了劣势。

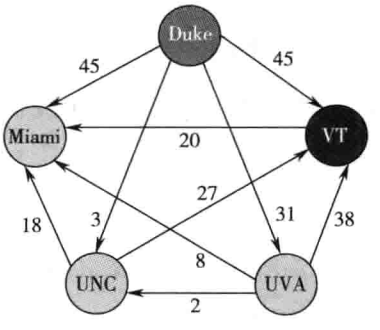


图6.2　分差投票矩阵的马尔可夫图

尽管这两幅图本质上是相同的，但这两种方法在原理上却大不相同。给定了链接的权值后，两种方法的目的都是找出各个结点的权重。马尔可夫法通过随机游走来找到结点权值。回忆一下那个在图中（或更准确来说，在随机化后的图中）进行随机游走的墙头草球

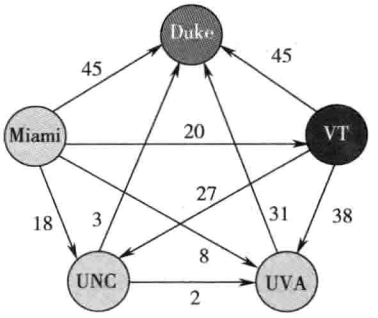


图 6.3　图 6.2 中同样 5 支队伍示例的梅西图

迷，他在长时期内耗费于任一结点 i 的时间所占的比重即为结点权值 r_i；另一方面，梅西法利用拟合方法来获取这些结点权值，即它找到的是能够最好地拟合给定的链接权值的结点权值 r_i。下面的比喻对你或许会有所帮助。

对马尔可夫法和梅西法的比喻

马尔可夫法驱使一名旅行者在其图上游走，而梅西法则给它的图裁制出了一件最为紧贴合身的服装。

杂谈：PageRank 和随机上网者

本节中的马尔可夫法，是根据谷歌用于网页排名的马尔可夫链而得到的。谷歌的马尔可夫链的稳态向量称为 PageRank 向量。解释 PageRank 含义的一种方法，就是设想一名跟我们的墙头草球迷一样的随机上网者在图上进行随机游走。作为 SIAM 的 WhydoMath? 计划的一部分，www. whydomath. org 网站上包括了一段 20 分钟的视频，用于为非数学领域的读者解释 PageRank 中跳转、悬挂结点、排名、超链接和随机上网者等概念。

杂谈：美国高校排名

每一年，《美国新闻》都会公布一份美国顶尖高校的名单。各高校的营销部门都非常谨慎地紧盯着这份名单，因为高排名就意味着大把的钞票进账。许多高中生和他们的家长将会根据这些排名来确定最终选择哪所高校。最近，围绕这些排名，发生了一些争议。在 2007 年 6 月，CNN 报导称，有 121 所私立和文科院校的管理者正式宣布退出了这一排名体系。这些退出的院校称该排名体系不够健全，因为排名是根据自行提交的数据以及来自各院校的校长和院长的主观评分而得出的。

围绕着《美国新闻》高校排名而展开的争论，促使我们去探究这些排名究竟是如何形成的。首先，根据卡耐基教学促进基金会每年所公布的分类法，参与排名的各所学院和大学被归入各个特定的类别。例如，卡耐基分类法中的研究 I 类包括了所有具有博士学位授予点且具有一定规模的院校。对于硕士 I 类和硕士 II 类学校也有着类似的分类。每所学校仅与同一个卡耐基分类中的其他院校进行比较。有关院校的数据则是通过《美国新闻》《彼得逊指南》和美国大学理事会所建立的 Common Data Set 获得。这些数据中约 75% 为数值型数据，它涵盖了 7 个主要方面：同行评价、在校生回返和毕业、师资力量、生源质量、学校资金、校友捐款以及毕业率。剩余 25% 的数据则是大学校长和院长的主观评分，他们将对同行院校即同一个卡耐基分类中的其他学校进行评价。《美国新闻》将这些数值和主观数据通过简单的加权平均结合起来，而权值则由教育研究分析者所确定。

在对《美国新闻》所用的排名方法之简单感到吃惊之余，我们不由得试着去将《美国新闻》的排名和一个如马尔可夫法这样更复杂方法的排名进行了比较。为示意起见，我们决定对 14 所同属一个卡耐基分类的院校进行排名。我们根据如下 7 个方面的 7 个矩阵产生了一个马尔可夫矩阵：学位点数目、平均公共财政支持、学费、师生比、录取率、新生回返率和毕业率。

表 6.6 给出了这 14 所院校的《美国新闻》的排名（2007 年 8 月）和马尔可夫排名。显然，这两个排名差别很大。例如，我们注意到，纽约州立大学奥尔巴尼分校由这些院校的《美国新闻》排行的最后一名跳到了马尔可夫排名的首位。由于很难说哪个排名更好些，因此我们在此想强调的是，不同的方法可以产生效果各异的排名。既然有如此多的东西（学校的名声以及因此而带来的收入）要和排名联系在一起，那么我们更应该采取一种复杂而缜密的方法。

表 6.6　《美国新闻》和马尔可夫法给出的高校排名

院　校	《美国新闻》排名	马尔可夫排名
哈佛大学	1	5
耶鲁大学	2	4
斯坦福大学	3	12
哥伦比亚大学	4	13
芝加哥大学	5	11
布朗大学	6	14
埃默里大学	7	10
范德比特大学	8	3
加州大学伯克利分校	9	7
加州大学洛杉矶分校	10	9
布兰迪斯大学	11	2
罗切斯特大学	12	8
加州大学圣芭芭拉分校	13	6
纽约州立大学奥尔巴尼分校	14	1

第 7 章　攻防评分法

一种想法十分自然的评分方法，是首先对每支队伍或每名参与者的各项单独属性进行评分，然后将这些评分结合为一个单一的数值，它反映了总的实力。特别地，要在大多数的对抗中获得胜利，需要强大的攻击力以及强大的防守力，因此一个合理的做法就是对这两方面分别进行评分，然后再决定谁是第一。

但说比做容易，而对于攻击实力和防守实力的评分而言尤其如此。现在的问题是，正如所有人都知道的那样，在对阵一支防守薄弱的队伍时，即使是最为乏力的进攻看起来也比其真实的实力要强；类似地，表面看来坚强的防守实际上是对手强劲（或弱）攻击力的直接后果。换而言之，在我们所体会到的攻击实力和防守实力之间，存在着一个循环关系，因此，当试图分别评判攻击能力或防守能力时，都必须考虑到另外的那一个因素。本章的重点就在于，这一评判是如何做到的。

攻防评分法的目的

攻防评分法（也称 OD 法）的目的是为每支队伍（或每名参与者）同时赋予一个*攻击评分*（对队伍 i 记为 $o_i > 0$）和一个*防守评分*（对队伍 i 记为 $d_i > 0$），这两个评分由以往比赛积累下来的统计数据而获得。给定任何一种体育项目或竞赛，通常都有许多统计数据需要考虑，但得分是最为基本的统计数据，而且实际上这一数据也相当有用。此外，一旦理解了 OD 法是如何利用得分来生成攻击评分和防守评分的话，就能很容易地了解如何利用比分之外的其他统计数据来构建你自己的体系。出于这些原因，我们将以得分为例来介绍 OD 评分法。

OD 法的前提假设

假设要对 m 支队伍评分，并使用得分来作为队伍表现的主要指标。如果队伍 j 在对手身上得了很多分，特别是在防守评分高的对手身上得了很多分，那它就被认为具有很强的攻击力，并应该得到高的攻击评分；类似地，如果队伍 i 让对手特别是攻击力强的对手只得了很低的分数，那么它就是一支防守优秀的队伍，并应该得到高的防守评分。因此，攻击评分和防守评分是交织在一起的，即每个攻击评分都是所有防守评分的某个函数 f，而每个防守评分也都是所有攻击评分的某个函数 g。

因此现在的问题便归结为如何确定函数 f 和 g。由于我们打算使用得分作为衡量队伍表现的主要指标，因此将以往的得分组织为一个得分矩阵 $\boldsymbol{A} = [a_{ij}]$，其中，$a_{ij}$ 为队伍 j 在对阵队伍 i 时获取的得分（如果这两支队伍交手了不止一次，则取平均值；如果这两支队伍未曾交锋，则置 $a_{ij} = 0$）。注意，a_{ij} 有一个对偶解释，即

$$a_{ij} = \begin{cases} j \text{ 从 } i \text{ 处获取的分数(攻击力描述)} \\ i \text{ 使 } j \text{ 未能超过的分数(防守力描述)} \end{cases} \tag{7.1}$$

换而言之，根据观点的不同，a_{ij} 可以同时反映相对的攻击和防守能力。这一对偶解释可用来给出每支队伍的如下 f 和 g 的定义。

OD 评分

给定一组防守评分 $\{d_1, d_2, \cdots, d_m\}$，定义队伍 j 的攻击评分为

$$o_j = \frac{a_{1j}}{d_1} + \frac{a_{2j}}{d_2} + \cdots + \frac{a_{mj}}{d_m} (\text{更大的 } o_j \text{ 值} \Rightarrow \text{更强的攻击力}) \tag{7.2}$$

类似地，给定一组攻击评分 $\{o_1, o_2, \cdots, o_m\}$，定义队伍 i 的防守评分为

$$d_i = \frac{a_{i1}}{o_1} + \frac{a_{i2}}{o_2} + \cdots + \frac{a_{im}}{o_m} (\text{更大的 } d_i \text{ 值} \Rightarrow \text{更弱的防守力}) \tag{7.3}$$

让我们在这些定义上稍作停留，以便让自己确信这些定义是合理的。事实上的确如此，理由如下。为了理解攻击评分，假设队伍 * 的防守很弱，因此 d_* 相对较大。如果队伍 j 在对阵队伍 * 时取得了很高的分数，则 a_{*j} 较大。但根据 o_j 的定义，这一较高的得分被大的 d_* 以 a_{*j}/d_* 的方式抵消掉了一部分。也就是说，在一个防守弱的对手上获取很高的分数，并不会对总的攻击评分有很大贡献。但如果队伍 * 的防守很强，则 d_* 相对较小，那么从 * 身上取得的高分将不会被小的 d_* 所抵消，而 a_{*j}/d_* 将明显提高队伍 j 的攻击评分。

对于防守评分，也有类似的分析。将攻击力强的对手的得分限制在一个低的水平，将使得队伍 i 的防守评分中的 a_{i*}/o_* 项减小，但让一个攻击力弱的对手得到很高分数将使得 a_{i*}/o_* 更大，从而造成队伍 i 更高的防守评分。

但先评定哪一个呢

由于 OD 评分是循环定义的，因此不可能在一组评分未知的情况下确定另一组评分。这意味着 OD 评分的计算只能通过逐步求精的方式来完成，其中一组评分初始时置任意值，然后通过逐步交替代换来更新式（7.2）和式（7.3），直至收敛。

交替求精过程

用正数初始化防守评分并保存为列向量 $\boldsymbol{d}_0$。这些初始评分可以任意指定——我们一般设为

$$\boldsymbol{d}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \boldsymbol{e}。$$

利用 $\boldsymbol{d}_0$，可以由式（7.2）计算攻击评分的第一个近似值，并将结果存在列向量 $\boldsymbol{o}_1$ 中。现在使用这些值，根据式（7.3）计算精化后的防守评分值，并将结果存放于列向量 $\boldsymbol{d}_1$ 中。重复这一交替代换过程将产生两个向量序列

$$\{\boldsymbol{d}_0,\boldsymbol{d}_1,\boldsymbol{d}_2,\boldsymbol{d}_3,\cdots\} \text{ 和} \{\boldsymbol{o}_0,\boldsymbol{o}_1,\boldsymbol{o}_2,\boldsymbol{o}_3,\cdots\} \tag{7.4}$$

如果这些序列收敛，则攻击评分和防守评分即为向量

$$\lim_{k\to\infty} \boldsymbol{o}_k = \boldsymbol{o} = \begin{pmatrix} o_0 \\ o_1 \\ \vdots \\ o_m \end{pmatrix} \text{和} \lim_{k\to\infty} \boldsymbol{d}_k = \boldsymbol{d} = \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_m \end{pmatrix}$$

中的元素值。为了更好地分析这一交替求精过程，我们将利用矩阵代数和一些不标准的记法来描述这一序列中的向量，这样将更为方便。给定一个各元素值均为正的列向量

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \text{定义} \boldsymbol{x}^{\div} = \begin{pmatrix} 1/x_1 \\ 1/x_2 \\ \vdots \\ 1/x_m \end{pmatrix} (\text{倒数向量})$$

如果 $\boldsymbol{A} = [a_{ij}]$ 为式（7.1）所定义的得分矩阵，则式（7.2）和式（7.3）中的 OD 评分可写为两个矩阵方程

$$\boldsymbol{o} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{d}^{\div} \tag{7.5}$$

和

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{o}^{\div} \tag{7.6}$$

进而对 $\boldsymbol{d}_0 = \boldsymbol{e}$，*OD* 序列，即式（7.4）中的向量由交替求精过程

$$\boldsymbol{o}_k = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{d}_{k-1}^{\div} \tag{7.7}$$

和

$$\boldsymbol{d}_k = \boldsymbol{A}\boldsymbol{o}_k^{\div} (k = 1,2,3,\cdots) \tag{7.8}$$

所定义。

分离

$\boldsymbol{o}_k$ 和 $\boldsymbol{d}_k$ 之间的相互依赖性可以被轻易打破，而计算过程也不必再在攻击评分和防守评分间换来换去。通过将式（7.8）简单地代入式（7.7）来产生如下两个独立的迭代，便能做到这一点。

$$\boldsymbol{o}_k = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{o}_{k-1}^{\div})^{\div} \quad (k = 1,2,3,\cdots) \tag{7.9}$$

$$\boldsymbol{d}_k = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{d}_{k-1}^{\div})^{\div} \quad (k = 1,2,3,\cdots) \tag{7.10}$$

式中，$\boldsymbol{o}_0 = \boldsymbol{e} = \boldsymbol{d}_0$。

但是，我们并不需要把这两个迭代都进行计算，因为如果能获得其中一个评分，则另一个评分可由式（7.5）或式（7.6）而得到。例如，如果攻击迭代式（7.9）收敛到了 $\boldsymbol{o}$，那么将它代入式（7.6）就能给出 $\boldsymbol{d}$；类似地，如果我们仅利用式（7.10）来进行迭代且收敛到 $\boldsymbol{d}$，则 $\boldsymbol{o}$ 可由式（7.5）获得。

结合两个 OD 评分

一旦对每支队伍或每名参与者都确定了一组攻击评分和防守评分，那么下一个问题就是如何将这两组评分结合起来，从而得到各队的一个单一的评分值？有许多方法可以用来结合或聚合评分值列表——实际上，第 14 章和第 15 章就是专门讨论这一问题的。但是对于 OD 评分，最简单（可能也是最合理）的做法，便是定义聚合评分或总评分，即对每个 i，有

$$r_i = \frac{o_i}{d_i} \tag{7.11}$$

按我们的向量记法，总评分向量可表示为 $\boldsymbol{r} = \boldsymbol{o}/\boldsymbol{d}$，其中，/表示逐元素进行除法。请记住，攻击力强的队伍的 o_i 值大，而防守力强的队伍的 d_i 值小，因此将攻击评分除以相应的防守评分是合理的，因为它使我们能够保持 $\boldsymbol{r}$ 中的总评分“越大越好”的这一解读方式。

再次回到范例

现在回到**第 5 页**上我们反复使用的 5 支队伍示例，将表 1.1 的得分放入一个得分矩阵中

$$\boldsymbol{A} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 52 & 24 & 38 & 45 \\ 7 & 0 & 16 & 17 & 7 \\ 21 & 34 & 0 & 5 & 30 \\ 7 & 25 & 7 & 0 & 52 \\ 0 & 27 & 3 & 14 & 0 \end{pmatrix} \end{array}$$

以 $\boldsymbol{d}_0 = \boldsymbol{e}$ 为初始值，并利用式（7.7）～式（7.8）的交替迭代方法，则需要经过 $k = 9$ 个循环才能使两个 OD 向量都收敛到 4 个有效数位。结果如下表所示。

攻击排名	队　伍	攻击评分	防守排名	队　伍	防守评分
1	Miami	151.6	1	VT	0.4104
2	VT	114.8	2	Miami	0.8029
3	UVA	82.05	3	UVA	0.9675
4	UNC	48.66	4	UNC	1.164
5	Duke	33.99	5	Duke	1.691

5 支队伍示例中的 OD 评分

由这些 OD 评分和排名可以明显看出，UVA，UNC 和 Duke 应该分别被排在总的第 3，4 和 5 位。但是，VT 和 Miami 的两项 OD 评分则正好换了一下位置，因此还不能断定它们在总排名中的位置。不过，利用式（7.11）中 OD 评分比可以解决这个问题。总的评分如下所示。

排　名	队　伍	评分（o_i/d_i）
1.	VT	279.8
2.	Miami	188.8
3.	UVA	84.8
4.	UNC	41.8
5.	Duke	20.1

5 支队伍示例的总排名　　(7.12)

对比得分和码数

正如在本章概述部分所提到的那样，竞赛中的几乎任何一种统计数据都可以用来取代得分。对于橄榄球，一个令人感兴趣的问题，就是看看当 OD 法中的得分被获得（或丢失）的码数所代替时，会发生什么事情。作为其查尔斯顿学院硕士学位论文的一部分，卢克·英格拉姆使用了 2005 年 ACC 橄榄球得分和码数数据对此问题进行了考察。我们反复使用的 5 支队伍示例中的得分数据便是卢克的实验中的一个子集。对于同样的 5 支队伍也使用与卢克的码数数据对应的子集，便可以得到*码数矩阵*

$$\boldsymbol{A}=\begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array}\begin{array}{c}\begin{array}{ccccc}\text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT}\end{array} \\ \begin{pmatrix}0 & 577 & 357 & 315 & 362\\ 100 & 0 & 188 & 452 & 167\\ 209 & 321 & 0 & 199 & 338\\ 207 & 500 & 270 & 0 & 552\\ 35 & 304 & 196 & 334 & 0\end{pmatrix}\end{array}$$

式中，

$$a_{ij}=\begin{cases}\text{队伍 } j \text{ 从 } i \text{ 处获取的码数}\\ i \text{ 丢失给 } j \text{ 的码数}\end{cases}$$

由 $\boldsymbol{d}_0=\boldsymbol{e}$ 开始，并利用式（7.7）~式（7.8）的交替迭代方法，则对于码数数据，需要经过 $k=6$ 个循环才能使两个 OD 向量都收敛到 4 个有效数位。结果如下所示。

攻击排名	队　伍	攻击评分	防守排名	队　伍	防守评分
1.	Miami	1608	1.	VT	0.6629
2.	UVA	1537	2.	Miami	0.7975
3.	VT	1249	3.	UNC	0.9886
4.	UNC	1024	4.	Duke	1.1900
5.	Duke	537.3	5.	UVA	1.4020

使用码数时 5 支队伍示例的 OD 评分

这一次，排名的定义就不如使用得分时那么显而易见了。由码数所得的总评分如下。

排　名	队　伍	评分（o_i/d_i）
1.	Miami	2016
2.	VT	1885
3.	UVA	1096
4.	UNC	1036
5.	Duke	451.6

使用码数时5支队伍示例的总评分

(7.13)

式（7.12）和式（7.13）中的结果相似，只不过Miami和VT换了一下位置。实验结果倾向于说明，在NCAA橄榄球比赛中，得分数据比码数数据能给出稍好一点的总体排名（根据最终的名次来判断）。也许这是由于获胜的总是得分最多的队，而不一定是总码数最多的队。但是，在赛季中期的特定时期内，基于码数的OD评分是反映总体实力的一个良好的指标。当然，在其他的体育项目或竞赛中，结果将有所不同。

NFL 2009—2010 赛季的 OD 评分

由于之前我们在本书其他部分的某些示例中使用了NFL 2009—2010赛季的数据，你也许想知道，对NFL 2009—2010赛季，它的OD评分是怎样的。我们使用了所有比赛（包括季后赛）的得分数据来计算攻击和防守评分，如果两支队伍交锋次数多于1次，则取平均得分。结果如下表所示。

排名	队　伍	攻击评分	排名	队　伍	攻击评分
1.	新奥尔良圣徒	560.99	17.	田纳西泰坦	281.58
2.	明尼苏达维京人	442.37	18.	丹佛野马	256.01
3.	圣迭戈闪电	393.23	19.	堪萨斯城酋长	255.60
4.	绿湾包装工	389.08	20.	卡罗来纳黑豹	249.47
5.	新英格兰爱国者	375.69	21.	辛辛那提猛虎	246.77
6.	印第安纳小马	369.65	22.	旧金山49人	238.40
7.	费城老鹰	362.99	23.	芝加哥熊	231.55
8.	休斯敦德州人	332.85	24.	杰克逊维尔美洲虎	224.24
9.	亚特兰大猎鹰	322.99	25.	华盛顿红皮	219.82
10.	亚利桑那红雀	322.05	26.	坦帕湾海盗	215.16
11.	巴尔的摩乌鸦	315.07	27.	布法罗比尔	214.38
12.	纽约巨人	306.60	28.	西雅图海鹰	210.83
13.	迈阿密海豚	304.32	29.	底特律雄狮	209.09
14.	匹兹堡钢人	304.24	30.	克里夫兰布朗	190.22
15.	纽约喷气机	298.27	31.	奥克兰突袭者	165.21
16.	达拉斯牛仔	297.71	32.	圣路易斯公羊	135.81

使用NFL 2009—2010赛季得分数据时的攻击评分

(7.14)

排名	队　伍	防守评分	排名	队　伍	防守评分
1.	纽约喷气机	0.64610	17.	绿湾包装工	1.0391
2.	巴尔的摩乌鸦	0.67875	18.	明尼苏达维京人	1.0526
3.	新英格兰爱国者	0.75685	19.	纽约巨人	1.0552
4.	旧金山49人	0.79343	20.	芝加哥熊	1.0560
5.	达拉斯牛仔	0.80562	21.	坦帕湾海盗	1.0769
6.	卡罗来纳黑豹	0.81649	22.	圣路易斯公羊	1.0815
7.	亚特兰大猎鹰	0.85158	23.	田纳西泰坦	1.0910
8.	辛辛那提猛虎	0.85385	24.	圣迭戈闪电	1.1005
9.	布法罗比尔	0.87673	25.	西雅图海鹰	1.1363
10.	丹佛野马	0.87783	26.	克里夫兰布朗	1.1395
11.	华盛顿红皮	0.92174	27.	亚利桑那红雀	1.1414
12.	费城老鹰	0.94134	28.	奥克兰突袭者	1.1474
13.	印第安纳小马	0.95219	29.	杰克逊维尔美洲虎	1.2207
14.	休斯敦德州人	0.95922	30.	堪萨斯城酋长	1.2661
15.	匹兹堡钢人	1.0113	31.	新奥尔良圣徒	1.2665
16.	迈阿密海豚	1.0143	32.	底特律雄狮	1.4548

(7.15)

使用 NFL 2009—2010 赛季得分数据时的防守评分

如果使用比率 $r_i = o_i/d_i$ 来聚合攻击和防守评分以得到每支队伍的单一总评分，则 NFL 2009—2010 赛季的结果如下。

排名	队　伍	$r_i = o_i/d_i$	排名	队　伍	$r_i = o_i/d_i$
1.	新英格兰爱国者	496.39	17.	丹佛野马	291.64
2.	巴尔的摩乌鸦	464.20	18.	纽约巨人	290.56
3.	纽约喷气机	461.64	19.	辛辛那提猛虎	289.01
4.	新奥尔良圣徒	442.95	20.	亚利桑那红雀	282.15
5.	明尼苏达维京人	420.25	21.	田纳西泰坦	258.09
6.	印第安纳小马	388.21	22.	布法罗比尔	244.52
7.	费城老鹰	385.61	23.	华盛顿红皮	238.48
8.	亚特兰大猎鹰	379.28	24.	芝加哥熊	219.26
9.	绿湾包装工	374.45	25.	堪萨斯城酋长	201.89
10.	达拉斯牛仔	369.54	26.	坦帕湾海盗	199.80
11.	圣迭戈闪电	357.32	27.	西雅图海鹰	185.53
12.	休斯敦德州人	347.00	28.	杰克逊维尔美洲虎	183.70
13.	卡罗来纳黑豹	305.53	29.	克里夫兰布朗	166.94
14.	匹兹堡钢人	300.85	30.	奥克兰突袭者	143.98
15.	旧金山49人	300.46	31.	底特律雄狮	143.73
16.	迈阿密海豚	300.03	32.	圣路易斯公羊	125.57

(7.16)

由攻击/防守比给出的 NFL 2009—2010 赛季总评分

然而，使用这个简单的比值并非唯一的（而且也不一定是最好的）聚合 OD 评分的方法。例如，这一方法给攻击和防守赋予了相同的重要性。如果希望实验一下某些更为复杂的策略，

那么你可以构建某种类型的回归模型，使得

$$\alpha(o_h - o_a) + \beta(d_h - d_a) = s_h - s_a$$

式中，o_h，d_h，o_a，d_a 分别为主队和客队的 OD 评分，而 s_h 和 s_a 为各自的得分。**警告！** 一旦你开始拿这样的点子做实验，你就会发现它们可是会成瘾的。实际上，它们就像流沙一样会把你吸进去，而你的人际关系和日常工作中的表现将可能因此而受到影响。

NFL 并非展示 OD 法优点的最佳例子，因为 NFL 极为均衡——远比其他大多数体育项目要表现得均衡。联盟花费了很大力气来保证这一点。因此，弱的进攻（防守）和强的进攻（防守）之间的差别根本没有 NCAA 比赛或其他体育项目中所表现出的程度那么大。这意味着哪怕仅使用原始的得分（而不用考虑比赛的水平高低）来评测 NFL 球队的攻击力和防守力，所得到的结果也不会太过离谱。但是在其他的体育项目或竞赛中，这一点可能就并非如此。例如，当我们使用给出了如上 NFL OD 评分的同样的得分矩阵 $\boldsymbol{A}$ 来计算

$$\text{每场比赛平均得分} = \frac{\text{总得分数}}{\text{比赛场数}}$$

和

$$\text{每场比赛平均失分} = \frac{\text{总失分数}}{\text{比赛场数}}$$

时，就得到了如下结果。

排名	队　伍	平均得分	排名	队　伍	平均得分
1.	新奥尔良圣徒	33.687	17.	纽约喷气机	21.107
2.	绿湾包装工	30.115	18.	旧金山 49 人	19.731
3.	明尼苏达维京人	29.200	19.	辛辛那提猛虎	19.615
4.	新英格兰爱国者	27.346	20.	丹佛野马	19.423
5.	费城老鹰	27.192	21.	卡罗来纳黑豹	19.038
6.	圣迭戈闪电	26.893	22.	芝加哥熊	19.000
7.	巴尔的摩乌鸦	25.808	23.	堪萨斯城酋长	18.615
8.	印第安纳小马	25.179	24.	杰克逊维尔美洲虎	18.042
9.	休斯敦德州人	24.577	25.	华盛顿红皮	17.538
10.	亚利桑那红雀	24.536	26.	西雅图海鹰	17.385
11.	匹兹堡钢人	24.269	27.	底特律雄狮	17.000
12.	纽约巨人	23.731	28.	坦帕湾海盗	16.958
13.	亚特兰大猎鹰	22.654	29.	克里夫兰布朗	16.654
14.	田纳西泰坦	22.375	30.	布法罗比尔	15.846
15.	达拉斯牛仔	21.893	31.	奥克兰突袭者	13.000
16.	迈阿密海豚	21.846	32.	圣路易斯公羊	11.692

NFL 2009—2010 赛季每场平均得分　　(7.17)

排名	队　伍	平均失分	排名	队　伍	平均失分
1.	纽约喷气机	15.346	17.	匹兹堡钢人	20.769
2.	巴尔的摩乌鸦	16.077	18.	费城老鹰	20.846
3.	达拉斯牛仔	17.357	19.	新奥尔良圣徒	21.250
4.	辛辛那提猛虎	18.808	20.	芝加哥熊	22.885
5.	明尼苏达维京人	19.133	21.	亚利桑那红雀	23.643
6.	印第安纳小马	19.250	22.	杰克逊维尔美洲虎	23.923
7.	旧金山 49 人	19.269	23.	克里夫兰布朗	24.154
8.	卡罗来纳黑豹	19.308	24.	迈阿密海豚	24.577
9.	休斯敦德州人	19.462	25.	奥克兰突袭者	24.692
10.	丹佛野马	19.731	26.	田纳西泰坦	24.808
11.	亚特兰大猎鹰	19.769	27.	圣路易斯公羊	25.192
12.	新英格兰爱国者	19.958	28.	坦帕湾海盗	25.423
13.	布法罗比尔	20.231	29.	堪萨斯城酋长	26.038
14.	华盛顿红皮	20.231	30.	西雅图海鹰	27.708
15.	绿湾包装工	20.346	31.	纽约巨人	28.542
16.	圣迭戈闪电	20.714	32.	底特律雄狮	30.346

NFL 2009—2010 赛季每场平均失分　　(7.18)

如果稍微花一点时间比较一下式（7.14）中的攻击评分和式（7.17）中的场均得分，就能发现，尽管它们并没有给出完全相同的排名，但相互之间总是多多少少有所吻合——这也理应如此。如果不是这样，那就肯定是有什么地方出错了，因为这两个排名都是由同样的数据所给出的。两者的区别在于式（7.14）中的攻击评分考虑了对抗的程度，而式（7.17）中的原始分数没有考虑这一点。类似地，式（7.15）中的防守评分也多多少少地与式（7.18）中的场均失分相吻合——这样便和谐了。

OD 法的数学分析

计算 OD 评分的算法确实非常简单——只需按**第 81 页**上所述的那样交替进行乘法，即式（7.7）和式（7.8）即可。在本书讨论的所有评分方法中，OD 法是实现最为容易者之一——一般的电子表格就可以很好地帮你完成计算，或者如果比赛中的竞争者不多的话，使用计算器甚至是纸笔就可以了。但是，实现的简单性却掩盖了围绕 OD 算法的理论上的某些深层数学问题。

主要的理论问题集中于 OD 评分的存在性。既然包含了 OD 评分的向量 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 被定义为极限

$$\boldsymbol{o} = \lim_{k\to\infty} \boldsymbol{o}_k \text{ 和 } \boldsymbol{d} = \lim_{k\to\infty} \boldsymbol{d}_k,$$

式中，

$$\boldsymbol{o}_k = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{d}_{k-1}^{\div}; \boldsymbol{d}_k = \boldsymbol{A}\, \boldsymbol{o}_k^{\div} \text{ 且 } \boldsymbol{d}_0 = \boldsymbol{e}。$$

因此确实可能出现 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 并不存在的情况——实际上，在很多情况下这些序列不能收敛。

我们需要一条定理来告诉我们，在什么情况下收敛能够得到保证，而在什么情况下则不可能收敛。在得到这样一条定理之前，我们需要若干术语、记法和背景知识。这其中的关键点在于认识到OD法与数据矩阵 $\boldsymbol{A}_{m\times m}$ 的逐次缩放或均衡化的过程密切相关。要理解这一点，假设首先对 $\boldsymbol{A}$ 的各列进行缩放（或均衡化），以使得各列元素之和为1，然后再将所得矩阵按行进行缩放来使得各行元素之和为1。这一过程可以用与对角缩放矩阵 $\boldsymbol{C}_1$ 和 $\boldsymbol{R}_1$ 的乘法来描述，其中 $\boldsymbol{AC}_1$ 使得所有列和为1，而随后的 $\boldsymbol{R}_1(\boldsymbol{AC}_1)$ 使得所有行和为1。换而言之，若

$$\boldsymbol{C}_1=\begin{pmatrix}\xi_1 & 0 & \cdots & 0\\ 0 & \xi_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \xi_n\end{pmatrix},\text{其中},\xi_i=\frac{1}{[\boldsymbol{e}^{\mathrm{T}}\boldsymbol{A}]_i}(\text{第}\,i\,\text{列元素之和的倒数})$$

且

$$\boldsymbol{R}_1=\begin{pmatrix}\rho_1 & 0 & \cdots & 0\\ 0 & \rho_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & \rho_n\end{pmatrix},\text{其中},\rho_i=\frac{1}{[(\boldsymbol{AC}_1)\boldsymbol{e}]_i}(\text{第}\,i\,\text{行元素之和的倒数})$$

则

$$\boldsymbol{R}_1\boldsymbol{AC}_1\boldsymbol{e}=\boldsymbol{e}(\text{即所有行和等于}1)$$

不幸的是，对 $\boldsymbol{A}\,\boldsymbol{C}_1$ 的各行进行缩放却破坏了前一步将 $\boldsymbol{A}$ 与 $\boldsymbol{C}_1$ 相乘所得到的列与列之间的均衡——即 $\boldsymbol{R}_1\boldsymbol{A}\,\boldsymbol{C}_1$ 的列和通常不等于1。$\boldsymbol{R}_1\boldsymbol{A}\,\boldsymbol{C}_1$ 的列和通常比 $\boldsymbol{A}$ 的列和更接近于1。之后，同样的交替缩放被再次执行，然后再次执行，然后又再次执行，直至数据得到了很好的均衡，即所有的行和与列和都等于1（或至少在某个给定的容许限下足够接近于1）。逐次列缩放以及之后跟着的行缩放，产生了一个对角缩放矩阵 $\boldsymbol{C}_k$ 和 $\boldsymbol{R}_k$ 的序列，满足

$$\boldsymbol{S}_k=\boldsymbol{R}_k\cdots\boldsymbol{R}_2\,\boldsymbol{R}_1\boldsymbol{A}\,\boldsymbol{C}_1\,\boldsymbol{C}_2\cdots\boldsymbol{C}_k\geqslant 0,\text{其中},\boldsymbol{S}_k\boldsymbol{e}=\boldsymbol{e}(\text{所有行和为}1)$$

所有元素非负且行和均为1的方阵 $\boldsymbol{S}$ 称为随机矩阵。当 $\boldsymbol{S}$ 的所有行和与列和都为1时，$\boldsymbol{S}$ 称为双重随机矩阵。

因此，对每个 $k=1,2,3,\cdots$，由逐次行列缩放得到的矩阵 $\boldsymbol{S}_k=\boldsymbol{R}_k\cdots\boldsymbol{R}_2\,\boldsymbol{R}_1\boldsymbol{A}\,\boldsymbol{C}_1\,\boldsymbol{C}_2\cdots\boldsymbol{C}_k$ 是一个随机矩阵。现在我们希望 $\boldsymbol{S}=\lim\limits_{k\to\infty}\boldsymbol{S}_k$ 存在且 $\boldsymbol{S}$ 为双重随机的。可惜的是，实际中可并非总是如此。不过我们并非一无所获，因为有一条定理告诉了我们，究竟何时会发生这一情况。要理解这条定理，需要一些关于方阵中的“对角线”的知识。

对角线

方阵 $\boldsymbol{A}$ 的主对角线为元素集合 $\{a_{11},a_{22},\cdots,a_{mm}\}$，这一点大多数人都已熟知。不过，还存在 $\boldsymbol{A}$ 中的其他“对角线”，他们就没有那么受重视了。一般而言，如果 $\sigma=\{\sigma_1,\sigma_2,\cdots,\sigma_m\}$

是整数 $(1,2,\cdots,m)$ 的一个排列，则 σ 对应的 $\boldsymbol{A}$ 的对角线定义为集合

$$\{a_{1\sigma_1},a_{2\sigma_2},\cdots,a_{m\sigma_m}\}$$

例如，在 $\boldsymbol{A}=\begin{pmatrix}1&2&3\\4&5&6\\7&8&9\end{pmatrix}$ 中，主对角线 $\{1,5,9\}$ 对应于自然序 $(1,2,3)$，而这一对角线以及由6个排列所给出的所有其他对角线如下所示。

排　列	对　角　线
(1,2,3)	$\{a_{11},a_{22},a_{33}\}$
(1,3,2)	$\{a_{11},a_{23},a_{32}\}$
(2,1,3)	$\{a_{12},a_{21},a_{33}\}$
(2,3,1)	$\{a_{12},a_{23},a_{31}\}$
(3,1,2)	$\{a_{13},a_{21},a_{32}\}$
(3,2,1)	$\{a_{13},a_{22},a_{31}\}$

辛克霍恩-诺普

1967年，理查德·辛克霍恩（Richard Sinkhorn）和保罗·诺普（Paul Knopp）[70]证明了一条定理，它可以被完美地应用于OD理论。他们的定理如下。

辛克霍恩-诺普定理

令 $\boldsymbol{A}_{m\times m}$ 为一个非负矩阵。

- 当且仅当至少存在 $\boldsymbol{A}$ 的一条严格为正的对角线时，交替对 $\boldsymbol{A}$ 的列和行进行缩放的迭代过程将收敛到一个双重随机极限矩阵 $\boldsymbol{S}$，此时称 $\boldsymbol{A}$ 具有*支撑*。
  这一命题保证了缩放过程能够收敛到一个双重随机矩阵，但它不能保证乘积 $\mathcal{C}_k=\mathbf{C}_1\mathbf{C}_2\cdots\mathbf{C}_k$ 和 $\mathcal{R}_k=\mathbf{R}_k\cdots\mathbf{R}_2\mathbf{R}_1$ 收敛——例如，对 $\boldsymbol{A}=\begin{pmatrix}1&1\\0&1\end{pmatrix}$ 进行逐次缩放将收敛到 $\boldsymbol{I}$（恒等矩阵），但 $\mathcal{C}_k$ 和 $\mathcal{R}_k$ 不收敛。
- 当且仅当 $\boldsymbol{A}$ 中的每个正元素都位于一条正对角线上时，对角阵的极限

$$\lim_{k\to\infty}\mathcal{C}_k=\boldsymbol{C},\lim_{k\to\infty}\mathcal{R}_k=\boldsymbol{R}\tag{7.19}$$

  存在且 $\boldsymbol{RAC}=\boldsymbol{S}$ 为双重随机的，此时称 $\boldsymbol{A}$ 具有*完全支撑*。

OD矩阵

如**第83页**上所讨论的，除了比分，其他统计数据（如橄榄球中的码数）也可用于计算

OD 评分。在有关 OD 法介绍的其余展开部分中，我们将假设 $\boldsymbol{A}$ 是一个具有非负统计数据（比分、码数等）的矩阵，而这样的矩阵称为 OD 矩阵。所得的主要结论如下：将辛克霍恩-诺普过程用于 OD 矩阵时，将给出 OD 评分的倒数。如下定理给出了这一结论。

OD 评分和辛克霍恩-诺普

在下面的讨论中，出于方便考虑，我们通过如下的映射，来用列向量 $\boldsymbol{x}$ 表示相对应的对角阵

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \Leftrightarrow \operatorname{diag}\{\boldsymbol{x}\} = \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_m \end{pmatrix}$$

注意，$\operatorname{diag}\{\boldsymbol{x}\}\boldsymbol{e} = \boldsymbol{x} = \boldsymbol{e}^{\mathrm{T}}\operatorname{diag}\{\boldsymbol{x}\}$。此外，对于所有元素非负的向量 $\boldsymbol{v}$ 以及所有对角元素非负的对角阵 $\boldsymbol{D}$，可以直接验证

$$\operatorname{diag}\{(\boldsymbol{D}\boldsymbol{v})^{\div}\} = \boldsymbol{D}^{-1}\operatorname{diag}\{v^{\div}\}, \operatorname{diag}\{(\boldsymbol{D}\boldsymbol{v}^{\div})^{\div}\} = \boldsymbol{D}^{-1}\operatorname{diag}\{\boldsymbol{v}\} \tag{7.20}$$

现在我们可以证明主要的 OD 定理了，它告诉我们 OD 评分向量何时存在，何时不存在。

OD 定理

对于 OD 矩阵 $\boldsymbol{A}_{m\times m}$，当且仅当 $\boldsymbol{A}$ 中的每个正元素都位于某条正对角线上时——即当且仅当 $\boldsymbol{A}$ 具有完全支撑时，OD 序列

$$\boldsymbol{o}_k = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{d}_{k-1}^{\div}, \boldsymbol{d}_k = \boldsymbol{A}\boldsymbol{o}_k^{\div} \quad (\text{取}\mathbf{d}_0 = \mathbf{e}) \tag{7.21}$$

分别收敛至 OD 的评分向量 $\boldsymbol{o}$ 和 $\boldsymbol{d}$。

证明：如在先前的讨论中那样，令 $\mathcal{C}_k$ 和 $\mathcal{R}_k$ 分别表示辛克霍恩-诺普过程所产生的对角缩放矩阵的乘积

$$\mathcal{C}_k = \boldsymbol{C}_1\boldsymbol{C}_2\cdots\boldsymbol{C}_k, \mathcal{R}_k = \boldsymbol{R}_k\cdots\boldsymbol{R}_2\boldsymbol{R}_1$$

现在的关键在于建立这些乘积与式（7.21）中 *OD* 序列之间的关系。要做到这一点，回忆一下辛克霍恩-诺普过程中的对角缩放矩阵是用如下方法产生的。

$$\begin{aligned} \boldsymbol{C}_1 &= \operatorname{diag}\{(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e})^{\div}\} & \boldsymbol{R}_1 &= \operatorname{diag}\{(\boldsymbol{A}\boldsymbol{C}_1\boldsymbol{e})^{\div}\} \\ \boldsymbol{C}_2 &= \operatorname{diag}\{((\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1)^{\mathrm{T}}\boldsymbol{e})^{\div}\} & \boldsymbol{R}_2 &= \operatorname{diag}\{(\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\boldsymbol{C}_2\boldsymbol{e})^{\div}\} \\ \boldsymbol{C}_3 &= \operatorname{diag}\{((\boldsymbol{R}_2\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\boldsymbol{C}_2)^{\mathrm{T}}\boldsymbol{e})^{\div}\} & \boldsymbol{R}_3 &= \operatorname{diag}\{(\boldsymbol{R}_2\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\boldsymbol{C}_2\boldsymbol{C}_3\boldsymbol{e})^{\div}\} \\ &\vdots & &\vdots \end{aligned} \tag{7.22}$$

注意到乘积 $\mathcal{C}_k$ 和 $\mathcal{R}_k$ 可以被表示为

$$\mathcal{C}_k = \operatorname{diag}\{\boldsymbol{c}_k\}, \mathcal{R}_k = \operatorname{diag}\{\boldsymbol{r}_k\} \quad (k = 1,2,3,\cdots) \tag{7.23}$$

式中，$\boldsymbol{r}_0 = \boldsymbol{e}$，且

$$\boldsymbol{c}_k = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}_{k-1})^{\div}, \boldsymbol{r}_k = (\boldsymbol{A}\boldsymbol{c}_k)^{\div} \quad (k = 1,2,3,\cdots) \tag{7.24}$$

通过写出最初的几次迭代便可以验证上述结果。例如，$\boldsymbol{c}_1 = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e})^{\div}$，因此由式（7.22）可知 $\boldsymbol{C}_1 = \mathrm{diag}\{\boldsymbol{c}_1\}$，且 $\boldsymbol{C}_1\boldsymbol{e} = \boldsymbol{c}_1$，因此

$$\boldsymbol{R}_1 = \mathrm{diag}\{(\boldsymbol{A}\boldsymbol{C}_1\boldsymbol{e})^{\div}\} = \mathrm{diag}\{(\boldsymbol{A}\boldsymbol{c}_1)^{\div}\} = \mathrm{diag}\{\boldsymbol{r}_1\}$$

利用这一结果以及式（7.20）可得

$$\begin{aligned}\boldsymbol{C}_2 &= \mathrm{diag}\{((\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1)^{\mathrm{T}}\boldsymbol{e})^{\div}\} = \mathrm{diag}\{(\boldsymbol{C}_1\boldsymbol{A}^{\mathrm{T}}\boldsymbol{R}_1\boldsymbol{e})^{\div}\} = \mathrm{diag}\{(\boldsymbol{C}_1\boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}_1)^{\div}\} \\ &= \mathrm{diag}\{(\boldsymbol{C}_1\boldsymbol{c}_2^{\div})^{\div}\} = \boldsymbol{C}_1^{-1}\mathrm{diag}\{\boldsymbol{c}_2\}\end{aligned}$$

因此 $\mathcal{C}_2 = \boldsymbol{C}_1\boldsymbol{C}_2 = \mathrm{diag}\{\boldsymbol{c}_2\}$。类似地，由这个结果以及式（7.20）可以给出

$$\begin{aligned}\boldsymbol{R}_2 &= \mathrm{diag}\{(\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{C}_1\boldsymbol{C}_2\boldsymbol{e})^{\div}\} = \mathrm{diag}\{(\boldsymbol{R}_1\boldsymbol{A}\boldsymbol{c}_2)^{\div}\} \\ &= \mathrm{diag}\{(\boldsymbol{R}_1\boldsymbol{r}_2^{\div})^{\div}\} = \boldsymbol{R}_1^{-1}\mathrm{diag}\{\boldsymbol{r}_2\}\end{aligned}$$

因此 $\mathcal{R}_2 = \boldsymbol{R}_2\boldsymbol{R}_1 = \mathrm{diag}\{\boldsymbol{r}_2\}$。推导这一过程便得到了式（7.23）。现在注意到式（7.24）中的序列不过就是式（7.21）中 OD 序列的倒数，换而言之，

$$\boldsymbol{c}_k = \boldsymbol{o}_k^{\div}, \boldsymbol{r}_k = \boldsymbol{d}_k^{\div} \tag{7.25}$$

再次地，通过写出如下最初几项，并推导下去，便可以容易地发现上述结论。

$$\begin{aligned}&\boldsymbol{c}_1 = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{e})^{\div},\ \boldsymbol{o}_1 = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{e}^{\div} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{e} \Rightarrow \boldsymbol{o}_1 = \boldsymbol{c}_1^{\div} \\ &\boldsymbol{r}_1 = (\boldsymbol{A}\boldsymbol{c}_1)^{\div},\ \boldsymbol{d}_1 = \boldsymbol{A}\boldsymbol{o}_1^{\div} = \boldsymbol{A}\boldsymbol{c}_1 \Rightarrow \boldsymbol{d}_1 = \boldsymbol{r}_1^{\div} \\ &\boldsymbol{c}_2 = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}_1)^{\div}, \boldsymbol{o}_2 = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{d}_1^{\div} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{r}_1 \Rightarrow \boldsymbol{o}_2 = \boldsymbol{c}_2^{\div} \\ &\boldsymbol{r}_2 = (\boldsymbol{A}\boldsymbol{c}_2)^{\div},\ \boldsymbol{d}_2 = \boldsymbol{A}\boldsymbol{o}_2^{\div} = \boldsymbol{A}\boldsymbol{c}_2 \Rightarrow \boldsymbol{d}_2 = \boldsymbol{r}_2^{\div} \\ &\quad\vdots \qquad\qquad \vdots \qquad\qquad \vdots\end{aligned}$$

第 89 页上的辛克霍恩-诺普定理指出，当且仅当 $\boldsymbol{A}$ 具有完全支撑时，极限

$$\lim_{k\to\infty}\mathcal{C}_k = \boldsymbol{C}, \lim_{k\to\infty}\mathcal{R}_k = \boldsymbol{R}$$

存在。因此，式（7.23）意味着当且仅当 $\boldsymbol{A}$ 具有完全支撑时，极限

$$\lim_{k\to\infty}\boldsymbol{c}_k = \lim_{k\to\infty}\mathcal{C}_k\boldsymbol{e} = \boldsymbol{C}\boldsymbol{e}, \lim_{k\to\infty}\boldsymbol{r}_k = \lim_{k\to\infty}\mathcal{R}_k\boldsymbol{e} = \boldsymbol{R}\boldsymbol{e}$$

存在，而由此，当且仅当 $\boldsymbol{A}$ 具有完全支撑时，式（7.25）保证了极限

$$\lim_{k\to\infty}\boldsymbol{o}_k = \lim_{k\to\infty}\boldsymbol{c}_k^{\div} = \boldsymbol{C}^{-1}\boldsymbol{e} \text{ 和} \lim_{k\to\infty}\boldsymbol{d}_k = \lim_{k\to\infty}\boldsymbol{r}_k^{\div} = \boldsymbol{R}^{-1}\boldsymbol{e}$$

的存在。∎

作一点弊

并非所有 OD 矩阵都能具有完全支撑的性质——即并不是所有正元素都位于某条正对角线上。如果没有完全支撑，那么 OD 评分向量就不存在，因为定义这些向量的序列将不能

收敛。所以自然的做法就是稍微作一点弊，以迫使序列得以收敛。一个相当明显的作弊方法就是简单地给 $\boldsymbol{A}$ 中的各个元素加上一个小数 $\varepsilon > 0$，即进行替换

$$\boldsymbol{A} \leftarrow \boldsymbol{A} + \varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}} > \boldsymbol{0}$$

这将迫使所有元素均为正，从而也自然具有了完全支撑性，OD 序列的收敛性也就得到了保证。在 $\boldsymbol{A}$ 的主对角线元素上加上些值，实际上就意味着队伍自己击败了自己，如果这一点使你感到困扰的话，可以进行如下的替换

$$\boldsymbol{A} \leftarrow \boldsymbol{A} + \varepsilon(\boldsymbol{e}\boldsymbol{e}^{\mathrm{T}} - \boldsymbol{I})$$

它同样能够保证完全支撑性。

如果 ε 相比于原始 OD 矩阵中的非零元素值而言为小，则原始数据中的核心信息就不会被过度损害，而由这个稍加修改的矩阵给出的 OD 评分，仍然能够不错地反映出攻击和防守的相对实力。但是，这个世界上没有免费的午餐，因此这个小骗术是有代价的。

尽管相对的攻击和防守实力并没有被小的扰动过度扭曲，但收敛率将可能变得非常慢。直观而言，当 ε 变得更小时，产生可行的结果所需要的迭代次数将变得更大。换而言之，你不得不在所希望达到的 OD 评分质量以及为获得这一评分花费的时间之间走钢丝。但是请记住，要产生有用的评分，可能并不需要很多有效位数的精度。

如式（7.25）所展示的那样，OD 序列不过（或完全）就是式（7.24）中辛克霍恩-诺普序列的倒数，因此 OD 序列的收敛率和辛克霍恩-诺普过程的收敛率相同。尽管对于辛克霍恩-诺普的收敛率已有很好的估计，但这个估计却并非直截了当。某些估计方法需要知道过程的最终结果，否则将仅能计算出一个先验估计，而其他的估计都仅为“序估计”，其中的序常数未知或无法计算。对收敛特性的详细讨论会偏离我们的主题太远，对于 OD 法的实现也没有多少帮助，因此在此将不进一步深入。对收敛性的技术性讨论请见参考文献［30，46，72，73］。有关 OD 法的更多内容和应用可在戈万的著作[34, 35]中可以找到。

杂谈：HITS


J. Kleinberg

我们的 OD 法是乔恩·克莱因伯格（Jon Kleinberg）称为 HITS（超文本归纳主题搜索）[45]算法的一个非线性情形下的类似算法。HITS 算法被设计用来对链接在一起的文档进行排名。HITS 是对网页这一最为常见的链接文档类型进行排名的一种关键技术。HITS 的思想已经被结合在了搜索引擎 Ask.com 之中。与许多搜索引擎模型类似，HITS 将万维网视为一个极其巨大的有向图。正如在本书中多处可以看到的那样，对抗性竞赛中的交互关系同样也可以被视为一幅很大的图。因此，可以对某些网页排名算法加以修改，以便用来产生排名（反之亦然）。关于网页排名的 HITS 算法的详细讨论可见参考文献［45］和［49］，而在此我们给出其基本思路的总结。在图 7.1 中以结点表示的每个网页都被赋予了一个“枢纽评分”和一个“权威评

分”。枢纽评分试图反映一个网页的出链的质量和数量，而权威评分则衡量了入链的质量和数量。直观的思路是，枢纽页面在网络中所扮演的，是为用户提供众多优秀链接的核心角色——例如，雅虎就有许多好的枢纽页面；权威页面则通常包含了很多信息，并且有许多网址都指向它们——例如，维基百科就有许多好的权威页面。一个页面的枢纽评分依赖于它所指向的页面的权威评分，而反过来，一个页面的权威评分则依赖于指向它的页面的枢纽评分。克莱因伯格称之为“相互强化”评分系统，因为一个评分依赖于另一个评分——你是否注意到了这与 OD 的相似性？为了计算枢纽评分和权威评分，我们从一个邻接方阵 $\boldsymbol{L}$ 开始，其中

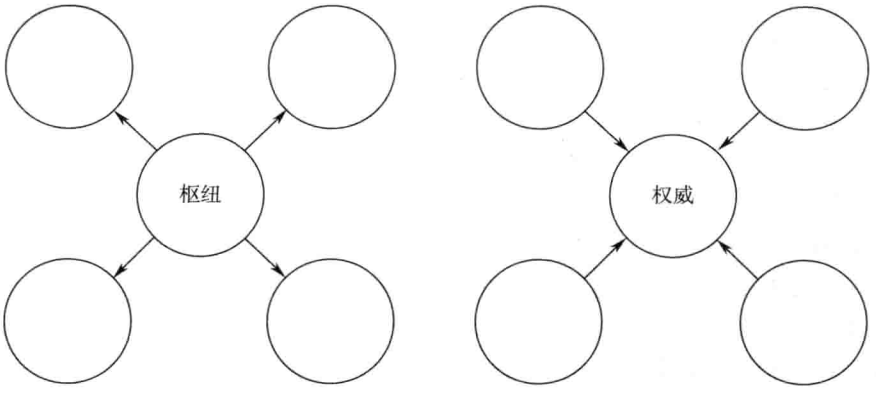


图 7.1　枢纽页面与权威页面

$$\boldsymbol{L}_{ij} = \begin{cases} 1, \text{若存在一条由结点 } i \text{ 至结点 } j \text{ 的边} \\ 0, \text{其他} \end{cases}$$

任意选择一个初始权威评分向量（通常使用均匀向量 $\boldsymbol{a}_0 = \boldsymbol{e}$），然后逐步通过

$$\boldsymbol{a}_{k+1} = \boldsymbol{L}^{\mathrm{T}} \boldsymbol{h}_k, \boldsymbol{h}_k = \boldsymbol{L}\,\boldsymbol{a}_k \tag{7.26}$$

对枢纽评分和权威评分进行精化。这些方程清楚显示了权威评分和枢纽评分之间的相互依赖性。利用**第 81 页**上所描述的完全相同的“分离”策略，可以打破这一相互依赖性，即将 $\boldsymbol{h}_k$ 的表达式代入定义 $\boldsymbol{a}_{k+1}$ 的方程，得到

$$\boldsymbol{a}_{k+1} = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\,\boldsymbol{a}_k, \boldsymbol{h}_{k+1} = \boldsymbol{L}\,\boldsymbol{L}^{\mathrm{T}}\,\boldsymbol{h}_k。$$

如果这两个序列收敛，则它们的极限就给出了所需的枢纽评分和权威评分。对于 HITS 的实现、编程和扩展等细节方面感兴趣的读者，请参见参考文献［49］中的第 11 章。

杂谈：OD 法的诞生


A. Govan

当我们在撰写那本关于 PageRank 算法的“谷歌之书”[49]的时候，我们常常在考虑将网页评分和排名的思路应用于体育项目的可能性。不管怎样，搜索引擎排名方法本身也不过是其他学科（如文献统计学）中一些基本原理的改进，因此，除了对网页进行排名，同样的这些思路为什么不能应用于体育领域之中呢？当然，的确可以。**第 67 页**上第 6 章中的马尔可夫法就是 PageRank 思想的直接应用。但对于其他网页排名方法而言又如何呢？2005 年的一个下午，我们漫不经心地谈论着这些想法，而这时克莱因伯格的 HITS 算法就冒了出来。如上所述，HITS 是一个优雅简单的想法，尽管它的扩展性不大好，但对于小的集合则仍然是有效的——甚至可能好过 PageRank。将克莱因伯格的“枢纽-权威”概念重新解读为“进攻-防守”的框架看来十分自然，但很快我们就搞清楚，直接将枢纽-权威转译为进攻-防守并不可行。不过，这一想法太过完美，因此我们无法舍弃，所以我们的车轮继续向前转动。没用多久我们就认识到，在式（7.26）中所述的枢纽和权威通过线性关系联系在一起，但进攻与防守之间的关系则本质上是非线性的，必须通过某种类型的倒数关系来加以描述。在这之后，思路便变得顺畅，而 OD 法也就此诞生。

之后不久，安吉拉·戈万作为一名研究生来到了北卡罗来纳州立大学，并开始研究数据分析中的马尔可夫方法，但在闲谈的过程中，她对评分和排名发生了兴趣。尽管有人劝她说，由橄榄球所促成的毕业论文并不可靠，但她仍然决定遵从自己的本意，并建立了 OD 法和辛克霍恩-诺普理论之间的联系，正如**第 87 页**上开始的讨论所描述的那样，这一联系提供了分析 OD 法数学性质的理论基础。此外，在提出和分析本书中若干马尔可夫方法以及其他几种未曾提及的方法等方面，她也发挥了作用。而其熟练的编程能力也为我们带来了重要而富有创新性的网络搜集软件，它能够自动地由不同的网站收集数据。仅此一项就已经是一个重要的成果，因为如果没有使用她开发的工具的话，之后所进行的不计其数的数值实验将变得更为困难。

杂谈：体育竞赛中的个人奖项

在赛季结束时，许多体育项目都会为球员颁发个人奖项。例如，大学橄榄球比赛会选出一名备受推崇的海斯曼奖获奖者。大学与职业篮球赛则会选出全进攻阵容和全防守阵容。这些选择通常是通过收集与综合来自教练、体育专栏作家以及有时甚至是球迷的投票来完成的。利用如本书中所给出的那些科学排名方法，可以很容易地为选择获奖者提供补充信息。不过，要为某个特定奖项选择一个最为合适的排名方法，这一工作可并非无足轻重。要对一个奖项选择一个合适的方法，需要缜密的思考和巧妙的数学建模。例如，第 6 章中的马尔可夫模型是选择全防守阵容的一个不错的选择，而对于年度最佳球员奖，本章的 OD 法可能是最为合适的，因为 OD 法同时结合了进攻和防守能力。有许多参数可以用来对模型进行调整和定制，这也成为了未来研究的一个令人激动的方向。

杂谈：更多的影片排名：当 OD 遇上 Netflix

在**第 25 页**的杂谈《影视排名——科利和梅西与 Netflix 相遇》中，我们介绍了 Netflix 奖。Netflix 赞助了 100 万美元的奖金来改进其当前的推荐系统。该奖金于 2009 年 9 月被 BellKor 研究小组斩获。在之前关于 Netflix 的杂谈中，我们叙述了科利法和梅西法是如何用于对影片进行排名的。而在现在的这个杂谈中，我们将说明如何修改 OD 法以达到同样的目的。实际上，由于此时的矩阵并非方阵，并且 OD 法将给出对偶的评分向量，因此我们将得到一个影片评分向量 $\boldsymbol{m}$ 和一个用户评分向量 $\boldsymbol{u}$。下面我们将只给出影片评分向量中排名前 24 位的影片。我们对最终的用户评分向量以及相应的用户排名没有多少兴趣，但对于诸如 Netflix 这样的公司而言，$\boldsymbol{u}$ 当然很有意义，因为它们将围绕这些用户信息来提出营销策略。Netflix OD 法基于如下的观点，即如果影片 i 在好的（即有鉴赏力的）用户那儿得到了高分，那么它就是部好影片，并应该获得高的评分 m_i；类似地，如果用户 j 给出的评分与影片的“真实”评分相吻合，那他或她就是名好用户，并应该得到高的评分 u_j。将这些语言描述并翻译为数学语言，就给出了如下的迭代过程，该过程用类似于 MATLAB 程序中的语言风格编写。有多种方法可以对“有鉴赏力的用户”进行定量化（例如，可参见 Netflix 获奖者的某些方法），但一个基本的方法，则是衡量用户的评分与真实评分之间相差多少。

```
u=e;
for i=1: maxiter
    m=Au;
    m=5 (m-min (m)) / (max (m) - min (m));
    u=1/ ( (R - (R>0) .* (em')) .^2) e
end
```

将这个算法用于 Netflix 数据子集，特别地，用于“超级用户”[⊖]——对 1000 部甚至更多影片给出了评分的用户——我们就得到了如下所列的排名前 24 位的影片。该列表与第 28 页上由科利法和梅西法所给出的清单有着明显差异。

排名	影　片	排名	影　片
1.	《指环王 III：王者归来》	13.	《星球大战 VI：杰迪归来》
2.	《指环王 II：双塔奇兵》	14.	《海底总动员》
3.	《肖申克的救赎》	15.	《第六感》
4.	《指环王 I：魔戒现身》	16.	《沉默的羔羊》
5.	《夺宝奇兵》	17.	《百万美元宝贝》
6.	《星球大战 V：帝国反击战》	18.	《超人总动员》
7.	《教父》	19.	《卢旺达饭店》
8.	《星球大战 IV：新希望》	20.	《夺宝奇兵 III》
9.	《迷失：第一季》	21.	《勇敢的心》
10.	《辛德勒的名单》	22.	《拯救大兵瑞恩》
11.	《非常嫌疑犯》	23.	《黑道家族：第一季》
12.	《兄弟连》	24.	《教父 II》

利用 Netflix OD 法的排名前 24 位的影片

⊖ 我们颇为吃惊地了解到在 Netflix 数据集中的超级用户是如此之多。有超过 13000 名用户对 1000 部甚至更多的影片进行了评分。

数字——

$ 3 142 000 000 = 2010 年内华达州在体育博彩上所下赌注的总和。

——The Line Makers

第 8 章　基于重新排序的排名方法

本书到目前为止所介绍的评分和排名方法的原理如图 8.1 所示。我们从与待排名对象有关的输入数据开始，然后执行算法来得到一个评分向量，进而再给出排名向量。我们关注的焦点是从图 8.1 中的左端到达右端，即将输入数据变换为排名向量。但是，如果我们反方向进行，并将焦点集中在最终产物即排名向量本身的话，那么就能够产生出一些新的思路。在本章中我们将介绍这些新思路的其中两种。这两种新的排名方法是按如图 8.2 所示的原理来进行的，也就是说，它们完全避开了评分向量。

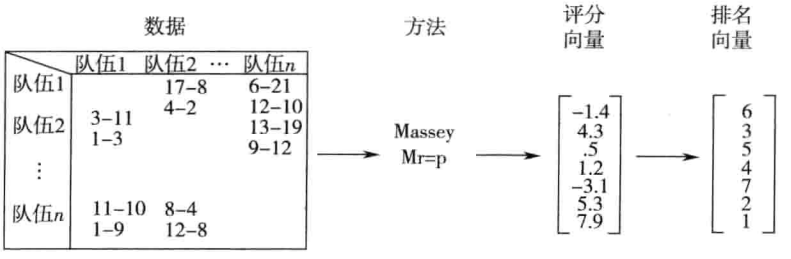


图 8.1　第 4 ~ 7 章排名方法的原理

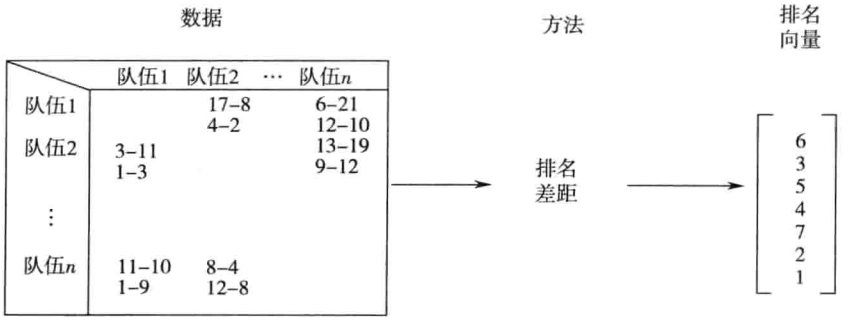


图 8.2　本章排名方法的原理

排名差距

排名向量最明显的一个特点，就在于它是一个排列，而这个即将变得清晰起来的特点正是本章的关键词。每个长度为 n 的排名向量都是整数 1 到 n 的一个排列，它为每个事物赋予了一个排位。例如，对于之前的 5 支队伍的示例，梅西排名向量为

$$\begin{matrix}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}5\\1\\4\\3\\2\end{pmatrix},$$

它为杜克大学赋予了排名位置 5，迈阿密大学为排名位置 1，等等。一个排名向量就产生了队伍的一个线性排序，不过它也可以用于考察成对比较关系。利用矩阵可以对成对比较关系进行最佳的描述。因此，以上的梅西排名向量就给出了如下的成对关系矩阵。我们称这一矩阵为 $\boldsymbol{R}$，因为它包含了排名差距的信息。

$$\begin{matrix}&\text{Duke}&\text{Miami}&\text{UNC}&\text{UVA}&\text{VT}\\ \text{Duke}\\ \text{Miami}\\ \boldsymbol{R}=\text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}0&0&0&0&0\\4&0&3&2&1\\1&0&0&0&0\\2&0&1&0&0\\3&0&2&1&0\end{pmatrix}。$$

在 $\boldsymbol{R}$ 的（UVA，Duke）这个元素位置上的 2，表明在梅西排名向量中，UVA 的排名比 Duke 要靠前 2 位。因此，该矩阵中的元素代表了对应的排名优势。

每个长度为 n 的排名向量都能产生一个 $n \times n$ 的排名差距矩阵 $\boldsymbol{R}$，它是如下*基本排名差距矩阵* $\hat{\boldsymbol{R}}$ 的一个对称重排（即对行与列以相同的方式进行排列）。

$$\hat{\boldsymbol{R}}_{n\times n}=\begin{matrix}\\1\\2\\\vdots\\n-1\\n\end{matrix}\begin{matrix}\begin{matrix}1&2&3&\cdots&n\end{matrix}\\\begin{pmatrix}0&1&2&\cdots&n-1\\&0&1&\cdots&n-2\\&&\ddots&\ddots&\vdots\\&&&\ddots&1\\&&&&0\end{pmatrix}\end{matrix}。$$

注意 $\hat{\boldsymbol{R}}$ 中那漂亮的整数上三角结构，这是因为 $\hat{\boldsymbol{R}}$ 是由相应的基本排名向量

$$\hat{r}=\begin{matrix}\text{第 }1\\\text{第 }2\\\text{第 }3\\\vdots\\\text{第 }n\end{matrix}\begin{pmatrix}1\\2\\3\\\vdots\\n\end{pmatrix}$$

所构建而得，其中 n 个对象按升序排列。总之，在排名向量和排名差距矩阵之间，存在着一个一一映射[⊖]。也就是说，每个排名向量 $\boldsymbol{r}$ 都生成了一个排名差距矩阵 $\boldsymbol{R}$，它能够被表示为

⊖ 如果在排名向量中没有相同排名出现的话，这一映射就是一一对应的。请见第 11 章中关于平手的完整讨论。

$\boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\hat{\boldsymbol{R}}\boldsymbol{Q}$，其中，$\boldsymbol{Q}$ 为一个排列矩阵，而反之亦然。

由于基本排名差距矩阵是一个队伍-队伍（或更一般来说，事物-事物）的矩阵，因此一个自然的做法，是把它和其他同样大小同样类型的数据矩阵进行比较。例如，我们可以将 $\hat{\boldsymbol{R}}$ 和第 6 章中的马尔可夫投票矩阵 $\boldsymbol{V}_{\text{pointdiff}}$ 进行比较。后者中包含了两支队伍之间交手的累积分差。得分差距和排名差距一样，也包含了两支队伍相对实力差距的信息。但是，马尔可夫投票矩阵中的元素对应于劣势。为了将 $\boldsymbol{V}_{\text{pointdiff}}$ 与优势关联起来，我们可以简单地比较 $\boldsymbol{R}$ 和 $\boldsymbol{V}_{\text{pointdiff}}$ 的转置。为清晰起见，我们将讨论一个一般性的数据差矩阵 $\boldsymbol{D}$。在大多数实验中，我们将使用 $\boldsymbol{D} = \boldsymbol{V}^{\mathrm{T}}_{\text{pointdiff}}$。

现在我们已经准备就绪，可以将排名问题打造成为一个重排或最近邻矩阵问题了，也就是说，给定一个数据矩阵 $\boldsymbol{D}$，其中包含了各支队伍以往成对交手的信息，要求找到 $\boldsymbol{D}$ 的一个对称重排，使得重排后的数据差矩阵和基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 之间的距离最小化。用严格的数学语言来描述，即是给定 $\boldsymbol{D}_{n\times n}$，希望找到一个排列矩阵 $\boldsymbol{Q}$，使得 $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q} - \hat{\boldsymbol{R}}\|$ 最小化。排列矩阵 $\boldsymbol{Q}$ 对恒等矩阵的各行进行了一次 1 到 n 之间整数的排列，因此，这一定义在排列空间的最优化问题受到如下约束。

$$\begin{aligned} &\min_{\boldsymbol{Q}} \|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q} - \hat{\boldsymbol{R}}\| \\ &\text{满足}\quad \boldsymbol{Q}\boldsymbol{e} = \boldsymbol{e} \\ &\qquad\quad \boldsymbol{e}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{e}^{\mathrm{T}} \\ &\qquad\quad q_{ij} \in \{0,1\} \end{aligned} \tag{8.1}$$

范例

我们使用得分差矩阵的转置作为数据差矩阵 $\boldsymbol{D}$。因此，

$$\boldsymbol{D} = \boldsymbol{V}^{\mathrm{T}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 45 & 0 & 18 & 8 & 20 \\ 3 & 0 & 0 & 2 & 0 \\ 31 & 0 & 0 & 0 & 0 \\ 45 & 0 & 27 & 38 & 0 \end{pmatrix} \end{array}。$$

对于 5 支队伍的示例，基本排名差距矩阵为

$$\hat{\boldsymbol{R}} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}。$$

我们的目标是找到整数 1 到 5 的排列，以及这 5 支队伍的相应排列，使得基本矩阵 $\hat{\boldsymbol{R}}$ 和重排后的 $\boldsymbol{D}$ 之间的误差最小化。为了处理两个矩阵的数据在尺度上的不同，我们首先将每个

矩阵都除以各自元素的总和以实现归一化。归一化后的矩阵分别为

$$\boldsymbol{D}=\begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 0.19 & 0 & 0.08 & 0.03 & 0.08\\ 0.01 & 0 & 0 & 0.01 & 0\\ 0.13 & 0 & 0 & 0 & 0\\ 0.19 & 0 & 0.11 & 0.16 & 0\end{pmatrix},\hat{\boldsymbol{R}}=\begin{pmatrix}0 & 0.05 & 0.10 & 0.15 & 0.20\\ 0 & 0 & 0.05 & 0.10 & 0.15\\ 0 & 0 & 0 & 0.05 & 0.10\\ 0 & 0 & 0 & 0 & 0.05\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}。$$

尽管 $n=5$ 的值很小，但在对 $\boldsymbol{D}$ 进行重排时，仍然有 $5!=120$ 种不同的排列需要考虑。通过强力法（即穷举），我们发现排列（5　2　4　3　1）是最优的，它给出了如下的排列矩阵 $\boldsymbol{Q}$ 和重排后的数据差矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}$。

$$\boldsymbol{Q}=\begin{pmatrix}0 & 0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 0\end{pmatrix},\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}=\begin{pmatrix}0 & 0 & 0.16 & 0.11 & 0.19\\ 0.08 & 0 & 0.13 & 0.01 & 0.16\\ 0 & 0 & 0 & 0 & 0.13\\ 0 & 0 & 0.01 & 0 & 0.01\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}。$$

重排矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}$ 与基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 最为接近。图 8.3 利用 MATLAB 的城市区图（cityplot）展示了重排后的归一化 $\boldsymbol{D}$ 和归一化 $\hat{\boldsymbol{R}}$ 矩阵之间不同侧面的对比。

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}=\begin{pmatrix}0 & 0 & 0.16 & 0.11 & 0.19\\ 0.08 & 0 & 0.13 & 0.01 & 0.16\\ 0 & 0 & 0 & 0 & 0.13\\ 0 & 0 & 0.01 & 0 & 0.01\\ 0 & 0 & 0 & 0 & 0\end{pmatrix},\hat{\boldsymbol{R}}=\begin{pmatrix}0 & 0.05 & 0.10 & 0.15 & 0.20\\ 0 & 0 & 0.05 & 0.10 & 0.15\\ 0 & 0 & 0 & 0.05 & 0.10\\ 0 & 0 & 0 & 0 & 0.05\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}$$

重排后的$\boldsymbol{D}$　　$\hat{\boldsymbol{R}}$

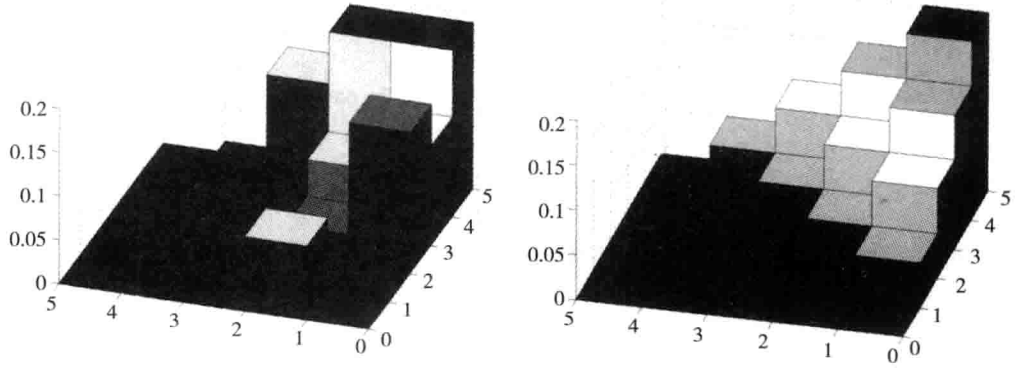


图 8.3　5 支队伍示例中重排后的 $\boldsymbol{D}$（即 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}$）和 $\hat{\boldsymbol{R}}$ 的城市区图

将这 5 支队伍按原来的排序（1　2　3　4　5）重排为最佳的排序（5　2　4　3　1），可以给出一个数据差矩阵，它与基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 的阶梯形状最为接近。

总之，排名差距方法与之前讨论的排名方法不同。具体来说，就是该方法是对给定的数

据矩阵，去搜索一个最接近的排名差距矩阵。既然每个排名差距矩阵都对应了一个排名，因此它为我们的弹药库中又增加了一种排名方法。

Q * bert 矩阵

现在是论功行赏的时候了。这一次的功劳归于大卫·哥特利勃（David Gottlieb），他是非常流行的、名为 Q * bert 的 1982 年街机游戏的发明人。图 8.4 展现了基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 的城市区图和 Q * bert 游戏界面之间清楚无误的相似性。为了纪念哥特利勃的游戏，我们有时又用一个更短且更容易记住的名称——Q-bert 矩阵——来指代基本排名差距矩阵 $\hat{\boldsymbol{R}}$。

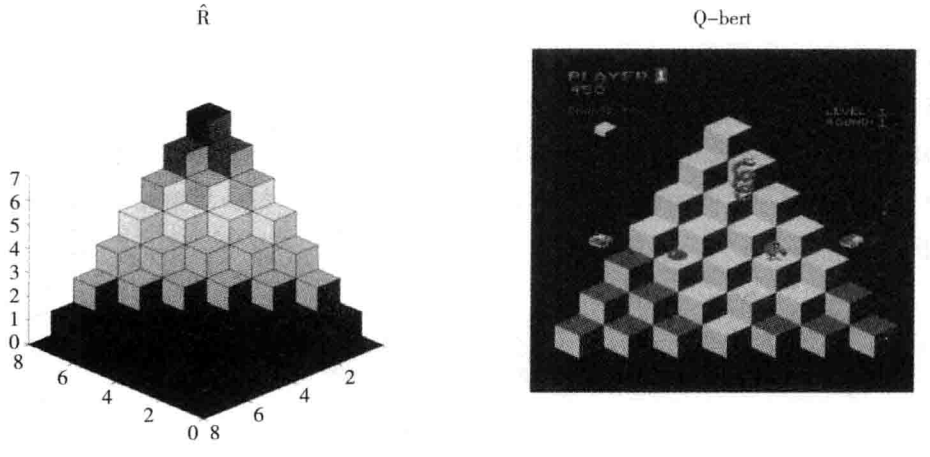


图 8.4　基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 和 Q * bert 游戏界面

求解最优化问题

让我们考察一下式（8.1）中的最优化问题。现在的目标是找到 $\boldsymbol{D}$ 的一个对称排列，使得重排后的数据矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}$ 和基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 之间的差别最小化。对于 n 个事物，有 $n!$ 种可能的排列方式需要考虑。当然，这意味着我们在前一节中所用的穷举法这样仅靠蛮力的途径是不可行的。关键之处在于：对于微型的 5 支队伍的例子，要找到最优排列，我们需要检查 5! = 120 种排列；如果将队伍数量增加一倍，那么需要考虑的排列数就变为 10! = 3 628 800。幸运的是，一个巧妙的分析告诉我们，可以不必考虑所有这 $n!$ 种排列。

该最优化问题的如下矩阵表达说明它是一个二值整数非线性规划问题。

$$\min_{Q} \|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q} - \hat{\boldsymbol{R}}\|$$

$$\text{满足} \quad \boldsymbol{Q}\boldsymbol{e} = \boldsymbol{e}, \boldsymbol{e}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{e}^{\mathrm{T}}, q_{ij} \in \{0,1\}$$

目标函数中的矩阵范数会影响到问题中非线性的类型。例如，当范数为弗罗贝尼乌斯型时，这一非线性是二次型的。尽管二次整数规划的求解颇有挑战性，但我们可以利用弗罗贝尼乌斯范数和 Q-bert 矩阵 $\hat{\boldsymbol{R}}$ 的某些性质，来显著减少计算量。例如，通过如下的代数计算可得

$$\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q} - \hat{\boldsymbol{R}}\|_F^2 = \mathrm{trace}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}) - 2\mathrm{trace}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{R}}) + \mathrm{trace}(\hat{\boldsymbol{R}}^{\mathrm{T}}\hat{\boldsymbol{R}})。$$

由于 $\mathrm{trace}(\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D})$ 和 $\mathrm{trace}(\hat{\boldsymbol{R}}^{\mathrm{T}}\hat{\boldsymbol{R}})$ 为常数，这就意味着原始的最小化问题等价于如下的最大化问题。

$$\max_{Q} \mathrm{trace}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{R}}) \tag{8.2}$$

$$\text{满足} \quad \boldsymbol{Q}\boldsymbol{e} = \boldsymbol{e}, \boldsymbol{e}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{e}^{\mathrm{T}}, q_{ij} \in \{0,1\}$$

目标函数的改变带来了计算上的极大好处。所有最优化算法都需要反复计算目标函数，而目标函数的计算通常是迭代算法中计算量最大的步骤。幸运的是，由于已知 $\hat{\boldsymbol{R}}$ 的结构，因此式（8.2）的这一等价表示特别高效。而更大的一个好处则是我们无需显式地构造出 $\hat{\boldsymbol{R}}$。实际上，如下的伪代码表明，我们可以如何非常高效地实现弗罗贝尼乌斯范数计算的编程。

高效计算 $\mathrm{trace}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{R}})$ 的伪代码

```
function [f] = trQtDQR (D, q);
输入：D = 大小为 n × n 的数据矩阵
      q = 大小为 n × 1 的排列向量
输出：f = trace(Q^T D Q R̂)
n = size (D, 1);
[sortedq, index] = sort (q);
clear sortedq;
f = 0;
for i = 2: n
    for j = 1: i -1
        f = f + (i -j) * D (index (i), index (j));
    end
end
```

松弛问题

Q-bert 问题的弗罗贝尼乌斯描述给它贴上了两个标签，即二次型和整数规划。在这两者中，整数规划这个标签更具有挑战性。通常情况下，整数规划问题是利用松弛技术来近似求解的，它将繁琐的整数约束放宽为连续约束。松弛的 Q-bert 问题由如下二次规划给出。

$$\max_{\boldsymbol{Q}} \operatorname{trace}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}\hat{\boldsymbol{R}}) \quad \text{满足} \quad \boldsymbol{Q}\boldsymbol{e} = \boldsymbol{e}, \boldsymbol{e}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{e}^{\mathrm{T}}, 0 \leqslant q_{ij} \leqslant 1 \tag{8.3}$$

如此一来，$\boldsymbol{Q}$ 就不再是一个排列矩阵，而是一个双重随机矩阵（即所有行和与列和都为 1 的矩阵）。因此，我们现在不再是在一个排列空间中进行最优化，而是在一组双重随机矩阵上进行最优化。松弛问题的最优解有时会取到整数值，此时它也是原二次整数规划问题的最优解。在这种情况下，松弛约束集的顶点就是原始约束集中的点。但不幸的是，目标函数的二次型本性使得最优解可能出现于可行域的内部，这时松弛问题的最优解对原始问题而言便可能不可行。一般情况下，松弛问题的非整数解可以被圆整到最为接近的整数解，从而给出原始问题的一个可行解。但同样不幸的是，没有什么直截了当的圆整方法，可以保证圆整后的解对于原问题而言是近最优的。实际上，你得到了另一个二次整数规划问题，即找到与所得的双重随机矩阵最为接近的一个排列矩阵。总之，对于我们的问题而言，松弛方法令人失望。

进化方法

虽然对二次整数规划问题进行松弛的方法没什么前途，但幸运的是，另一种最优化方法却前景光明。与其试图找到一个排列矩阵 $\boldsymbol{Q}$，倒不如等价地去尝试寻找一个排列向量 $\boldsymbol{q}$，它可以用于数据差矩阵 $\boldsymbol{D}$ 的重排。这一重排试图找到整数 1 到 n 的一个排列，使得 $\boldsymbol{D}(\boldsymbol{q}, \boldsymbol{q})$ 与 $\hat{R}$ 最为接近，即

$$\min_{\boldsymbol{q} \in p_n} \|\boldsymbol{D}(\boldsymbol{q}, \boldsymbol{q}) - \hat{\boldsymbol{R}}\|$$

式中，$\boldsymbol{q} \in p_n$ 意味着 $\boldsymbol{q}$ 必须属于所有 $n \times 1$ 长度的排列向量所构成的集合。例如，排列向量 $\boldsymbol{q} = (1 \quad 3 \quad 5 \quad 4 \quad 2)^{\mathrm{T}}$ 将

$$\boldsymbol{D} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.19 & 0 & 0.08 & 0.03 & 0.08 \\ 0.01 & 0 & 0 & 0.01 & 0 \\ 0.13 & 0 & 0 & 0 & 0 \\ 0.19 & 0 & 0.11 & 0.16 & 0 \end{pmatrix} \end{array}$$

对称重排[⊖]为

$$\boldsymbol{D}(\boldsymbol{q}, \boldsymbol{q}) = \begin{array}{c} \\ 1 \\ 3 \\ 5 \\ 4 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 3 & 5 & 4 & 2 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.01 & 0 & 0 & 0.01 & 0 \\ 0.19 & 0.11 & 0 & 0.16 & 0 \\ 0.13 & 0 & 0 & 0 & 0 \\ 0.19 & 0.08 & 0.08 & 0.03 & 0 \end{pmatrix} \end{array}$$

⊖ 对称重排是指使用相同的排列来对行的下标和列的下标进行重排。

注意，$\boldsymbol{D}$ 的元素在 $\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q})$ 中只是简单地被重新放置了。

矩阵重排已经得到了深入研究，并且在从线性系统的预调理到聚类算法的预处理等领域有着悠久的应用历史[9, 66, 7]。将问题考虑为一个矩阵重排问题并没有缩小搜索空间，而是仍有 $n!$ 种可能的排列能够对 $\boldsymbol{D}$ 进行重排，但与排列有关的最优化问题通常能够利用一种称为进化优化的方法加以解决。

正如进化优化这一名词所暗示的那样，这种方法借鉴了自然界的进化途径。方法的思路是首先从 p 个可能的解所构成的种群开始，在我们考虑的问题中，就是若干 1 到 n 的整数的排列。对种群中的每个成员，评价其适应度，此时解 $\boldsymbol{x}$ 的适应度即为 $\|\boldsymbol{D}(\boldsymbol{x},\boldsymbol{x}) - \hat{\boldsymbol{R}}\|$。种群中适应度最高的成员将交配产生后代，后代便包含了父代的最佳特性。继续这一进化的比拟，适应度不那么高的成员则通过无性的方式发生突变，而适应度最低的成员会被淘汰，并被其他外来群体所取代。新一代的 p 个排列的种群的适应度又被加以评价，而过程就如此继续下去。随着迭代的进行，我们将惊讶地观察到达尔文适者生存的原理。种群将朝着更加进化的、适应度更高的排列集前进。而更令人吃惊的，则是有定理证明进化算法在许多情况下可以收敛到最优解，而在一定的条件下，也能收敛到近最优解[55]。不幸的是，进化算法一般收敛较慢，这也说明了它通常被应用于非常具有挑战性或几乎无法求解的问题的原因。例如，由于能够求解规模越来越大的旅行商问题，所以进化算法变得声名显赫，而这一问题是众所周知的难以求解的问题。**第 104 页**上的杂谈对该问题进行了介绍。

如下的图 8.5 是进化算法的主要思路——特别是用于排名问题时的思路——的图形化表示。对这个算法的很棒的概括应归功于查尔斯顿学院的研究生凯思琳·佩丁斯。凯思琳在几次交流中都使用了这幅图，同时它还是一篇张贴论文[18]的组成部分。她是一位充满活力又很有趣的发言者，读者们无法亲身聆听凯思琳描述她的作品真的要算是相当令人遗憾的。作为一个微不足道的替代，我们尝试着在她的原图中加入文本标注，来模仿凯思琳对图 8.5 的讲解。读者只需按序数跟踪这些标注即可。

杂谈：旅行商问题

旅行商问题（TSP，Traveling Salesman Problem）是一个路径规划问题，一名推销商需要沿此路径造访 n 座城市，每座城市到达一次。该问题的某些变体要求推销商的出发点和终点为同一座城市，一般就是他的家乡。问题的目标是找到开销最小的路径，而开销可以通过不同方式加以定义，例如，总的旅行时间或旅行距离等。每条可能的路径都可以被视为这 n 座城市的一个排列。TSP 的更高级的变体中还加入了约束，例如，推销商在一天之内不能旅行超过 500 英里或接连访问同一个州内的两座城市，等等。尽管问题的表述简洁而优美，但 TSP 却是众所周知的难以解决的问题（实际上，对于熟悉下面这个术语的读者而言，该问题是 NP 困难问题[⊖]），这是由于所谓的状态空间爆炸所造成的。在其标准的无约束形式中，TSP 有 $n!/2$ 条需要考察的可能路径。即使 n 的增长十分缓慢，函数 $n!$ 也会爆炸式地迅速增加，参见表 8.1。在计算机出现之前，人们实现了启发式算法来获得近似最优解。在那时，一个 $n = 10$ 座城市的 TSP 问题就被认为是规模很大了。有了计算机后，对TSP的研究和应用迎来了飞跃。在最近一个时期内，进化

⊖ NP 困难问题，其中 NP 是指非确定性多项式（Non-deterministic Polynomial）。——编辑注

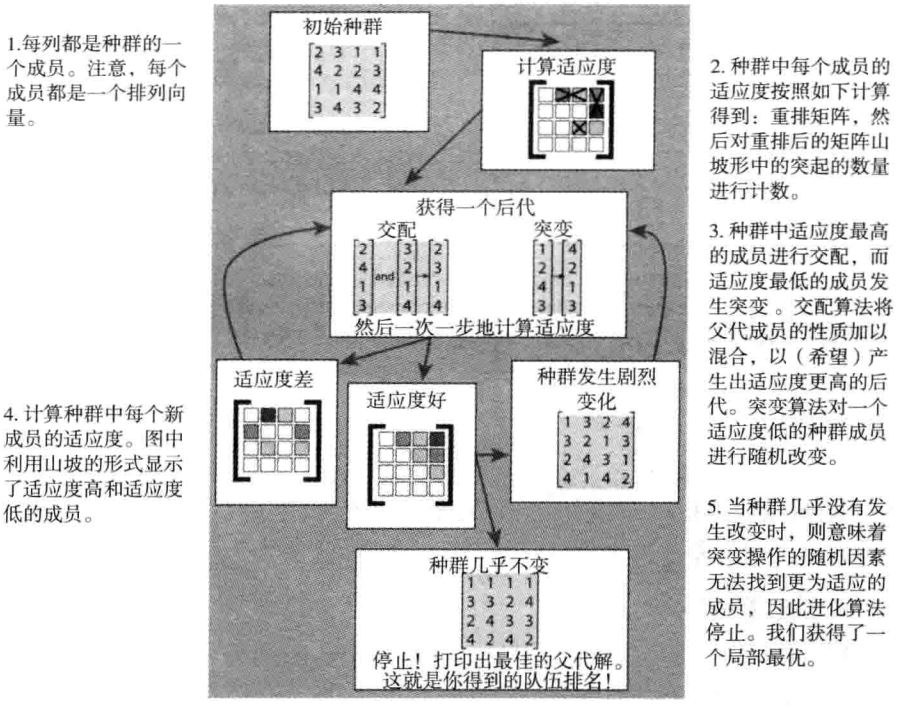


图 8.5　排名问题的进化算法步骤概述

算法引导了 TSP 研究的前进方向。现在可以求解的 TSP 问题的规模已经达到了数万座城市的数量级。实际上，在 2006 年解出了一个有着 $n = 85\,900$ 座城市的 TSP。可求解 TSP 问题规模的增长，使得它的应用也更为广泛。在今天，航空公司利用 TSP 来安排航线，而计算机硬件公司则利用 TSP 来确定电路板上铜线的布线路径。

高级排名差距模型

本书至今所介绍的大多数排名方法，都允许在产生排名的过程中考虑多种统计数据，排名差距法也同样如此。k 种统计数据可以被归入一个单独的数据矩阵 $\boldsymbol{D} = \alpha_1 \boldsymbol{D}_1 + \alpha_2 \boldsymbol{D}_2 + \cdots + \alpha_k \boldsymbol{D}_k$ 之中，其中为每个不同的矩阵 $\boldsymbol{D}_i$ 赋予了一个标量权值 α_i。之后，对相同的最优化问题 $\min_{\boldsymbol{q} \in p_n} \|\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q}) - \hat{\boldsymbol{R}}\|$ 进行求解。或者可以在最优化过程中分别处理这 k 个数据差矩阵如下

$$\min_{\boldsymbol{q} \in p_n} \alpha_1 \|\boldsymbol{D}_1(\boldsymbol{q},\boldsymbol{q}) - \hat{\boldsymbol{R}}\| + \alpha_2 \|\boldsymbol{D}_2(\boldsymbol{q},\boldsymbol{q}) - \hat{\boldsymbol{R}}\| + \cdots + \alpha_k \|\boldsymbol{D}_k(\boldsymbol{q},\boldsymbol{q}) - \hat{\boldsymbol{R}}\|。$$

我们仍然建议使用进化算法来求解这一最优化问题。注意，第二种问题形式的计算量要大得多，因为在每次迭代中计算种群中每个成员的适应度时，都需要进行 k 次矩阵范数的计算。

表 8.1　$n!$ 函数的爆炸式增长

n	$n!$
5	$= 120$
10	$\approx 3.6 \times 10^6$
25	$\approx 1.6 \times 10^{25}$
50	$\approx 3.0 \times 10^{64}$
100	$\approx 9.3 \times 10^{157}$

排名差距法的总结

记　法

$\boldsymbol{D}$　对象-对象数据矩阵，其中包含了对象的成对比较关系，如得分、码数或其他统计数据之间的差值。

$\hat{\boldsymbol{R}}$　基本排名差距矩阵

$\boldsymbol{Q}$　排列矩阵

$\boldsymbol{q}$　对应于 $\boldsymbol{Q}$ 的排列向量

$\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q})$　重排后的数据差矩阵 $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}$

算　法

1. 求解最优化问题

$$\min_{\boldsymbol{q}\in p_n} \|\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q}) - \hat{\boldsymbol{R}}\|$$

式中，p_n 为所有可能的 $n \times 1$ 排列向量的集合。我们建议使用进化算法求解。

2. 利用上述问题的最优解 $\boldsymbol{q}$ 来获得排名向量如下：将 $\boldsymbol{q}$ 按升序排序，并将排序后的下标值保存为排名向量。

排名差距法的性质

- 迄今所讨论的所有排名方法都同时给出了一个评分向量和一个排名向量。排名差距法是第一个仅给出排名向量的方法。
- 本方法寻找数据差矩阵 $\boldsymbol{D}$ 的一个重排，该重排与基本排名差距矩阵 $\hat{\boldsymbol{R}}$ 最为接近。尽管接近度可以用许多方式来加以衡量，但在**第 102 页**上已经说明，使用弗罗贝尼乌斯量度将特别高效。
- 由于这一最优化问题非常具有挑战性（实际上，它是 NP 困难问题），所以排名差距法比本书中的其他排名方法更慢，计算量也更大。因而，它可能仅限于在较

小的对象集上来使用，例如，对体育队伍所进行的排名还在这个方法力所能及的范围之内，除非发现了更加巧妙的实现算法，否则该方法对整个万维网上的网页进行排名将力不从心。

- 任何数据矩阵都可以作为排名差距矩阵的输入。在本章中，我们推荐使用马尔可夫法中的得分差矩阵，不过任意差值数据都可以使用。

杂谈：排名差距法和图的同构性

排名差距法与图和子图的同构性问题[⊖]紧密相关。此时，$\boldsymbol{D}$ 和 $\hat{\boldsymbol{R}}$ 表示两幅不同的图的（可能是经过加权的）邻接矩阵。现在的目的是确定是否可以将一幅图中的结点重新标号，从而给出另一幅图。如果可以，则 $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}-\hat{\boldsymbol{R}}\|=0$，而这两幅图称为同构的；如果不能，则最小化 $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{Q}-\hat{\boldsymbol{R}}\|$ 以获得 $\boldsymbol{D}$ 中结点的一个重标号，使得 $\boldsymbol{D}$ 变得与 $\hat{\boldsymbol{R}}$ 最为接近。我们的问题形式在两个方面上与这一同构性问题存在差别。首先，我们的问题更具一般性，即 $\boldsymbol{D}$ 和 $\hat{\boldsymbol{R}}$ 不必表示邻接矩阵；其次，我们的问题在另一方面又是更为特定化的，因为 $\hat{\boldsymbol{R}}$ 具有一个非常特别的结构。图的同构性在从化学成分识别到电子电路板设计等多个不同应用中都得到了成功运用。我们在本书中所取得的进展也可以实践于这些应用之中，以此凸显了我们的问题和图的同构性问题之间的这一联系的价值所在。

评分差距

第 4 章 ~ 第 7 章中的每一个排名方法都能给出一个评分向量，该向量随后被排序以给出排名向量。前一节中的排名差距法是第一种完全绕开评分向量而仅给出排名向量的方法。尽管在名字中用到了评分一词，但本节中介绍的评分差距法也同样跳过了评分向量而直接得到排名向量。尽管如此，我们将看到，方法名称中的评分一词是恰如其分的。

考虑 5 支队伍示例中的梅西评分向量 $\boldsymbol{r}$

$$\boldsymbol{r}=\begin{array}{c}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{array}\begin{pmatrix}-24.8\\ 18.2\\ -8.0\\ -3.4\\ 18.0\end{pmatrix},$$

它给出了如下的评分差距矩阵

$$\boldsymbol{R}=\begin{array}{c}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{array}\begin{array}{c}\begin{array}{ccccc}\text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT}\end{array}\\ \begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 43 & 0 & 26.2 & 21.6 & 0.2\\ 32.8 & 0 & 0 & 0 & 0\\ 21.4 & 0 & 4.6 & 0 & 0\\ 42.8 & 0 & 26 & 21.4 & 0\end{pmatrix}\end{array}。$$

⊖ 感谢奇姆埃尔·弗里德兰（Schmuel Friedland）指出这一联系。

我们已经直截了当地为这两个差距矩阵使用了同样的记号，这应当不致引起混淆。矩阵中的元素就指明了它是哪一种差距矩阵。排名差距矩阵仅包含整数值，而评分差距矩阵包含分数值。与排名差距中具有一个单一的基本排名差距矩阵的情况不同的是，并不存在一个单一的基本评分差距矩阵。不过却存在着一个基本结构。考虑一个已经排序的评分向量，例如

$$\boldsymbol{r}=\begin{matrix}1\\2\\3\\4\\5\end{matrix}\begin{pmatrix}0.9\\0.7\\0.3\\0.1\\-0.2\end{pmatrix}$$

给出了评分差距矩阵

$$\boldsymbol{R}=\begin{matrix}\\1\\2\\3\\4\\5\end{matrix}\begin{matrix}\begin{matrix}1&2&3&4&5\end{matrix}\\\begin{pmatrix}0&0.2&0.6&0.8&1.1\\0&0&0.4&0.6&0.9\\0&0&0&0.2&0.5\\0&0&0&0&0.3\\0&0&0&0&0\end{pmatrix}\end{matrix}$$

它具有一个非常有特点的结构。每个评分差距矩阵 $\boldsymbol{R}$ 都可以被重排以具有这一结构。

称一个评分差距矩阵 $\boldsymbol{R}$ 具有基本形（fundamental form），如果

$\forall i \geqslant j$，有 $r_{ij}=0$（严格上三角阵）

$\forall i, \forall j \leqslant k$，有 $r_{ij} \leqslant r_{ik}$（行中元素升序排列）

$\forall j, \forall i \leqslant k$，有 $r_{ij} \geqslant r_{kj}$（列中元素降序排列）

$\boldsymbol{R}(\boldsymbol{q},\boldsymbol{q})$ 的城市区图也形成了一个山坡形状，见图 8.6。因此，我们有时称这一基本形为

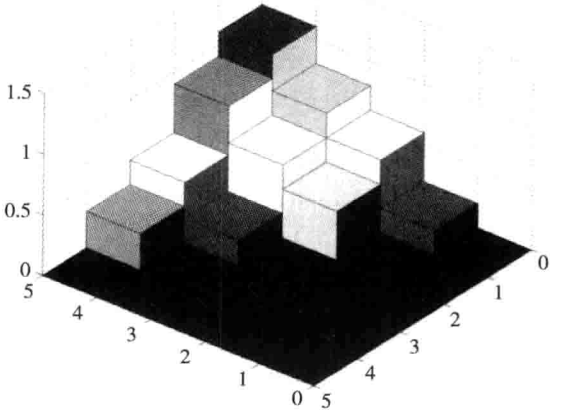


图 8.6　具有基本形或坡形的评分差距矩阵 $\boldsymbol{R}$ 的城市区图

坡形[⊖]。每个评分向量都将给出一个评分差距矩阵，它可以被重排为基本形。用于给出这一重排的对象的序，是 1 到 n 之间整数的一个排列，而这一排列正是排名向量。

机敏的读者现在可能会问：以上思路难道不是通过假设已经获得了一个评分向量来产生排名向量的吗？这种做法有什么价值呢？如果我们已经有了一个评分向量，那当然我们可以得到一个排名向量。下一节中的例子将说明，由这一看似逆向的思维所给出的定义，将如何带来一种新的、实用的排名方法，它不需要首先获得评分向量。

范例

评分差距法以一个队伍-队伍数据矩阵作为输入。出于在排名差距一节中所提到的相同原因，此时使用一个数据差矩阵 $\boldsymbol{D}$ 将更为合理。在这个例子中，我们将使用基于分差的马尔可夫投票矩阵 $\boldsymbol{V}$ 的转置 $\boldsymbol{V}^{\mathrm{T}}$，因为这样可以将投票数视为优势而非弱势的一个指标，而这是合理的。评分差距法的目的是找到 $\boldsymbol{D}$ 的一个重排，使之具有**第 108 页**上文本框中所给出的基本形，或尽可能地接近该形状。对于这个 5 支队伍的示例，

$$\boldsymbol{D}=\boldsymbol{V}^{\mathrm{T}}=\begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array}\begin{array}{c}\begin{array}{ccccc}\text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT}\end{array}\\ \begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 45 & 0 & 18 & 8 & 20\\ 3 & 0 & 0 & 2 & 0\\ 31 & 0 & 0 & 0 & 0\\ 45 & 0 & 27 & 38 & 0\end{pmatrix}\end{array}$$

现在有这 5 支队伍的 5！种排列可以对称地将矩阵重排。通过穷举我们发现，排列

$$\boldsymbol{q}=\begin{pmatrix}2\\5\\3\\4\\1\end{pmatrix}\Rightarrow\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q})=\begin{array}{c}\begin{array}{ccccc}2 & 5 & 3 & 4 & 1\end{array}\\ \begin{array}{c}2\\5\\3\\4\\1\end{array}\begin{pmatrix}0 & 20 & 18 & 8 & 45\\ 0 & 0 & 27 & 38 & 45\\ 0 & 0 & 0 & 2 & 3\\ 0 & 0 & 0 & 0 & 31\\ 0 & 0 & 0 & 0 & 0\end{pmatrix}\end{array},$$

而重排后的矩阵虽然并非基本形，但却是最为接近的，而且它违反基本形的约束的次数最少（仅 6 次）。对最优排列向量 $\boldsymbol{q}$ 进行排序，并跟踪各元素下标在排序前后的变动，便能得到排名向量

$$\begin{array}{c}\\ \text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{array}\begin{pmatrix}5\\1\\3\\4\\2\end{pmatrix}。$$

⊖ 实际上，我们有时称之为 R-bert 形。街机迷们可能还记得 1982 年的 Q-bert 游戏，其中的同名角色在一个阶梯状的山坡上跳来跳去，而这一等高台阶构成的山坡看起来就跟图 8.3 中的基本排名差距矩阵 $\boldsymbol{R}$ 一样。由于评分差距矩阵也具有不等高的台阶，因此我们开玩笑地称之为 R-bert 形而非 Q-bert 形。

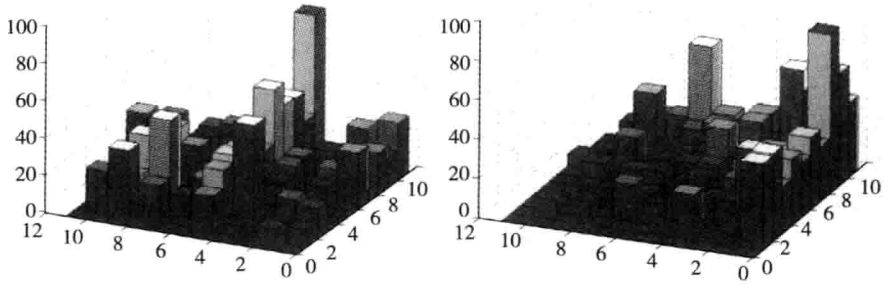
求解重排问题

评分差距问题可表述如下：找到一个排列 $\boldsymbol{q}$，使得重排后的数据差矩阵 $\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q})$ 违反**第 108 页**的阴影框中约束的次数最少。对于具有 n 个对象的集合，有 $n!$ 种可能的排列。由于状态空间巨大，同时我们需要寻找的是排列向量，因此进化最优化方法成为了一个有吸引力的而且是合适的选择。由于排名差距和评分差距方法是如此紧密地联系在一起，因此算法中的种群、交配和变异操作都是相同的，而仅需将适应度函数加以修改。在这种情况下，一个排列，即种群中的一个成员的适应度，由被违反的约束数量所给出。

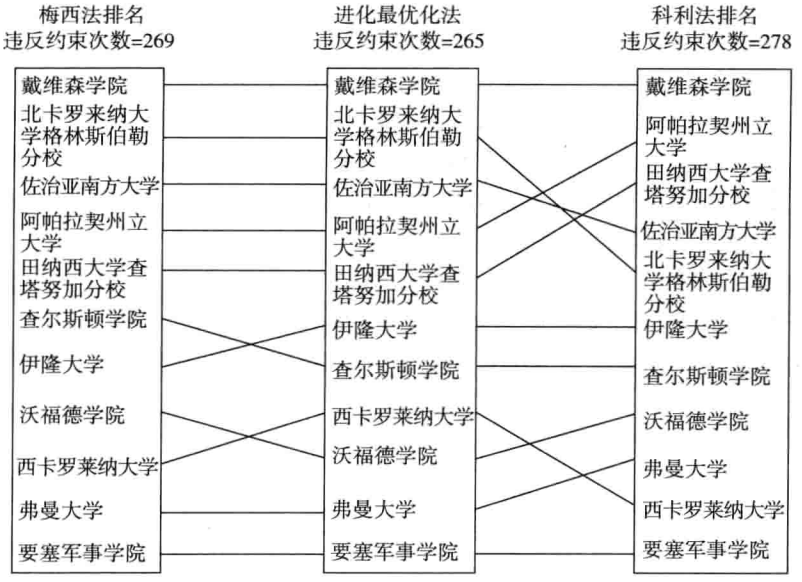
尽管可以使用求解评分差距问题的进化方法，但最好的方法还是将该问题转化为一个二值整数规划（BILP，Binary Integer Linear Program）问题，它可进一步被松弛为线性规划（LP，Linear Program）问题。我们不会在此介绍这一方法，而是将它推迟到第 15 章，因为可以证明，评分差距法实际上就是那一章中所介绍的排名聚合方法的一个特例。我们将在**第 194 页**对评分差距法再次加以讨论。

杂谈：更多的大学篮球赛内容

评分差距法是查尔斯顿学院研究生凯思琳·佩丁斯硕士论文中的一个主要部分。如同大多数研究生所做的那样，凯思琳在多个会议和多所大学中宣读或公开发表了她的工作。这些学术活动中最远的一个，发生在距离日本东京一小时路程的筑波大学，在那里，她使用了如下的示意图来说明她的工作。下图展示了 2007 年大学男子篮球南方联盟的 11 支球队的两个得分差矩阵。



左侧的城市区图显示了原有排序下的矩阵，其中看不出什么明显的结构。右侧的城市区图则是同一个矩阵经评分差距法重排后的结果，而山坡形的结构在其中显现了出来。接着，在如下的表格中，她比较了评分差距法（她称之为进化最优化法）的排名和梅西法以及科利法的排名。



注意，本节所介绍的评分差距法给出了违反约束次数最少的排名，如果这一点是需要优先考虑的因素的话，则显然评分差距法就是我们应该选用的方法，因为它能保证给出山坡形突起数量最少的排名。

评分差距法总结

— 记法

$\boldsymbol{D}$ 是对象-对象数据矩阵，其中包含了成对的比较关系，如得分、码数或其他统计数据的差值。

$\boldsymbol{q}$ 是定义了对 $\boldsymbol{D}$ 的重排的排列向量。

— 算法

1. 求解最优化问题

$$\min_{\boldsymbol{q}\in p_n} \boldsymbol{D}(\boldsymbol{q},\boldsymbol{q}) \text{ 违反第 108 页中约束的次数}$$

尽管进化算法是可行的，但我们仍推荐使用第 15 章中介绍的迭代 LP 法；

2. 上述问题的最优解 $\boldsymbol{q}$ 用来给出排名向量：将 $\boldsymbol{q}$ 按升序排序，然后将原有的下标值经排序后的结果保存为排名向量。

— 性质

- 之前各章所介绍的排名方法都同时给出一个评分向量和一个排名向量，而排名差距法和评分差距法都仅给出排名向量。
- 评分差距法的目的是找到数据差矩阵 $\boldsymbol{D}$ 的一个对称重排 $\boldsymbol{q}$，使得 $\boldsymbol{D}(\boldsymbol{q},\boldsymbol{q})$ 尽可能接近山坡形。

- 由于其中的最优化问题非常具有挑战性（实际上它是 NP 困难问题），因此相比本书中的其他排名方法，评分差距法的速度更慢，计算量更大，从而可能局限应用于较小的对象集合。例如，对体育队伍进行排名还在这个方法的能力范围之内，但对整个万维网上的网页进行排名则超出了这一范围。尽管仅限于较小的集合，但评分差距法却是一种全新的排名方法。
- 任何数据矩阵都可以作为评分差距法的输入。在本章中，我们推荐使用分差矩阵，不过其他数据差矩阵也同样适用。
- 如同排名差距法一样，评分差距法似乎对于拥有许多直接的成对比较关系的对象集合特别适用，即这些集合的对象-对象矩阵中存在着众多的连接。而因此一来，评分差距法在连通性很差的对象集上将可能表现得很糟糕。

数字——

18 = 1904 年在橄榄球比赛中丧生的球员人数。

——这一伤亡以及那数不清的断裂的骨头、破碎的头骨、折断的脖子、扭断的大腿和许许多多其他令人毛骨悚然的伤害，使得西奥多·罗斯福一度在 1905 年威胁将大学橄榄球运动宣布为非法。

——约翰-J. 米勒（John J. Miller）的《争球大战》（The Big Scrum）

第 9 章　分差

如果不对分差进行讨论，那么这本关于评分和排名的书就总是少了点什么，因为对于沉迷于博彩世界中的人们来说，猜中分差正是他们孜孜以求的圣杯。不过，科学的评分体系的目标，跟投注经纪人和赌徒们的目标却常常背道而驰。

一个好的科学评分体系，试图根据合理数量的对抗，准确地反映出每个队伍或竞争者总体实力之间的相对差别。它最基本的目标大概就是给出一个合理的排名，这个排名与专家给出的排名相吻合，并体现了长时期内取胜的能力。大多数流行的科学评分体系在求取与胜率相吻合的评分方面，都取得了不错的成功。但是，正如在**第 49 页**上对基纳评分法的讨论中所表明的那样，在估计胜率和估计分差之间，存在着一条鸿沟：某个体系在前者上做得不错，根本不足以说明它对后者也同样可以给出有用的信息。

对某些人而言，他们的终极目标是设计出一个体系，这一体系所给出的评分之间的相对差别，能够通过某种方式给出两支队伍相遇时所得分差的准确预测。尽管一般来说这是一项不可能完成的任务，但它还是一直占据着评分体系建立者的大部分注意力——话说回来，即使是艾萨克·牛顿，也从未失去对炼金术的热情。但是，即使你做到了评分上的点石成金，它也不能被真理似的传为“十拿十稳的赌注”，因为你的投注经纪人或赌场对于“分差”，有着不同的概念。

它是什么（又不是什么）

许多新手并没有理解一场给定比赛中的“分差”的真正含义，以及构建“分差”的机制。经纪人所给出的分差，根本不是指一支队伍优于另一支队伍的程度的精确体现或预测，而对分差做出准确预测也根本不是他们的目的。从某种意义上来说，他们的任务要更为简单一些。经纪人就某场比赛公布的分差的全部目的，就是使得对这两支队伍所下的赌注相等，如此一来就能确保两件事情。

1. 它有助于激发起最大的投注热情。
2. 它能够平衡经纪人的所得与所失的金钱。

为了理解以上两点是如何为经纪人的利润做出贡献的，我们需要确切知道经纪人到底是如何赚钱的。

抽头

投注经纪人一般不是通过跟你下不同的注并让你输掉赌注来赚钱的。一名优秀的经纪人并不在乎你是输是赢，因为他们的利润一般来自*抽头*（vig，vigorish[⊖]的简称，有时又称为 juice），它不过就是你为经纪人所提供的服务所付的一笔费用或佣金。当经纪人成功地在一场比赛中为两支队伍吸引了同样多的赌注时，那么不管哪一方获胜，经纪人的净支出都是 0，而他的抽头就成为了纯利润。

为什么不简单地赌胜负

如果经纪人并不是给出一个分差，而仅仅是对比赛中哪支队伍会获胜来下注的话，那么为了维持他所收取和支出的费用之间的平衡，赌赢的报酬就必须与胜者的赔率相符。因此，当一支强队和一支弱得多的队伍交手时，如果赌徒对强队下注，那么他可能只能赢很少的钱，却要冒着大输一笔的风险。此外，最有经验的赌徒们不会在一支很可能会输掉比赛的队伍身上去投下“大有赚头的注”，哪怕现在的赔率已经十分诱人。也就是说，如果只给出赔率，那么通常就会导致投注热情不高，而对于经纪人来说这就糟糕了，因为总的抽头是直接与所下的总注数相关的。

20 世纪 30 年代中期，查尔斯-K. 麦克尼尔（Charles K. McNeil）[⊜]认识到，如果能够让赌徒们对一个公开的分差下注，那么相比于仅有胜负的情况，经纪人就能够吸引到更多的投注（从而带来更多的抽头）。当时，赌博已经于不久前的 1931 年在内华达成为合法行为，因此当拉斯维加斯的经纪人们了解到麦克尼尔的新点子时，他们便马上意识到了它的好处，并立即采用了这一方法。从那以来，分差赌博就成为了体育博彩的标准方式。

分差是如何操作的

正如前述，经纪人的目的是最大化任何他想要设立的赌局中的总下注额，同时使双方的下注额均衡。他通过给出一个分差来对较强一方设置障碍，这时赌局不是简单地赌队伍 i *将战胜队伍* j，*而是赌队伍* i *将以超过（或不到）给定分差的优势战胜队伍* j。经纪人要做的就是确定分差，它能使得赌局中对两支队伍所下的赌注相等。

⊖ 来自俄语单词“vyigrysh”的意第绪（犹太人使用的国际语）俚语，意为赢取物或利润。

⊜ 查尔斯-K. 麦克尼尔（1903—1981）在芝加哥大学学习数学，并获得了硕士学位。之后，他至少从事了三个不同的职业——他在纽约市的一所贵族预科学校，即弗戴尔乡村学校（Riverdale Country School）中教授数学，也是芝加哥的一名证券分析师，并且在 20 世纪 40 年代还经营着自己的赌注登记买卖。他那个对分差下注的想法永久地改变了体育博彩。

表面上看起来，这个分差应该被如此计算得到，以使得下注双方具有均等的获胜概率。但是，对赌局投下的注码很少会是均等的，即使对一个“公平的”赌局也是如此，因此随着比赛的临近，分差将被调整。投注经纪是一门黑魔法，它可不适合脆弱的心灵，而成功的经纪人必须是只看得到数字和统计数据的人——而且他们还必须是相当不错的心理学家，并不断钻研人的本性，他们要对客户具有一定程度的了解。

赌中分差

那“赌中分差”这句话到底是什么意思呢？嗯，可以有两种方式来赌赢分差。如果你押宝在一支队伍上，而它获胜的优势又超过了所公布的分差，那你就赢了——你“赌赢了分差”！但还有另一种方式来赌赢分差。你可以下注给不被看好的队伍。如果它赢了，那你也赢了；不过如果它虽然告负，但是劣势却少于公布的分差，那你也还是赢了——你又“赌赢了分差”！让我们来看几个例子。

- 假设你的经纪人开设的赌局是野马队将以7分的分差击败闪电队，而实际比分为野马队 = 17分，闪电队 = 3分。如果你投注于野马队，那你就赢了，因为

  野马队 – 闪电队 > 7（野马队保住了分差）。

  如果你投注给了闪电队，那你就输了，因为闪电队输给对手的分数超过了7分。
- 假设你的经纪人开列的赌局是野马队将以3分的分差击败突袭者队，而实际比分为野马队 = 21分，突袭者队 = 20分。如果你投注于野马队，那你就输了，因为

  野马队 – 突袭者队 < 3（野马队没能保住分差）。

  但是，如果你赌的是突袭者队，那你就赢了，因为它并没有输给对手超过3分。

为了避免平局，经纪人常常将分差公布为带有0.5的小数，如7.5或3.5。这样，他们就可以避免平局以及因此而出现的退款。在某些场合，体育投注时称这种情况为“平局负”或“平局胜”。

超/欠赌

除了上述的赌分差之外，还有一种流行的赌局，称为赌超/欠（O/U）或赌总分。在这种赌局中，赌徒并不是对某支特定的队伍下注，而是对比赛的总比分下注。经纪人将设置一个“预期”总得分的值，而参赌者则可以选择赌实际比赛的总得分超过或不到经纪人的预测值。正如在赌分差的情况中一样，体育投注常常将设置的总得分表示为带有0.5的小数如28.5，以避免平局的情况。

例如，假设野马队将与闪电队交锋，而经纪人将总分设在了 42.5 分。如果最终的比分为野马队 = 24 分而闪电队 = 16 分，那么实际的总分就是 40 分，因此那些赌总分达不到的人将会赢钱；如果最终比分为野马队 = 30 分而闪电队 = 32 分，则实际总分为 62 分，则那些赌总分会超出的人将会赢钱。赌总分十分流行，因为它让赌徒们可以凭自己的感觉，来猜测比赛究竟是防守大战还是进攻大战，而无需挑出一个胜者。

正如分差的设置一样，经纪人要做的并不是真正地预测总比分，而是给出一个总得分，以使得对超/欠双方下的总注相等。类似于赌分差，惯常的抽头在 4.55% 左右。

体育投注常常允许联合使用超/欠赌和分差赌。例如，当野马队保住了分差，并且比赛总分高于经纪人的“预测”值时，赌“野马队以超总分获胜”将会赢钱（所赢的钱和所投的注的比率大概在 13 比 5 左右）。尽管大多数常见的超/欠赌都使用总得分，但如下所列的其他比赛统计数据也可以使用。

— 橄榄球队的总冲球码数或传球码数、拦截次数、三次触地攻防转换次数，等等；
— 篮球队员（或球队）的总助攻、盖帽、失误、抢断、中投命中率，等等；
— 棒球手（或球队）的总的本垒打次数、打点数，等等。

为什么使用评分难以预测分差

好了，我们已经看到，投注经纪人给出的分差并不等同于预测给定比赛的实际分差，但利用评分来预测实际分差的这个问题却仍然存在。通常而言，即使是最好的评分体系，也不能很好地预测分差。而在所有的体育项目中，通过评分来预测橄榄球赛的分差则尤为头痛。为什么呢？

呃，有这么两件事情限制了评分体系对分差进行预测的能力。一方面，两者的目的存在根本不同。一个好的评分体系的目的，是给出一支队伍相比于同一联盟中其他队伍的总体实力的一个量化描述。但在这一描述过程中，许多实际影响了比分的细节被过滤掉了。比如一支让人感到痛快淋漓的橄榄球队可能会打出一场爆发式的传球大战，但地面跑动进攻较弱，防守也不尽如人意，因此他们的总体评分倾向于中游的位置，但他们的比赛却倾向于具有井喷式或完全爆发式的得分；另一方面，一支乏善可陈的球队可能具有均衡但中庸的攻防，它同样可能具有中等的评分，但是他们在比赛中的分差将局限在一个更为狭窄的范围内。换而言之，仅靠单独一个评分数值，是无法把握住所有能够精确预测分差的微妙因素的。对分差进行估计，需要对比赛中各个单独的方面分别加以分析——你也许可以按照第 7 章中所讨论的那样，以攻击评分和防守评分的分离作为一个起点来做这件事情。

评分差不足以反映比分差的另一个原因，则是由于得分模式具有不规则性，而有时对于给定的项目而言，这一点是本性使然的。在橄榄球中，这一点尤为形象，其比赛过程中的得分倾向于按照3-4-3-4-…的间隔模式来累积。例如，图9.1中给出了NFL 2009—2010赛季的分差分布。

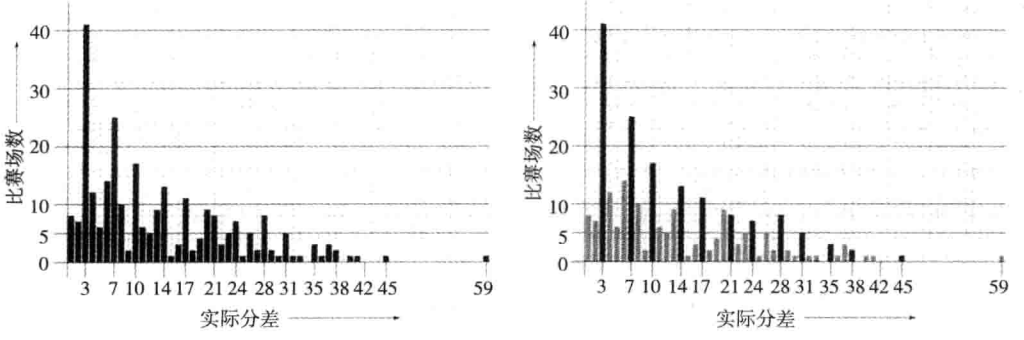


图9.1　NFL 2009—2010赛季的分差分布

图9.1中的右图清晰表明NFL分差的主要峰值出现在3分，7分，10分，14分，17分，21分等位置。如果了解了橄榄球中的得分规则的话，这一点就并不令人吃惊了。但是，分布中的这些尖峰使评分差难以被准确转换为对分差的预测。你可以合理期望的最好结果，就是让你的评分体系给出的差值能够产生那些最可能出现的尖峰——即假装除了3，7，10，14，17，21，…这些分差之外，其他的分差值都不可能出现，然后再试着去符合图9.1右图中那些黑条所形成的分布。即便成功做到了这一点，在整个赛季期间也仍然可能积累起大的误差。

利用分差来构建评分（并预测分差）

为了进一步说明构建能够反映分差的评分和排名体系中所涉及的问题，现在假设我们拥有一个能够看到未来的水晶球，因此我们便已经知道了在接下来的赛季中所有比赛的分差。类似于第9页上给出的梅西法，我们现在的目标是利用这一信息来建立一个最优的评分体系 $\{r_1, r_2, \cdots, r_n\}$，使得差值 $r_i - r_j$ 能尽可能准确地反映出队伍 i 与队伍 j 交锋时的分差。

要做到这一点，首先假设在一个绝对完美的世界中，存在着一个完美的评分向量

$$\boldsymbol{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

使得每个评分差 $r_i - r_j$ 都等于相应的得分差 $S_i - S_j$，其中，S_i 和 S_j 分别为队伍 i 和队伍 j 交锋时各自所得的分数。如果将得分差和评分差分别写成矩阵

$$\boldsymbol{K}=\begin{pmatrix} 0 & S_1-S_2 & \cdots & S_1-S_n \\ S_2-S_1 & 0 & \cdots & S_2-S_n \\ \vdots & \vdots & & \vdots \\ S_n-S_1 & S_n-S_2 & \cdots & 0 \end{pmatrix} \text{和} \boldsymbol{R}=\begin{pmatrix} 0 & r_1-r_2 & \cdots & r_1-r_n \\ r_2-r_1 & 0 & \cdots & r_2-r_n \\ \vdots & \vdots & & \vdots \\ r_n-r_1 & r_n-r_2 & \cdots & 0 \end{pmatrix},$$

则 $\boldsymbol{K}=\boldsymbol{R}=\boldsymbol{re}^{\mathrm{T}}-\boldsymbol{er}^{\mathrm{T}}$，其中，$\boldsymbol{e}$ 为全 1 列。

这意味着在理想情况下，得分差矩阵 $\boldsymbol{K}$ 是一个秩为 2 的斜对称阵——它是斜对称的，因为 $k_{ij}=-k_{ji}$ 隐含了 $\boldsymbol{K}^{\mathrm{T}}=-\boldsymbol{K}$，而秩为 2 是因为当 $\boldsymbol{r}$ 不是 $\boldsymbol{e}$ 的倍数（即 $\boldsymbol{R}\neq\boldsymbol{0}$）时，$\mathrm{rank}(\boldsymbol{re}^{\mathrm{T}}-\boldsymbol{e}\,\boldsymbol{r}^{\mathrm{T}})=2$。但我们不是生活在一个完美的世界里，因此，如参考文献［33］中所说的那样，我们能做到的最好程度，就是构建一个评分向量 $\boldsymbol{r}$，它使得得分差矩阵 $\boldsymbol{K}$ 和评分差矩阵 $\boldsymbol{R}=\boldsymbol{re}^{\mathrm{T}}-\boldsymbol{er}^{\mathrm{T}}$ 之间的距离为最小。这等价于寻找一个向量 $\boldsymbol{x}$，使得

$$f(\boldsymbol{x})=\|\boldsymbol{K}-\boldsymbol{R}(\boldsymbol{x})\|^2=\|\boldsymbol{K}-(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})\|^2 \tag{9.1}$$

对于某个矩阵范数为最小。范数的选择在一定程度上是任意的，因此让我们使用最为简单（也最为标准）的一个——弗罗贝尼乌斯范数，如下

$$\|\boldsymbol{A}\|_F=\sqrt{\sum_{i,j}a_{ij}^2}=\sqrt{\mathrm{trace}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}\text{（请见参考文献[54]，第279页）。}$$

式（9.1）中函数 $f(\boldsymbol{x})$ 的最小化可以通过初等微积分来完成——即对每个 i，令 $\partial f/\partial x_i=0$，并求解所得的方程组来得到临界点 x_1，x_2，…，x_n。矩阵的迹具有如下性质：

$$\mathrm{trace}(\alpha\boldsymbol{A}+\boldsymbol{B})=\alpha\mathrm{trace}(\boldsymbol{A})+\mathrm{trace}(\boldsymbol{B}),$$

且

$$\mathrm{trace}(\boldsymbol{AB})=\mathrm{trace}(\boldsymbol{BA})\text{。（请见参考文献[54]第90页）。}$$

利用这些性质可得

$$\begin{aligned} f(\boldsymbol{x}) &= \mathrm{trace}\,[\boldsymbol{K}-(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})]^{\mathrm{T}}[\boldsymbol{K}-(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})] \\ &= \mathrm{trace}\,\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}-\mathrm{trace}[\boldsymbol{K}^{\mathrm{T}}(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})+(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{K}]+ \\ &\quad \mathrm{trace}[(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{xe}^{\mathrm{T}}-\boldsymbol{ex}^{\mathrm{T}})] \end{aligned}$$

由 $\boldsymbol{K}$ 的斜对称性有

$$f(\boldsymbol{x})=\mathrm{trace}\,\boldsymbol{K}^{\mathrm{T}}\boldsymbol{K}-4\,\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ke}+2n(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})-2\,(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x})^2\text{。}$$

由于 $\partial\boldsymbol{x}/\partial\boldsymbol{x}_i=\boldsymbol{e}_i$，将 f 对 x_i 求导可得

$$\frac{\partial f}{\partial x_i}=-4\boldsymbol{e}_i^{\mathrm{T}}\boldsymbol{Ke}+4nx_i-4\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x},$$

令 $\partial f/\partial x_i=0$，得

$$nx_i-\sum_{j=i}^{n}x_j=(\boldsymbol{Ke})_i\quad(i=1,2,\cdots,n)\text{。}$$

这不过就是一个线性方程组，它可以写为一个矩阵方程

$$\begin{pmatrix} (n-1) & -1 & \cdots & -1 \\ -1 & (n-1) & \cdots & -1 \\ \vdots & \vdots & & \vdots \\ -1 & -1 & \cdots & (n-1) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} (\boldsymbol{Ke})_1 \\ (\boldsymbol{Ke})_2 \\ \vdots \\ (\boldsymbol{Ke})_n \end{pmatrix}$$

或等价地，

$$\left(\boldsymbol{I} - \frac{\boldsymbol{ee}^{\mathrm{T}}}{n}\right)\boldsymbol{x} = \frac{\boldsymbol{Ke}}{n}。 \tag{9.2}$$

由于 $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{Ke}$ 为标量，而 $\boldsymbol{K}^{\mathrm{T}} = -\boldsymbol{K}$，因此

$$\boldsymbol{e}^{\mathrm{T}}\boldsymbol{Ke} = (\boldsymbol{e}^{\mathrm{T}}\boldsymbol{Ke})^{\mathrm{T}} = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{K}^{\mathrm{T}}\boldsymbol{e} = -\boldsymbol{e}^{\mathrm{T}}\boldsymbol{Ke} \Rightarrow \boldsymbol{e}^{\mathrm{T}}\boldsymbol{Ke} = 0,$$

所以向量 $\boldsymbol{x} = \boldsymbol{Ke}/n$ 是式（9.2）的一个解。此外，可以证明

$$\left(\boldsymbol{I} - \frac{\boldsymbol{ee}^{\mathrm{T}}}{n}\right)\boldsymbol{e} = \boldsymbol{0} \text{ 且 } \operatorname{rank}\left(\boldsymbol{I} - \frac{\boldsymbol{ee}^{\mathrm{T}}}{n}\right) = n - 1,$$

因此式（9.2）的所有解都具有如下形式

$$\boldsymbol{x} = \frac{\boldsymbol{Ke}}{n} + \alpha\boldsymbol{e} \quad (\alpha \in \Re)。$$

通过加上一些合理的约束，例如，要求评分和等于0，我们可以计算出标量 α 的值，并确定一个唯一的解。实际上，如果我们要求 $0 = \sum x_i = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}$，则

$$0 = \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = \frac{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{Ke}}{n} + \alpha\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} = \alpha\boldsymbol{e}^{\mathrm{T}}\boldsymbol{e} \Rightarrow \alpha = 0,$$

因此

$$\boldsymbol{r} = \frac{\boldsymbol{Ke}}{n}(\text{即 } \boldsymbol{K} \text{ 中各列的重心或均值})$$

便是 $f(\boldsymbol{x})$ 的满足约束 $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = 0$ 的唯一临界点。请记住，一个点是临界点，仅是获得一个相对极小或极大的必要条件，因此我们还需要检验 $\boldsymbol{r} = \boldsymbol{Ke}/n$ 是否的确是一个最小点。一个临界点能给出极小值的一个充分条件，便是该点处的海森矩阵（Hessian matrix）[54，第570页] 为正定阵。但不幸的是，对于我们的问题而言，所得的海森矩阵仅为半正定的，因此还需要进行一些额外的工作。为了验证

$$\min_{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}=0} f(\boldsymbol{x}) = \min_{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}=0} \|\boldsymbol{K} - (\boldsymbol{xe}^{\mathrm{T}} - \boldsymbol{ex}^{\mathrm{T}})\|_F^2$$

在 $\boldsymbol{r} = \boldsymbol{Ke}/n$ 处取得，需要考虑在 $\boldsymbol{r}$ 的一个邻域内且满足约束的若干点，即考虑 $\boldsymbol{r} + \boldsymbol{\varepsilon}$，其中，$\boldsymbol{\varepsilon} \neq 0$ 且 $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\varepsilon} = 0$，并求得

$$\begin{aligned} f(\boldsymbol{r} + \boldsymbol{\varepsilon}) - f(\boldsymbol{r}) &= -4\,\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{Ke} + 2n\,\|\boldsymbol{r} - \boldsymbol{\varepsilon}\|_2^2 - 2n\,\|\boldsymbol{r}\|_2^2 - 2\,(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\varepsilon})^2 \\ &= -4n\,\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{r} + 2n(\|\boldsymbol{r}\|_2^2 + \|\boldsymbol{\varepsilon}\|_2^2 + 2\,\boldsymbol{\varepsilon}^{\mathrm{T}}\boldsymbol{r}) - 2n\,\|\boldsymbol{r}\|_2^2 - 2\,(\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\varepsilon})^2 \quad (9.3) \\ &= 2[n\,\|\boldsymbol{\varepsilon}\|_2^2 - (\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\varepsilon})^2]。 \end{aligned}$$

柯西-施瓦兹（Cauchy-Schwarz 或 CBS）不等式[54，第287页] 指出，$\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\varepsilon} \leqslant \|\boldsymbol{e}\|_2\,\|\boldsymbol{\varepsilon}\|_2$，当且仅当对某个 α 有 $\boldsymbol{\varepsilon} = \alpha\boldsymbol{e}$ 时，等号成立。在我们的问题里，等号不可能成立，因为 $\boldsymbol{e}^{\mathrm{T}}\boldsymbol{\varepsilon} = 0$ 而

$\boldsymbol{\varepsilon} \neq 0$，因此将 CBS 用于式（9.3）可得

$$f(\boldsymbol{r}+\boldsymbol{\varepsilon}) - f(\boldsymbol{r}) > 0。$$

所以，由分差给出的“最佳”评分向量为 $\boldsymbol{r} = \boldsymbol{Ke}/n$。下面是以上结论的总结。

分差给出的“最佳”评分

可以由分差给出的、满足 $\sum_i r_i = 0$ 的“最佳”评分集合 $\{r_1, r_2, \cdots, r_n\}$，也就是使得分差矩阵 $\boldsymbol{K} = [S_i - S_j]$ 和评分差矩阵 $\boldsymbol{R} = [r_i - r_j]$ 之间的（弗罗贝尼乌斯）距离最小的评分，由如下向量中的元素给出

$$\boldsymbol{r} = \frac{\boldsymbol{Ke}}{n} = \boldsymbol{K} \text{中各列的重心或均值} \tag{9.4}$$

今后，我们将称 $\boldsymbol{r} = \boldsymbol{Ke}/n$ 中的元素为分差评分或重心评分。对于式（9.1）中的函数 $f(\boldsymbol{x})$ 以及 $\boldsymbol{R} = \boldsymbol{re}^{\mathrm{T}} - \boldsymbol{er}^{\mathrm{T}}$，可以证明

$$\min_{\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}=0} f(\boldsymbol{x}) = \|\boldsymbol{K} - \boldsymbol{R}\|_F^2 = \|\boldsymbol{K}\|_F^2 - \|\boldsymbol{R}\|_F^2。$$

NFL 2009—2010 赛季的分差评分

为了描述式（9.4）中分差评分的使用，让我们看一看 NFL 2009—2010 赛季所有 267 场比赛中的实际分差。如果队伍 i 和 j 交手了不止一次，那么就令 S_i 和 S_j 为两队各自的平均得分来构建矩阵 $\boldsymbol{K}$。所得的分差评分如下。

排名	队　伍	评　分	排名	队　伍	评　分
1.	新奥尔良圣徒	6.2187	17.	卡罗来纳黑豹	-0.1094
2.	明尼苏达维京人	4.7187	18.	丹佛野马	-0.1250
3.	绿湾包装工	3.9687	19.	田纳西泰坦	-0.9531
4.	巴尔的摩乌鸦	3.6875	20.	纽约巨人	-1.0625
5.	新英格兰爱国者	3.6250	21.	华盛顿红皮	-1.0937
6.	纽约喷气机	3.0000	22.	迈阿密海豚	-1.1094
7.	印第安纳小马	2.8594	23.	芝加哥熊	-1.5781
8.	圣迭戈闪电	2.7031	24.	布法罗比尔	-1.7812
9.	费城老鹰	2.5469	25.	杰克逊维尔美洲虎	-2.9531
10.	休斯敦德州人	2.0781	26.	堪萨斯城酋长	-3.0156
11.	达拉斯牛仔	2.0156	27.	克里夫兰布朗	-3.0469
12.	匹兹堡钢人	1.4219	28.	西雅图海鹰	-3.3281
13.	亚特兰大猎鹰	1.1719	29.	坦帕湾海盗	-3.9687
14.	亚利桑那红雀	0.3906	30.	奥克兰突袭者	-5.1562
15.	辛辛那提猛虎	0.3281	31.	底特律雄狮	-5.4219
16.	旧金山 49 人	0.1875	32.	圣路易斯公羊	-6.2187

NFL 2009—2010 赛季分差评分

这些分差评分正确地选出了 NFL 2009—2010 赛季 267 场比赛中 191 场的胜者，其事后预测准确率为 71.54%。对于这样简单的方法而言，这个结果已经相当不错——我们所要做的不过是对分差求平均（即计算 $\boldsymbol{K}$ 中每一行的和然后再除以 32）。

那这些分差评分在估计实际分差方面表现如何呢？**第 116 页**上的讨论已经说明了我们为什么不应对此抱有太大希望，但不管怎样，还是让我们看一看吧。这些分差评分并没有考虑主场优势，因此，为了公平起见，我们应当允许将主场优势包括在内。利用线性回归来由评分得到以下线性模型的最优“主场优势”参数 α，β 和 γ，

$$\text{主队得分的期望值} - \text{客队得分的期望值} = \alpha + \beta r_h - \gamma r_a \tag{9.5}$$

式中，r_h 和 r_a 分别为第 i 场比赛中主队和客队的评分。参数 α、β 和 γ 可通过求取以下方程组的最小二乘解而得，

$$\alpha + \beta r_{h_i} - \gamma r_{a_i} = S_{h_i} - S_{a_i} \quad (i = 1,2,\cdots,267)$$

式中，r_{h_i} 和 r_{a_i} 分别为比赛 i 中主队和客队的评分；S_{h_i} 和 S_{a_i} 分别为比赛 i 中主队和客队的得分。利用矩阵，我们可以将这一问题归结为计算线性系统 $\boldsymbol{Ax} = \boldsymbol{b}$ 的最小二乘解，或等价地，求取 3 × 3 的正规方程 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 的解，其中

$$\boldsymbol{A}_{267\times 3} = \begin{pmatrix} 1 & r_{h_1} & r_{a_1} \\ 1 & r_{h_2} & r_{a_2} \\ \vdots & \vdots & \vdots \\ 1 & r_{h_{267}} & r_{a_{267}} \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} S_{h_1} - S_{a_1} \\ S_{h_2} - S_{a_2} \\ \vdots \\ S_{h_{267}} - S_{a_{267}} \end{pmatrix}。$$

得到的计算结果为 $\alpha = 2.3671$，$\beta = 2.4229$，$\gamma = 2.2523$（圆整到 5 位有效数字），因此，比赛 i 的估计分差由下式给出

比赛 i 中主队得分的估计值 − 比赛 i 中客队得分的估计值 $= 2.3671 + 2.4229r_{h_i} - 2.2523r_{a_i}$。

总分差的绝对误差为整个赛季中估计分差与实际分差之间差值的绝对值之和，计算可得

$$\sum_{i=1}^{267} \left| (2.3671 + 2.4229r_{h_i} - 2.2523r_{a_i}) - (S_{h_i} - S_{a_i}) \right| = 2827.6,$$

即每场比赛 10.59 分。为了对这一结果的好坏有一个直观感受，可以将它与其他受重视的体系以及实际的拉斯维加斯盘口进行比较。

几场决斗

最著名也最广受尊敬的两名评分者，恐怕就是自 1985 年以来为《今日美国》提供体育评分的杰夫·萨加林，以及 BCS 大学橄榄球评分系统的贡献者肯尼斯·梅西（第 2 章）。如下所示为这两人分别为 NFL 2009—2010 赛季给出的最终评分，数据来自在写作本

书时仍然有效的下面两个网址。

www. usatoday. com/sports/sagarin/nfl09. htm

www. masseyratings. com

排名	队　伍	评　分	排名	队　伍	评　分
1.	新奥尔良圣徒	32. 18	17.	辛辛那提猛虎	21. 07
2.	印第安纳小马	29. 86	18.	纽约巨人	20. 52
3.	明尼苏达维京人	27. 91	19.	丹佛野马	20. 27
4.	纽约喷气机	27. 69	20.	旧金山 49 人	20. 21
5.	圣迭戈闪电	25. 99	21.	田纳西泰坦	19. 76
6.	新英格兰爱国者	25. 89	22.	布法罗比尔	18. 18
7.	达拉斯牛仔	25. 65	23.	芝加哥熊	16. 48
8.	巴尔的摩乌鸦	25. 58	24.	杰克逊维尔美洲虎	16. 20
9.	亚特兰大猎鹰	24. 30	25.	坦帕湾海盗	13. 06
10.	绿湾包装工	24. 19	26.	华盛顿红皮	12. 84
11.	费城老鹰	23. 79	27.	克里夫兰布朗	12. 66
12.	卡罗来纳黑豹	22. 87	28.	奥克兰突袭者	12. 56
13.	休斯敦德州人	22. 49	29.	堪萨斯城酋长	11. 93
14.	迈阿密海豚	21. 43	30.	西雅图海鹰	11. 61
15.	亚利桑那红雀	21. 37	31.	底特律雄狮	6. 41
16.	匹兹堡钢人	21. 24	32.	圣路易斯公羊	3. 82

NFL 2009—2010 赛季的萨加林评分

排名	队　伍	评　分	排名	队　伍	评　分
1.	新奥尔良圣徒	1. 693	17.	田纳西泰坦	0. 110
2.	印第安纳小马	1. 416	18.	迈阿密海豚	0. 062
3.	圣迭戈闪电	1. 291	19.	纽约巨人	0. 036
4.	达拉斯牛仔	1. 002	20.	旧金山 49 人	-0. 093
5.	明尼苏达维京人	0. 909	21.	丹佛野马	-0. 182
6.	纽约喷气机	0. 897	22.	布法罗比尔	-0. 232
7.	费城老鹰	0. 814	23.	芝加哥熊	-0. 478
8.	绿湾包装工	0. 783	24.	杰克逊维尔美洲虎	-0. 621
9.	亚特兰大猎鹰	0. 671	25.	坦帕湾海盗	-0. 706
10.	新英格兰爱国者	0. 667	26.	华盛顿红皮	-0. 938
11.	卡罗来纳黑豹	0. 639	27.	克里夫兰布朗	-0. 947
12.	巴尔的摩乌鸦	0. 602	28.	奥克兰突袭者	-0. 986
13.	休斯敦德州人	0. 340	29.	堪萨斯城酋长	-1. 050
14.	亚利桑那红雀	0. 318	30.	西雅图海鹰	-1. 451
15.	辛辛那提猛虎	0. 230	31.	底特律雄狮	-2. 385
16.	匹兹堡钢人	0. 178	32.	圣路易斯公羊	-2. 591

NFL 2009—2010 赛季的梅西评分

萨加林和梅西没有提供多少细节（大多数评分者都不希望你知道他们的魔法药水中的所有神秘成分），但萨加林称他的评分是两个体系的“综合”——一个是如**第 53 页**第 5 章中所描述的、仅使用胜负信息的埃洛体系，以及另外某个未指明的体系，它考虑了分差。梅西则

称他的评分是一个最大似然估计，而且在一定程度上通过贝叶斯修正进行了平滑——他的描述过于含糊，无法实现。不过，之所以单单将萨加林和梅西挑出来进行对比，是由于他们所具有的高可见度，而且在许多个赛季之后，他们已经建立起了完备的记录。

在下面的式（9.6）中，我们用 NFL 2009—2010 赛季的结果来对比 6 个评分体系对比赛胜负的事后预测准确率（百分比），以及对分差的事后预测误差（每场比赛的平均分），这 6 个评分体系是——**第 120 页**上的分差或重心评分；**第 122 页**上的萨加林评分；**第 58 页**上的埃洛评分；**第 44 页**上的基纳评分；**第 122 页**上的梅西评分；此外出于好奇，我们还加入了由拉斯维加斯所定盘口给出的结果，这些盘口每天都公布在北卡罗来纳瑞利的《新闻与观察家》（News and Observer）上。为了根据萨加林评分获得事后预测，我们（按他的建议）为主队加上了 2.96 分。其余评分体系的事后预测通过**第 121 页**上的线性模型即式（9.5）来获得（尽管梅西为他自己的体系提出了更为复杂的方法）。

比赛胜负的事后预测准确率	平均分差预测误差（每场比赛的平均分）
分差评分 = 71.5%	分差评分 = 10.59 分每场
萨加林评分 = 70.8%	萨加林评分 = 10.53 分每场
埃洛评分 = 75.3%	埃洛评分 = 10.71 分每场
基纳评分 = 73.4%	基纳评分 = 10.54 分每场
梅西评分 = 72.7%	梅西评分 = 10.65 分每场
拉斯维加斯盘口 = 67.8%	拉斯维加斯盘口 = 11.44 分每场

(9.6)

6 种评分体系对 NFL 2009—2010 赛季预测结果的对比

如式（9.6）所示，所用的 5 种评分体系对胜负与分差的事后预测准确度都处在同一个水平上，而拉斯维加斯盘口则略微靠后。关键在于，分差（或重心）评分法在反映相对的总体实力方面可以与其他方法一较高下，但如果考虑到分差（或重心）评分法的计算极为简单（你在一个信封背后就可以做出来），那它就尤其具有吸引力，而正是这一点将重心法与其他评分体系区分开来。

表（9.6）最下面一行所体现的拉斯维加斯盘口的表现可能看起来有些奇怪，但如果从一个恰当的视角来看就不足为奇了。正如**第 113 页**上所说明的那样，经纪人所公布的分差并不是要反映或预测实际的分差或胜者。经纪人也不会去关心在赛季末为每支队伍相对的整体实力做出评价，这一点与那些可能需要进行 BCS 最终排名或挑选 NCAA 篮球锦标赛种子队的其他评分和排名体系正好相反。换而言之，所有的经纪人所关注的是对赛季中下一场比赛的展望，因此事后预测的准确度根本不是他们的目的。拉斯维加斯盘口的事前预测准确度通常要好过事后预测准确度，这并不令人吃惊，因为盘口一般都反映了许多形形色色的人的集体智慧，这些人仅仅关注于即将到来的那一场比赛。比方说，尽管式（9.6）中埃洛体系的事后胜负预测准确率明显好于拉斯维加斯盘口，但拉斯维加斯盘口的事前预测准确率却好过埃洛体系。请回忆一下**第 62 页**上的结果，对于 NFL 2009—2010 赛季，拉斯维加斯盘口的事前胜负预测准确率为 67.8%，而埃洛体系的事前预测准确率则为稍低一点的 65.9%。

在万维网上并不缺少可用的 NFL 评分（以及任何一种其他类型的评分），而用谷歌搜索一下所得到的东西，就比大多数人想要消化掉的内容还要多得多。万维网上提供的体育评分的麻烦之处在于，它们一般没有给出足够的技术细节来帮助了解其由来，或是对你设计自己的体系有所助益。在**第 225 页**的后记中所引用的许多文章，也体现了在体育评分方面已经投入的巨大努力，不过这些文章对于那些渴望更多细节的人来说将更为具体。如果你对除我们已经强调过的这些评分体系之外的其他体系比较感兴趣的话，请参见托德·贝克的预测跟踪网站

www. thepredictiontracker. com/nflresults. php

（在写作本书时仍可访问。）这个网站表明，较好的 NFL 评分体系之间的性能差别令人吃惊地小——也就是说，要想出人头地可绝非易事。

其他成对比较

如果你已经理解了产生分差（或重心）评分，即式（9.4）的推导过程，那么你也许会想到，对**第 118 页**的矩阵 $\boldsymbol{K}$ 中的“分差” k_{ij} 可以进行更广义的解释，而不仅仅是得分之差。例如，令

$$k_{ij} = \begin{cases} +1, \text{若队伍 } i \text{ 击败队伍 } j, \\ 0, \text{若队伍 } i \text{ 和 } j \text{ 战平}, \\ -1, \text{若队伍 } i \text{ 负于队伍 } j, \end{cases}$$

就是较之得分差更为基本的队伍 i 和 j 之间的成对比较，不过 $\boldsymbol{K}$ 仍然保持斜对称，而与之前完全相同的分析，也仍然能够给出一个更一般意义下的“分差评分” $\boldsymbol{r} = \boldsymbol{Ke}/n$。这一思路给出了如下的 NFL 2009—2010 赛季的“胜负”分差评分。

排名	队伍	评分	排名	队伍	评分
1.	新奥尔良圣徒	0.375	17.	卡罗来纳黑豹	0.03125
2.	圣迭戈闪电	0.25	18.	田纳西泰坦	0
3.	印第安纳小马	0.25	19.	纽约巨人	-0.03125
4.	明尼苏达维京人	0.21875	20.	杰克逊维尔美洲虎	-0.03125
5.	费城老鹰	0.15625	21.	旧金山 49 人	-0.03125
6.	绿湾包装工	0.15625	22.	芝加哥熊	-0.09375
7.	达拉斯牛仔	0.125	23.	迈阿密海豚	-0.09375
8.	纽约喷气机	0.125	24.	布法罗比尔	-0.15625
9.	巴尔的摩乌鸦	0.125	25.	克里夫兰布朗	-0.15625
10.	匹兹堡钢人	0.125	26.	华盛顿红皮	-0.15625
11.	休斯敦德州人	0.125	27.	堪萨斯城酋长	-0.21875
12.	新英格兰爱国者	0.09375	28.	奥克兰突袭者	-0.21875
13.	亚利桑那红雀	0.0625	29.	西雅图海鹰	-0.21875
14.	辛辛那提猛虎	0.03125	30.	坦帕湾海盗	-0.28125
15.	丹佛野马	0.03125	31.	底特律雄狮	-0.28125
16.	亚特兰大猎鹰	0.03125	32.	圣路易斯公羊	-0.34375

NFL 2009—2010 赛季的胜负分差评分

相比于由得分差给出的评分，这些胜负分差评分自然是一个更为粗略的评价，但它们仍然可以提供若干信息。例如，根据赢取比赛的能力来衡量，这些评分表明闪电队和小马队不相上下。类似地，老鹰队和包装工队难分伯仲，而牛仔、喷气机、乌鸦、钢人和德州人也都势均力敌。显然，还有其他队伍可以根据相同的取胜能力而归入到一组中。

根据所评分和排名的对象，可以有数不清的方式来进行成对比较。例如，令 $k_{ij} = S_i/S_j$ 甚至 $k_{ij} = \log(S_i) - \log(S_j)$，都有可能对某些场合更好地加以分析。唯一能够限制你的，仅仅是你自己的想象力。

结论

下次当你再看到盘口中所公布的分差时，请记住，设立赌局的那个人或是赌场或是不论什么，都不一定真的相信被看好的队伍就肯定比劣势一方强那么多分：他们不过是在赌下面这件事，即它们所“预测”的分差，能够为对阵双方吸引到相同数量的赌注。

而更大的教训也许在于，当我们对分差（甚至胜者）进行预测时，不要对任何人的评分体系抱有过度的信任——你也许是过于希望从一张数字清单中得到太多的东西了。如果你已经注意到了这一点，却仍然希望进行分差预测，那么你只需简单地如**第 120 页**的式（9.4）中那样，将以往的分差求取平均即可，这样得到的估计与更为复杂的方法所得出的结果相比，还是很有竞争力的。

数字——

24 =拉斯维加斯盘口制定者所曾给出的最大 NFL 分差。

——1976 年 12 月，匹兹堡胜坦帕湾。匹兹堡以 42-0 击败坦帕湾，保住了分差。

——sports. espn. go. com/nfl/news

第 10 章　用户偏好评分

第 120 页上式（9.4）中的“分差评分”的思路，可以应用于体育评分和排名之外的其他领域。近些年随着在线商务的发展，产生了一个主要问题，就是如何根据用户的偏好来对事物进行评分和排名。尽管 Amazon. com 和 Netflix 这样的公司可能并非是使用用户评分和排名体系来进行产品推荐的先行者，但它们却已经建立了经过高度改良的（且专有的）、基于产品推荐体系的在线营销方法。要探讨有关这些方法的所有细节，恐怕还得需要另外一整本书，不过我们还是可以在此说明一下，与分差评分有关的某些想法是如何能够应用于这一领域的。

假设我们是一家在万维网上出售商品的电子零售商，并且假设我们通过一种类似于 Amazon. com 的方式来收集用户的偏好评分——即由用户按五星等级制来给每件商品打分。

* * * * *　（最为推荐）
* * * *
* * *　　⋮
* *
*　　（最不推荐）

现在我们的目的，是将这些用户打分转化为一个评分体系，利用它来向购物者推荐与他们正在浏览或购买的商品相似的高评分商品。人们不大可能在网络浏览器中实实在在地翻阅一份很大的库存商品目录，特别是当他们并不确定到底想要什么或者目录中到底有什么的时候，就尤其如此。因此，在线销售能否取得成功，有很大一部分因素就在于公司能在何种程度上有效地为顾客提供可以信赖的推荐商品。其中的逻辑很简单，

建立一个更好的商品评分体系→赢得更多的信任→实现更多的销售！

要做到由用户推荐来构建一个简单的商品评分体系这一点很容易，我们只需将**第 118 页**上的斜对称分差矩阵 $\boldsymbol{K}$ 重新加以解释即可。例如，假设

$$\mathcal{P} = \{p_1, p_2, \cdots, p_n\}$$

是我们的库存商品集，并令 $\mathcal{U}_i$ 和 $\mathcal{U}_j$ 分别表示对商品 p_i 和 p_j 进行了评价的所有用户，所以 $\mathcal{U}_i \cap \mathcal{U}_j$ 便是对这两种商品都进行了评价的所有用户。在进行 p_i 和 p_j（$i \neq j$）的比较时，令 $n_{ij} = \#(\mathcal{U}_i \cap \mathcal{U}_j)$ 为对这两者都进行了评价的用户数，并定义 p_i 在 p_j 身上所取得的“分数”

为平均的评分星级

$$S_{ij} = \begin{cases} \dfrac{1}{n_{ij}} \displaystyle\sum_{h \in U_i \cap U_j} (\text{用户 } h \text{ 为 } p_i \text{ 所给出的星级}), n_{ij} \neq 0, \\ 0, n_{ij} = 0 \end{cases} \tag{10.1}$$

定义商品 p_i 和 p_j 之间的“分差”为

$$\begin{aligned} k_{ij} &= S_{ij} - S_{ji} \\ &= \frac{1}{n_{ij}} \sum_{h \in U_i \cap U_j} [(\text{用户 } h \text{ 为 } p_i \text{ 所给出的星级}) - (\text{用户 } h \text{ 为 } p_j \text{ 所给出的星级})] \end{aligned} \tag{10.2}$$

然后，一如在式（9.4）中那样，令

$$\boldsymbol{K} = [k_{ij}] = [S_{ij} - S_{ji}] \tag{10.3}$$

为分差的斜对称矩阵。由**第 120 页**上的推导可以给出如下结论，即最佳评分可通过对给定的“星级得分”求平均而得到。

最佳星级评分商品

对于一个给定的星级得分集，（在**第 118 页**所给出的意义下的）最佳评分集，可通过对星级得分求平均而得到，它由重心

$$\boldsymbol{r} = \boldsymbol{Ke}/n$$

给出，其中，$\boldsymbol{K}$ 是式（10.3）中给出的平均星级得分差所构成的斜对称矩阵。

例如，假设有 10 名用户 $\mathcal{U} = \{h_1, h_2, \cdots, h_{10}\}$ 对 4 种商品 $\mathcal{P} = \{p_1, p_2, p_3, p_4\}$ 进行了五星制的打分，他们各自给出的分数（星的数量）归总为矩阵

$$*\text{-}matrix = \begin{array}{c} \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \\ h_{10} \end{array} \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \\ \begin{pmatrix} 4 & 2 & 2 & \\ 3 & 1 & & 2 \\ 1 & & 2 & \\ 2 & 4 & 2 & \\ & 3 & 2 & 5 \\ 2 & 3 & & \\ & 4 & 1 & 3 \\ 3 & 1 & 1 & \\ & 3 & 2 & 5 \\ 2 & & 2 & 4 \end{pmatrix} \end{array}。$$

注意，在该矩阵中有一些元素位置为空，因为并非所有用户都对所有商品进行了评价。实际上，对这些丢失的元素进行外插，这正是发起著名的 Netflix 竞赛[56]的隐藏动机。当式（10.1）所定义的得分 S_{ij} 被排列为一个矩阵时，所得的得分矩阵为

$$\boldsymbol{S} = \begin{pmatrix} 0 & 14/5 & 12/5 & 5/2 \\ 11/5 & 0 & 17/6 & 11/4 \\ 9/5 & 10/6 & 0 & 7/4 \\ 6/2 & 15/4 & 17/4 & 0 \end{pmatrix}。$$

式（10.3）所定义的斜对称分差矩阵 $\boldsymbol{K}$ 为

$$\boldsymbol{K} = \begin{pmatrix} 0 & 3/5 & 3/5 & -1/2 \\ -3/5 & 0 & 7/6 & -1 \\ -3/5 & -7/6 & 0 & -10/4 \\ 1/2 & 1 & 10/4 & 0 \end{pmatrix}。$$

因此，由用户评价给出的星级评分为

$$\boldsymbol{r} = \frac{\boldsymbol{Ke}}{4} = \begin{pmatrix} 7/40 \\ -13/120 \\ -16/15 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 0.175 \\ -0.1083 \\ -1.066 \\ 1 \end{pmatrix},$$

所以这些商品的排名（由高到低）为 $\{p_4, p_1, p_2, p_3\}$。

直接比较

利用平均星级得分，即式（10.2）来定义两种商品之间的“分差”，是一种可以采取的自然的方法，但此外还有着无数种其他方式来定义分差。例如，零售商可能对于商品之间的直接比较更感兴趣，这一比较是通过将 S_{ij} 设为 $U_i \cap U_j$ 中相比于 p_j 而言更喜欢 p_i 的人数比例来给出的。如果用户的星级评分可用的话，就能从中得到这样的比较，不过通过老式的市场调查或零售点数据，常常可以获得直接比较数据。确切而言，令

$$\delta_{ij} = \begin{cases} 1, \text{当用户 } h \text{ 相比于 } p_j \text{ 而言更喜欢 } p_i, \\ 0, \text{当用户 } h \text{ 相比于 } p_i \text{ 而言更喜欢 } p_j \end{cases} \tag{10.4}$$

并定义

$$S_{ij} = \begin{cases} \dfrac{1}{n_{ij}} \displaystyle\sum_{h \in U_i \cap U_j} \delta_{ij}(h), \text{若 } n_{ij} \neq 0, \\ 0, \text{若 } n_{ij} = 0 \end{cases} \tag{10.5}$$

如果将 $\boldsymbol{K}$ 重新定义为直接比较分差的斜对称阵的话，有

$$\boldsymbol{K} = [k_{ij}] = [S_{ij} - S_{ji}]。 \tag{10.6}$$

则由**第 120 页**上式（9.4）中的结论，将给出如下有关最佳“直接比较”的商品评分的命题。

最佳直接比较评分

通过直接比较可以得到的（在**第 118 页**所给的意义下）最佳的商品评分为重心

$$\boldsymbol{r} = \boldsymbol{Ke}/n, \tag{10.7}$$

式中，$\boldsymbol{K}$ 为式（10.6）所给出的直接比较分差的斜对称阵。

例如，假设有 10 名用户 $\mathcal{U} = \{h_1, h_2, \cdots, h_{10}\}$ 对 4 种商品 $\mathcal{P} = \{p_1, p_2, p_3, p_4\}$ 进行了直接比较，并且将式（10.4）中定义的二值比较结果归总为矩阵

$$\delta = \begin{array}{c} \\ h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \\ h_{10} \end{array} \begin{array}{c} \begin{array}{cccc} p_1 & p_2 & p_3 & p_4 \end{array} \\ \begin{pmatrix} 1 & & 0 & \\ 1 & & & 0 \\ 0 & & 1 & \\ 0 & 1 & & \\ & 0 & & 1 \\ 0 & 1 & & \\ & 1 & & 0 \\ 1 & & 0 & \\ & 0 & & 1 \\ 0 & & & 1 \end{pmatrix} \end{array}。 \tag{10.8}$$

举例来说，这个矩阵意味着用户 h_1 对商品 p_1 和 p_3 进行了比较，并选择了 p_1；而用户 h_5 对 p_2 和 p_4 进行了比较，并选择了 p_4，等等。（这些用户可以是同一个人——例如，用户 h_1 与 h_5 可以是同一个人。）当式（10.5）所定义的得分 S_{ij} 排列为矩阵时，所得矩阵为

$$\boldsymbol{S} = \begin{pmatrix} 0 & 0 & 2/3 & 1/2 \\ 1 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \end{pmatrix}。$$

因此，直接比较分差的斜对称阵 $\boldsymbol{K}$ 为

$$\boldsymbol{K} = \begin{pmatrix} 0 & -1 & 1/3 & 0 \\ 1 & 0 & 0 & -1 \\ -1/3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

而由式（10.7）给出的直接比较商品评分为

$$\boldsymbol{r} = \frac{\boldsymbol{Ke}}{4} = \begin{pmatrix} -2/12 \\ 0 \\ -1/12 \\ 3/12 \end{pmatrix}, \tag{10.9}$$

换而言之，商品的排名（由高到低）为 $\{p_4, p_2, p_3, p_1\}$。

直接比较、偏好图和马尔可夫链

另一种根据商品间直接比较来获得评分的方法，就是构造一个偏好图（preference graph）。这是一个对边进行了加权的有向图，其中图的结点代表 $\mathcal{P} = \{p_1, p_2, \cdots, p_n\}$ 中的商品，而由 p_i 到 p_j 的加权边，在某种意义下代表了一个人目前正在“使用”商品 p_i 时，他更喜爱商品 p_j 的概率。尽管可以通过无数的方式来设计用户调查从而确定这些概率，但它们也可以根据式（10.8）中那样的直接比较来构造。

1. 对于每种商品 p_i，列出所有对 p_i 进行过评价的用户 $\mathcal{H}_i = \{h_{i_1}, h_{i_2}, \cdots, h_{i_k}\}$。
2. 如果在 $\mathcal{H}_i$ 中有一名用户更喜爱商品 p_j，则画一根由 p_i 到 p_j 的边。

3. 与 p_i 到 p_j 的边相关联的权值（或概率）q_{ij}，等于 $\mathcal{H}_i$ 中更喜爱商品 p_j 的人数比例。也就是说，若 n_{ij} 是 $\mathcal{H}_i$ 中更喜爱商品 p_j 的人数，则

$$q_{ij} = \frac{n_{ij}}{\#\mathcal{H}_i} = \frac{n_{ij}}{k}。$$

这个偏好图定义了一条马尔可夫链[54，第687页]，而对 $\mathcal{P}$ 中的商品进行评分时，可以在这幅图上进行随机游走，并分析确定在每个结点（或商品）上所停留的时间比例，由此来获得评分。这实际上也就是**第 67 页**上第 6 章中所用的同样的过程，其思路是从数学角度来观察一名“随机购物者”，他在这个偏好图中永无休止地由一种商品换到另一种商品——如果购物者目前停留在商品 p_i 处，那么他之后将以概率 q_{ij} 移动到商品 p_j 处。商品 p_i 的评分值 r_i 就是该随机购物者在该商品处停留的时间比例。

如果偏好图具有足够的连通性，那么随机购物者在商品 p_i 上耗费的时间比例（即评分值 r_i）就是满足如下方程的向量 $\boldsymbol{r}$ 中的第 i 个元素。

$$\boldsymbol{r}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{r}^{\mathrm{T}}, \boldsymbol{r}^{\mathrm{T}}\boldsymbol{e} = 1, \text{其中}, \boldsymbol{Q} = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix}。$$

这个给出了我们评分的向量 $\boldsymbol{r}$，称为该马尔可夫链（或随机游走）的稳态概率向量。实际上，这个评分和排名的方法正是谷歌的 PageRank 的基础——请见参考文献［49］。

例如，考虑式（10.8）的 $\boldsymbol{\delta}$ 矩阵给出的 10 名用户 $\{h_1, h_2, \cdots, h_{10}\}$ 对 4 种商品 $\{p_1, p_2, p_3, p_4\}$ 的直接比较。由 $\boldsymbol{\delta}$ 中的数据可见

$$p_1 \text{ 被用户} \{h_1, h_2, h_3, h_4, h_6, h_8, h_{10}\} = \mathcal{H}_1 \text{ 所评价}$$
$$p_2 \text{ 被用户} \{h_4, h_5, h_6, h_7, h_9\} = \mathcal{H}_2 \text{ 所评价}$$
$$p_3 \text{ 被用户} \{h_1, h_3, h_8\} = \mathcal{H}_3 \text{ 所评价}$$
$$p_4 \text{ 被用户} \{h_2, h_5, h_7, h_9, h_{10}\} = \mathcal{H}_4 \text{ 所评价}$$

此外，由 $\boldsymbol{\delta}$ 还可以看出，在对 p_1 进行了评价的 7 名用户中，有 3 名喜欢 p_1，2 名更喜欢 p_2，1 名更喜欢 p_3，还有 1 名更喜欢 p_4。因此，在如图 10.1 所示的偏好图中有四条发自结点 p_1 的边（路径），其权值分别为 $q_{11} = 3/7$，$q_{12} = 2/7$，$q_{13} = 1/7$ 和 $q_{14} = 1/7$，因此如式（10.10）中 $\boldsymbol{Q}$ 的第一行就可以被确定下来。发自第二个结点 p_2 的边（路径）以及 $\boldsymbol{Q}$ 中第二行的元素值也可以类似推得，此时对评价过 p_2 的 5 名用户而言，没有人更喜欢 p_1，3 人更喜欢 p_2，没有人更喜欢 p_3，还有 2 人更喜欢 p_4，因此 $q_{21} = 0$，$q_{22} = 3/5$，$q_{23} = 0$ 且 $q_{24} = 2/5$。

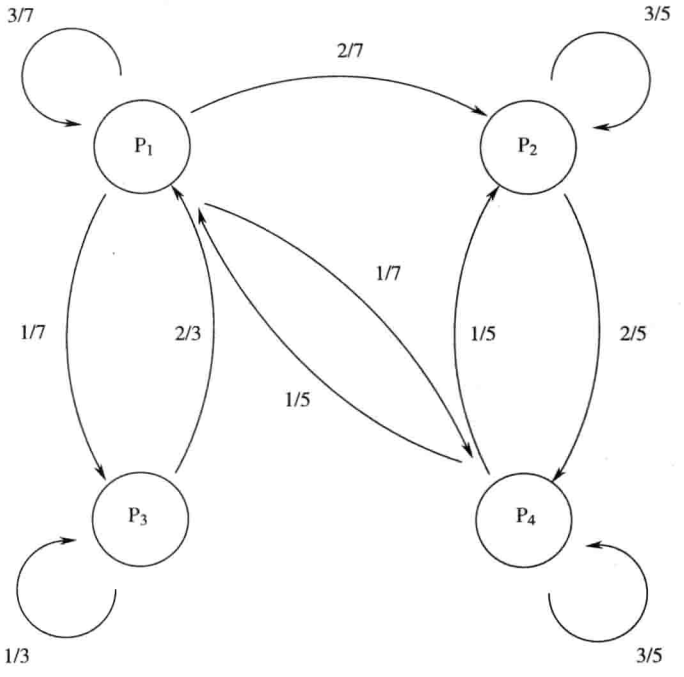


图 10.1　商品偏好图

如此继续，便完成了偏好图并获得矩阵

$$\boldsymbol{Q} = \begin{pmatrix} 3/7 & 2/7 & 1/7 & 1/7 \\ 0 & 3/5 & 0 & 2/5 \\ 2/3 & 0 & 1/3 & 0 \\ 1/5 & 1/5 & 0 & 3/5 \end{pmatrix} \tag{10.10}$$

的第三行和第四行。通过求解如下包含 4 个未知数 $\{r_1, r_2, r_3, r_4\}$ 的 5 个线性方程，便能获得马尔可夫评分向量 $\boldsymbol{r}$

$$\boldsymbol{r}^{\mathrm{T}}(\boldsymbol{I} - \boldsymbol{Q}) = 0, \boldsymbol{r}^{\mathrm{T}}\boldsymbol{e} = 1。$$

结果为

$$\boldsymbol{r} = \begin{pmatrix} 7/30 \\ 1/3 \\ 1/10 \\ 1/3 \end{pmatrix} \approx \begin{pmatrix} 0.2333 \\ 0.3333 \\ 0.1 \\ 0.3333 \end{pmatrix} \tag{10.11}$$

因此我们的商品排名（由高到低）为 $\{\{p_4, p_2\}, p_1, p_3\}$，其中 p_4 和 p_2 排名相同。

重心法与马尔可夫链方法对比

对于马尔可夫（或随机游走）评分方法而言，存在几个问题。首先，当状态较多（在现在的情况下就是商品种类较多）时，图中的连通性通常不充分，不能保证 $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{r}^{\mathrm{T}}$ 且 $\boldsymbol{r}^{\mathrm{T}}\boldsymbol{e} = 1$ 具有定义良好的解（即评分向量）。此时，要克服该问题，就需用将某些虚拟的信息或扰动，

以某种方式强加到图中，而这一虚拟信息必然对“真实”信息的纯洁性造成影响——请见参考文献［49］。因此，所得的评分和排名并非实际数据的真实体现。不过，这是使用马尔可夫链方法时必须接受的妥协；另一方面，重心法却并不需要数据具有特殊的结构，因此无需虚拟信息来产生评分。此外，即使能够通过某种方式对图进行扰动或修改，以保证存在一个定义良好的稳态概率向量，相比于相对极为简单的计算重心而言，计算 $\boldsymbol{r}$ 也是一件更繁重的任务。

结论

第 130 页上的重心法评分，即式（10.9）以及马尔可夫评分，即式（10.11）都来自相同的直接比较数据即式（10.8）。重心法将商品排名为 $\{p_4, p_2, p_3, p_1\}$，而马尔可夫法给出的排名为 $\{\{p_4, p_2\}, p_1, p_3\}$。仅根据这样一个人为的小型示例来作出结论，这种做法是危险的，但我们仍然可以看出重心法的排名和马尔可夫法的排名是具有相同水平的。由这个例子可以清楚看到，重心法排名显然更易于获得，因此，如果能够给出合适的信息，那就应该选用重心法。

数字——

24 =在 Netflix 为将其用户偏好预测改进 10% 而设立的百万美元赏金的竞赛中，BelKor 队险胜 Ensemble 队时比赛所剩余的时间（分钟）。

——世界标准时间 2009 年 7 月 26 日 18：18：28。

——netflixprize. com

第 11 章　处理平局

Amazon，Netflix，eBay 等公司征求了用户同意后，对他们的行为数据进行了采集，并由此累积了巨大的数据库，其中包含了用户对于商品和服务的评分记录。对这些数据进行分析的一个基本目的，就是要生成对象的排名，之后它可以被作为推荐系统的一部分。一般来说，产生排名的第一步，就是将用户评分转换为成对比较数据。有多种方法可以做到这一点，例如，请参见参考文献［33］以及第 10 章中的内容。每个转换都始于成对事物之间针锋相对的较量。例如，在用户对影片进行评分的数据库中，一个可能的成对比较矩阵中的元素 (i,j) 可以是影片 i 在较量中击败影片 j 的次数，而每当一名用户对这两部影片都进行了评价时，就会发生一次这样的较量。显而易见，在许多数据库应用中，将普遍存在平局的情况。

本章讨论输入数据中存在平手时的排名，而给出的结果中，有许多是参考文献［20］中的结果的汇编。在输出（最终排名）中同样可能出现平局——关于这一点的更多内容将在第 15 章中给出。尽管采用平局处理方式将对排名结果产生重要影响，但如何处理输入数据中的平局的问题却并未得到太多关注。许多排名方法简单地忽略掉平局。对于平局需要加以更为谨慎地考虑，对此有以下三个具有说服力的理由。

- 根据所考虑的应用，平分的情况可能会在数据中占据相当部分。对于大多数 NFL 赛季而言，有 9% ~13% 的比赛都以平局告终。在数据库应用中，平分的情况就更为普遍。例如，对用户评价的影片进行排名，当同一名用户对两部影片都进行了评价时，就进行了这两部影片的一次比赛，而所得评价较高的影片将赢得“比赛”，而在数据库中，常常可以发现有多达 50% 的“比赛”数据是平分。
- 诸如梅西法等某些方法已经考虑了平分的情况，但如科利法等其他方法则选择了忽略。因此，在比较不同方法给出的排名时，需要在所有方法中使用相同的平局处理方式，以使得比较更为公平。我们建议考虑而非忽略平局。当把平局考虑在内时，如果一支低排名的队伍与高排名的队伍战平，则低排名队伍将因为奋战而得到奖励，但高排名队伍则将受到惩罚。相比于将平局看做无事发生而言，这一方式直观上而言更具有吸引力。
- 在某些体育项目中，平局很少出现，甚至可能由于加时或决胜规则而根本不会出现平局。在这种项目中，有时可能会出现两支队伍在比赛的大部分时间里都相当势均力敌，但其中一支却可能由于最后一分钟的爆发而赢得比赛。在**第 142 页**上，我们引入了平局，来处理这一类接近平手的情况以及如分差很小这样的其他接近平手的

情况。我们将证明，这些引入的平局能够改善排名结果的预测能力。在诸如大学橄榄球比赛的 BCS 排名这样的应用中，排名结果的小幅变动也可能对各所大学产生重大影响，因为其收入与是否能获得碗系列赛的入场券密切相关。

输入平局与输出平局

一般来说，在有关排名的文献中，平局几乎不受关注。在此我们区分两类平局。输入平局指输入数据中的平局（即以平分告终的比赛），而输出平局指输出数据中的平局（即一个排名算法给出的结果中存在排名相同的情况）。对于输出平均，已经开展了若干工作，特别是在排名聚合的方面[3, 4, 23, 27, 50]。排名聚合与排名有关，但又有所不同。排名聚合又称总排名，它给出了一个单一的排名，该排名是其他数个排名的归总。更多有关排名聚合的内容请见第 14 章和第 15 章。相对地，在本章中，我们更感兴趣的则是输入平局以及它们对排名方法而非排名聚合方法的影响。

加入平局

对于一个数据中包含平局的排名应用而言，我们必须非常小心地对不同方法给出的排名进行比较。如梅西法和 OD 法等，某些方法能够十分自然地处理平局，而其他方法如科利法和各种马尔可夫方法的变体等，却都忽略了平局。所有这些方法要么已经结合了平局信息，要么可以经修改后利用平局信息。因此，为了能够进行方法之间公平的比较，对于每种方法，我们都必须给出两个模型：使用平局时的模型，以及忽略平局时的模型。

科利法

回忆一下第 3 章的内容，科利法可以用一个线性系统 $\boldsymbol{Cr} = \boldsymbol{b}$ 简洁地加以总结，其中，$\boldsymbol{r}_{n\times 1}$ 为未知的科利评分向量，$\boldsymbol{b}_{n\times 1}$ 为右向量，定义为 $b_i = 1 + \frac{1}{2}(w_i - l_i)$，而 $\boldsymbol{C}_{n\times n}$ 为科利系数矩阵，定义为

$$\boldsymbol{C}_{ij} = \begin{cases} 2 + t_i, i = j, \\ -n_{ij}, i \neq j。 \end{cases}$$

式中，标量 n_{ij} 为队伍 i 和 j 相互交手的次数；t_i 为队伍 i 进行的总比赛场数；w_i 为队伍 i 获胜的次数；l_i 为队伍 i 失利的次数。

既然科利法仅考虑了胜或负的情况，平局实际上被当做无效事件而忽略，因此，我们称上述模型为无平局的科利法。为了加入平局数据，我们必须对上述模型加以修改，来得到一个有平局的科利法。

在将平局结合到科利法之前，我们需要思索一下平局的含义。对平局的一种解释认为，

如果这两支队伍再次相遇，那么它们都有相等的机会能够获胜，或者我们可以互补地说，这两支队伍都有相等的可能来输掉比赛。因此，要包含平局，我们可以将队伍 i 跟队伍 j 之间的一场平局计为 2 场比赛，并将普通的 1 分的权值均分给这两场虚拟比赛。在一场虚拟比赛中，队伍 i 以权重 1/2 击败了队伍 j，而在另一场中则是队伍 j 击败队伍 i。对平局的这一解释不会改变向量 $\boldsymbol{b}$，同时以一种显然的方式修改了矩阵 $\boldsymbol{C}$（即对于队伍 i 和 j 之间的一场平局，c_{ij} 和 c_{ji} 均减 1，而 c_{ii} 和 c_{jj} 均加 1）。平局的这一解释所带来的一个很好的结果，便是科利法的守恒性仍然得到了保持，即所有科利评分之和为 $n/2$。

梅西法

回忆一下第 2 章的内容，梅西法可以被简洁地归纳为线性系统

$$\boldsymbol{Mr} = \boldsymbol{p},$$

其中梅西系数矩阵 $\boldsymbol{M}$ 可通过公式 $\boldsymbol{M} = \boldsymbol{C} - 2\boldsymbol{I}$ 与科利矩阵 $\boldsymbol{C}$ 联系起来，因此

$$\boldsymbol{M}_{ij} = \begin{cases} t_i, & i = j, \\ -n_{ij}, & i \neq j。 \end{cases}$$

梅西右向量 $\boldsymbol{p}$ 是一个累积分差向量，即 p_i 为队伍 i 在与所有对手交锋时所得的总得分与总失分之差。由于 $\operatorname{rank}(\boldsymbol{M}) = n - 1$，因此我们对 $\boldsymbol{M}$ 进行了修正，以保证它具有非奇异性。进行修正时，$\boldsymbol{M}$ 中的任意一行被替换为全 1 行，而 $\boldsymbol{p}$ 中的对应元素设为 0。这个新的约束使得评分和等于 0。按照梅西的建议，我们将这一非奇异性修正用于最后一行，给出了一个修正后的线性系统，记为 $\overline{\boldsymbol{M}}\boldsymbol{r} = \overline{\boldsymbol{p}}$。

上述梅西法自然地考虑到了平局的情况。对于队伍 i 和队伍 j 之间的一场平局，元素 m_{ij} 和 m_{ji} 都加 1，但分差向量 $\boldsymbol{p}$ 中的数据不变。因此，我们称上述方法为有平局的梅西法。

为能恰当地对各种方法进行比较，我们还需要有一个可以比较的无平局的梅西法，它在产生评分时不考虑平局。为了以最少的改动来做到这一点，我们让梅西法忽略所有的平局，即当出现队伍 i 和队伍 j 之间的平局时，我们就不对矩阵 $\boldsymbol{M}$ 和向量 $\boldsymbol{p}$ 进行更新。这类似于无平局的科利法，在那里，平局被视为无效事件。

马尔可夫法

第 6 章中的马尔可夫法的关键在于投票。一支较弱的队伍会为每支它所遭遇过的较强的队伍投出选票，而由于有几种投票方法，因此马尔可夫法有数个变体，在本章中，我们考虑如下三种。

1. 一支队伍为每个击败它的队伍投出一票。

2. 一支失利的队伍按落后的比分投出相应的票数。
3. 每场比赛中的胜者和负者都按各自的失分投出相应的票数。

不管采用何种投票方法，最后得到的结果都是一个队伍-队伍矩阵，并经过了行归一化以确保矩阵为随机矩阵，它给出了一条马尔可夫链的转移矩阵。这条链的稳态向量即为马尔可夫评分向量。正如在第 6 章中所讨论的，当在由该投票矩阵所定义的图上进行随机游走时，访问某支特定队伍所耗费的时间比例便对应了稳态向量中的元素。评分越高（即稳态向量中的值越高），对该队的访问就越频繁，而该队的相对重要性也越高。

投票方法是马尔可夫方法中讨论平局的一个焦点，这一点很快就将变得明了起来。因此，我们将依次探讨每种投票方法。

1. （*失利者按胜负投票*）正如最初所设计的那样，这种马尔可夫方法忽略了平局。但是，要将平局考虑进来也十分容易。实际上，我们将借用有平局的科利法中的处理手段——对于每个平局，我们令每支队伍都投出半票。
2. （*失利者按分差投票*）正如最初所设计的那样，这种马尔可夫方法同样忽略了平局。不幸的是，不像上面那种投票方式对应的马尔可夫方法，这时的平局并不容易处理。实际上，如果两支战平的队伍交手并未超过一次的话，平局的结果就无法保留。如果这两支打成平手的队伍相互较量了不止一次的话，我们就能够处理平局。假设队伍 i 和 j 交手了两次：在第一场比赛中，队伍 i 以 4 分的优势击败了队伍 j，而在第二场比赛中双方战平；那么两队之间多次交手的数据，通常可以用平均分差而非累积分差的形式放入马尔可夫矩阵中。因此，在这个例子里，对于有平局的马尔可夫法而言，矩阵中的元素 (i,j) 的值为 2，而在无平局的马尔可夫法中，这个值为 4。但是，由于各对队伍之间的比赛场数并不相等，因此我们建议不要使用这种投票方法，对于有平局的数据集而言，它给出的对比是不公平的。
3. （*胜负双方均按失分投票*）正如最初所设计的那样，这种马尔可夫方法可以自然地将平局情况包括在内。实际上，要忽略平局的话，我们必须将所有的平局从数据集中删除。

OD 法、基纳法和埃洛法

类似于本节中的这三者，某些排名方法可以自然地处理平局。第 7 章中的 OD 法为每支队伍计算两个评分，即一个攻击评分和一个防守评分，它们随后将被聚合为每队的一个单一评分值。正如式（7.7）和式（7.8）所描述的那样，这两个 OD 评分向量 $\boldsymbol{o}$ 和 $\boldsymbol{d}$ 是通过交替求精过程

$$\boldsymbol{o}_k = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{d}_{k-1}^{\div}, \boldsymbol{d}_k = \boldsymbol{A}\, \boldsymbol{o}_k^{\div} \quad (k = 1,2,3,\cdots)$$

计算而得，其中，$\boldsymbol{A} = [a_{ij}]$ 为“得分矩阵”；a_{ij} 为队伍 j 对阵队伍 i 时所得的分数（如果两队交手多次，则取平均值；而如果两队从未交手，则置 $a_{ij} = 0$）。收敛后，$\boldsymbol{o}$ 和 $\boldsymbol{d}$ 被聚合生成一个总体评分向量 $\boldsymbol{r}$。尽管有多种可选的聚合方法，但 $\boldsymbol{r} = \boldsymbol{o}/\boldsymbol{d}$ 这个简单规则在实际中效果不错。

由于 OD 法使用的是得分数据，因此它自然可以处理平局。因此，我们称上述方法为有平局的 OD 法。为了获得无平局的 OD 法，只需将所有平局收场的比赛比分置为 0。

类似地，第 4 章的基纳法的大多数变体也同样能够自然地处理平局，因为它们同样是构建于得分数据之上的。平局已经内建于第 5 章中的埃洛评分体系之中，考虑到在足球和国际象棋比赛这两个埃洛体系的主要应用中，平局的情况频繁出现，因此这一点十分重要。

扰动分析给出的理论结果

在本节，我们将使用扰动分析中的工具，来量化平局对于科利排名可能造成的影响。尽管对其他方法也可以导出类似结果，但为了避免赘述，这里仅用科利法这一种方法来说明我们的主要论点。

我们将使用一个赛季的数据来开始我们的分析，截至目前，数据中尚未出现平局。接下来，我们将比较无平局科利法和有平局科利法给出的排名结果，只不过现在新增了一场比赛，*而这场比赛的结果是场平局*。在这一扰动之前，两种方法都给出了相同的排名 $\boldsymbol{r} = \boldsymbol{C}^{-1}\boldsymbol{b}$。进行了扰动后，无平局的科利法忽略了这场新增的比赛，因此排名保持不变；另一方面，有平局的科利法的排名将因为这一扰动而发生改变。幸运的是，我们可以利用扰动分析精确地考察这一改变。特别地，该平局表现为对科利系数矩阵的一个秩一更新，而右向量 $\boldsymbol{b}$ 则保持不变。

令 $\tilde{\boldsymbol{C}}$ 表示扰动后有平局科利法的科利矩阵。这一矩阵 $\tilde{\boldsymbol{C}}$ 通过

$$\tilde{\boldsymbol{C}} = \boldsymbol{C} + (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^{\mathrm{T}}$$

与扰动前的矩阵 $\boldsymbol{C}$ 联系在一起。利用谢尔曼-莫里森秩一更新（Sherman-Morrison rank-one update）公式和科利矩阵的若干 M-矩阵的性质[54]，我们可以用 $\boldsymbol{C}$ 来表示 $\tilde{\boldsymbol{r}}$。具体而言，有

$$\begin{aligned}\tilde{\boldsymbol{r}} &= \tilde{\boldsymbol{C}}^{-1}\boldsymbol{b} \\ &= [\boldsymbol{C} + (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^{\mathrm{T}}]^{-1}\boldsymbol{b} \\ &= \boldsymbol{r} - \left(\frac{r_i - r_j}{1 + [\boldsymbol{C}^{-1}]_{ii} - 2[\boldsymbol{C}^{-1}]_{ij} + [\boldsymbol{C}^{-1}]_{jj}}\right)\boldsymbol{C}^{-1}(\boldsymbol{e}_i - \boldsymbol{e}_j)。\end{aligned} \tag{11.1}$$

由上式，我们可以得出两个结论：

- 仅仅一场平局，就可能影响到每支队伍的评分。这是由于 $\boldsymbol{C}^{-1}$ 的第 i 列和第 j 列几乎总是完全稠密的。
- 假设在扰动前，队伍 i 的排名在队伍 j 之前，因而 $r_i - r_j > 0$。那么在增加了队伍 i 与队伍 j 之间的这场平局后，队伍 i 的评分 $\tilde{r}_i$ 降低，而队伍 j 的评分 $\tilde{r}_j$ 升高。这是 $\boldsymbol{C}^{-1}$ 的对角优势性和非负性所带来的结果。

接着，我们将考虑队伍 i 在扰动后的评分下降，是否足以导致排名也发生下降。就数学方面而言，如果 $\tilde{r}_i < \tilde{r}_{i+1}$，则队伍 i 的排名将由于这场平局而下降一位。为了对队伍 i 扰动后排名下降的条件加以量化，我们定义 $\varepsilon = r_i - r_{i+1}$ 为扰动前队伍 i 和 $i+1$ 的评分差。进一步分析式（11.1），我们可以得出结论，即队伍 i 的排名将下降一位的条件为

$$\tilde{r}_i < \tilde{r}_{i+1} \Leftrightarrow \varepsilon < \frac{(r_i - r_j)([\boldsymbol{C}^{-1}]_{ii} - [\boldsymbol{C}^{-1}]_{ij} - [\boldsymbol{C}^{-1}]_{i+1,i} + [\boldsymbol{C}^{-1}]_{i+1,j})}{1 + [\boldsymbol{C}^{-1}]_{ii} - 2[\boldsymbol{C}^{-1}]_{ij} + [\boldsymbol{C}^{-1}]_{jj}}。$$

由这个式子，我们可以得出另外两个结论：

- 若队伍 i 和队伍 j 的评分相差很大（即 $r_i \gg r_j$），则队伍 i 的排名很可能下降。
- 另一方面，队伍 i 和队伍 j 的评分越接近，则由于平局而造成队伍 i 和 j 排名发生变化的可能性越小。

在本节中，我们量化了单独一场平局对于无平局和有平局的这两种科利法的影响。总的来说，我们发现，*单单一场平局的出现，就可能同时改变评分和排名向量*。在实际应用中，常常会出现远多于一场的平局，这也进一步强化了我们的理论发现，即对平局的处理可以对科利法所给出的评分和排名产生影响。对其他排名方法的扰动分析也给出了类似结论。本节中的理论，是对下一节中实验的合理性的一个事先检查，而实验结果也回应了我们的理论发现。

真实数据集上的结果

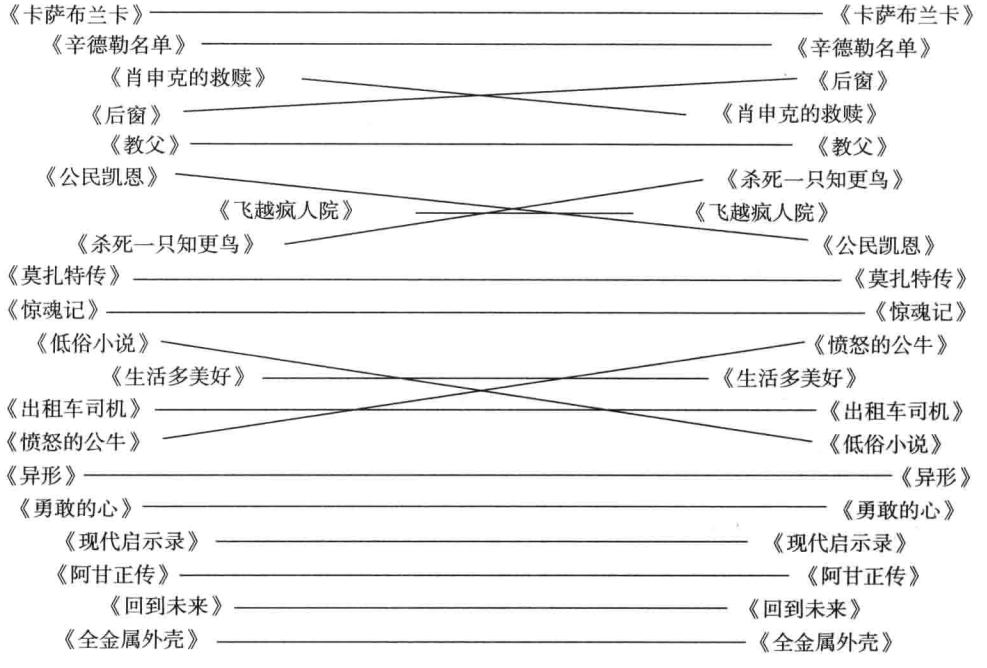
在本节中，我们将从扰动分析这一理论工具转到现实的应用。在此，我们有一个简单的目标——说明平局的处理方式能够对由实际数据所得的排名结果造成重大影响。

影片排名

我们的第一个应用是在影片上进行的。随着数据存储能力的增加，对于诸如影片、书籍或商品这类数据库数据进行排名的需求也在不断增加。如 Netflix 或 Amazon 的推荐系统，利用数据库中的以往购买情况以及用户评分来为用户提供推荐，而推荐列表则是通过聚类

和排名算法获得的[33]。由于我们已经知道，平分的情况在数据库应用中非常普遍[27]，因此对平局的恰当处理，是获得准确排名结果的一个关键。

在本节中，我们将采用一个称为 MovieLens，并得到了大量研究的数据集，其中包含了 943 名用户对 1 682 部影片给出的 100 000 个整数评分（1 到 5 分）。这一数据生成了一个 1 682 ×1 682 大小的矩阵。该矩阵十分稀疏，因为大多数用户都仅对这些影片中的一小部分进行了评价。当一名用户同时对两部影片 i 和 j 进行评分时，这两部影片之间就进行了一场“比赛”，而用户给出的评分就是这两部影片的“得分”。按照这些定义，我们发现在 MovieLens 数据集中，有 32% 的用户评分出现了平局的情况。我们对 MovieLens 数据集中所有 1 682 部影片进行了排名，并选出其中 20 部受欢迎的影片，得到了如下的结果图。图中左侧显示了考虑平局时的排名，而右侧的排名则忽略了所有的平局。由于平局数据差不多占了三分之一，因此如我们所预期的那样，这两个排名是不一样的。



20 部影片的科利排名：有平局与无平局对比

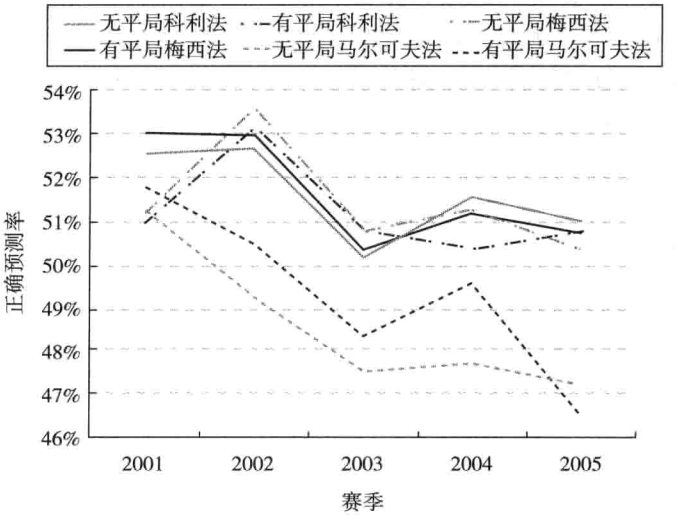
“忽略平局”策略的支持者声称，对平局进行建模并不值得，因为它们或者很少出现，或者仅影响排名较低者。但是，上图表明，平局可以对任何位置上的排名造成影响，包括前几位的排名。显然，平局的建模对于排名结果有着巨大影响。

NHL 冰球队排名

在第二个应用中，我们转向体育项目。在此，我们将对 NHL 冰球比赛进行研究，这是一项得分较低的运动，其中经常会出现平局。平局通常占到了一个赛季的比赛结果中的 9% ~13%。

正如 MovieLens 数据集一样，我们发现，处理平局或忽略平局，将给出两个不同的排名。

一个自然的问题便是：考虑平局的方法是否给出了更好的排名？这就要看你是如何定义“更好”了。通常而言，“更好”与预测能力有关——即是否具有先见之明。这时的问题是：在预测将来比赛的结果时，考虑平局的方法是否优于不考虑平局的方法？“更好”还可以与追溯能力有关——即是否具有事后之明。而这时的问题就变为：在与以往比赛结果进行比较时，考虑平局的方法是否优于不考虑平局的方法？下图给出了在 NFL 2001—2005 赛季中，考虑平局与不考虑平局的科利法、梅西法和马尔可夫法所给出的事前预测性能。



科利法、梅西法和马尔可夫法对 NFL 的事前预测性能：有平局与无平局对比

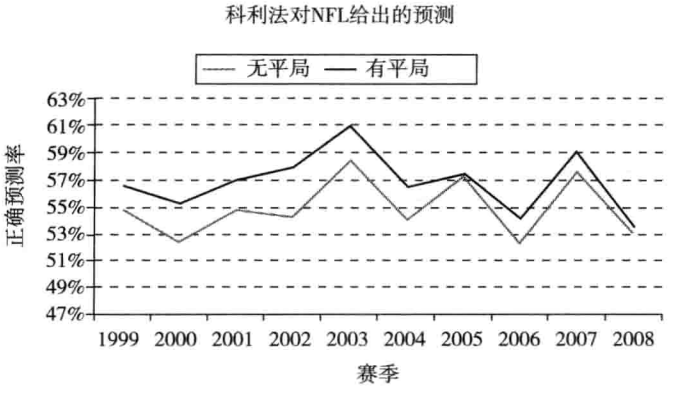
这幅图并未给出有平局对无平局的争论中的胜者[⊖]。我们在几个其他项目的不同赛季的数据上也进行了测试，包括篮球和橄榄球，并最终得到如下结论。*尽管在排名方法中考虑平局仅在某些时候能稍微提高预测性能，但至少它几乎不会带来任何坏处。*对于平局的这一不温不火的认可，促使我们对诱导平局的影响进行研究，这正是下一节的主题，并带来了这场有平局对无平局的争论中最为有趣的发现。

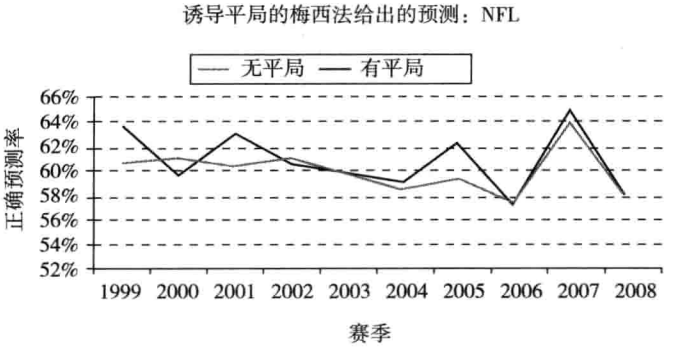
诱导平局

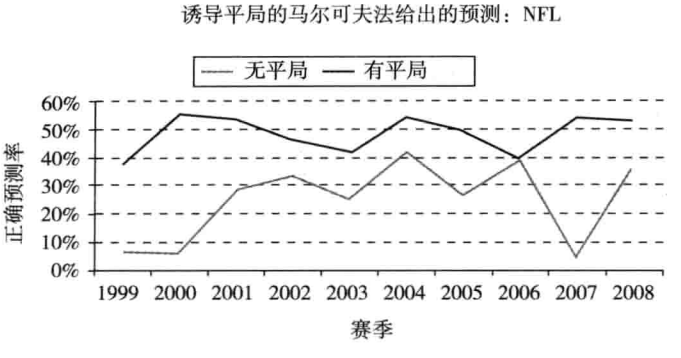
尽管如冰球或足球这样的某些体育项目允许出现平局，但其他项目却会利用加时或突然死亡法等规则，来打破常规比赛时间内的平局结果。此外，还有一些非加时阶段的事件，它们看起来简直就是完全随机地决定了获胜者：两支队伍在几乎整场比赛的时间内都在平局或几乎是平局的局面上僵持着，然后突然有某个最后一分钟的关键先生跳出来，并打破了均势。

⊖ 我们进行了更为深入的探究，考察了这两类方法在哪些排名位置上表现出了不同的预测能力，并试图去确定是否有某个方法在对排名最靠前的位置进行预测时做得更好却在较后的位置上表现较差，不过在这些方面，我们也未能就这场有平局对无平局的较量得到清晰的结论。

在篮球中，对于这样的英雄人物甚至还有一个专门的词语——“压哨制胜球”，指在比赛的最后一秒投出的、使球队获胜的那一球。在这种情况下，我们有理由认为，就这两支队伍的实力而言，平局将是更为贴切的衡量结果。因此，在这一节中，我们将在这些情况下诱导出平局，并观察这一改变所带来的影响。例如，在 NFL 橄榄球赛中，我们为每场最终分差为 3 分或更少的比赛诱导出平局。下面的三幅图给出了在 NFL 1999—2008 这 10 个赛季中，诱导平局对于科利法、梅西法和马尔可夫法这三种排名方法所造成的影响。

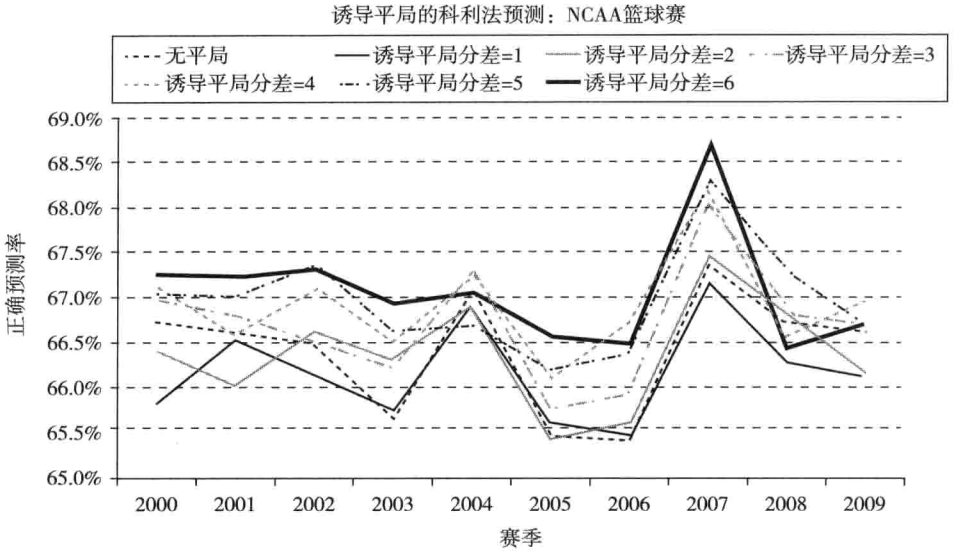






以上三幅图表明，诱导平局明显改善了这三种排名方法的预测性能。对于其他的排名方法以及如大学橄榄球赛等其他的数据集，也得到了类似的结果。

我们还研究了诱导平局对大学篮球赛的影响。虽然橄榄球赛有一个相当显然的、应该诱导出平局的分差（即橄榄球中的一个三分球），篮球比赛中的分差却不这么明显，因此我们将给出诱导平局的分差在 0 分到 6 分之间变动。科利法所得的结果如下。



无平局的结果用黑色虚线给出，而其余线条则用来表示所有诱导平局方法的结果。实线的颜色深浅不一，其中分差最小的（1 分）诱导平局法用浅灰表示，而分差最大的（6 分）用黑色表示。大多数分差所给出的诱导平局方法都好于对应的无平局版本。尽管难以确定哪个分差能给出最好的预测性能的改善，但我们的确发现了一个趋势，即大分差通常优于小分差。我们将在今后对诱导平局法的最优分差进行更精确的研究。

总之，本节的实验表明，在这场有平局对无平局的辩论中，有一个显然的胜者。就预测而言，诱导平局方法一般优于无平局的方法。

总结

由于在现代计算中，出现了大数据集的可用性以及对数据排名的频繁需求，也由于我们或许希望应用那些输入数据中可能包含平局的体育排名方法，并希望考虑平局对这些排名结果所具有的潜在影响，因此我们必须谨慎地考虑这些包含了平局处理的方法的使用。本章讨论了几种在常用的排名方法中包含平局的手段，此外，还介绍了诱导平局的概念，它能够改进排名的预测性能。

不过仍存在其他问题。例如，对于给定的体育项目而言，最有效的诱导平局的方式应该是怎样的？对于类似足球这样的得分较低的项目，诱导平局是合理的，在这样的比赛中，带来胜利的进球常常发生于常规比赛时间内的最后几分钟。当然，这又带来了一个问题，即到底是多少分钟应该被视为一场比赛的“最后几分钟”。在篮球比赛中，失利的球队会采用犯规战术，以希望对方错失罚球，从而在不消耗多少剩余比赛时间的情况下减少分差，当出现这种情况时，就可以认为比赛进入尾声了。你也许会问，在这种情况下，使用诱导平局是否仍然有利，而且可想而知地，应该怎样自动地对所有比赛作出这一进入尾声的判断，因为靠人工来一场比赛接一场比赛地加以评判显然是不现实的。同样也可以对其他的统计数据进行分析——例如，我们可以将每次控球时的得分而非比分结合到排名方法以及对平局的评判之中。

此外，是否存在着其他一些数据集，在这些数据集中，我们可以预计平局将不那么重要？是否对有些数据集，平局将对某些特定的排名位置造成影响？平局还会对哪些其他的应用和方法有用？

数字——

6 =一支 NFL 橄榄球队在一个赛季中最多的战平次数。

——芝加哥熊，1932 年。

24 =一支 NHL 冰球队在一个赛季中最多的战平次数。

——费城飞人，1969—1970。

2 =一支 NFL 冰球队在一个赛季中最少的战平次数。

——波士顿棕熊，1929—1930。

——en. wikipedia. org，couchpotatohockey. com

第 12 章　加入权重

“着眼于未来不仅能让我们避免囿于现在，也能帮助我们重新审视过去。”[⊖]——艾瑞克·霍弗（Eric Hoffer，1902—1983）

本章的主题就是重新审视过去。实际上，这也是本书中第一个更接近于文艺而非科学的章节。在确定排名的匠心之中，包括了根据专家或应用所特有的信息来定制某个方法的能力。考虑一下这个情况：队伍 A 在赛季初输了两场比赛，而在剩下的比赛中保持不败；而队伍 B 自赛季开始以来就保持不败，直到最后两场才出现失手。按照你的意见，哪支队伍的排名应该更高？大多数人都选择了队伍 A，因为“赛季开初并不重要”。这一直观的想法，便启发了本章的数学内容。

截至目前所介绍的所有模型中（埃洛法可能是一个例外），每场比赛都被赋予了平等的权重。尽管乍一看来，每一场比赛都一样重要的说法似乎很公平，但它却可能造成不合理的建模。例如，在体育项目中，巡回赛中的赛季末的比赛难道不应该比赛季前或赛季初的比赛具有更高的权重吗？状态正火的队伍难道不应该比不在状态的队伍更可能获胜吗？对于网页，更新得更为频繁的网页可能应该更加重要。因此，试图将这些时间上的考虑加入模型之中，是一个自然而然的想法。

加权的想法

尽管我们在本章中强调按时间加权，而任何类型的加权都是可能的。例如，我们可以对客场获胜进行加权，以超过在主场的胜利；或者我们可以对战胜某些已知的对手加权。这正是加权的一个特点——它十分灵活，足以包括任何你想象得到的进行加权的理由。

四种基本的加权方案

在大多数模型中，对比赛进行加权十分容易。在此我们介绍四种基本加权方案，尽管此外还有着无数的可能选择或这些方案的变体。这四种方案是线性函数加权、对数函数加权、指数函数加权以及阶跃函数加权。为了说明加权有多么容易和自然，我们利用马尔可夫模型

⊖ 译文引自 http：//article. yeeyan. org/bilingual/95241（2013 年 1 月 19 日）。——译者注

来对加权进行演示。此时，加权马尔可夫投票矩阵 $\bar{\boldsymbol{V}}$ 中的元素 $\bar{v}_{ij}$ 由

$$\bar{v}_{ij} = w_{ij}v_{ij}$$

计算得到，其中，w_{ij} 是对应于队伍 i 与 j 之间比赛权重的标量；而 v_{ij} 为队伍 i 向队伍 j 所投出的票数（根据第 6 章中马尔可夫模型的不同投票方式所决定）。权重 w_{ij} 由加权函数确定。我们给出的四种基本加权函数如图 12.1 所示，并依次介绍如下。再次地，我们将具体谈论时间加权，但任何加权方法或加权方法的组合都同样适用。

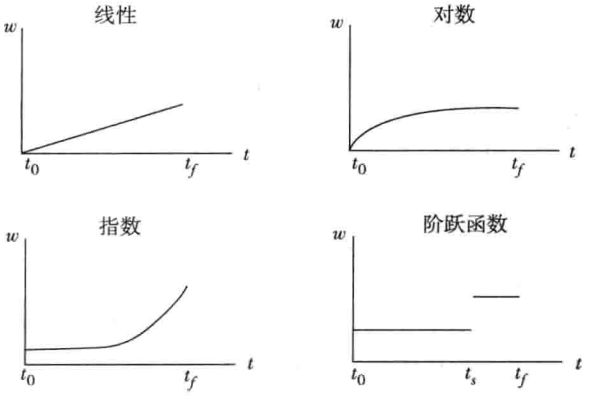


图 12.1　四种基本加权函数

对于线性加权方案，

$$w_{ij} = \frac{t - t_0}{t_f - t_0}。$$

分子表示相对于赛季开始的比赛时间，具体来说，就是比赛进行时，自 t_0 时刻，即赛季开始以来的天数；分母是整个赛季的总天数，即由赛季最后一场比赛的时间 t_f 减去赛季开始的时间 t_0 。

如果我们希望以一种缓慢上升的方式对后期的比赛进行加权，可以使用对数加权方案如下

$$w_{ij} = \log\left(\frac{t - t_0}{t_f - t_0} + 1\right)。$$

另一方面，如果我们希望急剧提高赛季末比赛的权重，则可以使用指数加权方案

$$w_{ij} = e^{\frac{t-t_0}{t_f-t_0}}。$$

本章所介绍的第四种加权方案使用了阶跃函数。一种可能的阶跃函数，便是对最后 2 周（或3周、4周或k周）中的比赛赋予比其他比赛更高的权值。一种给予赛季后期比赛以额外权

值的方法，就是简单地将对应的投票数乘以 2（3 或 4，等等）。这种方式的效果相当不错，特别对于篮球比赛更是如此，正如卢克·英格拉姆在其硕士论文[41]中的实验所表明的那样。此时，

$$w_{ij} = \begin{cases} 1, & \text{队伍 } i \text{ 和队伍 } j \text{ 之间的对决在 } t_s \text{ 之前进行}, \\ 2, & \text{其他}, \end{cases}$$

其中，t_s 为赛季中的某个特定时间，例如，联盟锦标赛之前的 1 个月。当然，同样可以很容易地实现多阶梯函数。这正是戴维逊学院本科生埃里希·克罗伊策在 2009 年三月疯狂锦标赛中所做的事情[19]。他给出了一个阶梯函数，其中比赛的权重随时间的推移，每两周将提高一次。他的双周间隔阶梯函数在科利法中取得的结果是如此之好，以至于获得了 1420 分的 ESPN 评分（最高可能得分为 1600 分），并在当年提交到 ESPN 锦标赛竞猜大赛的 4600 万个结果中，排在了 97.3 的百分位上。

时间加权方案中记法的总结

w_{ij}　赋予队伍 i 和队伍 j 之间比赛的权值

t_0　赛季开始的时间（例如，赛季的第一天为 $t_0 = 0$）

t_f　赛季最后一场比赛的时间

t_s　阶跃权值发生改变的赛季中的某个特定时间

t　当前比赛的时间

以下是对队伍 i 和队伍 j 在时刻 t 的比赛进行加权的四种简单方法。

线性　$w_{ij} = \dfrac{t - t_0}{t_f - t_0}$

对数　$w_{ij} = \log\left(\dfrac{t - t_0}{t_f - t_0} + 1\right)$

指数　$w_{ij} = e^{\frac{t - t_0}{t_f - t_0}}$

阶跃　$w_{ij} = \begin{cases} 1, & \text{队伍 } i \text{ 和队伍 } j \text{ 之间的对决在 } t_s \text{ 之前进行}, \\ 2, & \text{其他} \end{cases}$

现在我们讨论将加权方案应用于特定排名方法的细节。

加权梅西法

对于梅西模型，一个赛季中的每场比赛都表示为原始的梅西线性系统 $\boldsymbol{Xr} = \boldsymbol{y}$ 中的一行。每场比赛——因而也就是梅西系统中的每一行——可以利用任一选定的加权方案，根据它在

赛季中的发生时间来进行加权。此时，我们不再是将 $\boldsymbol{Xr} = \boldsymbol{y}$ 作为一个最小二乘问题来求解，而是作为一个加权最小二乘问题来求解。首先，产生一个每场比赛对应权值的向量 $\boldsymbol{w}$，同样的，这些比赛可能是按照相对于赛季开始日的比赛时间来进行加权的，以使得赛季后期的比赛更为重要。下一步，$\boldsymbol{w}$ 被变换为一个对角阵 $\boldsymbol{W}$，$\boldsymbol{w}$ 中的元素分布于这个矩阵的对角线上。最后，求解加权正规方程[36]

$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\boldsymbol{r} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y},$$

以得到加权后的原系统的唯一加权最小二乘解 $\boldsymbol{W}^{1/2}\boldsymbol{X}\boldsymbol{r} = \boldsymbol{W}^{1/2}\boldsymbol{y}$。

加权科利法

对科利法进行加权，将使得科利矩阵 $\boldsymbol{C}$ 的非对角元素不再是表示队伍之间交手次数的整数值，而是与每场比赛的重要性相应的权值。例如，在未加权的情况下，若 $c_{ij} = 3$，则意味着队伍 i 和队伍 j 相互交手了三次，而在加权的情况下，c_{ij} 则是这三次比赛的权值之和。类似地，右向量 $\boldsymbol{b}$ 也不再是总获胜次数减去总失利次数，而是加权后的值。从编程的角度来说，这一加权修改只是对算法的一个简单调整。

加权基纳法

一旦选中了一个加权方法后，加权的基纳矩阵 $\overline{\boldsymbol{A}}$ 就根据公式

$$\overline{a}_{ij} = w_{ij}a_{ij}$$

给出，其中，w_{ij} 是队伍 i 和队伍 j 之间比赛的权重；而 a_{ij} 为第 4 章中为基纳法所选择的统计数据。一旦得到这一加权基纳矩阵，则基纳法的剩余步骤便照原样进行。由于需要基纳矩阵为非负，因此我们必须使用非负的权值 w_{ij}。

加权埃洛法

埃洛法中的 K 因子是该方法内建的加权机制。请回忆一下在第 5 章中，对于足球比赛，K 因子被用来使世界杯决赛圈比赛的权值（$K = 60$）大于世界杯预选赛的权值（$K = 40$）大于其他锦标赛的权值（$K = 30$）。

加权马尔可夫法

一旦选定了一个加权方法后，加权马尔可夫投票矩阵 $\overline{\boldsymbol{V}}$ 就根据公式

$$\overline{v}_{ij} = w_{ij}v_{ij}$$

给出，其中，w_{ij} 是队伍 i 和队伍 j 之间比赛的权重；而 v_{ij} 是队伍 i 利用第 6 章中所介绍的投票方法中的一种来给队伍 j 投出的票数。一旦得到这一加权投票矩阵，则马尔可夫模型的剩余步骤就可按原样进行，即将加权投票矩阵 $\overline{\boldsymbol{V}}$ 进行行归一化，并且必要的话，对全胜的队伍进行处理。此时便得到了随机马尔可夫矩阵 $\boldsymbol{S}$，而其稳态向量即评分向量 $\boldsymbol{r}$ 便可求得了。

加权 OD 法

OD 法加权的实现几乎与马尔可夫法中完全一样，即在 OD 法的一个预处理步骤中完成加权。具体来说，OD 法的 $\boldsymbol{P}$ 矩阵中的元素 p_{ij} 经适当的加权方案所调整，然后在这个加权 $\boldsymbol{P}$ 矩阵上，OD 法照原样进行。

加权的差距法

时间加权可以很容易地被结合到第 8 章中的排名差距法和评分差距法当中去。实际的算法无需改变，而只需改变输入的数据矩阵 $\boldsymbol{D}$，即在规范化之前根据**第 147 页**上所讨论的方法进行加权。

杂谈：加权与三月疯狂锦标赛

我们在加权以及一年一度的三月疯狂篮球锦标赛上的工作始于 2006 年，在那段时间里，我们任其自行发展。随着每年即将离去的研究者们将火炬传给下一批新加入的研究者，我们的加权方法变得越来越复杂。我们在此按年代介绍这一工作中的各个亮点，并介绍一下主要的贡献者。

- 2006：作为查尔斯顿学院的一个课程设计，当时的研究生卢克·英格拉姆和约翰·麦康奈尔是最早着手解决三月疯狂锦标赛预测问题的人。他们的目标是利用数学方法来预测该锦标赛的各场比赛结果。是这两位发现了 ESPN 的锦标赛竞猜比赛（以及 10 000 美元的奖金），还有 ESPN 提供的在线预测自动提交和评分工具。在那一年，他们两人将马尔可夫模型的几个变体提交到了 ESPN。由于卢克和约翰相信球队状态的惯性可以具有很强的预测能力，因此所提交的结果之一就使用了时间加权，它使锦标赛前一个月内的比赛具有双倍的权值。但对于卢克和约翰这一对同学们戏称为“圣人门徒”的伙伴而言，2006 年不幸正好是一个颠覆的年份。许多读者可能还能够记得那一年的第 11 号种子乔治梅森大学的灰姑娘之旅，他们一路杀到了前四名。仅有很少的球迷（大多数都是乔治梅森大学队的校友）猜到乔治梅森队能走得这么远。实际上，当年赢得 ESPN 竞猜的球迷竟然是误打误撞地输入乔治梅森大学的，而他们的本意是想输入乔治·华盛顿大学。在发现了这个错误之后，拉塞尔·普莱曾特（Russell Pleasant）又提交了与他本意相符的看好乔治·华盛顿大学的结果。不过走运的是，他并未删除原来那个看好乔治梅森大学的结果，因为正是这个结果，使他赢得了当年的 10 000 美元奖金。
- 2008：2008 年春天，我（艾米·兰维尔）将同样的课题布置给了本科毕业班的学生尼尔·古德森和科林·斯蒂芬森。这两个人既是体育迷又是数学专业的学生，他们满怀热情地接过了火炬，并继承了卢克和约翰未竟的事业。到 ESPN 竞猜的最后期限时，全班帮助这两个人投出了超过 30 个结果，它们由科利法、梅西法、马尔可夫法和 OD 法

在不同加权下的排名结果所给出。尼尔和科林对他们的结果进行了深入研究，并发现了若干出人意料的走向。

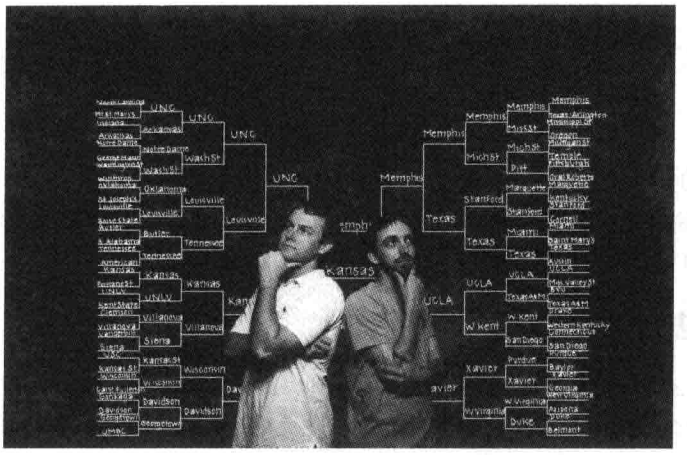


图 12.2　尼尔·古德森和科林·斯蒂芬森

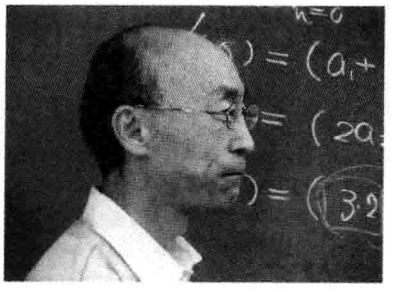
例如，大多数模型都预测到了第 11 号种子堪萨斯州立大学对第 6 号种子南加州大学的爆冷结果。同样的，由于大多数模型都预测堪萨斯大学将战胜孟菲斯大学而赢得锦标，所以尼尔和科林决定利用梅西法来预测这两队之间的最终比赛结果。他们给出的预测，是堪萨斯将在一场难解难分的比赛中以仅仅 3 分的优势胜出，这是一个非常好的预测，因为最终堪萨斯是在加时赛中取胜的。查尔斯顿学院的媒体业者们听说了尼尔和科林的工作，并在锦标赛的进行期间定期过来请求他们更新预测。受到在前几轮中所取得的巨大成功的推动，尼尔和科林发现他们身处一场媒体风暴的中央。然后就来了全国级的媒体，包括在 CBS 的《早间秀》（The Early Show）中的一次亮相。而最能扬名四海的邀请，可能得算是罗伯特·西格尔（Robert Siegel）和他的美国国家公共广播（NPR，National Public Radio）的《无所不虑》（All Things Considered）节目所发出的邀请。尼尔得到了一次现场连线采访的机会，其中他提到了几个预测，具体来说就是堪萨斯州立大学的大翻盘和堪萨斯大学的夺冠。尼尔是个天才——他恰好找到了数学技术内容的解读和激动人心的体育比赛结果与预测之间的平衡点。尼尔的采访片段在全国范围内得到了正面回应。实际上，这一采访是如此吸引人，以至于几名非体育领域的听众因为尼尔预言的最终准确性而目瞪口呆，并觉得必须在锦标赛结束之后给节目打来热线电话。一名听众称这两个人在他们的课程设计上做得如此之好，他们的教授理应给他们打个 A。他们的教授确实打了个 A，不过不是因为他们那个准确的预测，而是因为他们的积极性以及在整个学期中周复一周所体现出的职业道德。关于尼尔和科林的故事的更多内容，以及他们得分最高的竞猜之一，请见**第 212 页**上的杂谈。

- 2009：我（艾米·兰维尔）的运筹学课程每两年才在查尔斯顿学院开设一次。由于 2008 年的成功仍然鲜活地留在脑海之中，因此我肯定不会让这么一个小小的不利因素，来阻止我们的继续前进以及对 2009 年三月疯狂锦标赛中欢乐的追求。因此，我和好同事、戴维逊学院的蒂姆·夏提埃博士一同拉起了队伍来进行一个跨院校的排名

课题。我们的查尔斯顿学院小组由研究生凯思琳·佩丁斯和本科生赖恩·帕克组成，戴维逊小组则包括夏提埃博士和他的本科生埃里希·克罗伊策与马克思·温（Max Win）。后来队伍不断壮大，并纳入了斯坦福大学的戴维逊校友尼克·多维迪奥（Nick Dovidio）和日本筑波大学的山本洋右。我们的小组实际上遍布全美国、全北美乃至全球，如图 12.3 所示。



Kathryn Pedings



Yoshitsugu Yamamoto



图 12.3　我们的三月疯狂锦标赛小组遍布全球

2009 年，我们加入了更多的几种加权方法以及由数种排名聚合模型（请见第 14 章和第 15 章）所给出的排名。尽管 2009 年的比赛情况并不如具有历史性和宣传性的 2008 年那样利于竞猜，但我们仍然对所取得的进步感到非常高兴。有几个模型在当年提交到 ESPN 的 460 万个竞猜中进入了 95 和 97 百分位，而且得到了更多的新闻报道，包括在地方报纸上的新闻文章。

- 每年我们的模型都变得更为复杂，而且不管加入三月疯狂锦标赛小组的学生是否获得了 10 000 美元的 ESPN 奖金，他们都可以说是建模、编程和交流等动手实践方面的高手，并也因此带来了一些令人垂涎的工作职位。实际上，尼尔·古德森甚至在一次求职面试时听到了这样的话："嗨！你就是 NPR 上面那个疯狂三月排名的家伙，是吗？"

数字——

9 223 372 036 854 775 808 =（当所有种子队都已确定后）填写一张拥有64支队伍的“疯狂三月竞猜表”的不同方式的数目。

147 573 952 589 676 412 928 = 填写一张拥有68支队伍的“疯狂三月竞猜表”的不同方式的数目。

——这不过就是简单的算术：求出2^{63}和2^{67}。

第 13 章　“假如……会怎样”的问题与敏感性

教练和球迷们常常会在赛季的某些关键时间结点上，进行所谓的“假如……会怎样”的分析。例如，当赛季只剩下最后一场比赛要进行的时候，一名大学橄榄球队的教练可能会想，假如他的队伍在这最后一场比赛中以 β 分的优势获胜的话，那么球队的排名会有怎样的变化。当然，解答这名教练的问题的一种方式，便是计算出这场比赛之前的排名，然后假设最后一场比赛结果如教练所愿，并据此再计算排名。但是，对于规模非常大的应用（设想一下整个万维网大小的规模），对每个这类“假如……会怎样”的问题都进行完整排名的重新计算，不但不合理，而且有时甚至是不可能的。此外，完全重新计算跟使用强力法是一样的，对于一名数学家来说，实在是有些惭愧。因此，在本章中，我们将使用扰动分析的数学工具，用最少的额外计算，来对这类问题做出精确的回答。

秩一更新的影响

在本节，我们将使用扰动分析中的工具，来量化赛季比赛数据中的小变动对科利法排名可能造成的影响。随着赛季的展开，到某个时间点上，教练就要开始计算球队进入季后赛的机会了。在当前赛季的当前时刻，我们来计算一下科利评分；然后我们假设下一场比赛的结果与教练的设想相吻合，并重新计算评分向量。教练对下一场比赛的设想，可以用科利系统的系数矩阵中的一个扰动来表示。我们将研究一类特殊类型的扰动，它称为秩一更新。秩一更新是对矩阵进行的最简单类型的更新，因此，它将是我们的分析的一个良好起点。只需对本节中所推导出的公式稍加修改，便能适用于更高秩数的更新。总之，我们将对比扰动前和扰动后的评分。最后我们还将注意到，对于大多数其他方法，也能导出类似的结果。不过为了避免重复叙述，我们只选用科利法这一种方法，来给出有关“假如……会怎样”分析的重点。

扰动分析的工具使我们得以精确考察上述教练所设想中的局面能够造成的影响。令 $\boldsymbol{C}$ 和 $\boldsymbol{b}$ 为扰动前的科利矩阵和右向量，而 $\tilde{\boldsymbol{C}}$ 和 $\tilde{\boldsymbol{b}}$ 为扰动后的科利矩阵和右向量。假设现在教练想知道他的队伍 j 在下一场比赛中击败了队伍 i 时将要发生的事情。那么此时扰动前后的矩阵 $\boldsymbol{C}$ 和 $\tilde{\boldsymbol{C}}$ 之间便由公式 $\tilde{\boldsymbol{C}} = \boldsymbol{C} + (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^{\mathrm{T}}$ 联系在了一起。利用谢尔曼-莫里森秩一更新公式[54]和科利矩阵的若干 M-矩阵性质，我们可以用 $\boldsymbol{C}$ 来表示 $\tilde{\boldsymbol{r}}$。具体来说，就是

$$\begin{aligned}\tilde{\boldsymbol{r}} &= \tilde{\boldsymbol{C}}^{-1}\boldsymbol{b} \\ &= [\boldsymbol{C} + (\boldsymbol{e}_i - \boldsymbol{e}_j)(\boldsymbol{e}_i - \boldsymbol{e}_j)^{\mathrm{T}}]^{-1}\tilde{\boldsymbol{b}} \\ &= \boldsymbol{r} - \left(\frac{r_i - r_j}{1 + [\boldsymbol{C}^{-1}]_{ii} - 2[\boldsymbol{C}^{-1}]_{ij} + [\boldsymbol{C}^{-1}]_{jj}}\right)\boldsymbol{C}^{-1}(\boldsymbol{e}_i - \boldsymbol{e}_j)。\end{aligned} \tag{13.1}$$

由上式，我们可以得到三个结论。

- 哪怕仅仅只多了一场比赛，也会影响每支队伍的评分，因为 $\boldsymbol{C}^{-1}$ 的第 i 列和第 j 列几乎总是完全稠密的。
- 假设在扰动前队伍 i 的排名在队伍 j 之前，即 $r_i - r_j > 0$。则正如所预计的那样，如果在新的一场比赛中队伍 j 击败了 i，那么队伍 i 的评分 $\tilde{r}_i$ 将下降，而队伍 j 的评分 $\tilde{r}_j$ 将上升。这是由于 $\boldsymbol{C}^{-1}$ 具有的对角优势性和非负性所造成的。
- 如果 $\boldsymbol{C}^{-1}$ 存在，则我们只需挑出其中的少数元素，就能计算 $\tilde{\boldsymbol{r}}$ 了。有了 $\tilde{\boldsymbol{r}}$，我们就能确切知道每支队伍的评分将如何随扰动而改变，从而能够回答任意两支队伍之间的任何“假如……便怎样”问题。

要同时考虑多个“假如……会怎样”问题，我们可以利用秩 k 更新而非更简单的秩一更新来进行类似的分析。

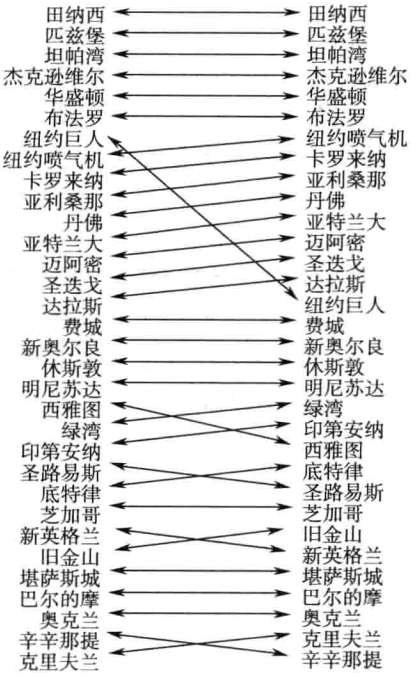
敏感性

“假如……会怎样”的问题，实际上与敏感性这个数学概念有关。在数学中，如果输入数据中小的变动将带来输出数据的大变化的话，我们便称该方法是敏感的。我们更愿意使用不敏感或鲁棒的方法。一个敏感的排名方法可能会在一次弱队战胜强队的比赛后，给出反复无常的排名。这正是参考文献［21］中得到的结论，这篇文献考虑了一种易于处理但实际中不大可能发生的“完美赛季”（没有任何爆冷的赛季）。换而言之，在一个完美的赛季中，某支队伍将保持不败，之后是仅输给前者但对其他所有队伍均保持不败的队伍，而第三名的队伍仅输给前两名但对剩下的队伍保持不败，等等。如果一个赛季的比赛数据符合这样的结构，那么各队的排名当然就很清楚。尽管完美赛季极其罕见，但它们给出的数据却能够通过封闭形式的表达来进行数学分析[21]。在许多方面，完美赛季都是所有输入赛季中性质最好的一个。如果一个排名算法在这种类型的赛季上都表现出了敏感性，那么在更为现实的不完美赛季上，这一敏感性问题只会变得更加糟糕。

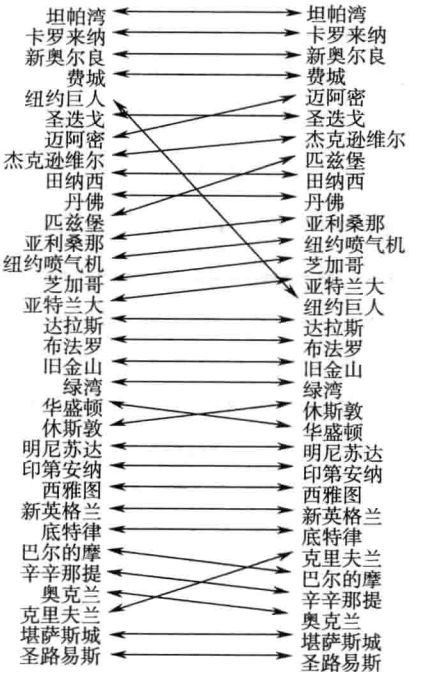
参考文献［21］中的研究对应用于完美赛季的科利评分法、梅西评分法和马尔可夫评分法进行了敏感性分析，分析结果对于科利法和梅西法而言是好消息。这两种方法对于系统中小的秩一改变不敏感，梅西法尤其如此。另一方面，马尔可夫法则对这样的改变表现出了敏感性。其中提到了 NFL 2008 赛季中发生的一件事，它明显地体现了马尔可夫法在不完美赛季中的敏感性。这件事发生于2008年10月13日星期一晚上的比赛中，当时排名第32位垫底的克里夫

兰队爆冷击败了排名第 17 位的纽约巨人队。在这个星期一夜晚的比赛前后的科利排名、梅西排名和马尔可夫排名如下所示。

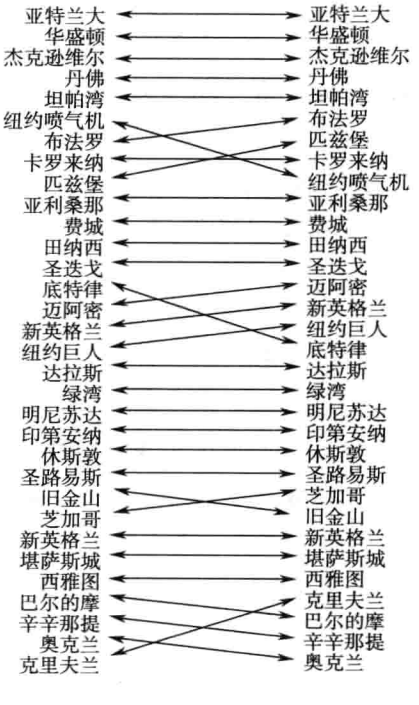
科利法



梅西法



马尔可夫法



正如所预计的那样，对于不敏感的科利法和高度不敏感的梅西法，在这场冷门之后，克里夫兰和巨人的排名分别上升和下降了几位。但是，尽管巨人输给了垫底的克里夫兰，马尔可夫法却还是使它的排名上升了一位。马尔可夫法在这个不完美的NFL赛季的这场比赛中表现出的奇怪举止，实际上与它在完美赛季上所得到的理论结果并不矛盾。更具体来说，参考文献［21］中的结论指出，马尔可夫法对于涉及排名拖尾的队伍的事件是敏感的——在我们的例子中，克里夫兰正是这个尾巴的最后一截。由于马尔可夫法的评分服从幂率分布[58]，因此在这个尾巴里会有许多队伍。有人也许会觉得这个方法的敏感性使其不大可靠，不过马尔可夫法却是谷歌最初用来进行网页排名的方法的核心。

杂谈：冷门和马尔可夫法的敏感性

尽管在数学领域，敏感性一般都被视为贬义，但至少还是有某些团体可以从马尔可夫法的敏感性中获益。如果马尔可夫法被用于球队排名，那么评分落在末尾的队伍的教练和球员可以寄希望于一场冷门来获得不成比例的好处。如果低排名的队伍能够爆冷，哪怕以1分的优势击败强队，那么就有可能使排名往前蹿好几位甚至是很多位。冷门越大（即对手排名越高），可能提高的排名位数就越高。

数字——

——有史以来最大的冷门。

1. “冰上奇迹”——1980年，美国奥林匹克冰球队击败前苏联队。
2. 1969年，超级碗III，纽约喷气机击败巴尔的摩小马。
3. 维拉诺瓦大学压倒乔治城大学，赢得1985年NCAA男子篮球冠军。
4. 1990年，巴斯特·道格拉斯击倒迈克·泰森赢得重量级拳击冠军。
5. 1919年，赛马Man O’War以100比1的赔率输给了Upset，这也是它唯一的一场失利。
6. 1994年，NBA季后赛中丹佛掘金队击败西雅图超音速队。
7. 1955年，在美国高尔夫公开赛上，杰克·弗莱克击败本·霍根。
8. 1969年，纽约大都会队在世界职业棒球大赛中击败巴尔的摩金莺队。
9. 2000年，鲁伦·加德纳击败亚历山大·卡列林赢得古典式摔跤金牌。
10. 北卡罗来纳州立大学击败休斯敦大学队，赢得1983年NCAA男子篮球冠军。

——espn. go. com/page2/s/list/topupsets/010525. html

第 14 章　排名聚合——第 1 部分

“总体大于各部分之和”这句名言，正是排名聚合方法的基本思想，它是本章以及下一章所要关注的内容。排名聚合的目的，是通过某种方式将数个排名列表加以合并，以构建一个单一的、新的、更好的排名列表。

将几个排名列表聚合为一个“超级”清单的需求很常见，也具有许多应用。例如，考虑一下如 Excite，Hotbot 和 Clusty 这样的元搜索引擎，它们将来自于主要搜索引擎的排名结果组合为一个（希望是）更好的排名列表。图 14.1 给出了排名聚合概念的图形描述。在此，l_1，l_2，…，l_k 表示 k 个单独的排名列表，通过本章中将要介绍的数学方法，它们将按一定的方式被组合起来，从而给出一个鲁棒的聚合列表。这一过程称为排名聚合，并在参考文献［28］，［25］和［24］中得到了深入研究。实际上，我们在**第 82 页**上就已经开始和排名聚合打交道了。在那里，我们只需要将区区两个排名列表加以组合，即由 OD 进攻评分向量 $\boldsymbol{o}$ 给出的攻击排名列表和 OD 防守评分向量 $\boldsymbol{d}$ 给出的防守排名列表。我们使用了简单的 $\boldsymbol{r} = \boldsymbol{o}/\boldsymbol{d}$ 的规则，来产生一个聚合评分向量 $\boldsymbol{r}$，再由它来生成“鲁棒的”排名列表。在本章中，我们将介绍几个当前的排名聚合方法，它们能够用于组合任意数量的排名列表，而且从这个意义上来说，它们也更具普遍性。在**第 167 页**和**第 172 页**上，我们将介绍在排名聚合领域我们自己的成果（两个新算法）。不过，在介绍这些方法之前，让我们再次回到本书最初的主题，即政治投票理论。

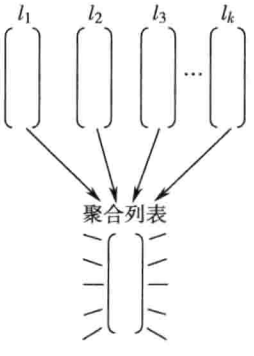


图 14.1　k 个排名列表的排名聚合

重温阿罗准则

本书最开始的章节中介绍了阿罗的不可能性定理[5]，其中指出了所有投票体系所固有的局限性。总的来说，阿罗的定理说明，没有一个投票体系可以同时满足无限制域、无关选择独立性、帕累托法则和非独裁性这四条准则。在本节中，我们将考虑我们和这些准则之间的牵连到底有多深，特别是在体育排名问题这个给定的情况之下。

尽管第 1 章指出了对政治体系中的候选人进行排名的问题和对体育赛事中的队伍或个人进行排名的问题之间的相似性，但是这两者之间的确存在若干差异，值得我们更为细致地加以考察。

阿罗定理所考虑的情境即政治投票，在两个重要的方面上不同于体育排名问题。首先，政治竞选主要考虑多数票。对于所有候选人来给出一个排名列表并不必要，因为只有占据榜首的人才会终有所得，而在如排球、网球和纳斯卡赛车这样的体育赛事中，冠军、亚军、季军乃至更后的名次通常都能得到积分或奖金；其次，尽管政治投票和体育排名都有需要排名聚合的一面，但两者涉及的参数大不相同。考虑一下图 14.2，其中显示了体育和政治排名问题在数据输入和处理方面的主要不同。

即使在政治选举中要求选民对所有候选人进行排名，对于大多数职位而言，都仅有相对较少的候选人，而选民的人数则十分巨大。在体育排名中，这种情况颠倒了过来，此时通常只有少数几个投票者（排名体系）对数百个候选者（队伍）进行排名。体育排名体系和政治排名体系之间的这两点差别，促使我们去重新检查要求阿罗准则仍然成立的正当性。

阿罗的第一条准则要求投票体系具有无限制域，这意味着每个选民都应该能够按任意他或她喜欢的方式来对候选人进行排名。例如，如果投票体系要求现任者必须被排到前五，那就违反了阿罗的第一条准则。在体育排名领域中，我们赞同阿罗的观点，即无限制域是一个需要的性质。

另一方面，阿罗的第二条准则即无关选择独立性，则可能是阿罗对于投票体系的要求中最具争议性的一个。这条准则指出，在候选人的子集中的相对排名，应该在对全体候选人进行排名时得到保持。对于社会实际上自然而然地遵循了这一公理的说法，许多投票理论家们争执不下，而 2000 年的美国大选则凸显了该准则的若干重要性。这场戏剧性的选举是如此难解难分，以至于直到最后一个投票完毕的佛罗里达州计票完毕后，才能确定胜者，而即使是在这个时候，随着数错选票和选票上打孔不彻底等问题的浮出水面，计票也不得不多次重新进行。许多分析家都认为，如果拉尔夫·纳德没有从艾伯特·戈尔那儿抢走某些选票的话，戈尔应该会在佛罗里达的对决中战胜乔治 W·布什。选民投给纳德的选票使布什赢得了实际上是三选一的一场胜利。在这种情况下，戈尔便是所谓的孔多塞胜者，而这种情况在社会选择领域中一般称为“败事者效应”。如果 2000 年的大选符合阿罗

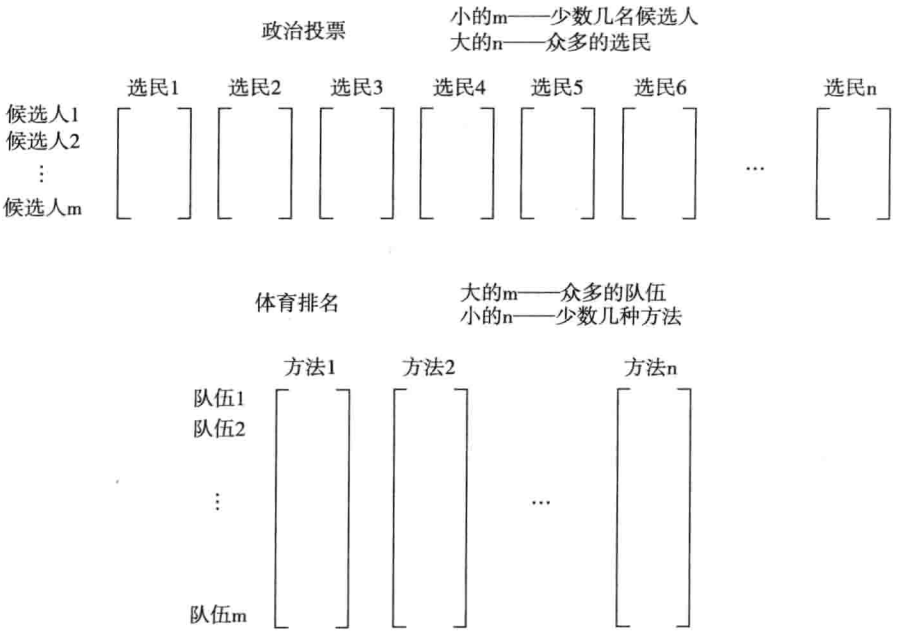


图 14.2　政治投票和体育排名中聚合排名列表的不同

的独立性准则的话，那么在子集 {戈尔，布什} 上的排名，不应该被扩展候选人集合 {戈尔，布什，纳德} 中第三者的到来而有所影响。2000 年大选表明，如果投票体系不满足阿罗的第二条准则，那么孔多塞胜者将可能失利。

既然政治和体育是两个非常不同的赛场，那么我们现在就来讨论无关选择独立性在体育队伍排名应用上的适用性。几乎所有的体育评分方法都是根据队伍之间的捉对比赛来操作的。注意“队伍”同样可以解读成“候选人”，以继续与政治投票体系的对比。我们已经提及，对于大多数体育项目，特别是高校体育比赛，一支队伍并不会在一个赛季里与所有其他队伍都交手，因为高校队伍的数量实在太多，而赛季的时间不够，使得这样的交手无法成为一种现实的可能性。（那些学生兼运动员有时还不得不去上课。）尽管如此，虽然 VT 可能在一个赛季里从未与 USC 交手，我们却仍然可以根据联盟间的赛事来推得这两支队伍之间的间接关系。因此，对手的对手可以揭示出队伍之间一度消失了的联系，即 USC 可能击败了 UNC，而 UNC 与 VT 对阵并击败了后者。因此，通过传递性，我们可以辩称 USC 应该排名在 VT 之前。进而我们可以宣称，将一支队伍从集合中移除，理应影响到整体排名。这一陈述使我们摈弃了阿罗对于评分体系的第二条准则。

反对使用无关选择独立性的另一个论据，可以通过一个对摔跤手排名的简单例子来说

明。对于包括摔跤在内的许多体育项目而言，一种常见的情况就是 j 在对抗中由于风格正好克制住对方等因素而战胜了 i，但并不矛盾的是 j 可能被大多数对手所击败，而 i 对大多数对手都表现很好。因此，如果摔跤手 i 在锦标赛中没有很早就碰上 j 的话，那他就将取得很好的成绩。如果扩展到锦标赛中所有选手的集合，只要 i 和 j 没有在竞赛的首轮就碰面的话，i 几乎总是会排名在 j 之前。而这就违背了独立性准则，该准则将要求在扩展到更大的摔跤手集合时，j 仍然排在 i 的前面。总之，相比于政治投票问题，阿罗的第二条无关选择独立性准则对于体育排名问题看来远不是那么有用。

但是，在将阿罗的第二条准则从体育排名应用中彻底划掉之前，让我们暂时扮演一下故意唱反调的人，并试图设想出一个体育比赛的局面，选择的独立性在其中仍然具有价值。我们使用我们的范例来构造出这样一个局面。回忆一下表 1.1 的内容，Miami，VT，UNC 和 UVA 都击败了 Duke。与无关选择独立性有关的一个问题就是：如果将 Duke 从表 1.1 中去除，是否真会影响排名？也就是说，在去掉了 Duke 后，由于 Miami 对 Duke 的优势比 UVA 对 Duke 的优势要多二十几分，那么 Miami 评分的增加是否会大于 UVA 的？阿罗将会回答“不会”，而上一段则指出“会的，在某些情况下。”这些都是棘手的问题。可能在面对着 Duke 这个弱的对手时，UVA 的教练派出了二线队员，而 Miami 的教练却没这么做。因此，对 Duke 所取得的胜利可能确实应该只具有很少的权值，也正因如此，如果从候选者集合中去除 Duke，那么排名应该不受影响，这就提供了一个支持阿罗的第二准则的论据。使用了比赛得分的体系必须在使用比分所带来的好处和强队有意在弱队身上取得大比分优势的问题之间做出协调。这一“大捞得分”的问题，看来正是 BCS 更喜欢使用不涉及比分而只考虑胜负记录的模型的主要原因。请回忆一下第 3 章中科利法所包含的无偏性的概念。这一局面同样突出表明了如下事实，即有关球员变动或伤病等信息，这些正是那些并未考虑在本书任何模型之中的信息。

上面几段给出了考虑独立性准则的两种方式。在第一种方式中，我们给出了反对使用该准则的理由，而在第二种方式中，我们则给出了支持的理由。但还有一种方式可以用来考察独立性准则，即从数学家的角度来看待这一问题。数学家们常常会考虑一个模型对于输入数据中小的改变或异常值的敏感性。如果一个系统在输入数据中存在小的改变时仍然能给出稳定的结果，那么系统就称为鲁棒的，这是一个我们希望的模型应具有的性质。另一方面，高度敏感的系统则是我们所不期望采用的。因此一个自然的做法，就是将阿罗的无关选择独立性的概念，与敏感性和鲁棒性的数学概念联系起来，即一个鲁棒的排名方法将能满足阿罗的第二准则。

在阿罗的第三条准则即帕累托原则上的争议就少得多了，该准则指出，如果所有选民都选择了候选人 A 而非 B，那么一个恰当的投票体系应该保持这一顺序。这显然是一个在政治投票和体育排名领域都应该被满足的有用的准则。例如，如果所有方法都将 Miami 排在了 Duke 之前，那么这些方法的聚合排名也应该将 Miami 排在 Duke 之前。

阿罗的第四条也是最后一条准则是非独裁性。尽管在政治领域中看起来很显然的是，没有哪个选民应该具有比其他选民更高的权重，在体育排名中这一点就不是那么显然了，其理由有二。首先，下一节中讨论的某些聚合方法需要有一个“打破僵局”的策略来处理十分接近的列表。此时，聚合函数通常会给予其中一个以更高的优先权。因此，获得优先权的列表就表现为一个独裁者或至少是部分的独裁者。在第二种状况下，我们可能有合理的理由来要求一个（部分）独裁者。例如，体育分析家们可能希望由教练和专家所给出的排名比计算机模型所给出的排名更为重要（或不重要）。如果我们愿意打破阿罗的第四条准则的话，这样的期待就能够得以实现。因此，阿罗的最后一条准则对体育排名而言，看来并不如在政治投票中那么重要。

排名聚合方法

排名聚合并不是一个新想法。正如前面一节中所暗示的那样，在政治投票领域它已经被研究了数十年。但是，网络搜索激发了对排名聚合的新兴趣。随着过去十年中以欺骗搜索引擎并不公平地令客户页面获得高排名的垃圾信息制造技术的急速发展，搜索引擎必须设计出反垃圾信息的反制措施。这正是排名聚合方法得到新的注意的原因和时机。例如，元搜索引擎便是基于聚合的思路，它利用多个搜索引擎给出的排名列表来生成一个超级父列表，该父列表保持了子列表的最好的性质。因此，如果一个比方说来自谷歌的子列表中包含了一个异常页面（一个由页面所有者对其页面进行了成功地优化，它能够欺骗谷歌给予它一个很高的但不公正的排名），则通过排名聚合后，这个异常造成的影响就应当被削弱。因此，排名聚合倾向于像一个“平滑器”那样工作，这意味着它将移除各个单独列表中的任何异常的排名突起或者说异常页面，并使得聚合结果更为平滑。排名聚合列表的另一个性质，就是它所能做到的程度，只能和用来构建它的列表一样好（或一样差），即我们无法希望由一组糟糕的输入列表得到一个非常准确的聚合列表。在本节中，我们将给出四种排名聚合方法。前两种是过去常用的方法，而后两种则是我们在排名聚合领域的新贡献。

杂谈：网络垃圾信息和元搜索引擎

这个杂谈的目的，在于强调网络垃圾信息的盛行及其无处不在的身影。我们通过两种流行的垃圾信息制造技术——谷歌炸弹和链接农场[49]——来进行说明。到杂谈结束时，我们将清晰地认识到具有垃圾信息抵抗力的方法的必要性。幸运的是，如本章将要讨论的那些利用排名聚合方法的元搜索引擎，已经在阻止垃圾信息方面取得了成功。

谷歌炸弹是一种垃圾信息制造方法，它利用了谷歌对于锚文本的重视。最为著名的谷歌炸弹，也许就是瞄准G. W. 布什的那一枚。图14.3说明了制造炸弹的机制。一个谷歌

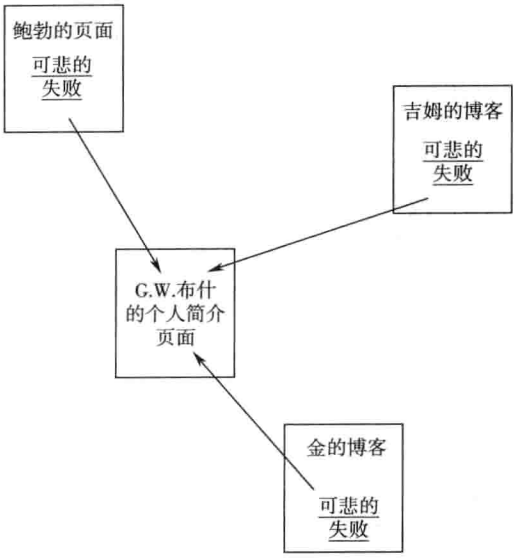


图 14.3　谷歌炸弹是如何工作的

炸弹的有效性与下面这一点紧密相关，即谷歌相信与一条超链接相关联的锚文本可以作为这个超链接所指向的页面的一个简短总结或摘要。如果多个页面（在图 14.3 的例子中，就是鲍勃的页面、吉姆的博客以及金的博客）都指向布什在白宫的个人简介页面，并且这些超链接都使用了“可悲的失败”这一相同的锚文本，那么谷歌就会认为布什的页面肯定与这一主题有关。如果有足够多的垃圾信息制造者合谋起来，那么这个垃圾信息制造策略就能够使得目标页面即布什的页面在查询项“可悲的失败”的结果中获得高排名。当谷歌于 2004 年公开募股时，他们的首次公开募股文档就提到了这一垃圾信息制造技术，并称之为投资者们必须加以考虑的一个风险。本章的排名聚合方法提供了谷歌炸弹的一种补救办法。将多个主要搜索引擎结果加以聚合的元搜索引擎将会发现，谷歌对于如“可悲的失败”这样的垃圾查询项所给出的结果与其他搜索引擎的结果并不一致，而利用下面几节中将要介绍的排名聚合过程，这样的异常结果将被抑制。

另一个流行的垃圾信息制造技术是链接农场。链接农场可以影响到任何在网页排名中使用基于链接分析的搜索引擎。大多数链接分析算法——特别是谷歌及其 PageRank 算法——都对获得了来自其他高排名页面的入链页面给予了奖励。链接农场将生产出常见主题中排名很高的页面，然后将这些页面的 PageRank 与客户页面共同分享。例如，考虑图 14.4 的情况。一个搜索引擎优化（SEO，Search Engine Optimization）公司可能会对于诸如美国前总统这样的主题产生互联密度很高的一组页面。这些页面包含了以上主题的合法的相关信息，并因此从一个搜索引擎那儿得到了与之相称的排名。如果一个客户想与 SEO 公司联系并希望提高他自己页面的排名的话，那只需从这个链接农场中引出一条链接连到客户的页面就行了。对于搜索引擎而言，链接农场难以检测，因此是一个颇为令人头疼的垃圾信息制造技术。正如在谷歌炸弹中一样，击退链接农场的一个可能方法，就是使用下面几节将要介绍的排名聚合方法。

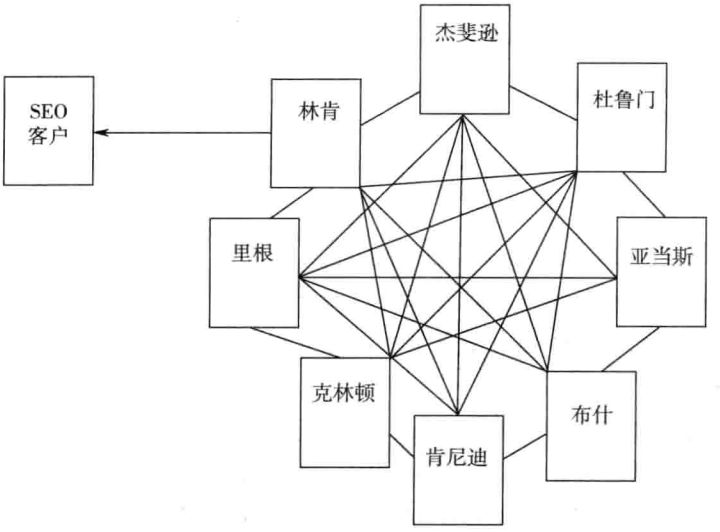


图 14.4　链接农场是如何工作的

波达计数法

一个常用的排名聚合方法就是波达计数法[12]，它可以追溯到让-查理·德·波达（Jean-Charles de Borda）在1770 年的工作。波达当时正试图将一场政治选举中各位候选人的排名列表加以归总。对于每个排名列表，每名候选人都将获得一个分数，它等于该候选人所击败的其他候选人的人数。每个列表中的分数相加，就为每名候选人产生出一个数，称为波达计数。在**第 17 页**的杂谈中介绍的 BCS 排名方法，就是说明波达计数使用方法的常用示例。

尽管波达的方法非常简单，但通过一个例子，还是能给出最好最简洁的说明。因此，我们再次回到我们的 5 支队伍的范例。我们选用波达计数，来聚合由 OD 法、梅西法和科利法所给出的三个排名列表，而计算所需的数据如表 14.1 所示。在 OD 一列中 VT 后列出的数字 4，表明 OD 法给出的排名列表中，VT 的排名在 4 支队伍之前。将各行中三种方法的数据相加，我们就得到了这三个列表的波达计数法聚合结果[⊖]。

这个例子展示了波达计数法进行排名聚合的简单性和计算的简易性。此外，还可以修改排名聚合的波达计数法以处理平局。例如，假设一个投票者将 5 支队伍由高到低排名为 {Miami，VT，UNC/UVA，Duke}，其中斜杠“/”表示排名相同，则此时，对于这一列表给出的波达分数为

$$\begin{matrix}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}0.0\\ 4.0\\ 1.5\\ 1.5\\ 3.0\end{pmatrix}。$$

⊖ 感谢卢克·英格拉姆制作这一数据和表格。

表 14.1　在 5 支队伍示例中对 OD 法、梅西法和科利法给出的排名进行聚合的波达计数法

	OD 法（$r = o/d$）	梅西法	科利法	波达计数法	波达排名
Duke	0	0	0	0	第 5
Miami	3	4	4	11	第 1
UNC	1	1	2	4	第 4
UVA	2	2	1	5	第 3
VT	4	3	3	10	第 2

注意，排名相同的队伍将平分这一排名位置所对应的波达分。因此，波达计数法可以处理存在平局的输入列表。此外，尽管不大可能发生，但波达计数法可能会给出一个包括平局的输出聚合列表。

不幸的是，波达法有一个严重的缺陷，即它可以被很容易地操控。也就是说，不诚实的投票者可以有意串通起来，以使得结果偏向于他们所期望的方向。如参考文献［16］这样的教科书中就给出了表明波达计数法的可操控性的经典示例。

杂谈：尤依计数



本章给出了波达计数和孔多塞胜者这两个数学概念，它们都得名于其发明者。但是，最近发现的手稿表明，这些数学概念实际上应该以拉蒙·尤依（Ramon Llull，1232—1315）的名字被命名为尤依计数和尤依胜者，他是一位极其多产的加泰罗尼亚作家、哲学家、神学家、占星家和传教士。尤依名下有超过 250 部作品。在 2001 年，随着三部遗失的手稿《标记之术》（Ars notandi）《选举之术》（Ars eleccionis）和《又一选举之术》（Alia ars eleccionis）的发现，历史学家们认识到他们应该在前面的那个作品总数上再加上 3，并在尤依的职业和头衔中加入数学家这个称号。由于尤依在选举理论上的工作比波达和孔多塞各自独立地发现要早了几个世纪，因此他已经被认可为发现本节所述的波达计数法和**第 160 页**上所述的孔多塞胜者的第一人。

平均排名

另一个非常简单的排名聚合方法是平均排名。这个方法求取几个排名列表中整数排名值的平均值，进而给出一个排名聚合列表。我们再次用一个例子来清楚说明这种方法。表 14.2 给出了由 OD 法、梅西法和科利法所得的三个排名列表的平均排名。

这个例子掩盖了平均排名法的一个常见缺陷——平均评分中会频繁出现平分现象。令 l_1 和 l_2 为两个完全排名列表，即包括了集合中所有元素的一个排列。假设队伍 i 在 l_1 中排名第一，而在 l_2 中排名第三；而队伍 j 在两个列表中都排名第二。那么，这时平均排名法就将给

表 14.2　在 5 支队伍示例中对 OD 法、梅西法和科利法给出的排名进行聚合的平均排名法

	OD 法（$r = o/d$）	梅西法	科利法	平均评分	平均排名
Duke	第 5	第 5	第 5	5	第 5
Miami	第 2	第 1	第 1	$1.\overline{3}$	第 1
UNC	第 4	第 4	第 3	$3.\overline{6}$	第 4
UVA	第 3	第 3	第 4	$3.\overline{3}$	第 3
VT	第 1	第 2	第 2	$1.\overline{6}$	第 2

出两者都排名第二的平局情况。使得情况更为复杂的是，平均排名法可以在多个位置以多种方式造成平局的出现。

有几种策略可以用来打破前段中所提到的这类平局，在此我们介绍其中两种。第一种策略是使用以往的数据来打破僵局。如果队伍 i 和队伍 j 相互交手，那么在这场比赛中的胜者（或在多场比赛中的“最佳”胜者）应该比负者排名靠前。当对超过两个列表进行平均时，通过追溯比赛结果来确定打破僵局的队伍的做法将变得更为困难，因为这时会出现若干难以解决的问题。比如说，假设使用三个或更多列表进行平均时，队伍 i，j 和 k 都被排在了同一个名次上，那么在用于产生这些列表的数据中，就有可能包含如下的比赛结果：i 击败了 j，j 击败了 k，而 k 又击败了 i。这样的循环平局，是在通过成对比较获得排名时常常被提到的一个难点。通常的解决方法是指定一个列表为打破僵局的列表，这一策略将在下一段中介绍。

如果打成平手的队伍实际上从未真正交手过，那么第二种打破僵局的策略便是指定一个排名列表为“打破僵局列表”，对于成对比较中的平局，这一策略颇为有效。例如，假设现在出现了先前那个平局，其中队伍 i 和 j 都在聚合列表中位居次席。但是，如果假设第一个排名列表被选为优先者，即打破僵局列表，那么这就意味着队伍 i 赢得了聚合列表中的第二名，而队伍 j 将落在第三位上。打破僵局列表也并非一定是任意指定的，特别是当我们有办法确定哪一个排名列表更好一些的时候，就更有合理的理由来指定打破僵局的列表。确定排名列表质量的指标将在第 16 章中讨论。

我们以有关平均排名的聚合方法的若干结论来结束本节。首先，要注意平均排名法只能在所有列表均为完全列表（即所有列表中的所有元素的交集包含了所有的队伍）的情况下才能适用；其次，平均排名的聚合方法可以给出一个不包含排名的聚合列表，这一点在表 14.2 中清晰可见。当然，解决的方法也很简单，对排名列表求平均就给出了一个平均评分向量，它可以接着被转化为一个排名向量。

模拟比赛数据

在这一节里，我们将给出一种新的排名聚合方法，这种方法来自于对排名列表的一个简单解释，请见图 14.5。这个解释是这样的：如果队伍 A 在一个排名列表中排在队伍 B 之前，那么在这两支队伍的一场较量中，A 应当会击败 B。此外，如果队伍 A 在排名列表的首位，

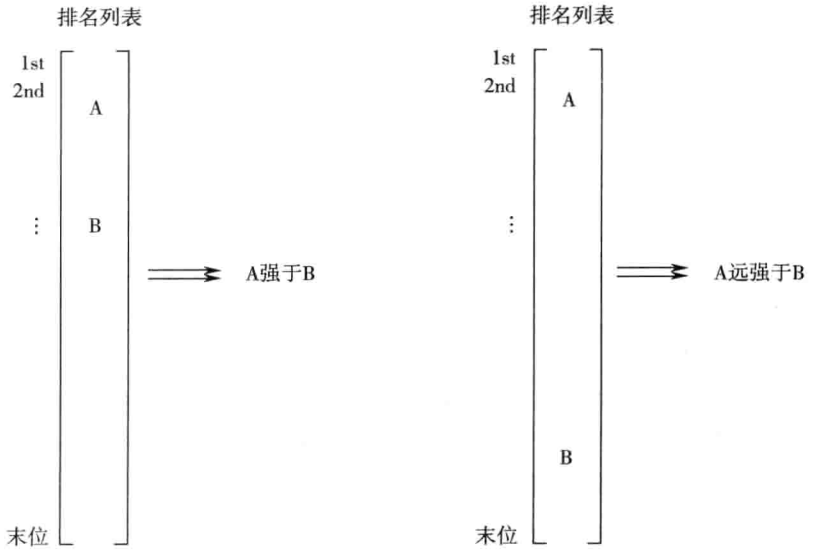


图 14.5　对象在排名列表中相对位置的解释

并与位居末席的队伍 B 比赛，那么我们预期 A 将以很大的优势击败 B。排名列表应该隐含地给出了未来的比赛结果信息，这一结论看起来十分自然，而从数学建模的角度来看，更聪明的一点在于我们可以使用排名列表，来产生所谓的模拟比赛数据，而且最终我们将用它来将多个列表聚合为一个列表。实际上，每个长度为 n 的排名列表都给出了 $\binom{n}{2} = n(n-1)/2$ 场模拟比赛的数据。让我们再次回到 5 支队伍的示例，以及用 OD 法、梅西法和科利法所得的那三个排名列表，如下所示。

	OD 法	梅西法	科利法
第 1	VT	Miami	Miami
第 2	Miami	VT	VT
第 3	UVA	UVA	UNC
第 4	UNC	UNC	UVA
第 5	Duke	Duke	Duke

由 OD 排名向量开始，我们可以得到这 5 支队伍之间 $\binom{5}{2} = 10$ 场成对较量的结果。如果 VT 与 Miami 交手，那么我们预计 VT 会获胜，这是根据 OD 向量中的信息所确定的。此外，如果简单地假设胜者的优势与两支队伍名次之差相关的话，我们可以给这场模拟比赛赋予一个获胜分差。在这一点上，建模者便有很多可以调整的地方来发挥创造性的想法。在此我们给出一种效果不错的基本思路。我们为每一个位次的名次差赋予一个 1 分的分差。因此，根据该 OD 向量，VT 应该以 1 分的优势击败 Miami，以 2 分的优势击败 UVA，以 3 分的优势击败 UNC，并以 4 分的优势击败 Duke。这个得分数据在后面的排名聚合步骤中十分有用。在为 OD 排名向量产生了模拟比赛数据之后，我们便着手处理下一个排名向量，如此类推，直至我们积累起一大堆模拟比赛数据。注意，我们可以对任意多的排名列表构建出完整的模拟

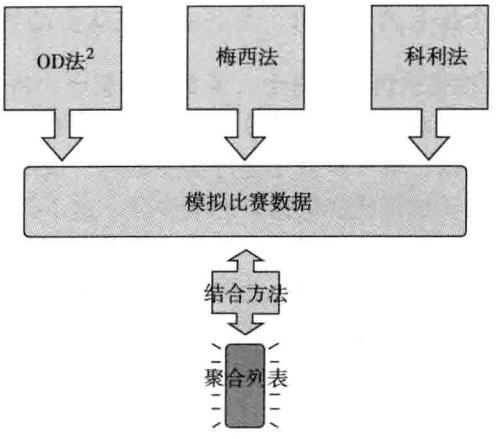
数据。在上例中，对于 3 个排名列表且每个都具有 5 支队伍的情况，我们产生了 $3\binom{5}{2} = 30$ 场模拟比赛。此外我们还注意到，*不同于大多数其他聚合方法，这些列表不必为完全列表*。某些列表可能为完全的，而某些仅为部分的，但仍然能够产生模拟比赛数据。

一旦完全产生了模拟数据，下一步就是进行实际聚合，说明如下。在这一步，我们选择任一排名方法，并将它用于模拟比赛数据以产生排名。这个我们称为*结合方法*的排名方法，可以是一种完全不同的方法，如未用于产生输入排名向量的方法，或者是那些产生输入排名向量的方法中我们较为偏爱的一个。此时按照阿罗的话，从某种意义上来说，这个被偏爱的同时用于输入和结合者的方法，就成为了一定程度上的独裁者。

表 14.3 给出了由 OD 法、梅西法和科利法排名向量构建的模拟比赛数据。注意，我们将失败者的得分设为 18 分。我们也可以选用 0 分、10 分或其他任意分数（为实验主义者们留下了更多发挥的余地），但由于 18 分是 2005 赛季输球队伍的平均得分，因此选用这一分数是合理的。还请注意，在这个 5 × 5 的表单的每个单元格中，都包含了 3 场比赛结果，分别对应了三个输入排名向量。具体举例来说，在（Duke，Miami）这个格子中的 18-21的结果表明，在 OD 评分向量中，Duke 的名次比 Miami 落后了三位，因此以 3 分之差输掉了模拟比赛。同一个格子中 18-22 的结果来自于梅西排名向量，其中 Duke 的排名比 Miami 落后了四位。

表 14.3　模拟比赛数据

	Duke	Miami	UNC	UVA	VT
Duke		18-21 18-22 18-22	18-20 18-19 18-19	18-20 18-20 18-19	18-22 18-21 18-21
Miami	21-18 22-18 22-18		20-18 21-18 20-18	19-18 20-18 21-18	18-19 19-18 19-18
UNC	19-18 19-18 20-18	18-20 18-21 18-20		18-19 18-19 19-18	18-19 18-20 18-21
UVA	20-18 20-18 19-18	18-19 18-20 18-21	19-18 19-18 18-19		18-20 19-20 18-20
VT	22-18 21-18 21-18	19-18 18-19 18-19	21-18 20-18 19-18	20-18 20-19 20-18	



排名聚合的模拟数据法

上图给出了排名聚合的模拟数据法的图示。我们用这一方法，将小型的 5 支队伍示例中的 OD 法、梅西法和科利法排名向量中的数据加以聚合。表 14.4 表明，根据所用结合方法的不同，聚合结果可能稍有出入。

表 14.4　使用 OD 法、梅西法和科利法作为结合方法的排名聚合的模拟数据法

	结合方法		
	OD 法（$r = o/d$）	梅西法	科利法
Duke	第 5	第 5	第 5
Miami	第 1	第 1	第 1
UNC	第 3	第 4	第 4
UVA	第 4	第 3	第 3
VT	第 2	第 2	第 2

我们现在指出排名聚合的模拟数据法的 4 个性质。

- 排名聚合列表只能做到跟输入列表一样好（或一样差）。
- 结合方法的作用就如同一个“平滑器”，即它使得异常值——即包含了看来与其他列表排名不符的异常数据的列表——的影响被最小化。这就回应了阿罗的第二条准则即无关选择独立性准则，以及相关的鲁棒性和敏感性问题。作为结合方法影响的一个例子，让我们看一下表 14.4。为了检验独立性，我们将 Duke 由数据集中去除，而使用梅西法或科利法作为结合方法时，这一移除就导致了 UNC 和 UVA 排名的互换。请对比表 14.5 和表 14.4。相比之下，当使用 OD 法作为结合方法时，得到的排名并没有受到移除最弱的Duke这一因素的影响。因此，对于这个非常小规模的例子，

我们称 OD 排名聚合方法是鲁棒的。

表 14.5　排名聚合的模拟数据法及其对于移除 Duke 的敏感性

结合方法	OD 法（$r = o/d$）	梅西法	科利法
Miami	第 1	第 1	第 1
UNC	第 3	**第 3**	**第 3**
UVA	第 4	**第 4**	**第 4**
VT	第 2	第 2	第 2

- 只要输入的列表满足阿罗的第三条准则帕累托法则，那么排名聚合列表就同样满足帕累托法则。
- 尽管这是一个小规模的例子，但表 14.4 表明，结合方法可以对聚合列表产生影响。此外，如果结合方法也是输入方法中的一种，那么用阿罗的第四条准则的话来说，这种方法在一定程度上可以被称为是独裁者。

杂谈：排名前 10 的书籍——聚合由人来给出的列表

大多数有史以来最伟大著作的列表都是由人来产生的。换而言之，便是首先进行问卷调查，然后将投票进行归总。书友会、出版商、杂志社等这一类机构，都会给出他们自己的有史以来最好的 10 本书的清单。2009 年 6 月，《新闻周刊》给出了他们的一个超列表，其中公布了最好的 100 本书。这个超列表是由《时代》、奥普拉的书友会、维基百科、纽约公共图书馆等来源所收集的 10 个十佳书籍的清单经聚合后所得。《新闻周刊》通过一个简单的加权投票法来聚合这些列表，不过，他们当然也可以使用本章（或下一章）中的任何一个排名聚合方法来做这件事。我们在此给出根据《新闻周刊》的超列表所列出的有史以来最好的 10 本书（见表 14.6），以飨各位喜爱读书的读者。

表 14.6　由《新闻周刊》超列表所列出的有史以来最好的 10 本书

排名	书名	作者
1	《战争与和平》	列夫·托尔斯泰
2	《1984》	乔治·奥威尔
3	《尤利西斯》	詹姆士·乔伊斯
4	《洛丽塔》	弗拉基米尔·纳博科夫
5	《喧哗与骚动》	威廉·福克纳
6	《隐形人》	拉尔夫·艾里森
7	《到灯塔去》	弗吉尼亚·伍尔夫
8	《荷马史诗》	荷马
9	《傲慢与偏见》	简·奥斯汀
10	《神曲》	但丁

排名聚合的图论方法

在本节中，我们将提出另一种聚合多个排名列表的方法。这一方法基于图论。考虑如图 14.6 所示的情况，其中每支队伍对应了图中的一个结点。此时，图中各边的权值并非比分或任何其他比赛统计数据，而是来自我们希望聚合的排名列表。例如，考虑如下建议的两种对 i 到 j 的链接赋予权值的方式。

$$w_{ij} = \text{将 } i \text{ 排名在 } j \text{ 之后的列表数量} \tag{14.1}$$

或

$$w_{ij} = \text{将 } i \text{ 排名在 } j \text{ 之后的列表中两者名次差的总和} \tag{14.2}$$

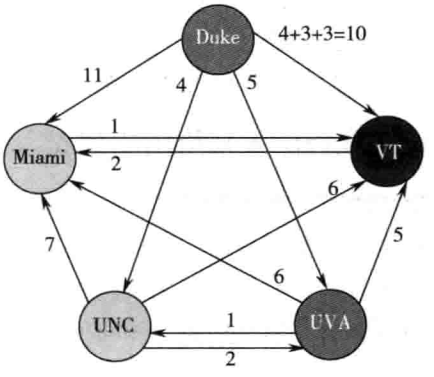


图 14.6　5 支队伍示例中排名聚合的图论方法所对应的图

这些权值可以被视为弱队投给强队的票数。而以上不过是进行投票的两种可能方式而已。

在**第 176 页**上，我们将对一种被称为评分聚合的扩展进行探讨，这种方法使用评分向量而非排名向量中的数值来给出聚合图中的权值。

一旦选中了对这一稠密的队伍图中的边赋予权值的方法，就可以在该图上应用所偏爱的结点排名算法。例如，可以使用 PageRank，HITS 或 SALSA 等任何来自网络搜索的流行的结点排名算法。有关这些算法的更多知识请见参考文献［49］。

我们对图 14.6 使用了 PageRank 法（一个流行的结点排名方法）[15]。这个图构建自**第 168 页**上由 OD 法、梅西法和科利法所给出的三个排名向量。图 14.6 中的权值由式（14.2）所确定。例如，由 Duke 到 VT 的链接具有权值 $4+3+3=10$，因为杜克大学在 OD 向量中落后 VT 4 位，在梅西向量中落后 3 位，在科利向量中也落后 3 位。将 PageRank 法用于这幅图，

就得到了一个 PageRank 向量，它给出了如下聚合排名向量

$$\begin{matrix}\text{第 1}\\ \text{第 2}\\ \text{第 3}\\ \text{第 4}\\ \text{第 5}\end{matrix}\begin{pmatrix}\text{Miami}\\ \text{VT}\\ \text{UVA}\\ \text{UNC}\\ \text{Duke}\end{pmatrix}$$

图论聚合排名

顺便提一句，这个排名与**第 170 页**的表 14.4 中以梅西法或科利法作为结合方法的模拟数据法的结果相吻合。

第 167 页和**第 172 页**上这两个新的排名聚合方法，比**第 165 页**和**第 166 页**上介绍的标准现有方法更为复杂。实际上，包括大学橄榄球队为之争得头破血流的 BCS 排名在内的严肃的排名聚合应用，都应该使用以上更为先进的方法来聚合多个排名列表。此外，在下一章中还将给出一个从数学角度而言甚至还要高级的排名聚合方法，该方法能够确保聚合列表与输入列表之间获得最优匹配。有关内容请见第 15 章。

杂谈：排名聚合和元搜索引擎

元搜索引擎将多个搜索引擎的结果加以结合，以产生对于特定查询的排名列表。在这个杂谈中，我们将说明元搜索引擎是如何使用本节中排名聚合的图论方法的。首先，我们将查询项“rank aggregation”输入谷歌、雅虎、MSN-bing 和 Ask.com 这 4 个搜索引擎，并获取 4 个前 10 位排名的列表。这 4 个列表的并集包含了 $n = 17$ 个不同的页面。运用图论方法，就得到了一个大小为 17×17 的聚合矩阵 $\boldsymbol{W}$。这幅图的 PageRank 向量给出了**第 174 页**上所示的排名聚合列表，它给出了我们的元搜索引擎的查询结果。

为了实现这个简单的元搜索引擎，请按照以下步骤进行。

1. 从你喜爱的 p 个搜索引擎获得排名前 k 位的列表；

2. 使用**第 172 页**上描述的方法（或本章及下一章中所介绍的任一排名聚合方法）产生一个聚合矩阵 $\boldsymbol{W}$。$\boldsymbol{W}$ 的大小为 $n \times n$，其中 n 为 p 个列表的并集中不同页面的数量；

3. 计算与 $\boldsymbol{W}$ 相关联的排名向量 $\boldsymbol{r}$。

由于 k 和 p 一般较小（$k < 200$，$p < 10$），所以 $\boldsymbol{W}$ 也较小，因此 $\boldsymbol{r}$ 可以实时计算得到。

[PDF] **Rank Aggregation** Methods for the Web
File Format: PDF/Adobe Acrobat - View as HTML
by C Dwork - Cited by 456 - Related articles - All 28 versions
ing the importance of **rank aggregation** for Web applications .... The ideal scenario for **rank aggregation** is when each judge ...
citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.8702&rep...

[PDF] **Rank Aggregation** Revisited
File Format: PDF/Adobe Acrobat - Quick View
by C Dwork - Cited by 28 - Related articles - All 14 versions
The **rank aggregation** problem is to combine many different rank orderings on the ... In this work we revisit **rank aggregation** with an eye toward reducing ...
citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.2507...

Rank Aggregation: Together We're Strong
by F Schalekamp - Cited by 3 - Related articles - All 7 versions
whereas in the case of the **rank aggregation** problem, we .... uate different algorithms for **rank aggregation** and re- lated problems. ...
www.siam.org/proceedings/alenex/.../alx09_004_schalekampf.pdf - Similar

[PDF] Efficient similarity search and classification via **rank aggregation**
File Format: PDF/Adobe Acrobat - Quick View
by R Fagin - Cited by 140 - Related articles - All 21 versions
The **rank aggregation** problem is precisely the problem of how ... and **rank aggregation**. As a simple but powerful motivating exam- ...
www.almaden.ibm.com/cs/people/fagin/sigmod03.pdf - Similar

BioMed Central | Full text | RankAggreg, an R package for weighted ...
by V Pihur - 2009 - Cited by 1 - Related articles
Two examples of **rank aggregation** using the package are given in the ..... The RankAggreg( ) function performs **rank aggregation** using either the CE algorithm ...
www.biomedcentral.com/1471-2105/10/62 - Cached - Similar

Supervised **rank aggregation**
by YT Liu - 2007 - Cited by 20 - Related articles - All 13 versions
This paper is concerned with **rank aggregation**, the task of combining the ranking results of individual rankers at meta-search. Previously, **rank aggregation** ...
portal.acm.org/citation.cfm?id=1242638

paper.dvi
Rank aggregation has been studied in many disciplines, most extensively in the context of social choice theory, where there is a rich literature dating from the latter half of the eighteenth century. We revisit **rank aggregation** with an eye toward meta-search.
citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.2...

[PDF] **Rank Aggregation** for Similar Items
213k - Adobe PDF - View as html
in which the **rankings** are noisy, incomplete, or even disjoint. ... **Rank aggregation** can be thought of as the unsupervised. analog to regression, in which the goal is to find an ...
www.eecs.tufts.edu/~dsculley/papers/mergeSimilarRank.pdf

[PDF] Unsupervised **Rank Aggregation** with Distance-Based Models
File Format: PDF/Adobe Acrobat - Quick View
by A Klementiev - Cited by 5 - Related articles
One impediment to solving **rank aggregation** tasks is the high cost associated with acquiring full or ... frame unsupervised **rank aggregation** as an optimiza- ...
www.cs.jhu.edu/~aklement/publications/icml08.pdf

CiteULike: Web metasearch: **rank** vs. score based **rank aggregation**
Search all the public and authenticated articles in CiteULike. Enter a search phrase. You can also specify ... Web metasearch: **rank** vs. score based **rank aggregation** methods Export ... **aggregation** metasearch search web www...
www.citeulike.org/user/pdlug/article/1167894

我们的简单元搜索引擎对于查询“rank aggregation”给出的结果

排名聚合后的一个精化步骤

当数个列表 l_1 ，l_2 ，…，l_k 通过**第 163 页**上的某个排名聚合方法被聚合为一个列表 μ 之后，便可以使用一个被称为局部凯梅尼化（Local Kemenization）[25] 的步骤，来进一步改进列表 μ。对于一个聚合列表 μ，如果将其中任意一对对象加以交换，都不会使任一输入列表 l_i（$i = 1,\cdots,k$）和 μ 之间的肯德尔 τ 度量之和降低，则称 μ 为*局部凯梅尼最优的*。（肯德尔 τ-度量将在**第 203 页**上被严谨定义。对于目前而言，我们只需要知道 τ 衡量了排名列表之间相隔距离的远近即可。）因此，局部凯梅尼化步骤考察 μ 中的每一对对象，并提出如下问题：将这对对象调换一下，是否能改进聚合后的肯德尔度量？我们用一个例子来说明这一求精步骤。考虑将 OD 法（l_1）、梅西法（l_2）和科利法（l_3）应用于 5 支队伍示例时得到的三个输入列表，以及利用**第 167 页**上的模拟数据聚合法得到的聚合列表 μ。

	OD 法 l_1		梅西法 l_2		科利法 l_3		模拟数据法 μ
第 1	VT	第 1	Miami	第 1	Miami	第 1	Miami
第 2	Miami	第 2	VT	第 2	VT	第 2	VT
第 3	UVA	第 3	UVA	第 3	UNC	第 3	UNC
第 4	UNC	第 4	UNC	第 4	UVA	第 4	UVA
第 5	Duke	第 5	Duke	第 5	Duke	第 5	Duke

局部凯梅尼化步骤由聚合列表 μ 的排名最高者（Miami）开始，并考虑如下问题：与之相邻的对象（VT）是否在大多数输入列表中，都排在它（Miami）前面？如果是，那么就交换这两者的排名，因为如此一来就能得到一个改善了的聚合列表，它减少了肯德尔距离。对于排序后的聚合列表中相邻的每一对，我们都考虑同样的问题。因此，对于长度为 n 的聚合列表，局部凯梅尼化需要 $n - 1$ 次检验，每个相邻对检验 1 次。

在 5 支队伍示例上的局部凯梅尼化

第 1 次检验
问题：第二名（VT）是否在大多数输入列表中击败了第一名（Miami）？
回答：否，VT 在 3 个列表中仅击败了 Miami 1 次。
动作：无。
第 2 次检验
问题：第三名（UNC）是否在大多数输入列表中击败了第二名（VT）？
回答：否，UNC 从未在这 3 个列表中击败过 VT。
动作：无。
第 3 次检验
问题：第四名（UVA）是否在大多数输入列表中击败了第三名（UNC）？
回答：是的，UVA 在 3 个列表中击败了 UNC 2 次。
动作：在 μ 中调换 UNC 和 UVA 的排名。
第 4 次检验
问题：第五名（Duke）是否在大多数输入列表中击败了第四名（UVA）？
回答：否，Duke 从未在这 3 个列表中击败过 UVA。
动作：无。

因此，局部凯梅尼化后的聚合列表 $\bar{\mu}$ 为 $\begin{matrix}\text{第 1}\\ \text{第 2}\\ \text{第 3}\\ \text{第 4}\\ \text{第 5}\end{matrix}\begin{pmatrix}\text{Miami}\\ \text{VT}\\ \text{UVA}\\ \text{UNC}\\ \text{Duke}\end{pmatrix}$，它与**第 170 页**表 14.4 中利用梅西法或科利法结合的模拟数据法的聚合列表相吻合。

评分聚合

至此，本章已经讨论了聚合排名列表的问题。但是，每个排名列表自然都来自于一个评分列表，这就带来了一个问题：评分列表是否也能同样地加以聚合？回答是肯定的，尽管有一些障碍需要克服，例如，尺度的障碍。我们应该怎样将评分值尺度相差悬殊的评分列表加以聚合？回忆一下，科利法给出的评分在 0.5 左右，而马尔可夫法给出了 0 到 1 之间的正评分。更为复杂的是，如梅西评分等某些评分可以包括正值和负值。图 14.7 展示了我们在本节将考虑的梅西法、科利法和 OD 法这三种方法给出的评分向量的数线图表示。我们将利用距离和百分数，来将这些不同的评分置于相同的尺度之下。一个长度为 n 的评分向量实际上包含了这总共 n 支队伍的 $\binom{n}{2} = n(n-1)/2$ 个成对比较。考虑如下的梅西评分向量 $\boldsymbol{r}_{\text{Massey}}$。

$$\boldsymbol{r}_{\text{Massey}} = \begin{matrix}\text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{matrix}\begin{pmatrix}-24.8\\ 18.2\\ -8.0\\ -3.4\\ 18.0\end{pmatrix}$$

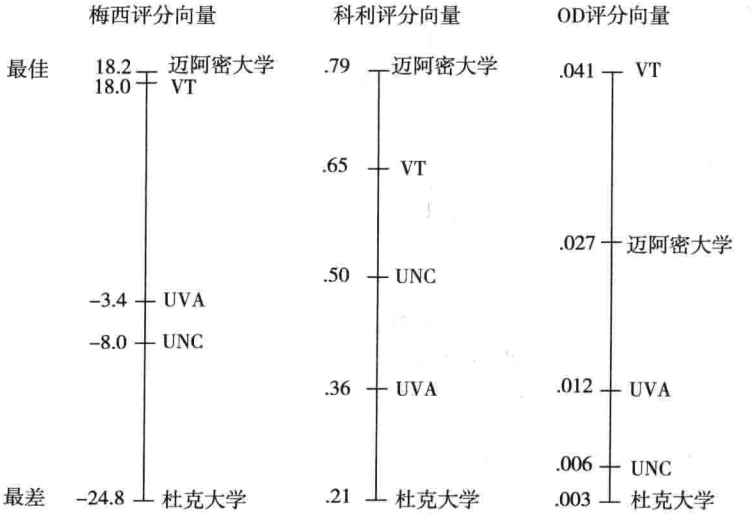


图 14.7　5 支队伍示例中梅西、科利和 OD 评分向量的数线图

由于 Miami 的评分大于 Duke 的评分，因此梅西法预测 Miami 将击败 Duke。这两队之间的评分差 43 相对而言是比较大的。当然，这肯定意味着些什么。实际上，我们更希望将这些评分差视为距离。Miami 和 Duke 相距遥远，而相对距离为 4.6 的 UNC 和 UVA 则相互接近。由于我们自然地假定距离为正值，因此我们可以用上述方式来获取距离，并为每个评分向量产生一个非对称的非负评分差矩阵。例如，上述梅西评分向量 $\boldsymbol{r}_{\text{Massey}}$ 可以给出如下的评分距离矩阵 $\boldsymbol{R}_{\text{Massey}}$。

$$\boldsymbol{R}_{\text{Massey}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 43 & 0 & 26.2 & 21.6 & 0.2 \\ 32.8 & 0 & 0 & 0 & 0 \\ 21.4 & 0 & 4.6 & 0 & 0 \\ 42.8 & 0 & 26 & 21.4 & 0 \end{pmatrix} \end{array}$$

对科利评分向量和 OD 评分向量计算评分距离矩阵可得

$$\boldsymbol{R}_{\text{Colley}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.58 & 0 & 0.29 & 0.43 & 0.14 \\ 0.29 & 0 & 0 & 0.14 & 0 \\ 0.15 & 0 & 0 & 0 & 0 \\ 0.44 & 0 & 0.15 & 0.29 & 0 \end{pmatrix} \end{array}$$

和

$$\boldsymbol{R}_{\text{OD}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.024 & 0 & 0.021 & 0.015 & 0 \\ 0.003 & 0 & 0 & 0 & 0 \\ 0.009 & 0 & 0.006 & 0 & 0 \\ 0.038 & 0.014 & 0.035 & 0.029 & 0 \end{pmatrix} \end{array}$$

我们一眼就能看出，在这些 $\boldsymbol{R}$ 矩阵中存在尺度问题。为了将这些距离置于同一尺度，我们可以使用归一化这个老方法。我们将原始矩阵 $\boldsymbol{R}$ 中的每个元素除以这个矩阵所有元素之和。记为 $\overline{\boldsymbol{R}}$ 的归一化矩阵如下。

$$\overline{\boldsymbol{R}}_{\text{Massey}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.1792 & 0 & 0.1092 & 0.09 & 0.0008 \\ 0.1367 & 0 & 0 & 0 & 0 \\ 0.0892 & 0 & 0.0192 & 0 & 0 \\ 0.1783 & 0 & 0.1083 & 0.0892 & 0 \end{pmatrix} \end{array}$$

$$\overline{\boldsymbol{R}}_{\text{Colley}} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.2 & 0 & 0.1 & 0.1483 & 0.0483 \\ 0.1 & 0 & 0 & 0.0483 & 0 \\ 0.0517 & 0 & 0 & 0 & 0 \\ 0.1517 & 0 & 0.0517 & 0.1 & 0 \end{pmatrix} \end{array}$$

以及

$$\bar{\boldsymbol{R}}_{\mathrm{OD}}=\begin{array}{c}\\ \text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{array}\begin{array}{c}\begin{array}{ccccc}\text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT}\end{array}\\ \begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 0.1237 & 0 & 0.1082 & 0.0773 & 0\\ 0.0155 & 0 & 0 & 0 & 0\\ 0.0464 & 0 & 0.0309 & 0 & 0\\ 0.1959 & 0.0722 & 0.1804 & 0.1495 & 0\end{pmatrix}\end{array}。$$

在归一化矩阵 $\bar{\boldsymbol{R}}_{\mathrm{Massey}}$ 中元素（Miami，Duke）的值为 0.1792，可以解释如下：在梅西模型所预测的所有成对比赛的距离总和中，迈阿密大学和杜克大学之间的距离所占比例为 17.92%。这个解释的一个很大的优点在于，*百分比可以用作不同方法之间比较的度量*。此外更加接近于本节重点的是，现在可以进行评分的聚合了。一个简单的评分聚合方法，是对若干归一化的矩阵 $\bar{\boldsymbol{R}}$ 求取（可能经过加权的）平均值。例如，对归一化的 $\bar{\boldsymbol{R}}_{\mathrm{Massey}}$，$\bar{\boldsymbol{R}}_{\mathrm{Colley}}$ 和 $\bar{\boldsymbol{R}}_{\mathrm{OD}}$ 求平均可得

$$\bar{\boldsymbol{R}}_{\mathrm{ave}}=\begin{array}{c}\\ \text{Duke}\\ \text{Miami}\\ \text{UNC}\\ \text{UVA}\\ \text{VT}\end{array}\begin{array}{c}\begin{array}{ccccc}\text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT}\end{array}\\ \begin{pmatrix}0 & 0 & 0 & 0 & 0\\ 0.1676 & 0 & 0.1058 & 0.1052 & 0.0164\\ 0.0841 & 0 & 0 & 0.0161 & 0\\ 0.0624 & 0 & 0.0167 & 0 & 0\\ 0.1753 & 0.0241 & 0.1135 & 0.1129 & 0\end{pmatrix}\end{array}。$$

由评分聚合矩阵得到评分向量

矩阵 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 中的每个非零元素，都给出了有关两队之间距离的信息，因此也给出了今后比赛时对可能的胜者的预测。只要再多做一点事情，就有可能从 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 中挤出更多的信息来给出分差预测，这将是今后工作的一个主题。目前，我们仅讨论将二维矩阵中的信息拆解为一个一维评分向量的问题。我们将给出由 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 得到评分向量 $\boldsymbol{r}$ 的三种方法。

方法 1

第一种方法是使用行和与列和。$\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 的行和给出了一支队伍进攻能力的一个度量，而列和则反映了防守的表现，即攻击评分向量为 $\boldsymbol{o}=\bar{\boldsymbol{R}}_{\mathrm{ave}}\boldsymbol{e}$，而防守评分向量为 $\boldsymbol{d}=\boldsymbol{e}^{\mathrm{T}}\bar{\boldsymbol{R}}_{\mathrm{ave}}$，其中，$\boldsymbol{e}$ 为全 1 向量。正如在第 7 章的 OD 法中那样，大的行和值对应了进攻得力的队伍，而低的列和值则对应了防守牢固的队伍。因此，总的评分向量可以由 $\boldsymbol{r}=\boldsymbol{o}/\boldsymbol{d}$ 计算得到。

方法 2

从矩阵 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 中抽取评分向量 $\boldsymbol{r}$ 的第二种方法，就是将马尔可夫法应用于 $\bar{\boldsymbol{R}}_{\mathrm{ave}}^{\mathrm{T}}$ 按行归一化所得的结果之上。这种方法之所以可行，是因为 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 中的元素代表了队伍之间的距离，即一支队伍投给另一支队伍的票数。

方法 3

从矩阵 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 中抽取评分向量 $\boldsymbol{r}$ 的第三种也是最后一种方法，是利用 $\bar{\boldsymbol{R}}_{\mathrm{ave}}$ 的主特征向量。

这个特征向量实际上有个专有的名字——它是 $\overline{\boldsymbol{R}}_{\text{ave}}$ 的佩隆向量，得名自佩隆定理，该定理告诉我们，一个不可约非负矩阵的主特征向量一定为非负。在实际中，我们并不会去检验 $\overline{\boldsymbol{R}}_{\text{ave}}$ 是否为不可约，而是通过在 $\overline{\boldsymbol{R}}_{\text{ave}}$ 上加上 $\varepsilon \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}$ 来迫使不可约性得以成立，其中，ε 为一个小的正标量。

表 14.7 给出了应用这三种方法得到的 $\overline{\boldsymbol{R}}_{\text{ave}}$ 所对应的三个评分向量。

表 14.7　三种评分聚合方法得到的评分向量

队伍	方法 1 $r = o/d$		方法 2 马尔可夫法 r		方法 3 佩隆向量 r	
Duke	0	第 5	0.020	第 5	0.27	第 5
Miami	16.4	第 2	0.465	第 2	0.58	第 2
UNC	0.4	第 3	0.025	第 3	0.34	第 3
UVA	0.3	第 4	0.024	第 4	0.33	第 4
VT	26.0	第 1	0.466	第 1	0.61	第 1

杂谈：对美国国家科学基金项目申请进行排名

美国国家科学基金会（NSF，National Science Foundation）是美国在科学研究方面最大的资助机构。因此，这一政府机构每年都会收到来自科学家们在各自研究领域中的数千份项目申请书。当然，NSF 的预算虽然很多，但终归有限，因此需要不断地决定哪些申请可以资助，而哪些申请要被拒绝。大多数资助都是通过一个称为小组审查的过程来决定的，这些审查有赖于所涉及领域中公认的专家所给出的评价。例如，NSF 的 FODAVA（数据与视觉分析委员会）在 2009 年收到了 32 份申请。大约 20 名小组审查专家（即与申请者之间不存在利益冲突或冲突很小的 FODAVA 专家）被要求审阅这些申请书。由于申请书数量很大，所以要每位审查专家对每份申请都加以审阅的做法并不现实，因此，部门主管为每名专家分配数量大致相等的申请书，例如，每位专家一般要对 6～7 份申请进行评审。专家评审一般包括给出评语以及与本书最为相关的一个评分。评分分为优异、很好、好、一般和差。由于并非每份申请都由所有专家进行了评审，而且有些专家给分比较松，而另一些专家却打分比较严，所以这一评审体系面临着一些固有的挑战。为了应对这些挑战，NSF 要求专家们集中起来进行一次为期两天的面对面的会议，在会上，每份申请都由那些对之进行了评审的专家做一个总结，然后由整个审查组加以讨论。这个讨论过程对于评审而言十分关键，但也非常费时。下面我们将提议一种方式，以使得这一为期两天的会议变得更为高效。我们的想法就是公平地对评分加以聚合，以给出一个全局性的总排名列表。有了这样一个聚合列表，就可以少花些时间放在排名位于后一半的申请上，而对于这些申请，审查组总体上是不同意给予资助的。

让我们考虑一个例子。假设有 6 份编号为 A 到 F 的申请需要进行排名，并且有 4 位评审专家，每人负责评阅 4 份申请。一个典型的审查组讨论会可能从如表 14.8 所示的清单开场。这个清单将评阅人的评分整齐地归总在一起，不过却掩盖了哪位专家给出了哪个评分

表 14.8　NSF 审查会上的申请评分示例

申　请	评　分
A	优异，很好，优异
B	一般，好
C	好，很好，很好
D	好，好，好
E	好，优异
F	差，很好，优异

这一重要信息。实际上，表 14.8 中的清单是由表 14.9 中的数据得来的，其中清楚表明，不同专家对申请打分的评分标准是有差异的。一位评阅人（如本例中的评阅人 3）打分很松，

表 14.9　各位评阅人给出的申请评分

评阅人 1		评阅人 2		评阅人 3		评阅人 4	
申请	评分	申请	评分	申请	评分	申请	评分
A	优异	A	很好	C	很好	B	好
B	一般	D	好	E	优异	F	优异
C	好	E	好	F	很好	D	好
D	好	F	差	A	优异	C	很好

大方地给出了很好或优异的高分，而另一位则严得多，他很少会打出优异的分数。因此，我们将使用在**第 176 页**评分聚合一节中所介绍的思路，将不同评阅人给出的评分变到一个相同的尺度之下。应用评分聚合方法可以得出 4 位评阅人的 $\overline{\boldsymbol{R}}$ 矩阵如下。

$$\overline{\boldsymbol{R}}_1 = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \\ \mathrm{F} \end{array} \begin{array}{c} \begin{array}{cccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} \end{array} \\ \begin{pmatrix} 0 & 3/9 & 2/9 & 2/9 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/9 & 0 & 0 & 0 & 0 \\ 0 & 1/9 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}, \overline{\boldsymbol{R}}_2 = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \\ \mathrm{F} \end{array} \begin{array}{c} \begin{array}{cccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 1/9 & 1/9 & 3/9 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2/9 \\ 0 & 0 & 0 & 0 & 0 & 2/9 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

$$\overline{\boldsymbol{R}}_3 = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \\ \mathrm{F} \end{array} \begin{array}{c} \begin{array}{cccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} \end{array} \\ \begin{pmatrix} 0 & 0 & 1/4 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}, \overline{\boldsymbol{R}}_4 = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \\ \mathrm{F} \end{array} \begin{array}{c} \begin{array}{cccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/7 & 0 & 1/7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2/7 & 1/7 & 2/7 & 0 & 0 \end{pmatrix} \end{array}。$$

$\overline{\boldsymbol{R}}_1$ 的元素（A，B）的分子表明评阅人 1 对申请 A 给出的评分比对申请 B 的评分高 3 个等级，而分母 9 则是评阅人 1 所评阅的 4 份申请的所有 $\binom{4}{2} = 6$ 种可能的成对比较的评分差总和。对 $\overline{\boldsymbol{R}}_1$，$\overline{\boldsymbol{R}}_2$，$\overline{\boldsymbol{R}}_3$ 和 $\overline{\boldsymbol{R}}_4$ 求平均可得平均距离矩阵

$$\overline{\boldsymbol{R}}_{\text{ave}} = \begin{array}{c} \\ \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \\ \text{F} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} & \text{F} \end{array} \\ \begin{pmatrix} 0 & 0.0833 & 0.1181 & 0.0833 & 0.0278 & 0.1458 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0635 & 0 & 0.0357 & 0 & 0 \\ 0 & 0.0278 & 0 & 0 & 0 & 0.0556 \\ 0 & 0 & 0.0625 & 0 & 0 & 0.1181 \\ 0 & 0.0714 & 0.0357 & 0.0714 & 0 & 0 \end{pmatrix} \end{array}$$

我们使用第 178 页评分聚合一节中的第三种方法，来由 $\overline{\boldsymbol{R}}_{\text{ave}}$ 得出评分向量，即计算 $\overline{\boldsymbol{R}}_{\text{ave}}$ 的佩隆向量，并得到了表 14.10 中的评分和排名向量。现在，审查组就可以根据这一评分聚合方法的结果，仅对前三名 A，E 和 F 进行讨论。

表 14.10　将评分聚合方法应用于 NSF 申请的例子

申请	佩隆向量 r	排名
A	0.88	第 1
B	0	第 6
C	0.08	第 5
D	0.15	第 4
E	0.40	第 2
F	0.20	第 3

聚合方法总结

- 本章中的示例展示了聚合方法的应用，它可以对多个计算机生成的列表加以组合。但是，这些聚合方法同样可以用来聚合由人所给出的列表，或甚至于合并来自于人以及计算机的数据。我们由此可以看到聚合方法的巨大潜力。
- 尽管还必须给出证明，但实验结果使我们猜想，$Q_{\text{worst}} \leqslant Q_{\text{aggregate}} \leqslant Q_{\text{best}}$，其中，$Q$ 是一个质量度量，它能用于对各种方法进行打分。（这一命题的证明，部分就在于设计一个恰当的质量度量 Q，而这并非轻而易举。请见第 16 章。）这一命题表明，一个聚合列表可能差于其中一个或多个单独方法所给出的结果，但另一方面，聚合列表又优于另外一些单独方法所给出的结果。因此，聚合列表对于一次性的预测问题是有帮助的。

杂谈：给疯狂三月锦标赛的建议

下面来考虑一下每年一度的疯狂三月锦标赛及其竞猜。对单独一次事件进行预测是很棘手的，你可能猜得中，也可能猜不中。本书中的各个单独（未聚合）方法“平均来说”能够给出不错的预测，即对一些相同队伍之间的多场比赛来说，预测效果不错。但是，竞猜却要

求方法能够对许多仅发生一次的事件给出良好的预测。这种“一次性”的预测问题与“平均”或“最好”预测问题不同，具体而言，在“一次性”事件中，由于诸如疯狂三月锦标赛这种“一劳永逸”式的锦标赛所具有的紧张气氛以及巨大赌注等因素，冷门将出现得更为频繁。因此，如果你必须投出一个且只有一个竞猜结果，那与其绑死在某一个方法上，倒不如选择更聪明（也更保守）的方法，把宝押在几种方法的聚合之上，因为这正是本章所介绍的排名聚合方法的长处。

数字——

47 000 000 000 ≈谷歌进行评分和排名的网页的数量。
12 000 000 000 ≈雅虎进行评分和排名的网页的数量。
11 000 000 000 ≈Bing 进行评分和排名的网页的数量。
——截至 2011 年 9 月。

——worldwidewebsize. com

第 15 章　排名聚合——第 2 部分

前一章中的排名聚合方法是一种启发式方法，即它们并不能保证聚合所得的排名为最优。另一方面，这些启发式方法的一个很大的好处在于速度，相比于本章中所给出的排名聚合的最优化方法而言，启发式方法的速度非常快。当然，如果排名精度至关重要的话，最优化方法所需的额外时间一般都是必要的。

现在我们来介绍一种最优排名聚合方法，它由日本筑波大学的山本洋右博士所提出[50]。该方法给出了一个聚合排名，使得输入排名之间的吻合程度或一致程度为最优。有多种方法可以用来定义列表间的吻合程度，一种可能的做法就是令常量

$$c_{ij} = (i \text{ 的排名在 } j \text{ 之前的列表数量}) - (i \text{ 的排名在 } j \text{ 之后的列表数量}) \quad (15.1)$$

如果在 k 个输入列表中共有 n 个对象，则 $\boldsymbol{C}$ 是上述定义的常量的一个 $n \times n$ 斜对称阵。$\boldsymbol{C}$ 也可以处理输入列表中包含相同排名的情况。此外，$\boldsymbol{C}$ 可以由完全或部分的输入列表所产生。

有了衡量一致性的常量所构成的矩阵 $\boldsymbol{C}$ 之后，我们现在的目的就是产生这 n 个对象的一个排名，使得一致性达到最大。为了实现这一点，我们定义决策变量 x_{ij}，它确定 i 是否排在 j 之前。具体来说，

$$x_{ij} = \begin{cases} 1, \text{如果 } i \text{ 排名在 } j \text{ 之前}, \\ 0, \text{其他}。 \end{cases}$$

为了理解矩阵 $\boldsymbol{X}$ 的用处，考虑一个小型的示例，其中有 $n = 4$ 个对象，标号为 1 到 4，并按此标号顺序排名。则与这一排名对应的 $\boldsymbol{X}$ 矩阵为

$$\boldsymbol{X} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array},$$

它表明对象 1 排名在对象 2、3 和 4 之前，对象 2 排名在对象 3 和 4 之前，而最后对象 3 的排名仅在对象 4 之前。在这个例子中，$\boldsymbol{X}$ 那整齐的阶梯状结构就体现了排名的情况。在其他例子中，这一点乍看上去可能并没有如此清晰。考虑这些对象的一个排名，由高到低为

[4,2,1,3]，对应的矩阵为

$$\boldsymbol{X}=\begin{array}{c} \\ 1\\2\\3\\4\end{array}\begin{array}{c}\begin{array}{cccc}1&2&3&4\end{array}\\\begin{pmatrix}0&0&1&0\\1&0&1&0\\0&0&0&0\\1&1&1&0\end{pmatrix}\end{array},$$

这不过是之前那个整齐的阶梯形的一个重排，也就是说，如果我们根据排名序[4,2,1,3]对 $\boldsymbol{X}$ 进行重排，则可得

$$\text{重排后的}\ \boldsymbol{X}=\begin{array}{c} \\ 4\\2\\1\\3\end{array}\begin{array}{c}\begin{array}{cccc}4&2&1&3\end{array}\\\begin{pmatrix}0&1&1&1\\0&0&1&1\\0&0&0&1\\0&0&0&0\end{pmatrix}\end{array}。$$

幸运的是，我们并不需要对最优化过程所得的矩阵 $\boldsymbol{X}$ 进行实实在在地重排，因为要从未进行重排的 $\boldsymbol{X}$ 得到排名是非常容易的，只需求出 $\boldsymbol{X}$ 的列和，并按升序排序，即可得到排名[⊖]。

充分理解了 $\boldsymbol{X}$ 后，我们现在便转向最优化问题。我们希望最大化输入列表之间的一致性，用常量 c_{ij} 和变量 x_{ij} 表示即为

$$\max\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij},\text{其中}\ x_{ij}\in\{0,1\}。$$

但是，我们必须加上某些限制，使得矩阵 $\boldsymbol{X}$ 具有我们在上面那个小型的 4 × 4 的例子中所观察和利用到的性质。通过加入两类约束便能做到这一点：

对所有不同的对(i,j)，有 $x_{ij}+x_{ji}=1$　　（类型 1——反对称性）

对于所有不同的三元组(i,j,k)，有 $x_{ij}+x_{jk}+x_{ki}\leqslant 2$　（类型 2——传递性）

第一类是反对称性约束，它指出 x_{ij} 和 x_{ji} 中有且仅有一个能被置为“开”状态（即设为 1）。这类约束反映了如下事实，即在描述 i 和 j 的关系时，只有两种可能选择：或者 i 排名在 j 之前，或者 j 排名在 i 之前；第二类约束是满足传递性的一种非常巧妙而紧凑的方式，传递性要求，若 $x_{ij}=1$（i 排名在 j 之前）且 $x_{jk}=1$（j 排名在 k 之前），则 $x_{ik}=1$（i 排名在 k 之前）。将类型 1 和类型 2 的约束结合起来，便得到了传递性。由于决策变量为二值变量，因此类型 2 的约束便能防止出现始于 i 又返回到 i 的长度为 3 的循环。类型 1 的约束则防止了长度为 2 的循环的出现。实际上，这两条约束结合起来，便能够防止任意长度的循环的出现。优势图（dominance graph）可以帮助说明这一点。

我们的 4 × 4 例子中的 $\boldsymbol{X}$ 为

$$\boldsymbol{X}=\begin{array}{c} \\ 1\\2\\3\\4\end{array}\begin{array}{c}\begin{array}{cccc}1&2&3&4\end{array}\\\begin{pmatrix}0&1&1&1\\0&0&1&1\\0&0&0&1\\0&0&0&0\end{pmatrix}\end{array}$$

⊖ 或者对行和按降序排序。

它还可以用图 15.1 中的优势图来描述。对应每个排名向量都可以画出这样的一幅图，它描述了对象对于所有排名在它之下的其他对象所拥有的优势地位。任何一个排名向量的优势图都不包含朝上的上行弧，因为上行弧将造成一个循环，即违反了类型 2 的约束。要理解类型 1 和 2 的约束是怎样结合起来以防止任何循环的出现，让我们考虑一下循环 $1 \rightarrow 3 \rightarrow 4 \rightarrow 1$，与之对应的约束是类型 2 约束 $x_{13} + x_{34} + x_{41} \leqslant 2$。由于对象 1 排名在对象 3 之前，因此 x_{13} 被置位（即 $x_{13} = 1$）；类似有 $x_{34} = 1$；然后根据类型 2 约束，x_{41} 必须为 0；再结合类型 1 约束可知，x_{14} 必须等于 1，传递性由此得到了满足。总之，所有这三类约束（类型 1 和类型 2 再加上对 x_{ij} 的二值性约束）结合起来，便能给出一个 $\boldsymbol{X}$ 矩阵的解，它只是阶梯形矩阵的一个重排。最后，由于 $\boldsymbol{X}$ 总是阶梯形矩阵的重排，因此它的行和与列和均各不相同，并由此给出了这 n 个对象的唯一排名。上述的完整二值整数线性规划问题（BILP，Binary Integer Linear Program）为

$$\max \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

对所有不同的对 (i,j)，有 $x_{ij} + x_{ji} = 1$　　（类型 1——反对称性）

对于所有不同的三元组 (i,j,k)，有 $x_{ij} + x_{jk} + x_{ki} \leqslant 2$　（类型 2——传递性）

$x_{ij} \in \{0,1\}$　　（类型 3——二值性）

这个 BILP 中包括 $n(n-1)$ 个二值决策变量，$n(n-1)$ 个类型 1 的约束，以及 $n(n-1)(n-2)$ 个类型 2 的不等式约束。$O(n^3)$ 阶次数量的类型 2 约束极大地限制了这一最优排名聚合法所能求解的排名问题的规模。幸运的是，可以采用多种策略来规避这一伸缩性问题——请见**第 190 页**和**第 191 页**上关于约束松弛和限界的章节。

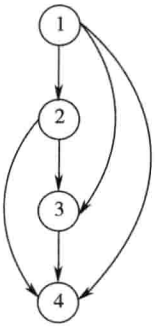


图 15.1　优势图

范例

在本章中，我们将抛开那个常用的 5 × 5 的范例，而使用一个稍大些的 12 × 12 的例子。这个例子来自于大学篮球联赛南方联盟（SoCon，Southern Conference）的 2008—2009 赛季，其中包含了若干5 ×5范例中未能包含的有趣性质。具体来说，这个SoCon示例中

包含了平局，而平局正是第187页上多最优解一节的主题，因此SoCon示例更适于演示之用。

在这一节里，我们将聚合由16个不同方法给出的排名。我们使用了采用不同加权方案的科利法、梅西法、马尔可夫法和OD法作为输入方法。利用式（15.1）给出的一致性定义，得到 $\boldsymbol{C}$ 矩阵为

$$\boldsymbol{C}=\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12\end{array}\begin{array}{c}\begin{array}{cccccccccccc}1&2&3&4&5&6&7&8&9&10&11&12\end{array}\\\begin{pmatrix}0&-16&-14&-12&-12&16&14&14&-2&16&-14&-14\\16&0&-12&-12&-12&16&16&16&14&16&12&0\\14&12&0&-12&-12&16&16&16&16&16&12&12\\12&12&12&0&-14&14&12&12&12&14&12&12\\12&12&12&14&0&16&12&12&12&12&12&12\\-16&-16&-16&-14&-16&0&12&12&-16&12&-16&-16\\-14&-16&-16&-12&-12&-12&0&-12&-14&6&-12&-14\\-14&-16&-16&-12&-12&-12&12&0&-14&16&-12&-14\\2&-14&-16&-12&-12&16&14&14&0&16&-12&-16\\-16&-16&-16&-14&-12&-12&-6&-16&-16&0&-16&-16\\14&-12&-12&-12&-12&16&12&12&12&16&0&-8\\14&0&-12&-12&-12&16&14&14&16&16&8&0\end{pmatrix}\end{array}。$$

求解BILP可得最优目标函数值870，而解矩阵 $\boldsymbol{X}$ 为

$$\boldsymbol{X}=\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12\end{array}\begin{array}{c}\begin{array}{cccccccccccc}1&2&3&4&5&6&7&8&9&10&11&12\end{array}\\\begin{pmatrix}0&0&0&0&0&1&1&1&0&1&0&0\\1&0&0&0&0&1&1&1&1&1&1&0\\1&1&0&0&0&1&1&1&1&1&1&1\\1&1&1&0&0&1&1&1&1&1&1&1\\1&1&1&1&0&1&1&1&1&1&1&1\\0&0&0&0&0&0&1&1&1&1&0&0\\0&0&0&0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&1&0&0&1&0&0\\1&0&0&0&0&1&1&1&0&1&0&0\\0&0&0&0&0&0&0&0&0&0&0&0\\1&0&0&0&0&1&1&1&1&1&0&0\\1&1&0&0&0&1&1&1&1&1&1&0\end{pmatrix}\end{array}。$$

对 $\boldsymbol{X}$ 的列和按升序排序，得SoCon的12支队伍排名为

$$\begin{array}{ll}\text{戴维逊学院}&5\\\text{查尔斯顿学院}&4\\\text{要塞军事学院}&3\\\text{沃佛德学院}&12\\\text{田纳西大学查塔努加分校}&2\\\text{西卡罗莱纳大学}&11\\\text{桑佛德大学}&9\\\text{阿巴拉契亚州立大学}&1\\\text{伊隆大学}&6\\\text{佐治亚南方大学}&8\\\text{弗尔曼大学}&7\\\text{北卡罗来纳州立大学格林波若分校}&10\end{array}。$$

求解 BILP

BILP一般通过一种称为分支限界法的方法来求解，这种方法利用问题的一系列线性规划（LP，Linear Programming）松弛来形成一棵树，以缩小在离散解空间中的搜索过程。由于离散优化比连续优化要难得多，所以问题的规模就成为了一个问题。请注意，问题的规模由待排名对象的数量n所决定。我们的排名聚合BILP中包含了$n(n-1)$个二值决策

变量，以及 $n(n-1)$ 个类型 1 的约束和 $n(n-1)(n-2)$ 个类型 2 的约束。$O(n^3)$ 阶次数量的约束极大地限制了这个最优一致性方法所能处理的排名问题的规模。相比之下，前一章中的启发式方法则不存在这类伸缩性方面的问题。幸运的是，有几种非常巧妙的策略可以用来绕开这个伸缩性问题，正如**第 190 页**和**第 191 页**中所说明的那样。

BILP 的多最优解

分支限界法过程终止于一个最优解 $\boldsymbol{X}$。正如在我们的小例子中所看到的那样，找出 $\boldsymbol{X}$ 的列和并将其按升序排序，就能给出事物的最优排名。问题搞定——我们也该收拾收拾东西回家了。但如果你是数学家的话，你就不会这么做了。数学家们喜欢问一些更深层次的问题，如：这个最优解是否唯一？如果不唯一，怎样找到其他的最优解？

事实表明，对这些问题的答案，揭示了有关上述可伸缩性问题的若干线索。通过一个简单的测试，可以确定 BILP 的最优解是否唯一。考虑最优排名列表中依次相邻的各对对象，并看一看是否可以将这一对对象 i 和 j 的排名调换，而不会改变目标函数值。我们只需考虑相邻对象的排名对调即可，因为这是唯一一种不违反约束的对调方式。如果 $c_{ij} = c_{ji}$，则目标函数值就不会发生改变。假设这一点成立，那么另一个最优解便是将这两个对象排名对调后给出的解。确实，在排名位置上，还有着比这样的二元平局更复杂的情况，例如，当排名相邻的对象 i、j 和 k 满足 $c_{ij} = c_{ji} = c_{ik} = c_{ki} = c_{jk} = c_{kj}$ 时，便会出现三元平局。按这种方式，从最优排名列表的首位开始，便可以一直往下检测每个位置上的所有二元或更多元的平局。我们将这个平局检测算法应用于 SoCon 示例。对于第 185 页上的范例，BILP 给出的最优排名为

$$
\begin{array}{ll}
\text{戴维逊学院} & 5 \\
\text{查尔斯顿学院} & 4 \\
\text{要塞军事学院} & 3 \\
\text{沃佛德学院} & 12 \\
\text{田纳西大学查塔努加分校} & 2 \\
\text{西卡罗莱纳大学} & 11 \\
\text{桑佛德大学} & 9 \\
\text{阿巴拉契亚州立大学} & 1 \\
\text{伊隆大学} & 6 \\
\text{佐治亚南方大学} & 8 \\
\text{弗尔曼大学} & 7 \\
\text{北卡罗来纳州立大学格林波若分校} & 10
\end{array}
\quad 。
$$

在这个平局检测算法的开始，我们对列表中最前两位的队伍进行比较。由于 $\boldsymbol{C}(5,4) \neq \boldsymbol{C}(4,5)$，所以这两支队伍不能交换名次。然后我们移到接下去的两支队伍，即队伍 4 和 3。由于 $\boldsymbol{C}(4,3) \neq \boldsymbol{C}(3,4)$，所以它们也不能换位。继续下移到队伍 3 和 12，它们仍然不能互换。接着，我们终于遇到了一对可以互换的队伍，即队伍 12 和 2。这意味着在一个最优排名列表中，队伍 12 和 2 可以按任意的相对先后顺序出现，或者如 BILP 算法那样在 12 之后是 2，或者是将 2 排在 12 之前，后者给出了另一个最优解，因为它并没有改变目标函数值，而解的可行性也仍然得到了保持。至此，我们就找到了一个队伍 12 和 2 之间的二元平局，不过还可能存在三元或更多元的平局，因此我们再来检查一下列表中再下一支队

伍，即队伍11，看它是否满足 $\boldsymbol{C}(12,2)=\boldsymbol{C}(2,12)=\boldsymbol{C}(2,11)=\boldsymbol{C}(11,2)=\boldsymbol{C}(12,11)=\boldsymbol{C}(11,12)$。不过这个条件未能满足。因此，在第4名这个位置上的平局，的确就只是队伍12和2之间的一个二元平局。继续顺着列表往下考虑11和9，然后是9和1，等等，而我们也并未再发现其他平局。因此，这个SoCon示例具有两个整数最优解，以及无穷多个在这两者之间的边界上的分数解。

总的来说，我们现在已经知道，我们可以（1）通过一个分支限界法来找到排名聚合BILP问题的一个最优解；（2）检查所得解的唯一性；并且（3）如果存在的话，通过上述简单的 $O(n)$ 次的测试找到多个可替代的最优解。归根结底，这一最优化方法给出了一个输出排名，其中可能包含了平局。这种方法从数学角度而言非常有吸引力，而且可以证明它是一个最优排名方法。但是，BILP比许多现有的评分和排名方法慢得多。实际上，由于具有 $O(n^3)$ 数量级的约束，在实际中，如DASH最优化软件或NEOS服务器等商业BILP求解软件只能解决大约1000个对象规模的问题。因此，对所有NCAA Ⅰ区的橄榄球或篮球队进行排名肯定是可以的，但对网络空间中的数十亿网页进行排名则超出了能力范围。不过，对数个流行的搜索引擎所给出的前50个结果进行排名，则不但是可能的，而且还能很快完成——即可以实时进行。幸运的是，在下一节中，我们将说明如何进一步提高实际应用中对 n 的限额，从而使得对数千个对象的聚合排名成为可能。

BILP的LP松弛

求解一个大规模的BILP难题的第一步，便是求解一个与之相关的问题，即LP松弛问题。就我们的具体情况而言，就是将类型3的约束进行松弛。类型3的约束要求变量 x_{ij} 必须取离散值，更具体来说就是 x_{ij} 必须为二值变量，因为 $x_{ij}\in\{0,1\}$。对这个约束进行松弛，将使得这些变量能够取连续值，即 $0\leqslant x_{ij}\leqslant 1$。实际上，界限约束 $0\leqslant x_{ij}\leqslant 1$ 的上界是多余的，因为这一限制已经被类型1的约束 $x_{ij}+x_{ji}=1$ 所包含在内。因此，简化后的排名聚合问题LP松弛为

$$\max\sum_{i=1}^{n}\sum_{j=1}^{n}c_{ij}x_{ij}$$

对所有不同的对 (i,j)，有 $x_{ij}+x_{ji}=1$　　（类型1——反对称性）

对于所有不同的三元组 (i,j,k)，有 $x_{ij}+x_{jk}+x_{ki}\leqslant 2$　（类型2——传递性）

$x_{ij}\geqslant 0$　　（类型3——连续形式）

当某些BILP被作为LP来求解时，LP松弛问题的最优解正好是二值解，从而显然也是BILP的最优解。这是最为理想的情况。稍差一些的情况则是LP的最优解中仅含有很少量的分数值。一般对这种情况，可以将这些分数值圆整为最接近的整数值，从而给出一个稍差的次优解，如果此解可行的话，便能成为严格最优整数解的一个足够好的近似。在这一节中，我们将证明上面的LP问题将给出非常有趣的结果。在很多时候，LP的解都是最优的，而且实际上这些解还能方便地告诉我们所有其他的最优解。

我们12支队伍的SoCon示例能很好地说明这一点。在第187页的多最优解一节中，我们发现该示例具有两个整数最优解。一个解给出的排名为（5　4　3　12　2　11　9　1　6　8　7　10），而另一个解则是将队伍12和2的位置互换。请注意这个平局的情况是

如何在下面这个 LP 解矩阵 $\boldsymbol{X}$ 中得到体现的。

$$\boldsymbol{X} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \end{array} \begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0.478 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0.522 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \end{array}$$

这个 LP 解在位置 $\boldsymbol{X}(12,2)$ 和 $\boldsymbol{X}(2,12)$ 上取分数值，这两个位置正对应着 $\boldsymbol{C}$ 中表明队伍 12 和 2 之间存在平局的元素位置。LP 的求解过程以一个分数最优解告终，而这个解正好位于 BILP 的两个整数最优解之间的边界上。由于 LP 终止于与 BILP 相同的最优目标函数值 870，因此我们可以知道，这个 LP 解对于原来的 BILP 问题是最优的。

尽管在上面的 SoCon 示例中，我们能够得到可以构造出最优二值解的分数解，但这一点并非总能保证成立。实际上，我们曾设计了一个包含 9 个对象的例子，它只有唯一一个分数最优解，因此当构造二值解时，对应的目标函数值要差于这个唯一的分数解所对应的值。但是，这类仅具有一个分数最优解的例子并不容易构建。实际上，为了找到这样一个例子，我们随机生成了数千个元素值均匀分布于 -1 到 1 之间的 $\boldsymbol{C}$ 矩阵，才碰到了这个具有唯一最优分数解的 9 个对象示例。因此，尽管唯一分数解是可能的，但在实际中，我们的排名问题的松弛 LP 形式更可能给出非唯一的分数解，由它可以构造出多个最优二值解。赖内尔特（Reinelt）等人进行的实验型研究[62~65]则提供了更多的证据，表明实际中排名问题的 LP 解好得出奇，而且常常是最优且二值的。

那么在什么情况下，我们才可以确定最优的 LP 解对于 BILP 也为最优呢？要记住，不管怎样，BILP 才是我们真正感兴趣的问题。当然，如果这个 LP 解本身是二值的，那么它对于 BILP 而言就是最优的。但即使当 LP 解是分数解时，则正如 SoCon 的例子所表明的那样，在某些实例中，这个解对于 BILP 也仍然是最优的。在参考文献［50］中，兰维尔、佩丁斯和山本洋右证明了 LP 解何时为 BILP 最优解的条件。他们的定理还将 BILP 多最优解的存在性与 LP 解中的分数值联系了起来，而多最优解就表明了排名中存在平局的情况。

他们的定理对于计算而言也具有重要意义。原始的 LP 求解方法即著名的单纯形法，并不是求解我们的 MVR（Minimum Violation Ranking，即最少违例排名的缩写）问题的选择。单纯形法是一个基于顶点的方法，也就是说它由一个顶点移动到另一个顶点，移动过程中目标函数值不断改善，直至到达一个最优解，而这个解自然也是一个顶点。与之不同的是，内点 LP 求解方法则是穿过可行解域的内部，直至收敛于一个最优解，根据穿越可行解域时所选择的路径，这个解可能是一个顶点，也可能是一个边界点。对于我们来说，正是边界最优解（其中包含了分数值）而非顶点最优解（仅包含整数值的解），才能给出

远为丰富的信息。因此，在求解排名问题所对应的 LP 松弛问题时，我们总是选择一个并非基于顶点的、非单纯形法的 LP 求解方法。

约束松弛

即使使用了 LP 松弛，$O(n^3)$ 数量级的类型 2 约束也仍然极大限制了可处理的排名问题的规模。要进一步提高可处理的问题规模，我们必须求助于另一类松弛技术即*约束松弛*。在 $n(n-1)(n-2)$ 个类型 2 的约束 $x_{ij}+x_{jk}+x_{ki}\leqslant 2$ 中，仅有极少部分的约束是真正必要的，而绝大多数约束都自然成立——问题在于我们不知道哪些是必要的，而哪些是不必要的。为了找到答案，我们首先假设所有约束均不必要，然后随着那些必要的约束被逐渐识别出来，我们再将它们添加回去。类型 1 和类型 3 约束（松弛的连续形式）并不会带来问题，因此我们不对它们进行变动。以下是这个约束松弛方法的步骤。约束松弛可同样成功地用于前述 BILP 或 LP 问题，我们在此仅就 BILP 来进行说明。

大规模排名 BILP 问题的约束松弛算法

1. 松弛所有类型 2 的约束，因此，必要的类型 2 约束的集合初始时为空。

2. 利用当前的类型 2 必要约束集合来求解 BILP，并生成所得解对应的最优排名。这个排名实际上是我们希望获得的、使用完整的类型 2 约束集合所得的原始问题真实排名的一个近似。

3. 确定第 2 步所得的解违反了哪些类型 2 的约束——它们便是必要的类型 2 约束。将这些类型 2 约束加入类型 2 必要约束集合，再返回第 2 步。如此重复直至没有类型 2 的约束再被违反。此时的 BILP 解便是具有完整约束集的原问题的一个最优排名。

在第 3 步中，要确定当前 BILP 解违反了哪些类型 2 的约束是很容易的，而且也并不需要逐条检查所有的约束。回忆一下，根据图 15.1，对类型 2 的传递性约束的违背，就对应了优势图中的上行弧。再从另一个角度来看，这些违反约束的情况，将表现为排名重排矩阵 $\boldsymbol{X}$ 中处于下三角部分的那些 1。接着，对于每条所发现的上行弧 (j,i)，我们就要找到所有满足 $x_{ik}=x_{kj}=1$ 的 k，并产生相应的类型 2 约束 $x_{ji}+x_{ik}+x_{kj}\leqslant 2$。利用矩阵 $\boldsymbol{X}$ 可以迅速找到这些元素 k：求出 $\boldsymbol{X}$ 的第 i 行和第 j 列的阿达马（Hadamard）乘积（即逐元素积），其中所有的非零元素均满足 $x_{ik}=x_{kj}=1$。由于重排的矩阵 $\boldsymbol{X}$ 具有上三角结构，因此要计算阿达马积（Hadamard product）并产生对应于每条上行弧的传递性约束，只需要远不到 $O(n)$ 次的计算。

此外，我们还能对可能存在的任一近似排名加以利用。如假设我们利用一个快速的启发

式方法得到了一个排名，而这个排名与某个上三角矩阵存在着一一对应关系，则我们称该矩阵为 $\overline{\boldsymbol{X}}$。对这个近似解矩阵 $\overline{\boldsymbol{X}}$ 计算目标函数值 $f(\overline{\boldsymbol{X}}) = \boldsymbol{C}. * \overline{\boldsymbol{X}}$，便得到了目标函数的一个有用的下界。一旦 BILP 的分支限界过程遇到某条分支的结点，其目标函数值会低于 $f(\overline{\boldsymbol{X}})$，那这时就不必再继续考察这一整条分支上的结点了。

总的来说，约束松弛方法是一种迭代方法，它对一系列 BILP 进行求解，这些问题的约束逐渐增多，直至找到所有必需的传递性约束。在每次迭代中得到的最优 BILP 解，都是具有完整约束集的原始问题的真正最优解的近似。这些近似解的质量在不断改善，直至达到最优排名。

敏感性分析

LP 相比于 BILP 的另一个优点，与求解线性规划问题时所固有的敏感性有关。此时，我们对于目标系数 c_{ij} 的变化更感兴趣。输入数据（具体来说，就是产生目标系数矩阵 $\boldsymbol{C}$ 的差值矩阵 $\boldsymbol{D}$）中的小变动，有可能会改变最优解以及由此得到的最优排名。利用如 XpressMP 这样的商业最优化软件，我们可以求出 12 支队伍的 SoCon 示例中各目标系数的敏感区间如图 15.2 所示。

```
      1          2          3          4          5          6          7          8          9          10         11         12
-------------------------------------------------------------------------------------------------------------------------------------
 1 : [   0,   0][-Inf,  16][-Inf,  14][-Inf,  12][-Inf,  12][ -16, Inf][ -14, Inf][ -14, Inf][-Inf,   2][ -16, Inf][-Inf,  14][-Inf,  14]
 2 : [ -16, Inf][   0,   0][-Inf,  12][-Inf,  12][-Inf,  12][ -16, Inf][ -16, Inf][ -16, Inf][ -14, Inf][ -16, Inf][ -12, Inf][  -0,  32]
 3 : [ -14, Inf][ -12, Inf][   0,   0][-Inf,  12][-Inf,  12][ -16, Inf][ -16, Inf][ -16, Inf][ -16, Inf][ -16, Inf][ -12, Inf][ -12, Inf]
 4 : [ -12, Inf][ -12, Inf][ -12, Inf][   0,   0][-Inf,  14][ -14, Inf][ -12, Inf][ -12, Inf][ -12, Inf][ -14, Inf][ -12, Inf][ -12, Inf]
 5 : [ -12, Inf][ -12, Inf][ -12, Inf][ -14, Inf][   0,   0][ -16, Inf][ -12, Inf][ -12, Inf][ -12, Inf][ -12, Inf][ -12, Inf][ -12, Inf]
 6 : [-Inf,  16][-Inf,  16][-Inf,  16][-Inf,  14][-Inf,  16][   0,   0][ -12, Inf][ -12, Inf][-Inf,  16][ -12, Inf][-Inf,  16][-Inf,  16]
 7 : [-Inf,  14][-Inf,  16][-Inf,  16][-Inf,  12][-Inf,  12][-Inf,  12][   0,   0][-Inf,  12][-Inf,  14][  -6, Inf][-Inf,  12][-Inf,  14]
 8 : [-Inf,  14][-Inf,  16][-Inf,  16][-Inf,  12][-Inf,  12][-Inf,  12][ -12, Inf][   0,   0][-Inf,  14][ -16, Inf][-Inf,  12][-Inf,  14]
 9 : [  -2, Inf][-Inf,  14][-Inf,  16][-Inf,  12][-Inf,  12][ -16, Inf][ -14, Inf][ -14, Inf][   0,   0][ -16, Inf][-Inf,  12][-Inf,  16]
10 : [-Inf,  16][-Inf,  16][-Inf,  16][-Inf,  14][-Inf,  12][-Inf,  12][-Inf,   6][-Inf,  16][-Inf,  16][   0,   0][-Inf,  16][-Inf,  16]
11 : [ -14, Inf][-Inf,  12][-Inf,  12][-Inf,  12][-Inf,  12][ -16, Inf][ -12, Inf][ -12, Inf][ -12, Inf][ -16, Inf][   0,   0][-Inf,   8]
12 : [ -14, Inf][ -32,   0][-Inf,  12][-Inf,  12][-Inf,  12][ -16, Inf][ -14, Inf][ -14, Inf][ -16, Inf][ -16, Inf][  -8, Inf][   0,   0]
```

图 15.2　目标系数 c_{ij} 的敏感区间

大多数 c_{ij} 系数的敏感区间都十分宽松。实际上，仅有的具有有限宽度的区间，就位于对应了队伍 2 和 12 这一对对象的两个位置上。对于这个数据集而言，这两个区间提醒我们，对于处在中游的这两支队伍的排名顺序，我们的把握相对更少一些，而实际上在**第 187 页**的多最优解一节中，这正是我们所发现的排名持平的两支队伍。如果目标系数的变化超出了这些区间，则将导致排名发生变化。

限界

限界法在最优化中极为有用，整数规划的求解尤其依赖于限界。在本节，我们将在 LP 迭代中应用限界法，以加速收敛并保证解的最优性。请记住，这个 LP 迭代法对传递性约束集进行了松弛。实际上，在第一次迭代中，LP 是在没有约束的情况下求解的。这个迭代所给出的最优解矩阵 $\boldsymbol{X}$ 对应了一个目标函数值，它是最优目标函数值 f^* 的一个上界，所以我

们记之为$\bar{f}$。这次迭代的解几乎肯定会违反某些传递性约束，因此不是原始 LP 问题的可行解。但是通过求取 $\boldsymbol{X}$ 的行和，可以得到一个非常有用的近似解，其中第 i 行的和便是一个很好的指标，说明第 i 支队伍能够击败多少对手。随着 LP 迭代法的进行，我们利用每次迭代的解矩阵 $\boldsymbol{X}$ 的行和来计算一个近似排名，当迭代次数增加时，这个近似排名将越来越接近于最优排名。由于包括近似解在内的每个排名都有一个一一对应的 $\boldsymbol{X}$ 矩阵，因此我们可以由每个近似排名计算出对应的目标函数值，我们将它记为 $\underline{f}$，因为它是f^* 的一个下界。于是可知

$$\underline{f} \leqslant f^* \leqslant \bar{f}。$$

下面我们求取近似排名的相对误差$\dfrac{\bar{f}-f^*}{f^*}$的界限。由于$\underline{f}$，f^* 和$\bar{f}$都具有相同的符号，因此把未知的f^* 用已知的$\underline{f}$和$\bar{f}$来替代，我们便可以将相对误差限制在

$$\frac{\bar{f}-f^*}{|f^*|} \leqslant \frac{\bar{f}-\underline{f}}{|f^*|} \leqslant \frac{\bar{f}-\underline{f}}{|\underline{f}|}。$$

进一步地，当目标系数矩阵 $\boldsymbol{C}$ 的所有元素均为整数且$\bar{f}-\underline{f} \leqslant 1$ 时，此时的近似排名便是原排名问题的最优解。即使最优性不能如此幸运地得到保证，但我们也可以确保这个近最优解所对应的误差落在某个范围之内。我们可以保证最优目标函数值落在$\underline{f}$和$\bar{f}$之间，而相对误差不会大于$\dfrac{\bar{f}-\underline{f}}{|\underline{f}|}$。这是使用限界的 LP 迭代法的另一个巨大优势。不仅因为它所需要的迭代次数更少，而且也因为用户可以在得到一个相对误差可接受的解之后，就能够立即终止迭代过程。

杂谈：排名在罗马

纵观本书，我们一直都在强调所介绍的排名方法的各色用途，在这些用途中，看似最不可能也最不为人所知者之一，便是它在考古学中的应用。设想一下在意大利庞贝或罗马的巨大发掘现场。对于那些希望正确确定考古发掘出的物品年代的考古学家来说，“到了罗马就排名”这句话是一个很好的建议。一次典型的发掘通常由几个挖掘点组成。在每个挖掘点所发现的文物，将根据各自出土的地层深度被贴上标签。各种风格的陶器或各种类型的文物将在多个挖掘点的多个地层中被发现，而考古学家的目标就是确定某类文物的相对年代。例如，在运水的时候，罗马人首先使用的是彩绘的黏土陶瓶，还是马赛克泥土陶罐？在一个挖掘点，陶瓶可能出现在陶罐的更下层，说明陶瓶首先被使用；但到了另一个挖掘点，这个次序可能又会反过来。实际上，每个挖掘点都根据出土深度给出了文物的一个排名，而现在的目的就是确定所有文物的一个整体排名，它与各个挖掘点给出的排名最为一致。数学上来说，就是要将各个挖掘点给出的排名聚合为一个最优排名，使它与这些挖掘点排名之间的吻合性达到最大。

最优化排名聚合方法总结

下面的阴影框总结了我们在介绍最优排名聚合法时所用的记法。

（利用最优化的）排名聚合方法记法

n　待排名的对象数量
$\boldsymbol{C}$　目标函数系数所组成的一致性矩阵

下面的清单总结了该排名聚合方法的步骤。

（利用最优化的）排名聚合方法

1. 根据某个一致性定义生成矩阵 $\boldsymbol{C}$。如
   c_{ij} =（i 的排名在 j 之前的列表数量）-（i 的排名在 j 之后的列表数量）。
2. 求解 LP 最优化问题

$$\max \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

   对所有不同的对 (i,j)，有 $x_{ij} + x_{ji} = 1$　（类型 1——反对称性）
   对于所有不同的三元组 (i,j,k)，有 $x_{ij} + x_{jk} + x_{ki} \leqslant 2$　（类型 2——传递性）
   $x_{ij} \geqslant 0$　（类型 3——连续形式）。
   对于 n 值较大的问题，使用第 190 页和第 191 页上各节中介绍的约束松弛和限界方法。
3. 将最优解矩阵 $\boldsymbol{X}$ 的列和按升序排序，便得到了最优排名。$\boldsymbol{X}$ 中具有分数值的元素位置给出了排名相同的位置。

我们通过说明该排名聚合方法的一系列性质来结束本节。

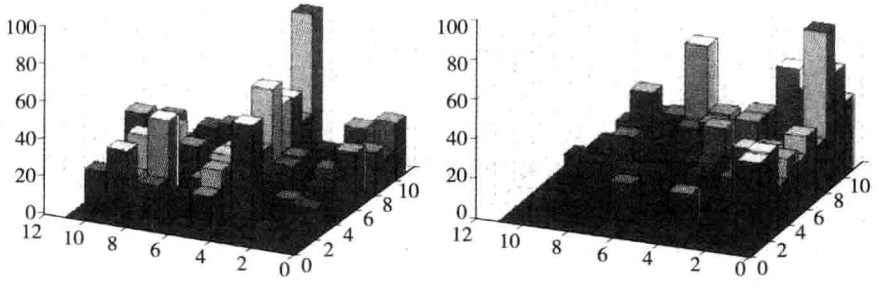
- 我们的实验表明，当需要聚合的列表从最初就在很大程度上相互符合时，那只需少数几次迭代便能完成聚合。如果你想要建立一个元搜索引擎，并将 5 个流行的搜索引擎所返回的前 100 个结果加以聚合，那上面所述的就是一个好消息。在最新的便携式计算机上，我们的排名聚合方法仅用几秒钟就能执行完毕。
- 简化的 LP 松弛的约束形成了一个多面体，这一多面体已得到了充分研究，它常被称为线性序多面体（linear ordering polytope）[65]。研究一下这个 LP 的多面体以及原始BILP问题的可行解域之间的关系，将对我们有所帮助。当然，BILP的可行

解域包含于 LP 的可行解域之中。一种最好的情况，便是 LP 的可行解域尽可能地紧贴 BILP 的可行解域，也就是说，LP 的可行解域是 BILP 可行解域中的点的凸包。对于我们的排名问题，好消息是所有的不等式约束（类型 2 的传递性约束和类型 3 的非负约束）就是定义了该线性序多面体若干表面的不等式，即这些不等式已经尽可能地紧。不过，LP 的约束集并没有涵盖线性序多面体的所有表面。如所谓的栅栏或莫比乌斯梯不等式这类复杂而合法的不等式，可以给出更强的 LP 松弛，不幸的是，要产生出这些不等式，所花费的代价过高[57,62,64]。

- 伴随 LP 而来的一个非常有用的能力，便是它能够进行敏感性分析，即 LP 可用于“假如……会怎样”这一类型的分析。当某些输入参数被改变时，我们可以确定最优解是否也会发生改变，以及改变的程度为多少。由于我们已经说明 LP 可以用于求解排名问题，因此它具有进行敏感性分析的潜力。在这种情况下，我们可以回答目标系数 c_{ij} 如何影响最优解的这个问题。这种敏感性分析是其他排名方法很少能传递的附加信息。

重温评分差距法

在第 8 章中，我们描述了对象排名的评分差距法。这个方法的目的在于找到对象的一个重排，当这个重排应用于对象-对象的数据差矩阵时，将生成一个尽可能接近于所谓“山坡形”的矩阵。如下用图示总结了这一思路。



左图是一个 11 × 11 的矩阵按对象原有顺序排列时的城市区图，而右图则是同样的数据经过新的最优重排后得到的城市区图。如果我们选择这个重排，以使矩阵尽可能接近于山坡形时，便出现了右图所体现出的山坡形状。

评分差距与排名聚合的对比

在第 8 章中，我们给出了一个正规的最优化问题，其目标在于，当所得的城市区图与第 108 页文本框中定义的坡形之间存在出入的话，该最优化问题将使（可能经过加权的）这类被违反的坡形约束的数量为最少。在那一章中，我们介绍了一种求解该优化问题的进化方法。我们将在本节说明，评分差距法中的最优化问题实际上可以利用本章的方法进行

最优而高效地求解。本节中的大部分工作出现在参考文献［50］中，其中我们称这种方法为最少违例排名（MVR，Minimum Violation Ranking）法，因为它给出了对坡形约束违反最少的排名。

实际上，评分差距法可以被严格地嵌入本章的 BILP 以及 LP 之中。让我们关注于分差构成的数据矩阵 $\boldsymbol{D}$，其中，d_{ij} 为胜队 i 在比赛中击败负者 j 时的分差，对其他情况则为 0。现在的诀窍在于将 $\boldsymbol{D}$ 的每一行和每一列都视为一支队伍对其对手给出的排名。例如，对于我们的 5 支队伍的例子有

$$\boldsymbol{D} = \begin{array}{c} \\ \text{Duke} \\ \text{Miami} \\ \text{UNC} \\ \text{UVA} \\ \text{VT} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Duke} & \text{Miami} & \text{UNC} & \text{UVA} & \text{VT} \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 45 & 0 & 18 & 8 & 20 \\ 3 & 0 & 0 & 2 & 0 \\ 31 & 0 & 0 & 0 & 0 \\ 45 & 0 & 27 & 38 & 0 \end{pmatrix} \end{array}$$

$\boldsymbol{D}$ 的第二行告诉我们，Miami 按照防守能力由高到低将对手排名为 UVA，UNC，VT 和 Duke。另一方面，$\boldsymbol{D}$ 的第一列则告诉我们 Duke 根据攻击能力将对手排名为 Miami/VT，UVA 和 UNC。由此，这些对 n 支队伍给出的攻击和防守排名可以被聚合为一个该赛季的总排名。相比于本章的排名聚合方法，其中唯一的调整便是一致性矩阵 $\boldsymbol{C}$ 的定义。此时，评分差距法的一致性矩阵并不是由式（15.1）定义的，而是

矩阵 $\boldsymbol{C}$ 的定义

令 $\boldsymbol{C} = [c_{ij}]$，对于任意 $i,j = 1,2,\cdots,n$ 定义为

$$c_{ij} = \#\{k \mid d_{ik} < d_{jk}\} + \#\{k \mid d_{ki} > d_{kj}\} \tag{15.2}$$

其中#表示对应集合的势。因此

$$\#\{k \mid d_{ik} < d_{jk}\}$$

为对阵 i 时的分差小于对阵 j 时的分差的队伍数量。类似地，$\#\{k \mid d_{ki} > d_{kj}\}$ 为对阵 i 时的分差大于对阵 j 时的分差的队伍数量。

注意：以上的矩阵 $\boldsymbol{C}$ 以二值的方式来计算违背坡形的次数，但我们还能弄得更复杂点。例如，我们可以考虑将出现违背坡形约束时的差值加以求和，来得到加权的违例度量。此时，$\boldsymbol{C}$ 定义为 $c_{ij} = \sum_{k:d_{ik}<d_{jk}} (d_{jk} - d_{ik}) + \sum_{k:d_{ki}>d_{kj}} (d_{ki} - d_{kj})$。

范例

这个矩阵 $\boldsymbol{C}$ 的定义对于 12 支队伍的 SoCon 示例给出了如下的一致性矩阵。

$$\boldsymbol{C}=\begin{array}{c} \\ 1\\2\\3\\4\\5\\6\\7\\8\\9\\10\\11\\12 \end{array}\begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array}\\ \begin{pmatrix} 0 & 15 & 15 & 14 & 17 & 7 & 4 & 4 & 9 & 2 & 10 & 11\\ 8 & 0 & 10 & 12 & 18 & 6 & 3 & 3 & 11 & 3 & 7 & 8\\ 5 & 11 & 0 & 9 & 14 & 6 & 2 & 4 & 9 & 2 & 5 & 9\\ 5 & 9 & 9 & 0 & 15 & 5 & 0 & 2 & 6 & 3 & 6 & 5\\ 2 & 2 & 5 & 3 & 0 & 2 & 1 & 2 & 0 & 1 & 1 & 2\\ 10 & 14 & 16 & 17 & 18 & 0 & 7 & 7 & 12 & 4 & 13 & 15\\ 15 & 18 & 18 & 20 & 20 & 13 & 0 & 8 & 16 & 10 & 15 & 15\\ 15 & 20 & 18 & 18 & 20 & 13 & 10 & 0 & 15 & 11 & 14 & 18\\ 10 & 9 & 11 & 14 & 19 & 7 & 4 & 7 & 0 & 2 & 10 & 9\\ 17 & 17 & 18 & 18 & 20 & 16 & 7 & 9 & 15 & 0 & 13 & 14\\ 10 & 14 & 14 & 10 & 18 & 8 & 4 & 4 & 12 & 7 & 0 & 12\\ 10 & 12 & 11 & 12 & 17 & 7 & 4 & 4 & 10 & 6 & 8 & 0 \end{pmatrix} \end{array}.$$

利用本章的 LP 形式求解这个示例，便得到了如下的最优 $\boldsymbol{X}$ 矩阵。

$$\boldsymbol{X}=\begin{array}{c} \begin{array}{cccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{array}\\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0.4951 & 0\\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1\\ 1 & 1 & 0 & 0.5896 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 1 & 1 & 0.4104 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0\\ 0.5049 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix} \end{array},$$

与之对应的队伍排名如下

$$\begin{array}{l} \text{戴维逊学院}\\ \text{查尔斯顿学院}\\ \text{要塞军事学院}\\ \text{桑佛德大学}\\ \text{田纳西大学查塔努加分校}\\ \text{沃佛德学院}\\ \text{阿巴拉契亚州立大学}\\ \text{西卡罗莱纳大学}\\ \text{伊隆大学}\\ \text{北卡罗来纳州立大学格林波若分校}\\ \text{弗尔曼大学}\\ \text{佐治亚南方大学} \end{array}\begin{pmatrix} 5\\4\\3\\9\\2\\12\\1\\11\\6\\10\\7\\8 \end{pmatrix}.$$

$\boldsymbol{X}$ 中的分数值表明其中存在数个平局。在第二名的位置上有一个队伍 4 和 3 之间的二元平局，而在第 6 名位置上还有一个队伍 11 和 1 之间的二元平局。

下面的文本框中对最少违例排名（MVR）方法进行了总结。

最少违例排名（MVR）法

1. 根据定义，由数据差矩阵 $\boldsymbol{D}$ 产生稀疏矩阵 $\boldsymbol{C}$[⊖]。

$$c_{ij} = \#\{k \mid d_{ik} < d_{jk}\} + \#\{k \mid d_{ki} > d_{kj}\}$$

2. 求解 LP 最优化问题

$$\min \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

对所有不同的对 (i,j)，有 $x_{ij} + x_{ji} = 1$　　（类型 1——反对称性）

对于所有不同的三元组 (i,j,k)，有 $x_{ij} + x_{jk} + x_{ki} \leqslant 2$　　（类型 2——传递性）

$x_{ij} \geqslant 0$　　（类型 3——连续形式）

对于 n 值较大的问题，利用第 190 页和第 191 页各节中介绍的约束松弛和限界方法求解。

3. 将最优解矩阵 $\boldsymbol{X}$ 的列和按升序排序以得到最优排名。$\boldsymbol{X}$ 中分数值的位置便给出了排名持平的位置。

杂谈：三月疯狂锦标赛和大规模 LP 问题

为了展示本章中的 MVR 排名法可以适用的 LP 问题的规模，我们对 2008—2009 赛季 NCAA 大学篮球赛的 347 支队伍进行排名。为了求解这个大规模的 MVR LP 问题，我们使用了迭代约束松弛技术。利用参考文献［50］中一条定理的条件以及我们的计算结果［即 $f(\text{LP}) = f(\text{BILP})$］，LP 迭代法给出了一个非唯一的分数解，它是原始 BILP 问题的一个最优解[⊜]。在这个最优 LP 解中，仅有 0.066% 的非零元素是分数值。此外，这些分数值以及平局都出现在较低的排名位置，具体来说就是排名第 252 到第 272 的位置上。

表 15.1 详细给出了 LP 迭代法在这个完整的 347 支队伍的例子上每次迭代所用的时间。例如，在第 1 次迭代中，求解 LP 问题需时 2.88s，给出的目标函数值为 1778224，而寻找必要的类型 2 约束只用了 0.73s，找到了 11885 个额外的约束，这些约束被加入到了第 2 次迭代所要求解的 LP 问题之中。总体而言，执行所有 23 次迭代并产生所有 20008 个约束，在便携式计算机上仅需 135.37s。由表 15.1 得到的另一个结论与第 190 页上所描述的约束松弛法的值得关注的优点有关。在原有的全体类型2约束中，仅有0.048% 为必需

⊖ 以上的矩阵 $\boldsymbol{C}$ 以二值的方式来对破坏坡形的情况进行计数，但我们还能采用更复杂的方式。例如，我们可以考虑通过将出现违背坡形时的差值加以求和，来得到加权的违例度量。此时，$\boldsymbol{C}$ 定义为 $c_{ij} = \sum_{k:d_{ik}<d_{jk}} (d_{jk} - d_{ik}) + \sum_{k:d_{ki}>d_{kj}} (d_{ki} - d_{kj})$。

⊜ 使用一个稍有不同的累积分差而非平均分差的输入矩阵 $\boldsymbol{D}$，将给出一个二值的 LP 解，它同时也是 BILP 问题的最优解。

的约束。这一开销的节约是十分巨大的，它使得方法能够处理更大规模的排名问题。由表15.1得到的最后一点结论与求解次序有关。注意到在第4次迭代后，LP迭代法已经达到了可行域的最优面上，但解本身不可行。在之后的各次迭代中，解的可行性而非最优性得到改进。换而言之，这些解仍然落在最优超平面上，但随着每次迭代而朝着可行解域越移越近。

表15.1　LP迭代法在347支队伍的示例上的计算明细

迭代	LP费时/s	目标函数值	约束生成费时/s	增加的约束数量
1	2.88	1778224.00	0.73	11885
2	3.94	1777829.00	0.64	6900
3	5.78	1777801.50	0.33	644
4	5.39	1777800.50	0.27	179
5	5.77	1777800.50	0.27	38
6	5.64	1777800.50	0.27	35
7	6.08	1777800.50	0.27	29
8	5.34	1777800.50	0.27	54
9	5.78	1777800.50	0.27	30
10	5.63	1777800.50	0.28	14
11	5.53	1777800.50	0.30	19
12	5.81	1777800.50	0.30	28
13	5.94	1777800.50	0.27	20
14	5.58	1777800.50	0.25	4
15	6.25	1777800.50	0.27	20
16	5.53	1777800.50	0.27	7
17	5.97	1777800.50	0.27	25
18	6.11	1777800.50	0.27	10
19	5.45	1777800.50	0.27	19
20	6.03	1777800.50	0.28	12
21	5.73	1777800.50	0.28	26
22	6.02	1777800.50	0.27	10
23	6.11	1777800.50	0.25	0
共计	128.28		7.09	20008

下面我们将展示使用限界法的LP迭代法所能达到的性能改善。

表 15.2　利用限界法的 LP 迭代法在 347 支队伍的示例上的计算结果

迭代	LP 费时/s	目标函数值	最佳排名	约束生成费时/s	增加的约束数量
1	1.19	1778224.0	1777474.0	0.28	11885
2	1.58	1777829.0	1777516.0	0.27	6900
3	2.24	1777801.5	1777712.0	0.16	644
4	2.13	1777800.5	1777779.0	0.13	179
5	2.22	1777800.5	1777787.0	0.13	38
6	2.17	1777800.5	1777787.0	0.14	35
7	2.33	1777800.5	1777793.0	0.13	29
8	2.09	1777800.5	1777793.0	0.13	54
9	2.22	1777800.5	1777793.0	0.14	30
10	2.19	1777800.5	1777793.0	0.13	14
11	2.14	1777800.5	1777793.0	0.14	19
12	2.27	1777800.5	1777793.0	0.14	28
13	2.30	1777800.5	1777793.0	0.13	20
14	2.16	1777800.5	1777793.0	0.13	4
15	2.39	1777800.5	1777793.0	0.13	20
16	2.16	1777800.5	1777793.0	0.14	7
17	2.30	1777800.5	1777793.0	0.14	25
18	2.36	1777800.5	1777793.0	0.14	10
19	2.13	1777800.5	1777793.0	0.13	19
20	2.34	1777800.5	1777793.0	0.14	12
21	2.22	1777800.5	1777793.0	0.13	26
22	2.34	1777800.5	1777793.0	0.13	10
23	2.34	1777800.5	1777793.0	0.13	0
共计	49.78			3.33	20008

注意，限界法明显减少了总运行时间（53.11s），并得到了一个解，虽然未能证明这是一个最优解，但可以证明它非常接近于最优解，因为相对误差不超过 0.000422%。我们最终能够发现，这个解的确是最优的，因为 BILP 迭代法给出了与 LP 迭代法相同的目标函数值。

数字——

第 8 = 获得 NCAA 篮球锦标赛冠军的最低种子排名

——1985 年，维拉诺瓦大学以 66-64 击败乔治城大学。

第 11 = 进入前四的最低种子排名

——1986 年，路易斯安那州立大学，在半决赛中负于路易斯维尔大学。

——wiki. answers. com，tourneytravel. com

第 16 章　比较排名的方法

第 2 章至第 8 章给出了一系列对象排名方法，这些方法的数量不断增多，但仍属有限。而当我们使用第 12 章中的加权方法和第 14 章与第 15 章中的排名聚合方法时，看起来似乎可以把之前那一系列方法，扩展为一个无限集合，因为将这些方法组合起来的方式，其数量是以指数级增长的。有了如此长长的一串可能的对象排名方法，一个问题便自然而然地冒了出来：哪个排名方法（包括排名聚合方法）最好？换而言之，我们如何比较这些对象排名方法？如往常一样，我们在最开始仍然是对已有文献进行回顾，而这将我们带回到了统计学家莫里斯·肯德尔和查尔斯·斯皮尔曼（Charles Spearman）所进行的工作。

两个排名列表间的定性差异

要确定许多排名列表中哪一个最好，这是一个非常困难的问题，它简直就是一座圣杯。但还有一个稍微容易回答一点的相关问题。这个问题是，任意两个排名列表之间隔得有多远？这个问题的一个答案将最终帮助我们回答之前那个排名列表中何者最佳的更困难的问题。确定两个排名列表之间数值偏差[⊖]的问题，已经由统计学家和数学家进行了一段时间的研究。在第 203 页上关于肯德尔的 τ 测度和第 206 页上关于斯皮尔曼简捷测度的小节中，我们将介绍这类经典数值偏差中的两个，之后，我们会分别在其中加入自己的改进。

尽管两个排名列表之间的数值偏差能够为我们给出比较两个列表的精确定量度量，但在某些时候，一个不那么精确的定性评估也同样有用。因此，在讨论定量度量之前，我们将介绍两个排名列表间差异的一种可视化显示方法。

⊖ 细心的读者可能已经注意到，在这一节中我们将新的度量 ϕ 称为偏差度量而非距离度量。我们是有意如此的。一个度量 ϕ 要成为距离，必须对任意列表 l_1，l_2 和 l_3 满足以下三个性质：

- 恒等性：当且仅当 $l_1 = l_2$ 时，有 $\phi(l_1, l_2) = 0$。
- 对称性：$\phi(l_1, l_2) = \phi(l_2, l_1)$。
- 三角不等式：对于任意 l_1，l_2 和 l_3，$\phi(l_1, l_2) + \phi(l_2, l_3) \geqslant \phi(l_1, l_3)$。

我们的度量 ϕ 能够满足恒等性和对称性，但不满足三角不等式。因此，我们必须称 ϕ 为一个偏差测度，而我们也很乐意去这么称呼，因为这一度量是专门为体育排名问题而量身打造的。换而言之，对我们来说，它所具有的突出性质比三角不等式更加重要。

考虑如下两个具有不同长度的部分排名列表 l 和 $\tilde{l}$ 。

l

第 1	A
第 2	B
第 3	C
第 4	E
第 5	H

$\tilde{l}$

第 1	C
第 2	B
第 3	D
第 4	A

图 16.1 给出了这两个列表的一个二分图表示。

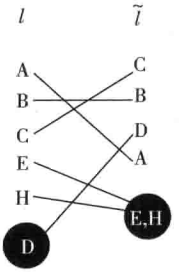


图 16.1　两个排名列表的二分图对比

由于这些列表是部分的，如 D，E 和 H 等某些对象仅出现于一个列表之中。为了处理这些对象，我们在每个排名列表末尾加入了一个虚结点，这个结点的行事方式有如一个黑洞。尽管我们知道哪些对象属于哪个黑洞，我们却不知道每个黑洞中这些对象的相对次序。左侧的每一项和右侧的对应项之间用一条线连起来。这个图形化表示十分简单，但却是一个有力地评价两个排名列表间差异的工具，因为图中水平或接近水平的线意味着两个排名列表对于某个特定对象的地位有着类似的评价，而对角线以及陡峭的对角线则表明两个列表对于该对象的地位有着非常不同的评价。例如，上面的二分图中，在接近顶部有许多陡峭的对角线，意味着这两个列表十分不一致，而且这些不一致多与排名靠前的对象有关。如果在图中存在对角线，那么我们更希望它们出现在靠近图中底部的位置。例如，图 16.2 显示了由马尔可夫法和梅西法给出的 NCAA 男子篮球赛Ⅰ组前 20 位排名列表分别与 RPI 列表进行比较时所表现出的差异。这个计算是在 2005 赛季末完成的。RPI 是一个得到广泛认可的衡量球队实力的标准，也是 NCAA 用来确定哪些队伍将被邀请参加三月疯狂锦标赛的因素之一。这个线图表示使我们很容易地看出，相比于马尔可夫法的列表，梅西法的列表与 RPI 之间有着更多的共同之处。

一个更加有帮助的可视化显示方法，是将这个二分图与评分向量的数线图表示合并起来。此时，排在首位和末位的队伍之间的距离被固定为某一个常值，而剩下的队伍则根据它们各自对应的评分值分布在两者之间。这种表示方法体现了队伍之间的相对距离。我们将每个列表都通过这种方式加以表示，而两个列表的队伍之间则相互连线以生成二分图。图 16.3 就对南方联盟 2007—2008 赛季的梅西排名和马尔可夫排名进行了如此处理。

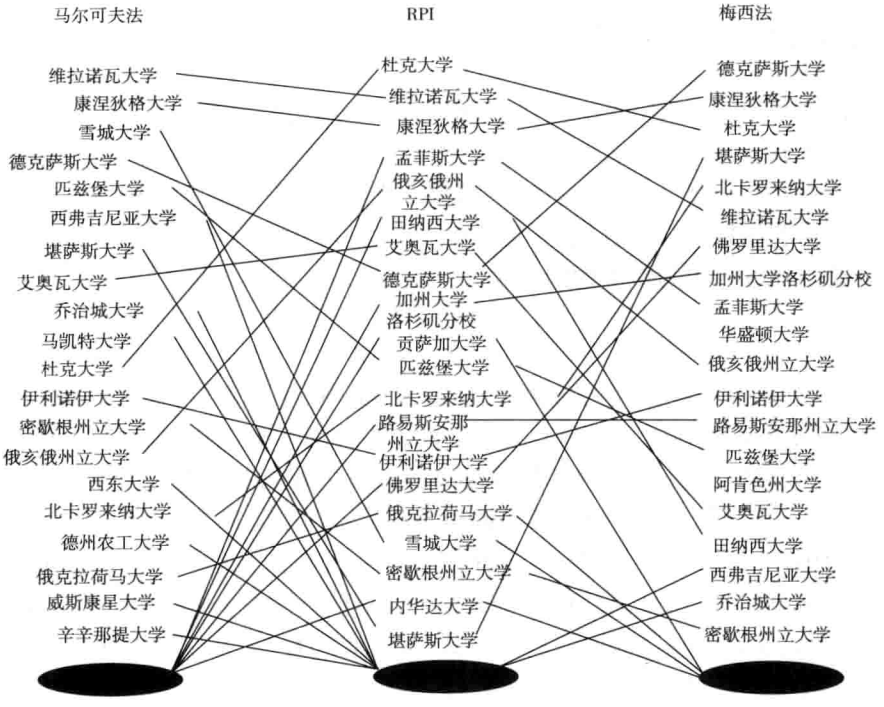


图 16.2　马尔可夫法与 RPI 和梅西法与 RPI 之间的二分图对比

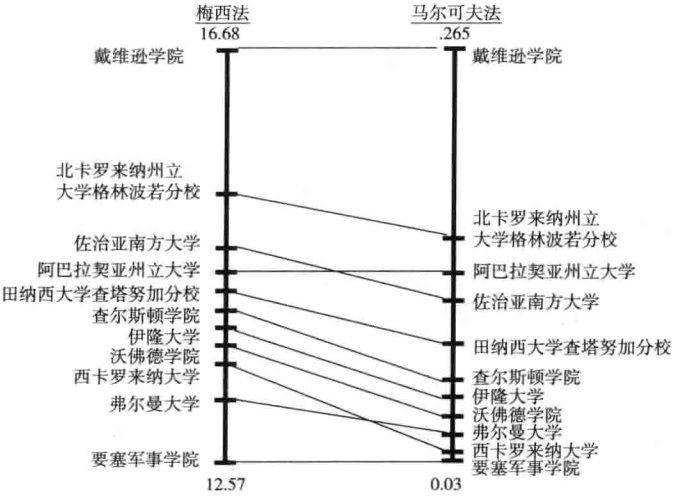


图 16.3　梅西法和马尔可夫法的二分图加线图对比

肯德尔测度 τ

尽管图 16.1 和图 16.2 中的图形表示给出了两个排名列表间的粗略定性比较，但还需要通过定量度量来弥补精确度上的不足。在本节中，我们将讨论一种定量测度，它归功于统计学家莫里斯·肯德尔。肯德尔最初是针对完全列表定义这个测度的，因此要使其能够用于部分列表，还需要进行一些修改。所以我们将本节分为两个部分，即完全列表的情况和部分列表的情况。

完全列表上的肯德尔测度 τ

1938 年，莫里斯·肯德尔提出了一种测度，如今我们称之为肯德尔 τ 秩相关测度，它确定了两个长度相等的完全列表之间的相关性[43]。肯德尔的相关性测度 τ 给出了一个列表与另一个列表之间吻合（或不吻合）的程度。实际上，计算该测度中所需的部分数据的方法之一，就是利用冒泡排序法将一个列表排列为另一个，并计算此时所需要的对调次数。不过肯德尔测度 τ 更为常见的定义是

$$\tau = \frac{n_c - n_d}{n(n-1)/2}。 \tag{16.1}$$

式中，n_c 为两个列表中先后顺序一致的对象对的个数；n_d 为排名不一致的个数；分母 $n(n-1)/2$ 为列表中 n 个对象所构成的所有对象对的总数。对于列表中的每个对象对（即每场比赛），我们需要确定，这两个列表各自给出的两者的相对排名是否吻合。更确切地说，对于对 (i,j)，如果队伍 i 在两个列表中都排名在队伍 j 之前（或之后），那么这个对就称为一致的；否则，它就是不一致的。τ 的分数表示体现了它作为两个列表之间吻合性度量的直观想法。由于这 $n(n-1)/2$ 个对中的每一个都被标明为一致或不一致，因此显然有 $-1 \leqslant \tau \leqslant 1$。如果 $\tau = 1$，则两个列表完全吻合；如果 $\tau = -1$，则它们的顺序完全颠倒。

例 考虑如下三个排名列表 l_1，l_2 和 l_3，它们是包含对象 A 到 D 的完全列表。

	l_1		l_2		l_3
第 1	A	第 1	A	第 1	B
第 2	D	第 2	C	第 2	D
第 3	C	第 3	D	第 3	C
第 4	B	第 4	B	第 4	A

列表 l_1 和 l_2 间的测度 $\tau(l_1,l_2) = 4/6$，而 $\tau(l_1,l_3) = -4/6$。这与我们的直观感受一致，因为可以容易地看出，相比于列表 l_3，列表 l_1 更接近于列表 l_2。此外，列表 l_3 是按与 l_2 完全相反的顺序排列的，因此肯德尔距离 $\tau(l_2,l_3) = -6/6 = -1$。

肯德尔测度 τ 是比较两个排名列表间偏差的一个很好的出发点。但是，它有几个缺陷。首先，当列表很长时，其计算开销很大[28]；其次，两个列表都必须是完全的，从而限制了该测度的适用性。我们常常会希望比较所谓的前 k 位的列表。例如，我们可能希望比较 BCS 给出的大学橄榄球队排名前 10 的列表和我们的排名方法之一所给出的排名前 10 的列表。此时，由于这些列表是部分列表，所以它们几乎不可能正好都包含相同的球队，而标准的肯德尔 τ 测度也因此无法使用。对于这种情况已经提出了若干变体[25, 24, 47]。实际上，我们将在下一节中给出我们自己的变体；第三点——并且也是与我们的体育排名问题特别有关系的一点——便是在排名列表靠后和靠前的位置上的不一致性所造成的影响相同。对于大多数排名应用而言，我们通常更关注靠近排名列表最顶部的那些结果，换而言之，在列表底部出现的错误相对不那么重要一些。肯德尔测度 τ 的最后一个弱点，促成了在第 206 页关于斯皮尔曼简捷测度一节中新距离度量的提出。

部分列表上的肯德尔测度 τ

为了应用于部分列表以及长度不同的列表，需要对肯德尔测度 τ 进行一定的修改。例如，考虑如下的列表 l 和 $\hat{l}$。

$$\begin{array}{c} l \\ \text{第 1} \\ \text{第 2} \\ \text{第 3} \end{array}\begin{pmatrix} \text{A} \\ \text{E} \\ \text{B} \end{pmatrix} \qquad \begin{array}{c} \hat{l} \\ \text{第 1} \\ \text{第 2} \end{array}\begin{pmatrix} \text{C} \\ \text{D} \end{pmatrix}$$

此时，存在某些如（A，B）这样的对象对，我们没有足够的信息来确定它是否为一致或不一致。对于这样的对，我们需要一个新的“未知”标签。实际上，在上例中存在着 4 个这样的未标记对：（A，B），（A，E），（B，E）和（C，D）。针对这一情况，我们将完全列表的肯德尔测度 τ 修改以处理部分列表，这里就包括长度不同的部分列表以及包含平局的列表。

$$\tau_{\text{partial}} = \frac{n_c - n_d - n_u}{\binom{n}{2} - n_u}, \tag{16.2}$$

式中，n_c 和 n_d 与正常定义中一样（即分别为一致和不一致的对的数量）；n 是两个列表中不同对象的总数（即 $|l \cup \hat{l}|$）；n_u 是未标记对的数量。对于上例，

$$\tau_{\text{partial}} = \frac{0 - 6 - 4}{10 - 4} = -10/6 = -1.\overline{6},$$

它是这个例子中偏差的一个上界，因为

$$\frac{-\binom{n}{2}}{\binom{n}{2} - n_u} \leqslant \tau_{\text{partial}} \leqslant \frac{\binom{n}{2}}{\binom{n}{2} - n_u}。$$

τ_{partial} 的一个优点在于它对互不相交的部分列表对给出了比顺序正好颠倒的部分列表对更高的惩罚，而这与我们的直观感受十分吻合。

例　给定如下 4 个部分列表

$$\begin{array}{c} l_4 \\ \text{第 1} \\ \text{第 2} \\ \text{第 3} \end{array}\begin{pmatrix} \text{A} \\ \text{E} \\ \text{B} \end{pmatrix} \qquad \begin{array}{c} l_5 \\ \text{第 1} \\ \text{第 2} \\ \text{第 3} \end{array}\begin{pmatrix} \text{B} \\ \text{E} \\ \text{A} \end{pmatrix} \qquad \begin{array}{c} l_6 \\ \text{第 1} \\ \text{第 2} \\ \text{第 3} \end{array}\begin{pmatrix} \text{C} \\ \text{A} \\ \text{F} \end{pmatrix} \qquad \begin{array}{c} l_7 \\ \text{第 1} \\ \text{第 2} \\ \text{第 3} \end{array}\begin{pmatrix} \text{C} \\ \text{D} \\ \text{F} \end{pmatrix}$$

直接使用上述改进的肯德尔测度，可得如表 16.1 所示的值。这些值相当混乱，因为看起来尽管 l_4 和 l_6 两者间只有一个共有对象，但 l_6 比 l_5 更接近于 l_4。这个例子突出表明了使用 τ_{partial}

表 16.1　不正确的 n 值计数方式时的肯德尔测度 τ

列表	n	τ_{partial}
l_4, l_5	3	-1
l_4, l_6	5	-0.5
l_4, l_7	6	$-1.\overline{6}$

测度时的一个需要注意的地方——当比较超过两个部分列表时，只有将 n 定义为所有列表中不同对象的总数时，n ，n_c ，n_d 和 n_u 的数值才是有意义的。如果加上这个修改，那么 τ_{partial} 就能与我们的直观感受很好地吻合。对于列表 l_4 ，l_5 ，l_6 和 l_7 而言，$n = 6$ ，而 τ_{partial} 值在表 16.2 中给出。

表 16.2　正确的 n 值计数方式时的肯德尔测度 τ

列表	n	τ_{partial}
l_4, l_5	6	0.25
l_4, l_6	6	$-0.\overline{7}$
l_4, l_7	6	$-1.\overline{6}$

注意，如果输入使用的是完全列表，那么未标记对的数量就等于 0（即 $n_u = 0$），因此第 205 页上的 τ_{partial} 公式，即式（16.2）就简化为第 204 页上的完全 τ 公式即式（16.1）。因此，如下的定义便足以涵盖以上两种情况。

比较排名列表的肯德尔测度 τ

$$\tau = \frac{n_c - n_d - n_u}{\binom{n}{2} - n_u}。$$

式中，n_c 为两个列表中先后顺序一致的对象的数量；n_d 为排名不一致对象的数量；n_u 为未标记对的数量；n 为所有进行比较的列表中不同对象的数量。

我们已经修改了肯德尔测度，以同时处理完全和部分列表。这个新的改进定义还处理了诸如长度不同的部分列表以及包含平局的列表等特殊情况。不过仍有一个问题尚未解决。肯德尔测度 τ 仅使用了简单的计数值（n_d 和 n_u）来对两个列表间的不吻合之处加以惩罚，因此，惩罚中仍未考虑这些不吻合之处出现的相对位置。沿着这个思路，便带来了一个新的测度，它根据不一致性在列表中的出现位置来对其进行加权。让我们先考虑完全列表的情况，之后再来处理更为棘手的部分列表的情况。

完全列表上的斯皮尔曼加权简捷测度

我们在这一节中将定义的新测度，是历史悠久的斯皮尔曼简捷测度（Spearman footrule measure）的自然延伸，该测度用于比较两个排名列表[41]。斯皮尔曼的简捷测度 ρ 不过就是长度为 k 的两个排序后的完全列表 l 和 $\tilde{l}$ 之间的 L_1 距离，即

$$\rho = \|l - \tilde{l}\|_1 = \sum_{i=1}^{k} |l(i) - \tilde{l}(i)|,$$

其中，$l(i)$ 为队伍 i 在列表 l 中的排名；$\tilde{l}(i)$ 为队伍 i 在列表 $\tilde{l}$ 中的排名。我们所进行的修改，便是针对每个不一致之处所受到的惩罚值进行加权，这个称为加权简捷测度的新测度定义于式（16.3）中。

比较两个完全排名列表的加权简捷测度 ϕ

两个长度为 k 的完全排名列表 l 和 $\tilde{l}$ 之间的加权简捷测度 ϕ 为

$$\phi = \frac{\sum_{i=1}^{k} |l(i) - \tilde{l}(i)|}{\min\{l(i), \tilde{l}(i)\}}。 \tag{16.3}$$

注意，这一新偏差测度的提出是基于如下想法。给定两个排名列表，靠近列表顶部的不一致性比靠近底部的不一致性更加令人不安。因此，在计算偏差测度时，我们根据不一致性在列表中的出现位置进行相应的惩罚。上述 ϕ 的定义对高排名项之间的不一致性进行了很重的惩罚。

例　假设队伍 i 在 l 中排名第三而在 $\tilde{l}$ 中排名第二，则

$$\min\{l(i), \tilde{l}(i)\} = 2,$$

而如果队伍 i 在 l 中是第 13 名而在 $\tilde{l}$ 中是第 12 名，那么 $\min\{l(i), \tilde{l}(i)\} = 12$，因此出现于前 k 位列表后部的不一致性所受到的惩罚 ϕ 被显著减小了。

显然，对于所有同时出现于两个列表的队伍 i，有 $1 \leqslant \min\{l(i), \tilde{l}(i)\} \leqslant k$。注意到 $\phi \geqslant 0$，因此 ϕ 越小，两个被比较的列表就越接近。

部分列表上的斯皮尔曼加权简捷测度

对于完全列表，加权简捷测度的计算是直截了当的。但是，我们必须考虑更为一般（也更为复杂）的部分列表的情况，这种情况又可以进一步细分为长度相同的部分列表和长度不同的部分列表。我们从较简单的问题开始，考虑两个长度同为 k 的部分列表。

对于部分列表的分析是较为棘手的，因为一个对象可能并不会同时出现在两个列表中。为了计算 ϕ，我们必须考虑在两个列表 l 和 $\tilde{l}$ 中所有对象的并集或论域，看看每个对象各自在其中的贡献。对于每个对象 i，它或者属于两个列表中对象的交集 $l \cap \tilde{l}$（因此同时出现在两个列表中），或者属于这个交集的补集 $(l \cup \tilde{l})/(l \cap \tilde{l})$（因此仅出现于一个列表之中）。交集中的队伍 i 对于 $\phi = \sum_i \phi_i$ 的贡献 ϕ_i 较为容易计算——我们只需使用加权简捷测度式（16.3）就可以求得。但是，在补集中的队伍 i 的贡献则更难计算，因为 $l(i)$ 和 $\tilde{l}(i)$ 中正好有一个为未知。例如，假设队伍 i 在列表 l 中排名在第 j 位，但并未出现在列表 $\tilde{l}$ 中。那么 $\phi_i = |j - \tilde{l}(i)|/j$ 和 $\tilde{l}(i)$ 显然就是缺失的信息。我们暂时称这个丢失数据为 x，因为它是一个我们希望最终通过数学分析来得到的未知量。于是在这种给定的情况下，$\phi_i = \frac{|j - x|}{j}$。一旦知道了 x，我们就能够将各个 ϕ_i 惩罚值相加得到偏差测度 ϕ，即 $\phi = \sum_{i=1}^{m} \phi_i$，其中，$m$ 为两个列表中所有队伍的

总数，即 $m = |l \cup \tilde{l}|$ ⊖。对这个测度的解释很简单：ϕ 越小，两个列表就越接近。

现在我们通过寻找未知数 x，来将这幅 ϕ 的拼图中缺失的那一块填补进去。我们将按类似于参考文献［47］中的思路来进行数学分析，这个思路认为，一个好的测度应当满足如下三个性质：

1. 两个完全相同的列表的偏差度量应该为 0，即 $\phi = 0$。
2. 两个互不相交的列表应该具有测度所能给出的最大偏差。
3. 如果两个列表具有相同的对象，但排名次序正好完全颠倒，那么这两个列表间的偏差应该等于性质 2 所给出的偏差的一半。

我们利用性质 2 和性质 3 来巧妙地确定 x，从而计算等长部分列表的测度 ϕ。

首先，我们从性质 2 开始，并计算 $\phi(l,l^c)$，即列表 l 和与之不相交的 l^c 之间的偏差。由于 x 未知，因此这个值将是 x 的一个函数。假设 l 和 l^c 的长度均为 k。由于 l 和 l^c 没有共同元素，因此需要计算 $2k$ 个单独的测度 ϕ_i。对于第一个列表 l，ϕ_i 距离为

$$\begin{aligned}\phi_1 &= (x-1)/1\\ \phi_2 &= (x-2)/2\\ &\vdots\\ \phi_k &= (x-k)/k\end{aligned}$$

而列表 l^c 中的距离也一样。因此，

$$\phi(l,l^c) = 2\sum_{i=1}^{k}(x-i)/i = -2k + 2x\sum_{i=1}^{k}1/i。$$

这个依赖于 k 和 x 值的和，便是两个长度为 k 的列表间的最大距离。因此，$\phi(l,l^c)$ 可用于最后一步，来对偏差值进行归一化处理。总之，现在我们已经得到了性质 2 中测度所能给出的最大偏差。

现在我们转向性质 3，并计算 $\phi(l,l^r)$，即一个列表 l 和它的逆序列表 l^r 之间的偏差。我们的策略是首先得到用 k 和 x 来表达的 $\phi(l,l^c)$ 和 $\phi(l,l^r)$，然后根据方程 $\phi(l,l^c) = 2\phi(l,l^r)$ 来求解未知的 x。

由于列表 l 及其逆序 l^r 包含了相同的 k 个对象的集合（即 $l \cup l^r = l \cap l^r$），所以 $\phi(l,l^r)$ 不是未知数 x 的函数，因而 $\phi(l,l^r)$ 可以直接计算得到。我们利用向下取整运算 $\lfloor\ \rfloor$ 将它写为

$$\phi(l,l^r) = 2\sum_{i=1}^{\lfloor k/2\rfloor}(k-(i-1)-i)/i = -4\lfloor k/2\rfloor + 2(k+1)\sum_{i=1}^{\lfloor k/2\rfloor}1/i。$$

与 x 有关的最后一块拼图，便是利用方程 $\phi(l,l^c) = 2\phi(l,l^r)$ 来获得 x 的表达式，尽管这是个

⊖ 原文误作 $m = |l(i) \cup \tilde{l}(i)|$。——译者注

有点乱糟糟的表达式。通过少许代数运算可得（搞定!）

$$x = \frac{k - 4\lfloor k/2 \rfloor + 2(k+1)\sum_{i=1}^{\lfloor k/2 \rfloor} 1/i}{\sum_{i=1}^{k} 1/i}。$$

因此，对于比如说 $k = 3$，有 $\phi(l,l^c) = 8$，$\phi(l,l^r) = 4$，且 $x = 42/11 \approx 3.82$。由于归一化测度更易于解释，所以通过除以 $\phi(l,l^c)$ 所给出的最大可能偏差，可以使 ϕ 取值在 0 到 1 之间。

比较长度为 k 的部分列表的加权简捷测度 ϕ

两个长度为 k 的部分列表 l 和 $\tilde{l}$ 之间的加权简捷测度 ϕ 由各个单独的 ϕ_i 值构造而得，并经归一化处理使得 $0 \leqslant \phi \leqslant 1$。$\phi$ 越接近于下界值 0，两个排名列表就越接近。

$$\phi = \frac{\sum_{i=1}^{k} \phi_i}{\phi(l,l^c)},$$

式中，

$$\phi(l,l^c) = -2k + 2x\sum_{i=1}^{k} 1/i。$$

每个对象 i 都属于以下两种情况之一，并可相应求得它对于 ϕ 的贡献 ϕ_i。

- 对于对象 $i \in l \cap \tilde{l}$（即同时在列表 l 和 $\tilde{l}$ 中出现的对象），

$$\phi_i = \frac{|l(x) - \tilde{l}(x)|}{\min\{l(x), \tilde{l}(x)\}}。$$

- 对于对象 $i \in (l \cup \tilde{l})/(l \cap \tilde{l})$（即仅在一个列表中出现的对象，为不失一般性，我们假设它出现于 l），

$$\phi_i = \frac{|l(i) - x|}{\min\{l(i), x\}},$$

其中，x 定义为

$$x = \frac{k - 4\lfloor k/2 \rfloor + 2(k+1)\sum_{i=1}^{\lfloor k/2 \rfloor} 1/i}{\sum_{i=1}^{k} 1/i}。$$

例　考虑第 205 页上的 4 个列表 l_4，l_5，l_6 和 l_7，我们将它们再次列举如下。

	l_4	l_5	l_6	l_7
第 1	A	B	C	C
第 2	E	E	A	D
第 3	B	A	F	F

表 16.3 给出了列表 l_4 和其他列表之间的 τ_{partial} 以及加权简捷测度 ϕ。相比于规模更大的例子，在这个小示例中，这两个偏差测度之间的差别并不那么明显。实际上，下一个例子将非常清楚地表明 τ 和 ϕ 之间的差别。

表 16.3　4 个部分列表 l_4，l_5，l_6 和 l_7 的肯德尔测度 τ 和加权简捷测度 ϕ

列表	肯德尔测度 τ_{partial}	加权简捷测度 ϕ
l_4, l_5	$3/12 = 0.25$	$4/8 = 0.5$
l_4, l_6	$-7/9 = -0.\overline{7}$	$29/44 = 0.66$
l_4, l_7	$-15/9 = -1.\overline{6}$	1

例　考虑如图 16.4 所示的例子，其中包含三个列表 l_8，l_9 和 l_{10}，长度均为 12。表 16.4 表明 $\tau(l_8, l_9) = \tau(l_9, l_{10})$，但列表 l_8 肯定比 l_{10} 更接近于 l_9，因为它前半部分的排名与 l_9 中的排名完全一致。另一方面，$\phi(l_8, l_9) < \phi(l_9, l_{10})$，因此加权简捷测度准确把握住了我们希望加以反映的这个情况。

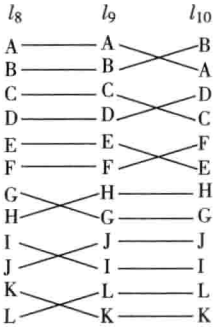


图 16.4　肯德尔测度 τ 和加权简捷测度 ϕ 之间的差别

表 16.4　3 个部分列表 l_8，l_9 和 l_{10} 的肯德尔测度 τ 与加权简捷测度 ϕ

列表	肯德尔测度 τ_{partial}	加权简捷测度 ϕ
l_8, l_9	0.95	0.02
l_9, l_{10}	0.95	0.09

不等长的部分列表

前一节说明，利用一些数学技巧，我们能够将加权简捷测度应用于等长的部分列表。不幸的是，要使 ϕ 适用于不等长部分列表，其中所需的数学技巧将更为费解难懂，因此不适于

在此进行介绍。在实际中，很少会频繁碰到比较两个不等长列表的应用。加权简捷测度的这一局限性与肯德尔测度 τ 形成了鲜明对比，我们可以修改后者以适用于所有类型的列表：完全列表、等长部分列表和不等长部分列表。另一方面，对于如列表 l_8，l_9 和 l_{10} 的示例所示的完全列表和前 k 位列表而言，加权简捷测度 ϕ 比肯德尔测度 τ 更为直观，也更为适用。

评判标准：与已知的标准相比较

现在我们已经得到了一个测度 ϕ，并且用它比较两个排名列表的效果令人满意（特别是当列表具有相同长度时），所以我们再回到本章开始时提出的那个问题，即到底哪个排名列表最好？如果有了一个大家都能接受的衡量成功的评判标准，那这个问题便有了一个简单的回答。再次考虑大学橄榄球赛的情况，这时 BCS 的排名列表常常被作为成功与否的标准。此时，我们就计算出 BCS 排名和任何我们希望的排名方法的结果之间的偏差 ϕ(BCS,梅西)，ϕ(BCS,科利) 或 ϕ(BCS,某种方法) 等，并将给出最小 ϕ 值的方法宣布为最佳方法。当然，这时我们假定 BCS 列表是“正确的”，而正如我们之前已经指出的，对这一点，有着无休无止的争论。由于使用一个如 BCS 这样的富有争议性的排名列表来作为金科玉律并不能让人满意，所以另外一个列表可能更配得上这个金灿灿的地位，这就将我们带到了下一节。

评判标准：与聚合列表相比较

使用第 14 章中某个聚合方法得到的排名聚合列表，看起来是我们的黄金标准或评判准绳这顶桂冠的一个出众的候选者。这个列表包括了许多单独的排名列表的最好性质，并移除了异常值的影响。有了这个标准，我们就可以将与该聚合标准列表之间的偏差最小的排名列表宣布为最佳列表。当然，第 203 页上的肯德尔测度 τ 或第 206 页上的加权简捷测度 ϕ 都可以用于衡量与聚合列表之间的偏差。

例　图 16.5 显示了一个聚合列表和 OD 列表及马尔可夫列表之间的偏差。定性与定量度量都已在图中给出。二分图表示从视觉上表明 OD 列表比马尔可夫列表更接近于聚合列表。这个聚合列表是通过第 176 页上评分聚合一节中所介绍的评分聚合法得到的。两个数值测度 τ 和 ϕ 精确地量化并进一步确认了定性观察的结果。回忆一下，对于肯德尔测度，τ 取越大的正值，则列表越吻合；而对于加权简捷测度，越接近于 0 的 ϕ 值说明两个列表中的吻合之处越多。在这个数据来自 NCAA 男子篮球 2004—2005 赛季的例子中，我们可以认定 OD 法是更好的方法，因为它距离聚合列表最为接近。

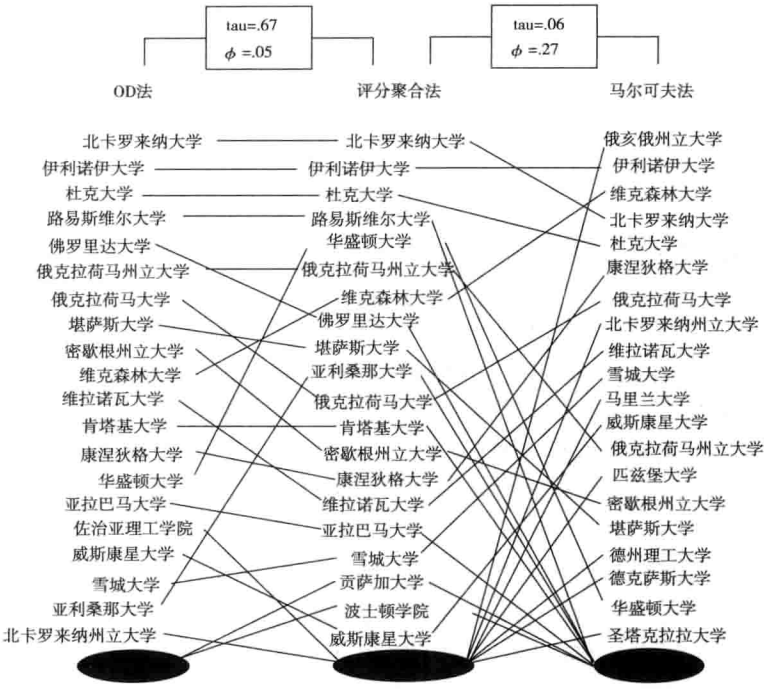


图 16.5　（OD，评分聚合）和（马尔可夫，评分聚合）之间距离的对比

回溯打分

回溯打分利用一个排名列表来预测以往比赛的胜者。根据这个思路，可以根据一个排名列表得到不同的评分。例如，每个与以往比赛结果相吻合的预测都将为相应的排名方法积累 1 分，而得分最高的方法将获胜，并被宣布为年度最佳方法。如果考虑比赛分差的话，还能得到更为高级的评分方法。对于以往比赛数据可用的应用而言，回溯打分给出了确定排名列表质量的一种直接的方法。

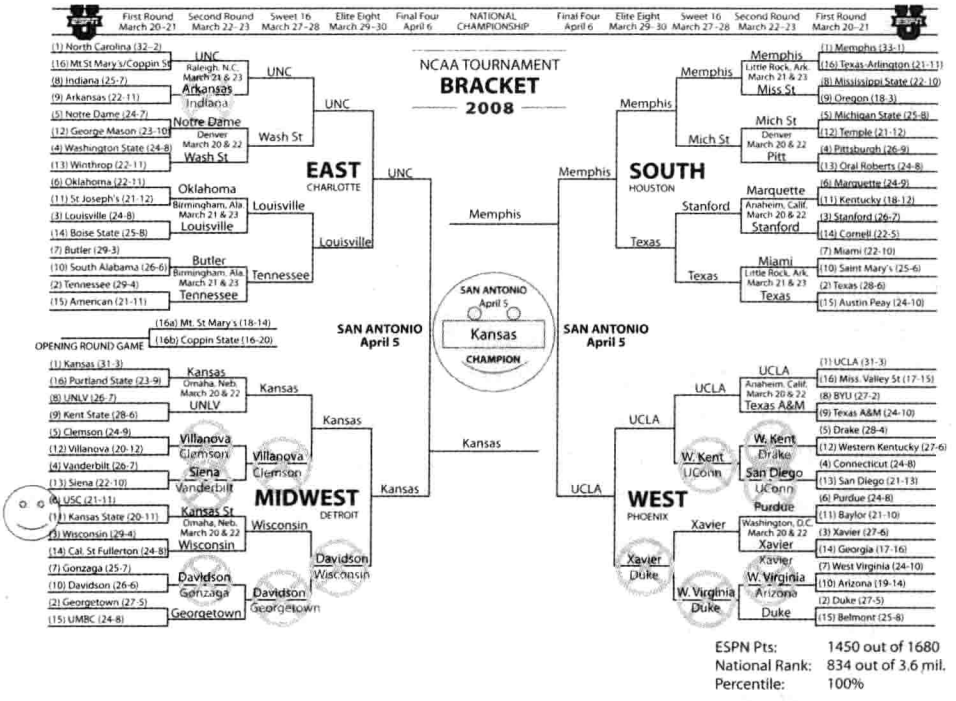
事前预测

事前预测在思路上非常类似于回溯打分，只不过此时排名列表的质量由它预测未来比赛结果的能力所决定。最常见的情况是利用赛季末的排名列表来预测年度的压轴锦标赛的结果。正如下面的杂谈所要说明的那样，大学篮球三月疯狂锦标赛就是一个完美的例子。

杂谈：竞猜学

每年三月，NCAA 大学篮球赛 I 组的赛季都将在三月疯狂锦标赛中达到高潮。若干支球队被邀请或保送参加这个赛季末尾的单淘汰制锦标赛。锦标赛以总冠军赛结束，这场比赛将决定当年的全国冠军。三月疯狂锦标赛总是在新闻报道、竞猜、收入、关注度、公路交通和营销活动方面搞得沸沸扬扬。球迷得以参与三月疯狂锦标赛的众多途径之一，便是进行竞猜。在锦标赛正式开

始之前，球迷们可以填写一份锦标赛竞猜表，首先选出第一轮中获胜的队伍，然后再由这些预测的胜者进行下一轮比赛，再选出这些比赛的胜者，如此继续，直至对所有 65 场比赛的胜者都进行了预测。球迷可以将这些竞猜提交到特定的投注池或甚至更大规模的竞赛中去，包括 ESPN 一年一度的锦标赛竞猜，在这个竞猜中，结果最为准确的球迷将获得 10 000美元的奖金。随着锦标赛的推进，正确的预测被赋予越来越高的权值。实际上，ESPN 使用下面的规则，来计算每张提交上来的竞猜表的累积得分：每正确预测 1 场第 1 轮的比赛得 10 分；第 2 轮，20 分；第 3 轮，40 分；第 4 轮，80 分；第 5 轮，120 分；以及第 6 轮，160 分。全部猜对的竞猜表将获得最高的 1 680 分。如下便是一个例子，它是（两名查尔斯顿学院前本科生）尼尔·古德森和科林·斯蒂芬森在 2008 年锦标赛期间提交到 ESPN 竞猜的一张竞猜表。



这个竞猜表是根据第 7 章中的对数加权 OD 模型给出的排名来填写的。为了填写给定比赛的结果，这两名学生比较了这两支队伍在排名向量中的名次，名次较高的队伍被预测为可以进入下一轮的队伍。当锦标赛结束时，科林和尼尔的竞猜表获得了 1450 分的 ESPN 评分。在所提交的总共超过 360 万张竞猜表中，他们的竞猜表位居第 100 个百分位，排名为 834 名。错误的预测在表中以“X”表示，而两个微笑符号则表示预测正确的冷门，即大多数体育分析家都未能预测出获胜队伍的比赛。其他几个由本书中表现良好的模型给出并提交到 ESPN 竞猜的结果列于表 16.5 中。

我们很高兴地告诉大家，由于尼尔和科林表现得非常出色，因此他们获得了数量惊人的全国性报道。这哥俩甚至还在美国国家公共广播（NPR）的《无所不虑》和 CBS 的《早间秀》中亮了相。

表 16.5　2008 年锦标赛竞猜中所使用的排名方法与 ESPN 得分

方　法	ESPN 得分
线性加权梅西法	1450
对数加权梅西法	1450
对数加权 OD 法	1450
阶跃加权梅西法	1420
指数加权梅西法	1400
阶跃加权 OD 法	1320
未加权梅西法	1310
线性加权 OD 法	1310
未加权 OD 法	1310
线性加权科利法	1100
对数加权科利法	1010

学习曲线

评判排名列表质量的另一个方面，与产生该列表的排名方法的智能程度有关。如果一种排名方法仅需要一小段时间的热身就能开始做出正确的预测，那么这种方法就被认为是“聪明的”。下面是一个典型的情况。随着 NFL 赛季一周周地过去，一群群的朋友们逐渐组织起了小规模的非正式竞猜活动，来预测某一周比赛的胜者。每一周，预测最为准确的参与者都将赢得当周的奖金。当然，本书的排名方法也可以用来预测胜者，但为了赢得最多的钱，在选择排名方法时，必须同时考虑到方法的准确度和它的智能程度。一个非常准确的方法可能需要很多个星期的数据才能使得预测结果优于其他方法，因此，它未必就是能够积攒起最丰厚回报的方法。北卡罗来纳州立大学的博士生安吉拉·戈万参与了类似的活动（除了金钱之外），并成为了其博士论文[34]中的一部分。

与坡形之间的距离

评估排名列表质量的最后一种方法，就是利用第 8 章中的重排思想。利用排名对类似马尔可夫分差矩阵 $\boldsymbol{V}$ 这样的数据差矩阵进行对称重排，并计算重排后的矩阵与第 108 页上定义的坡形之间的距离。这一距离可以通过违背坡形约束的（可能经过加权的）次数来衡量。能够使破坏次数达到最少的方法就被认为是最佳方法。

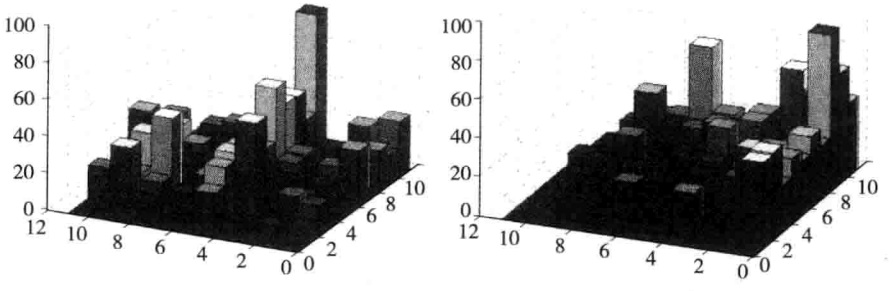
实际上，查尔斯顿学院研究生凯思琳·佩丁斯在南方联盟 11 支篮球队 2007—2008 赛季上比较数种排名方法时，就正是这样做的。她发现，在表 16.6 给出的方法中，重排法认为评分差距法给出的排名对坡形的破坏次数最少。

这一结果并不令人吃惊，因为评分差距法本身就是一个最优化方法，它被设计用来使上

表 16.6　排名方法以及它们对坡形约束的违反次数

方法	违反约束的次数
评分差距法	265
OD 法	268
梅西法	269
科利法	275

述距离目标函数为最小。如下所示是按原顺序排列的数据差矩阵以及根据评分差距法进行重排后的矩阵的城市区图。



数字——

$ 3 000 000 000≤ 每年在 NCAA 男子篮球锦标赛上所下的总赌注。

——在这项赛事上所下的赌注比美国的其他任何一项赛事都要多。

——据 FBI 估计（基瓦姆和索科尔[48]）

第 17 章　数据

本书中的每个体育评分模型都需要某些比赛统计数据来作为模型输入。感谢万维网，使我们找到这些数据变得不再那么困难。但是，将它们整理为一种对计算机算法而言友好的格式，则并不容易。本书给出的每种评分模型的核心都是一个矩阵，一旦将它构建出来，便能对它进行分析。即使对于很小的例子而言，手工输入数据来产生这个矩阵也是十分枯燥乏味的。因此，我们必须（a）构建一个如 perl 脚本语言编写的网络搜集程序这样的工具，来自动将网页中的数据转为算法友好的格式，或者（b）获取已经是这种格式的数据。当然，如果可以选择的话，我们大多数人都会选择（b）。感谢第 2 章梅西模型的发明者肯尼斯·梅西，我们的确能做到这一点。当把这些评分模型用于其他，如政治、娱乐或科学等非体育领域时，我们就会明白，拥有这样的数据是一件多么奢侈的事情。

梅西的体育数据服务器

梅西模型（第 2 章）的发明人肯尼斯·梅西已经创建了一个非常了不起的资源，从这儿可以获取与体育相关的数据。他的网站 www. masseyratings. com 包含了非常多的信息。特别是数据网站 www. masseyratings. com/data. php 可以用于生成本书中给出的模型。在这个页面的底部，你可以找到指向许多体育项目数据的链接，对每个项目又根据级别和赛季进行了归类。例如，梅西目前拥有棒球、篮球、橄榄球、冰球和曲棍球等项目的数据，而且他似乎还打算在不久后加入其他项目的数据。大多数项目都被进行了进一步细分，如冰球类就包括了美国冰球联盟、大学男子冰球联赛和大学女子冰球联赛等。某些类别中有多个赛季的数据，例如，美国职业棒球大联盟就有着从 1990 年以来各赛季的数据。

图 17. 1 显示了由梅西的工具所给出的数据的屏幕截图，这一数据请求获取的是 2007 赛季 NCAA 大学男子篮球赛 I 组的比赛数据。我们在图中加入了文本标注，以指明各列所给出的数据。当然，这个数据服务器总是以相同的格式给出数据，对于编写程序来说，这一点非常理想。梅西的数据服务器的另一个了不起的特点，在于众多对数据进行“子集化”的功能。例如，对于大学男子篮球的数据请求，我们可以选择数据源自 NCCAA，NJCAA，NAIA 的两个组以及 NCAA 的所有三个组（或者是当年所有大学篮球比赛的组合数据）；或者可以请求某个特定的联盟如 NCAA I 组的东北联盟；或者可以请求 I 组所有联盟间的比赛；或者……你自己想吧。肯尼斯·梅西已经试着去将差不多所有想象得到的数据请求都包含在内，因为可以用来驾驭这个系统的功能的数量实在是令人震惊。梅西的服务

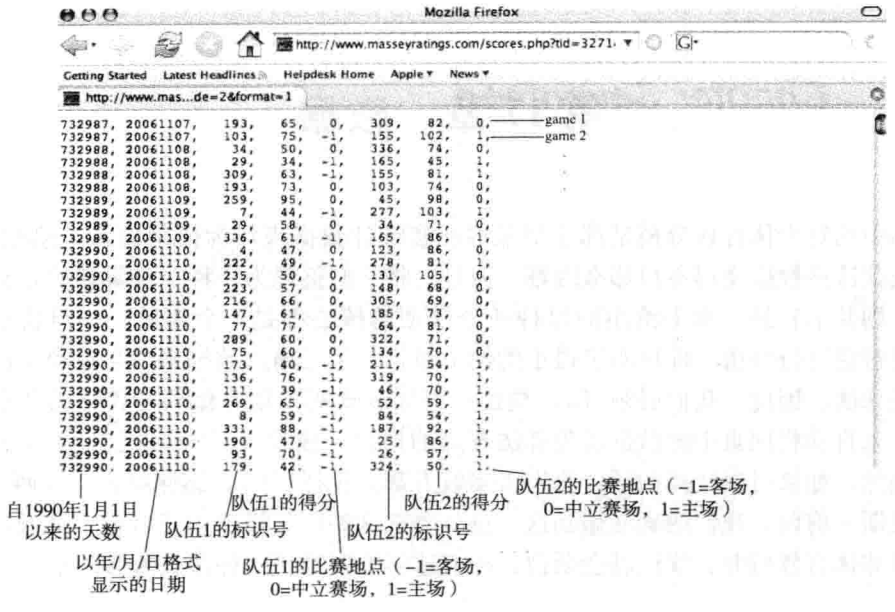


图 17.1　梅西的体育数据服务器的屏幕截图

器是我们在本书中每个体育排名例子所用数据的中转站。

波默罗伊的大学篮球数据

如果你对大学篮球感兴趣，那么肯·波默罗伊（Ken Pomeroy）提供的数据就是你的理想数据来源。他对大学篮球队进行评分，并且他的网站中埋藏着一笔数据宝藏，其中有各种具体的统计数据如进攻篮板和防守篮板、罚球、两分球和三分球命中率以及失误等。波默罗伊用逗号分隔的文本文件格式来提供数据，这些文件可以从其网站 www. kempom. com 下载。

搜集你自己的数据

尽管梅西的数据服务器可以响应非常多的体育数据需求，但还有一些项目的数据是它无法给出的。例如，OD 模型和马尔可夫模型的某些高级变体，需要除简单的比分之外的其他体育数据。如果希望利用篮板球数量来构建一个马尔可夫模型，那么很不幸地，你需要从梅西服务器之外的其他来源去获取这一数据。最开始，你可能会固执于一个无益的希望，去寻找提供这些额外统计数据的类似的服务器；等到你放弃了找到拥有一切所需数据的完美服务器的梦想之后，你就应该评判一下这些额外的统计数据的重要性；如果你认定这些附加的数据对于模型而言必不可少，那就必须自己来收集它们。这就是*自己搜集数据*的技术——编写一个脚本程序，在特定的网页上定位你所需要的数据，然后将它采集到一个你所选定的格式的文本文件中去。一般情况下，你需要让这个工具定期地从一个选定的

网页列表中搜集数据，比如说当你在采集橄榄球数据时，你可能要在赛季进行过程中每周搜集一次。当然，所有这些都是假设你已经找到了可用于搜集的网页，而这需要投入相当多的时间。尽管如此，搜集自己的数据可以提供最大程度的自由度和灵活性。对于一个熟练的搜集者来说，这些可能性的确是无穷无尽的，也就是说这是一门值得我们去获得的、有价值的编程技术（我们也希望经过这一路走来，我们已经学会了这门手艺）。

现在，如果你在读书的日子里错过了数据搜集的那些课的话，我们给出的建议，就是找一个没有错过它们的学生。我们很幸运，能够找到具有编程能力的数学专业学生。北卡罗来纳州立大学的学生安吉拉·戈万编写了她自己的 perl 脚本程序来搜集 NFL 比赛数据，以作为其博士学位论文中的一个部分[34]。查尔斯顿学院的赖恩·帕克是一名专业的数据搜集者——这个评价非常名副其实，请见第 219 页上的杂谈。他是个 NBA 球迷和疯狂粉丝。他编写了程序，每晚都从诸如 ESPN. com 和 NBA. com 这样的网站搜集数据。赖恩拥有所有我们想象得到的统计数据。为了对他的数据库存的庞大有一个认识，赖恩能够给出在本赛季的第十一场比赛中的第十二分钟内是哪十名球员在场上比赛这样细致的数据。对于赖恩的研究来说，细节层次十分重要，因为他所研究的内容，是根据每次控球的情况来预测得分多少（0 分、1 分、2 分、3 分、4 分或 5 分）。有关他的研究和数据，请见其网站 www. basketballgeek. com。总之，如果你需要 NFL 数据，请联系安吉拉·戈万；如果你需要 NBA 数据，那赖恩就是你要找的人。

杂谈：球队极客

在最近的几十年中，职业体育队伍的从业人数发生了巨大的增长。首先，教练组的人数增多了。球队现在有一个主教练、助理教练、第二助理教练、第三助理教练、第四助理教练，等等；有进攻专家、防守专家、投篮教练，等等。球队外出比赛时的随行人员甚至更多，包括球队的训练员、营养师和按摩治疗师。而最能使我们振奋的是最近加入的成员。NBA 最新的一个趋势，似乎就是雇佣一名“球队极客”。这个新的成员对数学和统计数据加以研究，以找出对于球队可能有用的任何走向或历史趋势。已有数支 NBA 球队与查尔斯顿学院毕业班学生赖恩·帕克取得了联系。赖恩创办了他自己的网站 www. basketballgeek. com，其中收集了大量的 NBA 数据。如果你询问赖恩有关任何一支队伍的任何一名球员的任何一项统计数据，他都会把你要找的海底捞针似的数据用电子邮件发给你。赖恩对这些数据进行分析，以发现可能对教练组有用的信息。例如，作为其毕业班论文的一部分，赖恩分析了球队的控球情况，并试图找出最有可能在一场比赛的最后一次控球中取得 3 分或 4 分的球员组合。对于处在一场比赛最后几秒钟的教练来说，这是一个重要的信息。尽管尚未成为现实，但下面这种情况便是我们为赖恩和他的那些球队极客们所设想的场景。比赛只剩下 8s，他的球队还落后 3 分，对手的 5 名球员都在对方后场，而己方的后卫正准备从中线发球，这时教练叫了次暂停，并转头问赖恩：“根据你的统计和数学分析，我们的最佳策略是什么?”

生成成对比较矩阵

在本节，我们将强调不同类型的对比数据之间的差异。本书中的许多应用都与体育有关。在这种情况下，球队在比赛中一决高下，而这些比赛就给出了直接的成对比较。例如，在2008年10月13日纽约巨人队击败克里夫兰布朗队，等等。但是对其他应用而言，这样的直接成对比较就不是那么显然，而必须由原始数据中推算或产生出间接对比。Netflix的用户-影片评分矩阵或亚马逊的用户-商品购买矩阵就是这类数据的例子。为了用本书中的方法产生影片的排名，我们需要将用户-影片评分矩阵转换为影片之间的成对比较矩阵。

在第11章中，我们使用科利法来对影片进行排名，其中使用了一种非常初等的变换，来将原始的用户-影片评分矩阵转为成对比较数据。例如，假设用户1同时对影片 i 和 j 进行了评分，给影片 i 打了5颗星，而影片 j 是3颗星。我们便将这次评分视为这两部影片之间的一次比赛，而影片 i 以2分的优势胜出影片 j。每名同时对这两部影片评了分的用户，就对应进行了一场比赛。

还有许多其他方式来完成由用户-影片评分数据到影片-影片数据的变换。相比于将每个同时对两部影片进行了评分的用户视为一场单独的比赛，我们可以用另外一种方式，将所有这样的比赛全部累加起来，并视其为一场超级比赛。在此我们给出参考文献［33］中所提及的三种变换。本书中的排名方法随后便可用这些影片-影片比赛数据作为输入，其中每部影片与另一部影片最多比赛一次。

1. *所有同时对两部影片进行评分的用户所给分数的算术平均值*：例如，假设仅有3名用户同时对影片 i 和 j 进行了评分。第一名用户给 i 的分数比 j 多2分，第二名用户给 i 的分数比 j 多1分，而第三名用户给 i 的分数比 j 少1分。于是，在影片-影片对比矩阵中，(i,j) 元素的值为 $(2+1-1)/3 = 2/3$，而 (j,i) 元素的值为 $(-2-1+1)/3 = -2/3$，这样便给出了一个斜对称的比较矩阵。
2. *所有同时对两部影片进行评分的用户所给分数的几何平均值*：此时，对数几何平均值更容易计算。假设上例中三名同时给影片 i 和 j 进行评分的用户给出的分数为：
   - 用户1：影片 i——5颗星，影片 j——3颗星
   - 用户2：影片 i——5颗星，影片 j——4颗星
   - 用户3：影片 i——3颗星，影片 j——4颗星

   则影片-影片对比矩阵的 (i,j) 元素为 $\{[\log(5)-\log(3)]+[\log(5)-\log(4)]+[\log(3)-\log(4)]\}/3 = 0.065$。与算术平均值类似，几何平均值也能给出斜对称比较矩阵。
3. *i 击败 j 的概率减去 j 击败 i 的概率*：这个变换仅考虑胜负而非得分。例如，假设4名用户都对影片 i 和 j 进行了评分，其中3人更喜欢 i 而1人更喜欢 j。则影片-影片对比矩阵的 (i,j) 元素为 $3/4-1/4 = 1/2$。（平局可以考虑在内或排除在外。）

杂谈：Netflix 的最后一个杂谈：Netflix 2 竞赛暂停

在遍布书中的多个杂谈中，我们已经一再介绍了 Netflix 数据和有名的百万美元赏金。这些杂谈说的是第一次 Netflix 竞赛的故事，它始于 2006 年 10 月，结束于 2009 年 7 月，当时 BellKor 队赢得了奖金。Netflix 认为他们设立的奖金是一笔很好的投资，于是在第一次竞赛颁奖仪式之后不久的 2009 年 8 月，他们宣布计划组织第二次竞赛。奖金仍是 1 百万美元，而目的也仍是利用公司所提供的数据来改进推荐系统。但是，这一次，数据将比之前涵盖更广的范围。具体来说，除了原来的那个简单的用户在某个时间给影片打的评分列表之外，Netflix 2 的参赛者还将能够访问许多其他信息，如租借者的年龄、性别、邮政编码和影片类型评分等。就像全世界范围内的许多数据分析者一样，我们都激动地投入竞赛并开始尝试各种算法的新点子。但是，5 个月后，Netflix 将这个 Netflix 奖的续集叫停，并称这是出于隐私权和合法性方面的考虑。联邦贸易委员会已经要求他们考虑这个竞赛对其会员的隐私会造成何种影响，而 KamberLaw 有限责任公司也对他们提起了一项诉讼。博客作家、隐私权专家以及法律教授保罗·欧姆（Paul Ohm）是最早敲响这一警钟的人。他的疑虑是基于拉坦娅·斯维尼（Latanya Sweeney）的一项统计结果，后者在 2000 年证明，仅靠三个信息，就能确定美国 87% 的人，这三个信息是：出生日期、性别和邮政编码。尽管 Netflix 没有给出租片者的出生日期，但欧姆称年龄、性别和邮政编码这三个因素就足以将一个人的身份缩窄到仅仅数百个可能。

Netflix 幸运地避开了一次尴尬的泄露隐私的事件。但在 2006 年，AOL 就没有如此幸运了。AOL 对大众公开了网络搜索数据，希望能帮助学术界的研究者推进这一领域。具体来说，这些数据中包含了 2 千万个经过审查的网络搜索，即对于和搜索查询有关的任何 AOL 用户信息都使用一个用户标识号来代替，如第 4417749 号。在三个月内，用户 4417749 进行了从“手指发麻”到“60 名单身汉”到“在什么东西上面都撒尿的狗”等主题的数百次搜索。其中有一些诸如“佐治亚里尔伯恩的庭园设计师”和“佐治亚格威内特县影湖区的房屋出售”等查询，它们包含了暗示用户 4417749 的所在地的信息。一名记者仅仅花了 3 天时间，就根据这些数据来到了用户 4417749 的家门口。这名记者的直觉告诉他，这些经过审查处理的 AOL 数据，并不像表面看起来那么无害。出于好奇，他随机挑选了用户 4417749，并试着看能不能找出她的身份。三天之后，他的搜寻结束了，他敲响了一扇门，而开门的是塞尔玛·阿诺德（Thelma Arnold）和她那有撒尿癖的宠物狗杜德利。开门的这两位看来都被这名造访者以及他对自己那来历不明而又无所不知的了解惊得目瞪口呆。AOL 立刻从网站中将这些数据删除，并为这两位隐私权的无意侵犯表示了歉意。我们中没有几个人对搜索查询信息所揭示的有趣的心理和个人特征进行过认真考虑，但阿诺德女士的故事却带来了一个骇人的发现。你喜爱的搜索引擎对于你最深层的、最内心的欲望的了解，可能比你最好的朋友还要多。

尽管如此，Netflix 确实希望有朝一日能够通过另一次竞赛来延续他们与国际研究团体的合作，而这一次当然会对隐私进行足够的保护。

数字——

4 = 一名女演员获得奥斯卡奖的最多次数。

——凯瑟琳·赫本（全部都是最佳女主角奖）。她曾获 12 次提名。

4 = 一名男演员获得奥斯卡奖的最多次数。

——克林特·伊斯特伍德，但不是因为表演而获奖（2 次最佳导演奖和 2 次最佳影片奖）。

——wiki. answers. com

第 18 章　后记

在写作本书的时候，我们常常不得不做出一些决断，其中频繁面对的一个就是：是否应当把方法 X 放在书中？最后，我们不得不在某个时候放弃某些写作，这也意味着略去某些有趣的方法。作为部分地补救，我们撰写了这个后记。因此，尽管下列方法并未在书中的其他章节中出现，但我们仍然将它们推荐给渴望了解更多也希望本书不是就此结束的读者。但是，本书还有出第二版的可能，因此我们欢迎读者给出反馈和建议。

层次分析法（AHP）

20 世纪 70 年代，托马斯·萨蒂（Thomas Saaty）发明了层次分析法（AHP，Analytic Hierarchy Process），以帮助决策者做出复杂的多准则决策[68, 69]。由于这种方法的广泛使用和影响，同时由于它对全球的政府和军方的影响，INFORMS（the Institute For Operations Research and Management Science，即运筹学与管理学研究协会）在 2008 年为萨蒂博士和他的 AHP 方法颁发了声誉卓著的影响奖（Impact Prize）。

AHP 方法的核心是成对比较倒数矩阵，通过计算该矩阵的主特征向量可以得到一个评分向量。从这个意义上讲，AHP 与第 4 章中的基纳法有着密切的联系。从另一个方面来说，AHP 也与第 2 章的梅西法有着紧密联系。在一个非常巧妙的分析中，大卫·格莱克证明，将标准 AHP 中的算术平均值改为几何平均值后所得的几何 AHP 法，在数学上与梅西法等价[31]。在参考文献［11］中，AHP 法被应用于大学橄榄球赛，而在参考文献［71］中则被用于以色列足球联赛。

雷德蒙法

在一篇 2003 年发表的《数学杂志》文章[61]中，查尔斯·雷德蒙（Charles Redmond）提出了一个评分方法，它是胜负评分体系的一个自然推广。雷德蒙法始于队伍的平均优势这一想法，它是对队伍所得的正负分差进行求和，然后再除以该队的比赛场数而得。

略去雷德蒙法对我们来说是一个艰难的选择，因为其中涉及了若干有趣的线性代数的内容。但是它落入了 YAMM（又一种矩阵方法）的框框里。这个方法给出的结果常常和其他的 YAMM 方法处于同一个水平，不过雷德蒙法的局限性更大，因为它要求所有队伍（或竞争者）要进行相同场数的比赛。

朴-纽曼法

在参考文献［59］中，朴洙勇（Juyong Park）[⊖]和 M. E. J. 纽曼（M. E. J. Newman）采取了一种网络法，来对美国大学橄榄球队进行排名。他们的方法同时考虑了*直接获胜*和*间接获胜*，来计算出每支队伍的一个获胜得分和一个失利得分。队伍 i 对队伍 j 的一次间接获胜，是指队伍 i 击败了队伍 k，而队伍 k 击败了队伍 j。因此，即使队伍 i 和队伍 j 没有在一次直接的对抗中碰头，但根据这一长度为 2 的间接关系也仍然能够推算出某些信息。同样可以考虑长度为 3，4 乃至更长的关系，尽管它们的权重将不断减小。朴-纽曼法使用了非常优美的数学，来考虑各种长度的关系。用户可以通过设置折扣参数，来控制权重如何随长度的增加而减少。这个方法与第 6 章的马尔可夫法和第 7 章的 OD 法之间存在着有趣的联系。

逻辑斯谛回归/马尔可夫链法（LRMC）

LRMC 评分法由索科尔和基瓦姆提出[48]，用于根据得分信息以及主场优势来对大学篮球队进行排名。他们的方法在预测三月疯狂锦标赛中取得了成功，同时还使得许多球迷在办公室的打赌中赢了钱。

LRMC 法中的马尔可夫链部分，在某些方面类似于第 6 章中的马尔可夫法。它们的终极目标是一样的——计算马尔可夫转移矩阵的稳态向量或主特征向量。两者的一个区别在于，LRMC 法使用了逻辑斯谛回归，来巧妙地估计马尔可夫转移矩阵中的元素，并考虑了主场优势。LRMC 的发明者还证明了 LRMC 法和科利法以及梅西法之间存在一个美妙的联系，而后两种方法都是基于赛程强度的思想展开的。

霍赫鲍姆法

最优化理论方面的专家多里特·霍赫鲍姆（Dorit Hochbaum）利用网络优化方法构造了几个排名方法[2,37,38,39]。霍赫鲍姆从计算量、复杂度和对于操纵的敏感性等角度分析了她的方法。这些方法可以适用于不同情况，因为它们可以根据需要来定制目标函数。当满足某些性质时，某些这类最优化排名方法的计算时间与基于线性代数的排名方法相比具有可比性。

蒙特卡洛仿真

仿真是许多技术人员特别是对分析棒球比赛感兴趣的技术人员所喜爱的一种方法。如 Accuscore. com 这样的商业体育预测公司，常常将仿真作为他们的主要工具。通过由以往表现所归总的统计数据，可以运行一个马尔可夫链来模拟两支队伍之间的比赛，马尔可夫链中的状态对应了比赛的不同情况（例如，对于给定投手的击球成功率、高飞球、将球击向左场时一垒上的跑垒员在二垒被封杀出局，等等），而转移概率则由以往的统计数据来构建。一种得

⊖ 根据英文音译。——译者注

到评分并进行预测的方法，就是模拟出两支队伍之间数千场的比赛，然后再对结果加以平均。在棒球上，仿真的应用效果相当不错，但是在应用于其他项目特别是 NFL 橄榄球赛时，它与本书所涵盖的许多不那么复杂的方法的效果相比而言是平分秋色的。仿真是一个有趣但较深的主题，它本身便足以写成一本书。对此感兴趣且知识程度更高的读者，可以通过简单的谷歌搜索找到丰富多样的讨论。

纯粹统计分析

我们决定放弃纯粹的统计方法，对于真正的统计学家而言，这可能令人失望。统计分析是一个切实可行的方法，特别是当有丰富的统计数据可用时更是如此，而且有一系列数量很多的统计方法可资运用。但正如仿真一样，统计分析本身就是一个足以写出若干本著作的研究领域，因此我们决定不打开这个潘多拉魔盒。尽管如此，我们可以尝试进行一下有趣的对比，在进行评分和排名这个任务时，去比较书中某些代数方法和基于根据观测数据拟合出分布的方法。梅西在他的网站上暗示他现在更多地依赖于统计方法，而非第 2 章中给出的代数方法。

还有更多更多

要对所有已经提出的评分和排名模型进行完整的综述，恐怕需要许多本书才能做到。仅仅是对于橄榄球项目的模型数量就已经让人难以置信了。以下所列的，是由大卫·威尔逊（David Wilson）所总结的数量巨大的信息源中的一部分。这些信息源中有许多都可以在如下网站中找到，当写本书时，该网站仍然可以访问。

homepages. cae. wisc. edu/ ~ dwilson/rsfc/rate/biblio. html

- I. Ali, W. Cook, and M. Kress. On the minimum violations ranking of a tournament. *Management Science*, 32(6):660–672, 1986.
- B. Amoako-Adu, H. Marmer, and J. Yagil. The effeciency of certain speculative markets and gambler behavior. *Journal of Economics and Business*, 37, 1985.
- L. B. Anderson. Paired comparisons. *Operations Research and the Public Sector: Handbooks in Operations Research and Management Science*, S. M. Pollock, M. H. Rothkopf, and A. Barnett, eds., 6(Chapt. 17):585–620, 1994.
- David H. Annis and Bruce A. Craig. Hybrid paired comparison analysis, with applications to the ranking of college football teams. *Journal of Quantitative Analysis in Sports*, 1(1), 2005
- David H. Annis. Dimension reduction for hybrid paired comparison models. *Journal of Quantitative Analysis in Sports*, 3(2), 2007.
- David H. Annis and Samuel S. Wu. A comparison of potential playoff systems for NCAA I-A Football.

- Gilbert W. Bassett. Robust sports ratings based on least absolute errors. *The American Statistician*, May:1–7, 1997.
- R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika* (39):324–45, 1952. The idea is believed to have been first proposed in E. Zermelo in "Die berechnung der Turnier-Ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung" *Mathematische Zeitschrift*, 29:436–60, 1929.
- Hans Bhlmann and Peter J. Huber. Pairwise comparison and ranking in tournaments. *The Annals of Mathematical Statistics*, 501–510, 1962.
- Thomas Callaghan, Mason A. Porter, and Peter J. Mucha. Random walker ranking for NCAA Division I-A football. Georgia Institute of Technology, 2003.
- Thomas Callaghan, Peter J. Mucha, and Mason A. Porter. The bowl championship series: A mathematical review. *Notices of the AMS* (September)887–893, 2004.
- C. R. Cassady, L. M. Maillart, and S. Salman. Ranking sports teams: a customizable quadratic assignment approach. *Interfaces* 35(6): 497–510, 2005.
- P. Y. Chebotarev and E. Shamis. Preference fusion when the number of alternatives exceeds two: Indirect scoring procedures. *Journal of the Franklin Institute: Engineering and Applied Mathematics*, 336(2):205-226, 1999.
- G. R. Conner, and C. P. Grant. An extention of Zemelo's model for ranking by paired comparison. *European Journal of Applied Mathematics*, 11(3):225–247, 2000.
- W. Cook, I. Golan and M. Kress. Heuristics for ranking players in round robin tournaments. *Computers and Operations Research*, 15(2):135–144, 1988.
- Morris L. Eaton. Some optimum properties of ranking procedures. *The Annals of Mathematical Statistics*, July:124–137, 1966.
- Arpad E. Elo. *The Rating of Chess Players Past and Present*, 2nd Ed., Arco Publishing, 1986.
- L. Fahrmeir and G. Tutz. Dynamic stochastic models for time-dependent ordered pair-comparison systems. *Journal of the American Statistical Association*, 89:1438–49, 1994.
- Christopher J. Farmer. Probabilistic modelling in multi-competitor games. Univ. of Edinburgh, 2003.
- John A. Fluek and James F. Korsh. A generalized approach to maximum likelihood paired comparison ranking. *The Annals of Statistics*, 3(4)846–861, 1975.
- L. R. Ford Jr. Solution of a ranking problem from binary comparisons. *American Mathematical Monthly*, 64(8):28–33, 1957.
- Mark E. Glickman and H. S. Stern. A state-space model for National Football League Scores. *Journal of the American Statistical Association*, 93:25–35, 1998.
- S. Goddard. Ranking in tournaments and group decission making. *Management Science*, 29(12): 1384–1392, 1983.

- Clive Grafton. Junior college football rating systems. *Statistics Bureau of the National Junior College Athletic Association,* 1955.
- S. S. Gupta and Milton Sobel. On a statistic which arises in selection and ranking problems. *The Annals of Mathematical Statistics,* pp. 957–967, 1957.
- David Harville. The use of linear-model methodology to rate high school or college football teams. *Journal of the American Statistical Association*. June:278–289, 1977.
- David Harville. Predictions for NFL games via linear-model methodology. *Journal of the American Statistical Association*. September:516–524, 1977.
- David Harvilleand M. H. Smith. The home-court advantage: How large and does it vary from team to team? *The American Statistician,* 48(1):22–28, 1994.
- David Harville. College football: A modified least-squares approach to rating and prediction. *American Statistical Proceedings of the Section on Statistics in Sports,* 2002.
- David Harville. The selection and/or seeding of college basketball or football teams for postseason competition: A statistician's perspective. *American Statistical Proceedings of the Section on Statistics in Sports,* pp. 1–18, 2000.
- David Harville. The selection or seeding of college basketball or football teams for postseason competition. *Journal of the American Statistical Association,* 98(461):17–27, 2003.
- Dorit S. Hochbaum. Ranking sports teams and the inverse equal paths problem. Department of Industrial Engineering and Operations Research and Walter A. Haas School of Business, University of California, Berkeley.
- Tzu-Kuo Huang, Ruby C. Weng, and Chih-Jen Lin. Generalized Bradley-Terry models and multi-class probability estimates. *Journal of Machine Learning Research,* 7:85–115, 2006.
- Peter J. Huber. Pairwise comparison and ranking: Optimum properties of the row sum procedure. *The Annals of Mathematical Statistics,* pp. 511–520, 1962.
- Thomas Jech. The ranking of incomplete tournaments: A mathematician's guide to popular sports. *American Mathematical Monthly,* 90(4):246–66, 1983.
- Samuel Karlin. *Mathematical Methods & Theory in Games, Programming, & Economics,* Dover Publications, March 1992.
- L. Knorr-Held. Dynamic rating of sports teams. *The Statistician,* 49:261–76, 2000.
- R. J. Leake. A method for ranking teams: with an Application to college football. *Management Science in Sports,* ed. R. E. Machol et al., North-Holland Publishing Co., pp. 27–46, 1976.
- J. H. Lebovic and L. Sigelman., The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting,* 17(1):105–120, 2001.

- Joseph Martinich. College football rankings: Do computers know best? *Interfaces*, 32(5):85–94, 2002.
- William N. McFarland. *An Examination of Football Scores*, Waverly Press, 1932.
- David Mease. A penalized maximum likelihood approach for the ranking of college football teams independent of victory margins. *The American Statistician*, November, 2003.
- Joshua Menke and Tony Martinez. A Bradley-Terry artificial neural network model for individual ratings in group competitions. Computer Science Department, Brigham Young University, 2006.
- D. J. Mundfrom, R. L. Heiny, and S. Hoff, Power ratings for NCAA division II football. *Communications in Statistics Simulation and Computation*, 34(3):811–826, 2005.
- Juyong Park and M. E. J. Newman. Network-based ranking system for U.S. college football. *Journal of Statistical Physics*, 2005.
- Michael B. Reid. Least squares model for predicting college football scores. University of Utah, 2003.
- Jagbir Singh and W. A. Thompson Jr. "A Treatment of Ties in Paired Comparisons." The Annals of Mathematical Statistics, 39(6):2002–2015, 1988.
- Z. Sinuany-Stern. Ranking of sports teams via the AHP. *Journal of the Operations Research Society*, 39(7):661–667, 1988.
- Warren D. Smith. Rating systems for game players, and learning. *NEC*, July, 1994.
- M. S. Srivastava and J. Ogilvie. The performance of some sequential procedures for a ranking problem. *The Annals of Mathematical Statistics*, 39(3):1040–1047, 1968.
- Raymond T. Stefani. Football and basketball predictions using least squares. *IEEE Transactions on Systems, Man. and Cybernetics*, 7, 1977.
- Raymond T. Stefani. Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics*, pp. 116–123, 1980.
- Hal Stern. A continuum of paired comparisons models. *Biometrika*, 77(2):265–73, 1990.
- Hal Stern. On the probability of winning a football game. *The American Statistician*, August:179–183, 1991.
- Hal Stern. Who's number one? - rating football teams. *Proceedings of the Section on Statistics in Sports*, pp.1–6, 1992 .
- Hal Stern. Who's number 1 in college football?... and how might we decide? *Chance*, Summer:7–14, 1995.
- H. S. Stern and B. Mock. College basketball upsets: will a 16-seed ever beat a 1-seed? *Chance* (11):26–31, 1998.

- H. S. Stern. Statistics and the college football championship. *American Statistician*, 58(3):179–185, 2004.
- H. S. Stern. In favor of a quantitative boycott of the bowl championship series. *Journal of Quantitative Analysis in Sports*, 2(1), 2006.
- Daniel F. Stone. Testing Bayesian updating with the AP top 25. John Hopkins University, October, 2007.
- I. B. Thomas. *Method of ranking college football teams*, Allen, Lane and Scott, 1922.
- Mark Thompson. On any given Sunday: Fair competitor orderings with maximum likelihood methods. *Journal of the American Statistical Association*, 70:536–541, 1975.
- Y. L. Tong. An adaptive solution to ranking and selection problems. *The Annals of Statistics*, 6(3):658–672, 1978.
- John A. Trono. Applying the overtake and feedback algorithm. *Dr. Dobb's Journal*, February:36–41, 2004.
- John A. Trono. An effective nonlinear rewards-based ranking system. *Journal of Quantitative Analysis in Sports*, 3(2), 2007.
- Brady T. West and Madhur Lamsal. A new application of linear modeling in the prediction of college football bowl outcomes and the development of team ratings. *Journal of Quantitative Analysis in Sports*, 4(3), 2008.
- R. Wilkins. Electrical networks and sports competition. *Electronics and Power*, 29(5):414–418, 1983.
- R. L. Wilson. Ranking college football teams: A neural network approach. *Interfaces*, 25(16):44–59, 1995.
- R. L. Wilson. The "real" mythical college football champion. *Operation Research / Management Science Today*, pp. 24–29, 1995.
- R. A. Zuber, J. M. Gander, and B. D. Bowers. Beating the spread: Testing the efficiency of the gambling market for National Football League games. *Journal of Political Economy*, 93, 1985.

数字——

7.3 = White Russian 鸡尾酒所得到的评分（按 1 到 10 打分）。

——它是（100 种）鸡尾酒配方中的第 1 名，得到了 324 票。

6.2 = Sex on the Beach #2 所得的评分。

——它是排名最后的鸡尾酒（第 100 名），仅得到了 29 票。

——www. drinknation. com/drinks/best

词 汇 表

阿罗不可能性定理 肯尼斯·阿罗关于任何投票体系（以及评分体系）必须满足的四个条件的著名定理。定理中的不可能性来自他所证明的“没有任何投票体系能够同时满足所有这四个条件”的事实。

平均排名法 一种排名聚合方法，其中表示某个对象在数个排名列表中排名的整数值被进行平均，以给出一个排名聚合列表。

权威评分 根据结点的入链的质与量赋予有向图中一个结点的评分。

无偏性 这一描述是指一种评分方法能够不受如下问题的影响，即强队在弱队身上掠取大的比分。

波达计数法 一种常用的排名聚合方法，由让-查理·德·波达于1770年提出，其中每名候选者都获得一个分数，对于每个排名列表而言，这个分数等于排名在该候选者之后的人数。对各个排名列表给出的每名候选者的这一分数进行求和，便可以得到一个单一的值。而之后候选者便根据其总的波达数进行排名。

BCS（Bowl Championship Series，即碗系列锦标赛） NCAA大学橄榄球赛的管理机构，它为每支球队给出一个BCS评分，并根据这些评分挑选出各个碗系列锦标赛的参赛队。

竞猜 在锦标赛正式开始之前，体育迷们填写一份竞猜表来预测锦标赛的结果。

重心法 一种计算量非常小的评分体系，它使用分差矩阵的重心作为评分向量。

科利法 由威斯利·科利博士建立的一种评分体系，他对以胜率作为各队评分的这一最简单也最古老的评分体系进行了改进。这个小改动的核心是方程 $r_i = (1 + w_i)/(2 + t_i)$，这一改动弥补了胜率法的若干缺陷。

科利化的梅西法 一种体现了梅西法和科利法之间联系的评分体系，通过由公式 $\boldsymbol{C} = 2\boldsymbol{I} + \boldsymbol{M}$ 给出的这两种方法之间的联系，将两种方法的性质结合了起来。

结合方法 一种排名聚合方法，能够使得看起来与其他列表中的排名不一致的异常值造成的影响最小化。

一致对 给定两个排名列表，称一对同时出现于两个列表中的对象为一致的，如果这两个对象在两个列表中的相对排名是一样的话；也就是说，如果对象 i 在两个列表中都排名在对象 j 之前（或之后），那么对 (i,j) 就是一致的。一致对用于排名相关性的肯德尔测度 τ 之中。

孔多塞胜者 当第三名候选人退出角逐时剩下两者中的获胜者。例如，艾伯特·戈尔就是2000年大选时在佛罗里达州的孔多塞胜者。如果拉尔夫·纳德没有从戈尔手中分走选票并令乔治 W. 布什获胜的话，戈尔就将是最终的胜者。

悬挂结点 没有出链的结点。在体育比赛中就是从未落败的队伍。

防守评分向量 OD 法给出的一个向量，其中包含了所有队伍的防守评分。

不一致对 给定两个排名列表，如果一对同时出现于两个列表中的对象在各个列表中的相对排名不符合，那么这一对对象就称为不一致的；也就是说，如果对象 i 在一个列表中排名在对象 j 之前，而在另一个列表中排在对象 j 之后，那么对 (i,j) 就是不一致的。不一致对是排名相关性的肯德尔测度 τ 的一个组成部分。

埃洛法 一种基于线性更新法则的评分体系，由阿帕德·埃洛提出，用以对全世界的国际象棋棋手进行评分和排名。该体系还被修改并运用于包括足球在内的其他应用之中。

进化优化 一种利用达尔文进化原理（适者生存、交配和变异）来求解困难的最优化问题的方法。

完全排名列表 包含了所有待排名对象的一个排列的列表。

HITS 超文本归纳主题搜索的简称，由乔恩·克莱因伯格提出，用以对链接文档进行排名。该算法是最为流行和成功的网页排名方法之一，并被用于现在的搜索引擎 Ask. com 中。

枢纽评分 根据结点出链的质与量赋予有向图中结点的评分。

无关选择独立性 阿罗对于投票体系的第二个要求，它指出，在所考虑子集之外的其他候选者的排名变化，不应影响到子集内两个候选者之间的相对排名先后。

基纳法 由詹姆斯·基纳所建立的一种评分体系，使用（可能经过修正的）得分矩阵的主特征向量作为评分向量。

凯梅尼化 一个计算量较小的求精步骤，能够用于一个排名聚合过程给出的结果，以提高聚合排名和各输入排名之间的吻合度。

肯德尔测度 τ 排名相关性的一个测度 τ，由莫里斯·肯德尔提出，以确定两个排名列表之间的偏差，$-1 \leqslant \tau \leqslant 1$。如果两个列表完全吻合，则 $\tau = 1$；如果两个列表的排名顺序完全颠倒，则 $\tau = -1$。

MAD 赛季中一支队伍的胜率，及其与该队累积失分和累积得分的比值之间的绝对平均偏差。

马尔可夫法 利用马尔可夫链的稳态向量来进行对象评分和排名的一种评分体系。

梅西法 由肯尼斯·梅西所建立的一种评分体系，使用最小二乘法来对队伍进行评分和排名。梅西法围绕方程 $r_i - r_j = y_k$ 进行，该方程表达了如下想法，即两支队伍之间的评分差，应能理想地预测两者对抗时胜方的优势分差。

OD 法 一种改进的 HITS 算法，它生成了一个有向图，其中的结点代表球队，而链接代表两支队伍之间的交互。每条有向链接可以具有两种解释：一个与球队的攻击能力有关，另一个则与球队的防守能力有关。

非独裁性 阿罗对于投票体系的第四个要求，它指出，没有任何一个投票者可以对选举拥有不成比例的控制力。

进攻评分向量 OD 法给出的一个向量，包含了所有球队的进攻评分。

PageRank 由谷歌及其用来进行网页排名的马尔可夫法所给出的向量。

帕累托法则 阿罗对于投票体系的第三个要求，它指出，如果所有的投票者都选择了 A 而非 B，那么一个合理的投票体系应当总是将 A 排名在 B 之前。

部分列表 包含了所考虑的对象的一个子集的列表。它不同于完全列表，后者包含了所有对象。

排列 长度为 n、包含整数 1 到 n 的一个向量。对于排名问题而言，这些整数的顺序就对应了对象的排名位置。

多数投票 一种投票方法，其中每名投票者都为他们的最优先选择投出唯一一张选票。

分差 两支队伍在一场给定的比赛中的得分之差。

偏好列表 一种投票方法，其中每名投票者都提交候选者的一个排名列表。

排名列表 n 个对象的一个列表，其中包含了整数 1 到 n 的一个排列，每个整数值便给出了对应对象的排名位置。

排名聚合 使用数学方法来由多个单独的排名列表生成一个鲁棒的聚合列表的过程。

评分列表 一个标量的列表，这些标量代表了对应对象的评分。当对评分列表进行排序后，便给出了一个排名列表。

回溯打分 一种比较不同排名方法的方式。回溯打分利用排名方法所得出的排名列表来预测以往比赛的胜者。

智能排名方法 只需经过短暂的热身期便能进行准确预测的方法，即在给出好的预测之前不需要大量数据的方法。

前 k 位列表 一种长度为 k 的部分列表，其中包含了某个排名中仅前 k 位即第 1 位到第 k 位的对象。

无限制域 阿罗对于投票体系的第一个要求，它指出，每名投票者都应该能够按他或她希望的任何排列方式来对候选人进行排名。

抽头 投注经纪人的赢利手段。它不过就是经纪人因其所提供的服务而收取的一笔费用。

胜率法 一种利用对象在所有成对比较中获胜次数的比例来进行排名的非常简化的方法。

参考文献

[1] Ralph Abbey, John Holodnak, Chandler May, Carl Meyer, and Dan Moeller. Rush versus pass: Modeling the NFL, 2010. Preprint. Available for download from `meyer.math.ncsu.edu/Meyer/REU/REU2009/REU2009Paper.pdf`.

[2] R. K. Ahuja, Dorit S. Hochbaum, and J. B. Orlin. Solving the convex cost integer dual of minimum cost network flow problem. *Management Science*, 49:950–964, 2003.

[3] Iqbal Ali, Wade D. Cook, and Moshe Kress. On the minimum violations ranking of a tournament. *Management Science*, 32(6):660–672, 1986.

[4] Ronald D. Armstrong, Wade D. Cook, and Lawrence M. Seiford. Priority ranking and consensus formation: The case of ties. *Management Science*, 28(6):638–645, 1982.

[5] Kenneth Arrow. *Social Choice and Individual Values,* 2nd Ed. Yale University Press, 1970.

[6] James R. Ashburn and Paul M. Colvert. A bayesian mean-value approach with a self-consistently determined prior distribution for ranking of college football teams, 2006. `arxiv.org/pdf/physics/0607064`.

[7] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. Van der Vorst. *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*. SIAM, 2nd edition, 1994.

[8] Gely P. Basharin, Amy N. Langville, and Valeriy A. Naumov. The life and work of A. A. Markov. *Linear Algebra and Its Applications*, 386:3–26, 2004.

[9] Michael W. Berry, Bruce Hendrickson, and Padma Raghavan. Sparse matrix reordering schemes for browsing hypertext. In J. Renegar, M. Shub, and S. Smale, editors, *Lectures in Applied Mathematics (LAM)*, volume 36, pages 99–123. American Mathematical Society, 1996.

[10] Dimitris Bertsimas and John N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.

[11] Vladimir Boginski, Sergiy Butenko, and Panos M. Pardalos. Matrix-based methods for college football rankings, 2005. Preprint.

[12] J. C. Borda. Mémoire sur les élections au scrutin. *Histoire de l'Académie Royale des Sciences*, 1781.

[13] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika*, 39:324–45, 1952.

[14] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 33:107–17, 1998.

[15] Sergey Brin, Lawrence Page, R. Motwami, and Terry Winograd. The PageRank citation ranking: Bringing order to the Web. Technical Report 1999-0120, Computer Science Department, Stanford University, 1999.

[16] Edward B. Burger and Michael Starbird. *The Heart of Mathematics: An Invitation to Effective Thinking*. Wiley, 2005.

[17] Thomas Callaghan, Peter J. Mucha, and Mason A. Porter. Random walker ranking for NCAA division I-A football. *American Mathematical Monthly*, 114:761–777, 2007.

[18] Timothy P. Chartier, Erich Kreutzer, Amy N. Langville, Kathryn Pedings, and Yoshitsugu Yamamoto. Mininum violations sports ranking using evolutionary optimization and binary integer linear program approaches. In Anthony Bedford and Matthew Ovens, editors, *Proceedings of the Tenth Australian Conference on Mathematics and Computers in Sport*, pages 13–20. MathSport (ANZIAM), 2010.

[19] Timothy P. Chartier, Erich Kreutzer, Amy N. Langville, and Kathryn E. Pedings. Sports ranking with nonuniform weighting. *Journal of Quantitative Analysis in Sports*, 2010. `www.bepress.com/jqas/vol7/iss3/6/`.

[20] Timothy P. Chartier, Erich Kreutzer, Amy N. Langville, and Kathryn E. Pedings. Accounting for ties when ranking items, 2011. In preparation.

[21] Timothy P. Chartier, Erich Kreutzer, Amy N. Langville, and Kathryn E. Pedings. Sensitivity of ranking vectors. *SIAM Journal on Scientific Computing*, 2011. To appear.

[22] Wesley N. Colley. Colley's bias free college football ranking method: The colley matrix explained, 2002. `www.Colleyrankings.com`.

[23] Wade D. Cook and Lawrence M. Seiford. Priority ranking and consensus formation. *Management Science*, 24(16):1721–1732, 1978.

[24] Cynthia Dwork, Ravi Kumar, Moni Naor and D. Sivakumar. Rank aggregation methods for the Web. In *The Tenth International World Wide Web Conference.* ACM Press, 2001.

[25] Cynthia Dwork, Ravi Kumar, Moni Naor, and D. Sivakumar. Rank aggregation revisited. `citeseerx.ist.psu.edu`.

[26] Nadav Eiron, Kevin S. McCurley, and John A. Tomlin. Ranking the Web frontier. In *The Thirteenth International World Wide Web Conference.* ACM Press, 2004.

[27] Ronald Fagin, Ravi Kumar, Mohammad Mahdian, D. Sivakumar, and Erik Vee. Comparing and aggregating rankings with ties.
`www.almaden.ibm.com/cs/people/fagin/bucketwb.pdf`.

[28] Ronald Fagin, Ravi Kumar, and D. Sivakumar. Comparing top k lists. In *ACM SIAM Symposium on Discrete Algorithms*, pages 28–36, 2003.

[29] William Feller. *An Introduction to Probability Theory and Its Applications,* Vol. 1, 3rd Edition. John Wiley, 1968.

[30] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra And Its Applications*, 114/115:717–734, 1989.

[31] David Gleich. Geometric ahp, 2010. Preprint.

[32] David Gleich, Leonid Zhukov, and Pavel Berkhin. Fast parallel PageRank: A linear system approach. In *The Fourteenth International World Wide Web Conference*. ACM Press, 2005.

[33] David F. Gliech and Lek-Heng Lim. Rank aggregation via nuclear norm minimization, 2010. Preprint, Submitted to KKD 2011.

[34] Anjela Y. Govan. *Ranking Theory with Application to Popular Sports*. Ph.D. thesis, North Carolina State University, December 2008.

[35] Anjela Y. Govan, Amy N. Langville, and Carl D. Meyer. Offense-defense approach to ranking team sports. *Journal of Quantitative Analysis in Sports*, 5(1):1–17, 2009.

[36] William H. Greene. *Econometric Analysis*. Prentice Hall, 1997.

[37] Dorit S. Hochbaum. Ranking sports teams and the inverse equal paths problem. *Lecture Notes in Computer Science*, 4286:307–318, 2006.

[38] Dorit S. Hochbaum and A. Levin. Methodologies for the group rankings decision. *Management Science*, 52:1394–1408, 2006.

[39] Dorit S. Hochbaum and E. Moreno-Centeno. Country credit-risk rating aggregation via the separation-deviation model. *Optimization Methods and Software*, 23:741–762, 2008.

[40] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[41] Luke Ingram. Ranking NCAA sports teams with linear algebra. Master's thesis, College of Charleston, April 2007.

[42] James P. Keener. The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, 35(1):80–93, 1993.

[43] Maurice Kendall. A new measure of rank correlation. *Biometrika*, 30, 1938.

[44] Maurice G. Kendall. Further contributions to the theory of paired comparisons. *Biometrics*, 11, 1955.

[45] Jon Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, 1999.

[46] Philip A. Knight. The Sinkhorn-Knopp algorithm: Convergence and applications. *SIAM Journal of Matrix Analysis*, 30(1):261–275, 2008.

[47] Ronald Fagin, Ravi Kumar, and D. Sivakumar. Comparing top-k lists. *SIAM*, 17, 2003.

[48] Paul Kvam and Joel S. Sokol. A logistic regression/Markov chain model for NCAA basketball. *Naval Research Logistics*, 53(8):788–803, 2006.

[49] Amy N. Langville and Carl D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. Princeton University Press, Princeton, 2006.

[50] Amy N. Langville, Kathryn Pedings, and Yoshitsugu Yamamoto. A minimum violations ranking method, 2009. Preprint.

[51] Chris Pan-Chi Lee, Gene H. Golub, and Stefanos A. Zenios. A fast two-stage algorithm for computing PageRank and its extensions. Technical Report SCCM-2003-15, Scientific Computation and Computational Mathematics, Stanford University, 2003.

[52] Kenneth Massey. Statistical models applied to the rating of sports teams. Bachelor's thesis, Bluefield College, 1997.

[53] Fabien Mathieu and Mohamed Bouklit. The effect of the back button in a random walk: Application for PageRank. In *The Thirteenth International World Wide Web Conference*, pages 370–71, New York, 2004. Poster.

[54] Carl D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.

[55] Zbigniew Michalewicz and David B. Fogel. *How to Solve It: Modern Heuristics*. Springer, New York, 1998.

[56] Netflix.com. The netflix prize, 2006. `www.netflixprize.com/rules`.

[57] Alantha Newman and Santosh Vempala. Fences are futile: On relaxations for the linear ordering problem. *Lecture Notes in Computer Science*, 2081:333–347, 2001.

[58] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–255, 2003.

[59] Juyong Park and M. E. J. Newman. A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment*, October 2005.

[60] Ronald L. Rardin. *Optimization in Operations Research*. Prentice Hall, 1998.

[61] C. Redmond. A natural generalization of the win-loss rating system. *Mathematics Magazine*, 76(2):119–126, 2003.

[62] Gerhard Reinelt. *The Linear Ordering Problem: Algorithms and Applications*. Heldermann Verlag, 1985.

[63] Gerhard Reinelt, M. Grötschel, and M. Jünger. Optimal triangulation of large real world input-output matrices. *Statistical Papers*, 25(1):261–295, 1983.

[64] Gerhard Reinelt, M. Grötschel, and M. Jünger. A cutting plane algorithm for the linear ordering problem. *Operations Research*, 32(6):1195–1220, 1984.

[65] Gerhard Reinelt, M. Grötschel, and M. Jünger. Facets of the linear ordering polytope. *Mathematical Programming*, 33:43–60, 1985.

[66] Yousef Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, 2003.

[67] T. L. Saaty. Rank according to Perron: a new insight. *Mathematics Magazine*, 60(4):211–213, 1987.

[68] Thomas L. Saaty. *The Analytic Hierarchy Process*. McGraw-Hill, 1980.

[69] Thomas L. Saaty and L. G. Vargas. *Decision Making in Economic, Political, Social and Technological Environments with the Analytic Hierarchy Process*. RWS Publications, 1994.

[70] Richard Sinkhorn and Paul Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21, 1967.

[71] Zilla Sinuany-Stern. Ranking of sports teams via the AHP. *Journal of Operational Research Society*, 39(7):661–667, 1988.

[72] Warren D. Smith. Sinkhorn ratings, and newly strongly polynomial time algorithms for Sinkhorn balancing, Perron eigenvectors, and Markov chains, 2005. `www.yaroslavvb.com/papers/smith-sinkhorn.pdf`.

[73] George W. Soules. The rate of convergence of Sinkhorn balancing. *Linear Algebra And Its Applications*, 150:3–40, 1991.

[74] G. W. Stewart. *Matrix Algorithms*, volume 2. SIAM, 2001.

[75] William J. Stewart. *Introduction to the Numerical Solution of Markov Chains*. Princeton University Press, 1994.

[76] Gilbert Strang. *Introduction to Linear Algebra,* 4th Edition. Wellesley Cambridge Press, 2009.

[77] Marcin Sydow. Random surfer with back step. In *The Thirteenth International World Wide Web Conference*, pages 352–53, New York, 2004. Poster.

[78] John A. Tomlin. A new paradigm for ranking pages on the World Wide Web. In *The Twelfth International World Wide Web Conference*, ACM Press, 2003.

[79] Sebastiano Vigna. Spectral ranking, 2010. Preprint. Can be downloaded from the site `vigna.dsi.unimi.it/papers.php`.

[80] Philipp von Hilgers and Amy N. Langville. The five greatest applications of markov chains. In Amy N. Langville and William J. Stewart, editors, *Proceedings of the Markov Anniversary Meeting*, pages 155–368. Boson, 2006.

[81] T. H. Wei. *The algebraic foundations of ranking theory*. Ph.D. thesis, Cambridge University, 1952.

[82] Wayne L. Winston. *Operations Research: Applications and Algorithms*. Duxbury Press, 2003.

[83] Wayne L. Winston. *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football.* Princeton University Press, 2009.

[84] Laurence A. Wolsey. *Integer Programming.* Wiley-Interscience, 1998.

[85] Laurence A. Wolsey and George L. Nemhauser. *Integer and Combinatorial Optimization.* Wiley-Interscience, 1999.

索　引

Ali, I. , 等, 225
Amoako-Adu, B. , 等, 225
Anderson, L. B. , 225
Annis, David H. , 等, 225
AOL 数据, 221
Ask. com, 92

Bassett, Gilbert W. , 225
BCS, 6, 9, 21, 136, 162, 204, 211
BCS 排名, 17
BelKor, 134
Bhlmann, Hans 和 Huber, Peter J. , 226
Bing, 网页数量, 182
Bradley, R. A. 和 Terry, M. E. , 225

Callaghan, Thomas, 等, 226
Cassady, C. R. , 等, 226
Chebotarev P. Y. 和 Shamis, E. , 226
Colvert, Paul. M, 53
Conner, G. R. 和 Grant, C. P. , 226
Cook, W. , 等, 226

Eaton, Morris L. , 226
eBay, 135
Ensemble 队, 134
ESPN, 151
ESPN 锦标赛竞猜, 213

Facemash, 65
Fahrmeir, L. 和 Tutz, G. , 226
Farmer, Christopher J. , 226
FBI, 216
FIDE, 53
Fischer, Bobby, 66
Fluek, Jong A. and Korsh, James F. , 226

Ford, L. R. , Jr. , 226

Glickman, Mark E. , 226
Goddard, S. , 226
Grafton, Clive, 226
Gupta, Shanti S. and Sobel, Milton, 227
Harville, David, 227
HDI, 6
Huang, Huang, Weng, and Lin, 227
Huber, Peter J. , 227

IMDb, 6

Jech, Thomas, 227

Karlin, Samuel, 227
Knorr-Held, L. , 227

Leake, R. J. , 227
Lebovic J. H. and Sigelman L. , 227
Line Makers, 96

MAD, 45, 233
Martinich, Joseph, 227
MATLAB, 16, 95
McFarland, William N. , 228
Mease, David, 228
Menke, J. and Martinez, T. , 228
MovieLens, 141
Mundfrom D. J. , 等, 228

NCAA 竞猜表, 数目, 154
NCAA 男子篮球, 种子排名最低的获胜者, 200

NCAA 男子篮球, 总赌注, 216

Netflix，6，25，95，135，140，220，221
Netflix 奖，134
NFL，143
NFL 的分差误差，123
NFL 分差，有史以来最大，126
NFL 平局，最多，146
NHL，141
NHL 平局，最多和最少，146
NPR，152

PageRank，6，14，51，67，68，77，172，233
Park，J. and Newman，M. E. J.，228

Q-bert 矩阵，101

Reid，M. B.，228
RPI，6，202

SALSA，172
Sex on the Beach #2，230
SIAM，77
Singh，J. and Thompson W. A. Jr.，228
Sinuany-Stern，Z.，228
Smith，M. H.，227
Smith，Warren D.，228
Srivastava，M. S. 和 Ogilvie，J.，228
Stefani，Raymond T.，228
Stern，Hal.，等，228
Stone，Daniel F.，229

Thomas，I. B.，229
Thompson，Mark，229
Tong，Y. L.，229
Trono，John A.，229

USCF，53

West，B. T. and Lamsal，M.，229
White Russian，230
whydomath ?，77
Wilkins，R.，229
Wilson，R. L.，229

Zermelo，E.，226
Zuber，R. A.，等，229

《阿凡达》，28
阿罗，肯尼斯，4，231
阿罗的不可能性定理，4，160～163，231
阿什伯恩，詹姆斯·R，53
埃洛，阿帕德·E，53，226
埃默特，马克，19
艾伯特·戈尔，160
安德森，杰夫，17
奥巴马，巴拉克，19
奥科涅夫斯基，迈克尔，xvi
奥斯卡奖，获得的最多次数，222

巴沙贝，伊拜，xv
贝萨妮，v
比林斯利，理查德，17
毕达哥拉斯定理
　棒球，45
　橄榄球，45
　最优指数，46
冰球，141
波达，让-查理·德，3
波达计数，165，166，231
波达计数，BCS，17
波尔，芭芭拉，xv
伯曼，克里斯，19
不可能性定理，4，231
不可约马尔可夫链，73
不留破绽，BCS，18
不一致，204，205
不一致对，232
布拉德利-特里模型，18
布朗，埃米特，博士，48
布林，谢尔盖，51
部分列表，205，233

参与，大学橄榄球，8
层次分析法（AHP），223
超级比赛，220
超级列表，171
成对比较，124，135，220

成对比较倒数矩阵，223
成功的评判标准，211～215
承续法则，30
城市区图，100
橙碗，19
抽头，234
聪明的排名方法，214，234

打破平局，167
大 10 联盟，19，20
大 12 联盟，19
大东部联盟，19
大西洋沿岸联盟，19
大学橄榄球，观众，8
大学橄榄球，伤亡，112
大学排名，6
道格拉斯，埃米，xv
得分差距矩阵，118
电影《社交网络》，64
东南联盟，19
赌中分差，115
赌注，NCAA 男子篮球，216
对称性，201
对称重排，98
对客场获胜进行加权，147
对女孩排名，64
对女孩评分，64
对体育赛事所下的赌注，96
对战胜某些已知的对手加权，147
多数投票，3，160，233
多维迪奥，尼克，153
多最优解，187

二分图，202
二值整数线性规划，185

反垄断与 BCS，19
防守评分，11，79，232
非独裁性，4，160，163，171，233
非周期性，73
分差，11，14，233
分差，NFL 中最大，126
分差评分，117
　基于胜负的，124
　与主场优势，121
　NFL 2009—2010，120
　根据得分差，120
　总绝对误差，121
弗罗贝尼乌斯，基奥里，51
弗罗贝尼乌斯范数，102，118

橄榄球教练员调查，17
高斯-马尔可夫定理，47
高斯消元法，23
格莱克，大卫，xv，223
格雷斯体育年鉴，48
古德森，尼尔，xiv，xvi，151，213
谷歌，51
谷歌，收入，52
谷歌，网页数量，182
谷歌炸弹，163
规范化，基纳法，32
国际象棋评分，有史以来最高分，66

哈佛辣妹，64
哈里斯大学橄榄球调查，17
海森矩阵，119
河中落叶，51
赫本，凯瑟琳，222
赫斯特，克里斯，17
黑洞，202
恒等性，201
后退按钮，74
霍赫鲍姆，多里特，224，227

鸡尾酒，评分最高的，230
基本方程，基纳法，34
基本排名差距矩阵，98
基本形，108
基纳，詹姆斯，18，29，51
基纳法的不可约性，35
基纳法的非负性，35
基纳法的约束，35
基瓦姆，保罗，224
几何平均值，220
计算机生成的列表，181

加权邻接矩阵，16
加权正规方程，150
加权最小二乘法，150
加时，142
假如……会怎样，155～158
节日碗，19
结合方法，169，231
《今日美国》，17，53
进攻评分，11，79，233
进化方法，194
进化优化，104，232
竞猜，213，231
竞猜表，数量，154
酒，评分最高的，230
局部凯梅尼化，175
矩阵的对角线，88
距离度量，201
聚合的图论方法，172
聚合列表，211
绝对平均偏差，45

卡里塔，马利·让·安托万·尼古拉，3
卡耐基分类，78
卡斯帕罗夫，加里，66
凯梅尼化，175，232
考古发掘，192
柯西-施瓦兹（或 CBS）不等式，119
科利，威斯利，17，21
科利法和 LRMC，224
科利化的梅西法，25，231
可悲的失败，164
可扩展性，16
克莱因伯格，乔恩，92
克雷洛夫法，23
肯德尔，莫里斯，51，201，203
孔多塞胜者，160，166，232
控球，219

垃圾信息，16，163
拉普拉斯的承续法则，22，30
拉斯维加斯盘口，NFL 2009—2010，123
辣妹，评分与排名，64
篮球，143
朗尼安，鲍勃，53
雷德蒙，查尔斯，223
雷德蒙法，223
冷门，史上最大，158
联合国，6
脸谱，64
链接农场，163
两个排名列表间的定量偏差
　肯德尔测度 τ，203
　　公式，206
　　部分列表，205
　斯皮尔曼简捷测度，206
　　完全列表的未加权公式，206
　　完全列表的加权公式，206
　　部分列表的加权公式，207～211
两个排名列表间的定性偏差，201
两个排名列表之间的偏差，201
邻接矩阵，16，93
路易斯安那州立大学，200
路易斯维尔，200
罗杰斯，克莱尔，xv
罗斯福，西奥多，112
逻辑斯谛参数，56
逻辑斯谛函数，54
旅行球迷，50
旅行商问题，104

马尔可夫，A·A，67
马尔可夫链和 LRMC，224
马尔可夫链和偏好评分，130
马尔可夫链和重心评分，132
马尔可夫投票矩阵，99
码数，71
麦康奈尔，约翰，xiv，151
麦克弗莱，马蒂，48
麦克尼尔，查尔斯·K，114
锚文本，163
冒泡排序法，204
玫瑰碗，19
梅西，肯尼斯，xv，9
梅西的网站，9
梅西法和 LRMC，224
梅西化的科利法，25

梅耶，贝萨妮·B，v
《美国新闻》，77
《美国新闻》高校排名，6
美国大学排名，77
美国国际象棋联合会（USCF），53
美国国家科学基金会，179
美国科学技术奖章委员会，xvi
美国司法部，19
米勒，约翰·J，112
幂法，38
敏感性，155～158
　克里夫兰对纽约巨人，156
模拟比赛数据，167
　性质，170
莫兰，帕特里克，xv
莫雷，达里尔，46
莫里斯，赖恩·K，xvi

纳德，拉尔夫，160
纽曼，M·E·J，224
诺普，保罗，89
女孩，评分与排名，64

帕克，赖恩，xv，153，219
帕累托法则，4，160，162，171，233
排列，98，233
排名，定义，5
排名差距法，98～107
　高级模型，105
　算法，106
　记法，106
　最优化，101
　范例，99
　总结，106
排名差距矩阵，98
排名聚合，136，159～182，201，233
　定义，159
　方法
　　模拟比赛数据，167
　方法，163～176
　　平均排名法，166
　　波达计数法，165
　　图论方法，172
　　模拟比赛数据，170
　多最优解，187
　OD 法，159
通过优化的方法，183～199
　　算法，193，196
　　记法，193
　　性质，193
　　总结，193
　求精，175
　总结，181
排名聚合后进一步求精，175
排名列表，233
佩丁斯，凯思琳，xiv，xvi，104，110，153，214

佩奇，拉里，51
偏好列表，3，233
偏好评分，127
偏好图，130
偏好函数，31
平局，26，135～145，166，205，220
平局，在 NFL 和 NHL 中最多的数量，146
平均排名，166，231
平均优势，223
评分，定义，5
评分差距法，107～112，194
　算法，111
　基本形，108
　坡形，108
　记法，111
　最优化，110
　性质，111
　范例，109
　总结，111
评分差距矩阵，107，118
评分聚合，176～182
　总结，181
评分聚合方法，211
评分列表，234
评分体系
　聚合，176
　重心法，120
　科利法，21
　由分差导出，120

直接比较，129
埃洛法，53
HITS 法，92
基纳法，29
马尔可夫法，67，130
梅西法，9
攻防法（OD 法），79
PageRank，14，51，77
排名差距法，98，106
用户偏好，127
加权，147
坡形，108，194
朴洙勇，224
朴-纽曼法，224
普希金，A·S，67

前 k 位列表，204，234
前四，最低种子排名，200
墙头草球迷，68，73，75
乔列斯基分解，23
乔治 W. 布什，160，163
乔治城大学，200
乔治梅森大学，151
权威，16，92
权威评分，93，231
全国冠军赛，BCS，18

扰动分析，139，155
人类发展指数，6
人所给出的列表，181

萨蒂，T. L.，51
萨蒂，托马斯，223
萨加林，杰夫，17，53
NFL 2009—2010 评分，122
萨维林，爱德华多，64
赛程强度，22
三角不等式，201
三月疯狂锦标赛，xiv，181，197，202
三月疯狂锦标赛，竞猜表数量，154
山本洋右，xvi，153，183
伤亡，大学橄榄球，112
上行弧，185，190
圣母大学规则，18
胜负分差评分，124
胜率，21，234
时间加权，147～153
科利法，150
差距法，151
埃洛体系，150
指数，148
基纳法，150
线性，148
对数，148
马尔可夫法，150
梅西法，149
OD 法，151
阶跃函数，148
实力
基纳法中的绝对实力，33
基纳法中的相对实力，33
世界国际象棋联合会（FIDE），53
事后胜负预测准确率，123
事后预测，48，142
事前预测，48，142
守恒性，24
枢纽，16，92
枢纽评分，92，232
数线图表示，11，202
摔跤，162
双重随机矩阵，88
司法部，19
斯蒂芬森，科林，xiv，xvi，151，213
斯皮尔曼，查尔斯，201
斯皮尔曼简捷测度，206
完全列表的未加权公式，206
完全列表的加权公式，206
部分列表的加权公式，207
搜集数据，218
素性
强迫满足，39
基纳法中的，36
算术平均值，220
随机的，68，73
随机矩阵，88
随机上网者，77

随机游走，69
索科尔，乔尔，224

《泰坦尼克》，28
坦能，比夫，48
糖碗，19
体育博彩，96
跳转矩阵，73
投票，67
凸组合，73
图的同构性，107

瓦尔尼，克里斯汀，19
完全排名列表，166，204，232
碗系列锦标赛，6，9，231
网页，谷歌，雅虎，Bing，182
网页流量，14
威尔逊，大卫，225
韦塞尔，查尔斯·D，xv
韦斯，查理，19
维格纳，塞巴斯提亚诺，52
维基百科，93
维拉诺瓦大学，200
未曾落败的队伍，68，71
魏，T. H，51
温斯顿，韦恩，45
稳态向量，68，69
沃尔夫，彼得，17
无关选择独立性，4，160，170，232
无偏，23～25，162，231
无限制域，4，160，234

夏提埃，蒂莫西，xv
线性规划，186
相关系数，47
相互强化评分体系，93
项目申请评分，171，179
斜对称，220
谢尔曼-莫里森，139，155
辛克霍恩，理查德，89
辛克霍恩-诺普定理，89
星级评分商品，128
休斯敦火箭队，46

悬挂结点，68，232

雅虎，93
雅虎，网页数量，182
亚马逊，135，140，220
《叶甫盖尼·奥涅金》，67
一致，204，205
一致对，232
一致性，183
伊斯特伍德，克林特，222
异常值，211
隐私权，221
英格拉姆，卢克，xiv，67，83，149，151，165
影片，总票房最高，28
优势图，184，190
尤依，拉蒙，166
尤依计数，166
尤依胜者，166
诱导平局，136，142
预测得分，142
元搜索引擎，159，163，173

在排列空间中的最优化，99
扎克伯格，马克，64
詹姆斯，比尔，45
正规方程，10，47
直接对比评分，129
秩一更新，139，155
种子队，获胜者的最低排名，200
种子队，进入前四的最低排名，200
重心法，231
重心评分，120
重心评分和马尔可夫链，132
主场优势，121
主特征向量，68，69
筑波大学，110
追溯得分，142，234
总绝对分差误差，121
总票房最高的影片，28
足球，142
足球，埃洛体系，53，56
最小二乘法，9，47