

近期还将推出以下书籍，届时可以在 www.verycd.com 或 <http://ishare.iask.sina.com.cn/> 上找到，也可以加入科学心理学与书籍交流群，群号 42506864，或到 <http://blog.sina.com.cn/u/1408815407> 查阅更新。

1.进化心理学 by 巴斯 2. 社会生物学:新的综合 by wilson 3.谁会认错？我们为什么会有如此愚蠢的信念错误的决定和让别人受伤的行为 4.社会心理学-阿伦森版 5. 布鲁克林有棵树 6. 学习的艺术 维茨金 著 7. 疯狂实验史 8. 哪来的天才 9.为什么选错的总是我 10.思维改变生活 11. 积极心理学：关于人类幸福和力量的科学 12.欲望之源 13. 吃的真相 14. 直觉——你所不知的潜力与危害 by 迈尔斯 15.进化的大脑 16. 态度改变与社会影响 by 津巴多 17.害羞心理学 by 津巴多 18. 简捷启发式 19. 学习乐观 by 塞利格曼 (现有的版本是超星的，不太清晰) 20. 认知心理学 第五版 艾森克著 21. 实验心理学：通过实例入门（第七版） 22. 生命的心流 23.幸福的真意 24. 别太苛求：摆脱完美主义的束缚 25. 批判性思维:思维、写作、沟通、应变、解决问题的根本技巧 26. 为什么大猩猩比专家高明(How we decide) 27. 心理和脑 脑与心智历程 100 项 28. 心理和脑与生活 训练脑与心智的 75 项窍门 29. 结构方程模型: AMOS 的操作与应用 30.颠倒思维 31.动机心理学 皮特里著 32.决策的艺术 33. 改变: 问题形成和解决的原则 34.象与骑象人 35.心理学研究方法 by 舒华 36.美丽圣经 by 宝拉·培冈 37.撬动幸福 by 奚恺元

图片来自读秀网，我用的是简单的办法——咨询下载，但每次只能咨询 50 页，咨询时间选在晚上 22 点以后可以免受咨询不得超过 80%的限制。具体咨询下载的方法我会公布在我的博客上。

快捷的办法是直接从掌握读秀、超星漏洞的那些人手里购买，价格基本在 1 元/100 页左右（taobao 上搜索读秀或超星 咨询 下载即可，可以讲价的），或者在读书宝库、诺贝尔学术网等论坛上通过论坛币来求助。

以上书籍之外，再推荐些绝对值得一看的书籍：

心理学类:社会性动物 少有人走的路 社会心理学(迈尔斯版) 决策与判断 与众不同的心理学 亲密关系 影响力 20 世纪最伟大的心理学实验 人类的性存在

其他类:如何阅读一本书 学会提问:批判性思维指南 把时间当做朋友 神奇的睡眠 中国健康调查报告，别让不懂营养学的医生害了你 异类 营养圣经-最佳营养学指南 外语学习的真实方法及误区

Fateholder 2010-10-25

QQ 群见上面，博客 <http://blog.sina.com.cn/fateholder>
豆瓣 <http://www.douban.com/people/fateholder/>

哥德·吉戈伦尔
彼得·M·托德 著
及ABC研究组

C934
J17

刘永芳 译
杨治良 审定

简捷启发式

——让我们更**精**明



A1024638

14 METHODS
THAT MAKE US
SMART

华东师范大学出版社

新加坡
PDG

图书在版编目(CIP)数据

简捷启发式——让我们更精明/[德]哥德·吉戈伦尔,[德]彼得·M·托德著;刘永芳译;杨治良审定.上海:华东师范大学出版社,2002.9

ISBN 7-5617-3054-3

I. 简… II. ①哥…彼…②刘… III. 决策-方法 IV. C934

中国版本图书馆 CIP 数据核字(2002)第 066497 号

简捷启发式

——让我们更精明

编 著 哥德·吉戈伦尔
彼得·M·托德及 ABC 研究组
译 者 刘永芳
审 定 杨治良
责任编辑 彭呈军
责任校对 邱红穗
封面设计 卢晓红
版式设计 蒋 克

出版发行 华东师范大学出版社
市场部 电话 021-62865537
传真 021-62860410

<http://www.ecnupress.com.cn>

社 址 上海市中山北路 3663 号
邮编 200062

印 刷 者 华东师范大学印刷厂
开 本 890×1240 32 开
印 张 16.75
字 数 481 千字
版 次 2002 年 9 月第一版
印 次 2002 年 9 月第一次
印 数 5 100
书 号 ISBN 7-5617-3054-3 /B·167
定 价 25.00 元

出 版 人 朱杰人

(如发现本版图书有印订质量问题,请寄回本社市场部调换或电话 021-62865537 联系)



目录

Mu Lu

THAT MAKE US SMART

1	ABC 研究组及其成员
第一部分 研究的总体框架	
3	第 1 章 快速节俭启发式——适应工具箱
第二部分 基于无知的决策制定	
47	第 2 章 再认启发式——无知如何让我们精明
74	第 3 章 无知能在股市赢利吗
第三部分 单一理由决策制定	
93	第 4 章 为一个好理由下赌注——“采纳最佳”启发式
120	第 5 章 简捷启发式有多好
148	第 6 章 为什么单一理由决策制定能够生效 ——生态理性的个案研究
175	第 7 章 人们何时采用简捷启发式以及如何识别
205	第 8 章 快速节俭启发式的贝叶斯基准
第四部分 超越选择：记忆、估计和分类	
233	第 9 章 事后通偏向——值得为快速节俭记忆付出的代价
255	第 10 章 快速估计——让环境来做这项工作
287	第 11 章 排除归类法——用少量线索做选择

第五部分 社会智能

- 315 第 12 章 动作如何揭示意图——社会互动归类
- 350 第 13 章 从骄傲、偏见到规劝
——寻找配偶过程中的“满意性”
- 378 第 14 章 借助于简单决策规则的父母投资

第六部分 总结、回顾与展望

- 399 第 15 章 人工智能、行为生态学及经济学领域的
全能神灵与启发式
- 437 第 16 章 迄今我们知道了什么
- 449 参考文献
- 483 人名索引
- 493 主题索引



译者序

Yi Zhe Xu

THAT MAKE US
SMART

译《简捷启发式——让我们更精明》一书是我在德国柏林的马克斯·普朗克人类发展研究所(Max Planck Institute for Human Development, 简称 MPI)工作期间完成的主要任务之一。翻译此书的过程,正如本书作者在其《序》中所说,仿佛经历了一次艰难而有趣的旅行!说它艰难,是因为该书是“适应行为与认知”(Adaptive Behavior and Cognition, 简称 ABC)研究组历经三年时间集体协作完成的一项跨学科研究方案的总结,其内容涉及心理学、进化生物学、计算机科学(包括人工智能)、经济学、统计学和数学等众多学科领域,各章作者的行文风格和措辞习惯各异。翻译过程中,我必须调用所有知识储备,力求准确理解自己并非十分熟悉的问题,同时必须适应不同作者的表达习惯,以形成连贯一致的风格。好在与 ABC 研究组的同事们朝夕相处,可以随时与他们讨论有关问题,这大大易化了我的工作!说它有趣,是因为本书提出的观点极富创意和挑战性,用来论证其观点的实验资料 and 实际例证也十分生动有趣,翻译过程中我常常情不自禁地被一种观点、一种方法,甚至一种妙不可言的隐喻所深深吸引!

对待理性的观点和态度是衡量科学家们(尤其是人文和社会科学家)理论倾向的重要尺度之一。从哲学、经济学(包括管理科学)等传统学科,到心理学、计算机科学等新兴学科,学者们无不因为在这个问题上的观点不同而被归入不同派别。单就心理学领域来说,从行为主义范式的产生和衰落,到现代认知心理学范式的兴起,无不贯穿着反叛理性和回归理性之间的斗争。弗洛伊德的精神分析理论不啻于对心理学家们坚守已久的传统观念的致命一

让我们更精明

击,其影响波及到了心理学之外的众多学科领域,但其极端的非理性主义色彩又总是令人望而却步!可以看到,心理学家们并非不愿意在理性和非理性两个极端之间取得平衡,但却始终未能找到理解和看待理性的第三条路线。笔者认为本书朝这个方向迈出了坚实的一步!

在本书中,作者开宗明义地指出,人类和动物的理性是有限的(既不是非理性的,又不是纯理性的),但这种有限理性已足以使人类和动物在现实环境中作出合理判断和决策。现实环境并不苛求人类和动物,也就是说并不要求人类和动物时时处处都作出最优化选择和决策,所以任何人都不必为自己理性资源的有限而忐忑不安。相反,那种奢望通过无限理性实现最优化目标的理想主义者反而是不合时宜的!这并不是说本书作者希望无限制地降低人类理性的目标,将人类还原到低等动物的理性水平。他们对理性的最低要求是:能够与现实环境(包括自然和社会环境)的要求相匹配!当理性能够做到这一点时,从生态学角度看就已经足够了。这种理性被称为“生态理性”。一个重要的假设是:有机体是否有理性或其作出的判断和决策是否合理,应该用现实的外在标准来判断,而不是用唯智论者所推崇的不切实际的理想标准来判断。不存在与现实环境相脱离的不着边际的抽象理性,如果有,它对有机体适应现实环境也是毫无价值的。只有当与环境的现实要求结合起来考察人类和动物理性时,才能找到评判理性的合理标准,这个标准就是“生态合理性”。

提出一种观点是容易的,展开和证明一种观点却是困难的。难能可贵的是,ABC研究组从上述关于人类理性的基本观点出发,形成了连贯严密的研究方案,并从诸多学科领域为其观点累积了丰富证据。既然只有当理性被付诸实际使用时,才具有适应价值,才能提供评判它的机会,那么本书作者首先从两个方面对有机体需要使用理性资源加以解决的现实问题作了分类:首先,他们按照有机体通常面对的认知问题将有关任务分为选择、记忆、数量

估计、归类等类别；其次，他们按照现实生活中有机体面对的现实适应性问题将有关任务分为食物选择、配偶选择、父母投资以及从动作推断意图等类别。如果说人类和动物的理性是有限的，那么当面临上述诸多现实任务时，他们就应该且必须采用简捷而“精明”的方法来配置和使用其资源；如果说人类和动物的理性具有生态合理性的话，那么也只有在他们“精明地”配置和使用其有限理性资源时，才能够做到这一点。所以，“有限理性”和“生态理性”这两个术语成了 ABC 研究方案的中心概念，它们携手使得有机体在现实环境中采用“快速节俭的”简捷启发式成为必要。有机体是否采用简捷启发式呢？他们可能采用哪些简捷启发式呢？他们何时采用简捷启发式呢？简捷启发式在现实环境中的成效如何呢？弄清楚了这些问题，关于有限理性和生态理性的命题也就得到了检验。正由于如此，对这些问题的回答构成了本书的主要内容。

值得强调的是，启发式的概念并不新颖。在当代认知心理学先驱赫伯特·西蒙的有关著作中它甚至是一个最基本的概念。本书的创新之处在于一方面拓展了西蒙关于有限理性的思想，另一方面创造性地提出了许多行之有效的具体启发式规则。大体上说，这些启发式规则可以分为四类：基于无知的决策规则（如，再认启发式）、单一理由决策规则（如，采纳最佳启发式、采纳最近启发式、最少化启发式）、排除规则（如，排除归类法）以及满意性规则（如，抱负水准终止规则）。其中最著名的是“再认启发式”和“采纳最佳启发式”，前者仅根据能否再认有关对象作出选择和决策，后者仅将判断和决策建立在单一理由基础之上。虽然这些简捷规则分别适用于不同问题和任务，但它们有一个共同特点，那就是：能够使有机体快速、节俭地作出推断和决策。本书报告的有关研究分别采用实验设计、计算机模拟、数学分析等手段证明：尽管是快速的、节俭的、毋需审慎权衡的，但上述多种简捷规则在现实环境中取得的成效一点也不比复杂多元回归和贝叶斯模型的成效差！正是由于本书在判断、推理和决策制定领域提出了一些全新的思

想,将关于有限理性的研究推进了一大步,拓宽了研究的范围和内容,所以自从 1999 年出版以来,已经在相关领域引起了巨大反响。有限理性之父、诺贝尔奖获得者赫伯特·西蒙指出:“《简捷启发式——让我们更精明》一书对认知科学中有限理性革命的分析 and 讨论是令人着迷的,不啻于对传统无限理性观的致命一击。”另一位诺贝尔奖获得者莱因哈德·泽尔腾也指出:“本书是对有限理性理论的重大贡献之一。它表明快速简捷方法取得令人惊奇成效的原因在于它们拟合了环境结构。对此种‘生态理性’的强调是一种巨大的进步,开辟了一个富有前景的新的研究方向。”

翻译此书使我得以了解包括心理学在内的许多学科领域的最新发展趋势和前沿性研究课题。在 MPI 工作期间我还作为 ABC 研究组的一名正式成员参与了有关的研究活动,掌握了有关的研究思路和技术路线,亲眼目睹了世界上一流研究机构的运作规范和模式。这一切对我来说都是极其宝贵的!

值得指出的是,本书英文版中的个别专有名词(如,个别公司名、动物名、酒的品牌等)和缩略语在《远东英汉大辞典》之类的大型词典中也找不到其对应的中文名。鉴于不同的人对此类专有名词可能有不同译法,加之保留原文不影响理解有关内容,所以译者未对此类术语加以翻译。

在本书翻译和出版过程中得到了许多人的支持和帮助!哥德·吉戈伦尔教授自始至终对本书的翻译工作给予了极大关心和支持,不仅为我提供了访问学者的工作位置,还提供了完成翻译工作所需要的一切便利条件!彼得·M·托德博士随时接受我的咨询,耐心解释有关疑难问题,本书其他作者也给了我许多帮助,这使我得以将犯错误的可能性降至最低!在译稿的编辑和后期技术处理阶段,彼得·M·托德博士多方联络,安排有关人士帮助我排除技术上的困难!Yvonne Misun 和 Ulrich Kuhnert 两位先生为我修订了所有插图,使我能够非常容易地将其制成中文格式!Andreas Wilke 帮助我扫描并修订了“人名索引”和“主题索引”,免

序

Xu

HEURISTICS THAT MAKE US SMART

本书邀请您到一个未知领域去做一番观光旅游。这次旅行将涉险进入一个全新的人类理性王国，它不同于传统认知科学和经济学向我们展示的那个人人皆知的理性王国，在那里人们生活拥有无限时间和无穷知识的理想世界里，耀眼的理性之光射下逻辑和概率的光束。相反，我们即将出发去探索的这个人类理性新大陆布满了杂草荆棘，笼罩着暗淡的不确定性的光线。在这个世界里，人们只有极其有限的时间、知识和运算能力，不得不利用这些有限资源对发生在周围世界某个角落里的预想不到的事件进行推断。

在视野受限，时间紧迫，而又无现成经验可供利用的情况下，一个人怎样才能确保自己决策制定的合理性？在本书中，我们指出人们可以运用一些快速节俭启发式做到这一点，其推断机制是简捷的，却是精明的。这些简捷规则为人们适应周围环境装备了一个功能齐全的工具箱，在这里，传统的逻辑和概率规则派不上用场。这些简捷启发式所以有效，原因在于从生态学角度看它们是合理的，即适合于它们应用于其中的环境的信息结构。

我们是作为一个群体开始这次旅行的，这个群体由来自通常毫无联系的多种领域的彼此迥异的人所组成，这些领域包括心理学、数学、计算机科学、经济学和进化生物学。共同的爱好把我们带到了一条船上，在这里大家不得不相互交流，一起工作，彼此学习对方的语言和技能。这条船能够顺利地航行，得益于马克斯·普朗克协会(Max Planck Society)长期的经费资助，它允许我们确定一个延续多年的远景目标，去实施一项没有确定结果的富有挑

让我们更精明

战性的研究方案。本书所描述的多数研究是在慕尼黑的马克斯·普朗克心理学研究所(Max Planck Institute for Psychological Research)完成的,而对以前未曾涉足过的更加广阔的生态理性领域的进一步探索则完成于我们的新家——柏林的马克斯·普朗克人类发展研究所(Max Planck Institute for Human Development)。我们的航行才刚刚经历了不到三年时间,但是我们已经看到了一条新的地平线,并在启发式领域获得了一些新发现,这使得我们顺利地赢得了对该领域的优先报道权。

我们研究组的名称——ABC 研究组——是“适应行为与认知中心”(The Center for Adaptive Behavior and Cognition)的缩写。它还具有对本书主题更加重要的第二层含义:我们研究的是决策制定启发式的“ABCs”,即这些推断机制得以形成的基本构筑块。

本书各章在一定程度上叙说了我们探险的故事,这些故事反映了我们在长途的艰难跋涉过程中所体验到的情感和所经历的激动人心的时刻。在漫长而痛苦的困惑不解和失败的推测之后,我们常常为多种简捷启发式的令人惊奇的作用以及隐含在其背后的原因而踌躇迷茫。我们甚至开始怀疑我们研究结果的真实性(我们的许多同事很快就产生了这种思想情绪,而且比我们放弃得要晚!),只是到了最后才认识到我们有充分理由相信这些发现是千真万确的。为了反映我们从迷惘怀疑到确信不疑的进步过程,本书前面的有关章节将报告我们当初的一些令人迷惑不解的发现,而后面的有关章节将把这些零散发现综合在一起形成一幅完整画面。

本书意欲激励和鼓舞那些热爱和敢于探索的智力探险家,而不钟情于那些追求四平八稳和困守既得学问的庸碌之士。这并不是说我们仅仅提供了一些大胆的、不明智的结果——我们已经整理了来自实验的、现实情景模拟的大量数据,形成了可靠的证据库。不过我们也的确进行了一些大胆的探索,如将我们的启发式规则之一运用于股市,看是否能够赚到钱。

我们涉猎了生态理性这个全新领域的几个不同的方面,从股票市场的混沌无序到选择伴侣的错综复杂。但最重要的是我们的指导方针,它为我们指明了前进的大方向。也许我们仅仅看到了这个广袤领域的区区一角,也许我们把一些启发式规则错误地标在了我们的航行图上,但是我们坚信快速节俭启发式王国是我们航行的正确方向。

本书各章均是由多名作者共同完成的,反映了这些研究者之间精诚合作的态度,他们形成了一个探险小分队。即使在我们这个跨学科研究组刚刚组建的那些日子里,完全没有分歧的合作也是很少见的东西。我们的成功部分来源于我们的研究者所持有的这样一种态度:在最初的迷惑和误解(我能从数学家那里学到什么?)之后,对知识的好奇心和相互间的信任——一个人可能与我产生争论和分歧,但仍然受到我的尊重,并成为我的朋友——最终将会占据上风。同样重要的是,我们营造了一种上述态度得以形成和发展的环境:我们每个人的办公室紧密相连,办公室的门敞开着,随时欢迎同伴的到来,而且每天下午4点钟是大家一起喝咖啡和饮茶的时间。这使我们的旅行既充实又愉快。

对新领域的任何一种探索都必须建立在前人过去探索的基础之上,并借鉴当代有关专家的知识和经验。多纳德·诺曼的《让我们精明之事》一书,无论其书名还是其思想,都对我们产生了重要影响。感谢阅读了本书个别章节草稿并提出宝贵反馈意见的下列同事:格列高里·阿瑟伯、彼德·艾登、塔利亚·本·齐夫、吉姆·巴特曼、格雷戈·布拉克、阿恩德特·布鲁德、艾得·巴科萨、尼克·查托、埃德加·埃德费尔德、克劳斯·费尔德勒、布鲁诺·弗雷、威廉姆·M·古德斯腾、卡尔·格拉姆、尼格·哈维、里德·哈斯贴、沃尔夫冈·海尔、奥斯瓦德·休伯、赫尔默特·荣格门、彼德·加西林、蒂姆斯·凯特拉、杰克·肯特齐、阿瑟·考瑞特、简·兰卡斯多、帕特·昂利、巴巴拉·梅勒、里查德·尼斯伯特、里查德·奥森、约翰·帕恩、斯托德·鲁塞尔、彼德·塞多梅

简捷启发式

尔、绍尔·西格瓦塔桑、汤姆·斯蒂瓦特、里查德·彻勒、瑞安·托尼、吉姆·威森特、汤姆·韦尔斯腾、X·T·王、艾克·韦伯、昆·温夫特。

特别感谢阅读了本书全部手稿的瓦勒瑞·M·查尔斯、艾克·科兹、卡纯·路德和威廉姆·威姆萨特,以及帮助我们搜集资料和编辑手稿的唐纳·亚历山大、安德鲁斯·德塔斯、马丁·笛尔伦格、蒂莫·科勒、尤里齐·库内特、图尔斯腾·莫尔巴奇、玛丽安娜·穆勒-布雷托、布拉第·里查德斯、罗迪格·斯帕、安尼塔·托德、安·维沃和吉尔·怀斯。

本书是一项处于不断发展中的研究方案的起点,关于这项研究方案的进一步发展及其结果,请您访问我们中心的网址: <http://www.mpib-berlin.mpg.de/abc/>。

欢迎您加入我们的旅行!

哥德·吉戈伦尔
彼得·M·托德
柏林



去了重新输入它们的麻烦！我指导的研究生胡琳丽、杜秀芳、李艳红、孙江涛、高鑫、付秋芳、王大伟帮助我初步翻译了有关章节，使我的工作更加容易，并提前完成了全部书稿的翻译任务！杨丽霞博士和唐洪女士先后从头至尾认真校对了我的译稿，X. T. Wang 博士也审校了部分译稿，他们均提出了一些很有价值的建议！我的前博士导师杨治良先生在百忙之中抽出时间，最终审定了我的译稿！华东师范大学出版社的彭呈军编辑和朱杰人社长对本书的出版给予了大力支持，在此谨向上述两位先生表示衷心感谢！

由于译者水平有限，译稿中的缺点、错误在所难免，恳请专家、学者和广大读者批评指正！

刘永芳

2001 年 11 月 26 日 于柏林



ABC 研究组及其成员

适应行为与认知中心(The Center for Adaptive Behavior and Cognition,简称 ABC)是组建于 1995 年的一个跨学科研究群体,主要研究有限理性的相关心理学问题,以及在变化莫测的现实世界里怎样才能作出有效决策。

Patricia M. Berretty
Department of Psychology
University of Illinois
603 E. Daniel Street
Champaign, IL 61820, USA
email: pberrett@s. psych. uiuc. edu

Philip W. Blythe
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: blythe@mpib-berlin. mpg. de

Jean Czerlinski
Department of sociology
University of Chicago
Social science 306—310
1126 East 59th Street
Chicago, IL 60637, USA
email: jfczerli@midway. uchicago. edu

Jennifer Nerissa Davis
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: davis@mpib-berlin. mpg. de

Bernhard Borges
c/o Coopers and Lybrand
830 Post Road East
Westport, CT 06880, USA
email: borges@ibm. net

Seth Bullock
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: bullock@mpib-berlin. mpg. de

Ralph Hertwig
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: hertwig@mpib-berlin. mpg. de

Ulrich Hoffrage
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: hoffrage@mpib-berlin. mpg. de

Kathryn Blackmond Laskey
Department of Systems Engineering

Gerd Gigerenzer
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: gigerenzer@mpib-berlin.mpg.de

Daniel G. Goldstein
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195, Berlin, Germany
email: goldstein@mpib-berlin.mpg.de

Adam S. Goodie
Department of Psychology
University of Georgia
Athens, GA 30602-3013, USA
email: goodie@egon.psy.uga.edu

Andreas Ortmann
Department of Economics
Bowdoin College
Brunswick, ME 04011, USA
email: aortmann@bowdoin.edu

Jorg Rieskamp
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: rieskamp@mpib-berlin.mpg.de

George Mason university
Fairfax, VA 22030-4444, USA
email: klaskey@gmu.edu

Laura Martignon
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: martignon@mpib-berlin.mpg.de

Geoffrey F. Miller
Center for Economic learning and Social
Evolution(ELSE)
University College London
Gower Street, London WC1E 6BT, UK
email: uctpgfm@ucl.ac.uk

Peter M. Todd
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany
email: ptodd@mpib-berlin.mpg.de

Gregory M. Werner
School of Cognitive and Computing Science
University of Sussex
Brighton, Falmer BN1 9QH, UK
email: gwerner@ix.netcom.com



第一部分 研究的总体框架

THE FACTS
THAT MAKE US
SMART

数字科学
PDG

第1章

快速节俭启发式——适应工具箱

哥德·吉戈伦尔 (Gerd Gigerenzer)

彼德·M·托德 (Peter M. Todd)

真理一向是以简明朴素的方式,而不是以令人眼花缭乱的复杂方式被发现的。虽然这个世界向我们的肉眼呈现了纷繁多样的事物,但当我们用哲学眼光来审视它的时候,其内部结构显然是相当简单的,惟其如此,它才能更好地被理解。

——艾萨克·牛顿 (Isaac Newton)

一个由于心脏病发作而感到胸闷的男人被急急忙忙送进了医院。医生需要迅速决定他是否应该被看成是一个高危病人。如果他的生命确实受到了威胁,需要给予最高级别和最细心的监护,那么他就是一个高危病人。尽管这种决策人命关天、意义重大,但医生却没有进行广泛审慎考虑的奢侈特权:他或她必须在极短时间内,使用仅有的可以利用的线索作出决定,而且每条线索至多只是病人危险程度的不确定指标之一。例如,在加利福尼亚大学的圣地亚哥医学中心,有多达 19 项这样的线索,包括血压、年龄等。心脏病病人一旦被收入医院,就要逐项对其进行检测。通常认为,作出决策的最好方法是看一看每项指标检测的结果,按照其重要性程度排出顺序,再将它们整合在一起作出最后结论,最好是能运用一些花样翻新的统计软件包来完成这项任务。

相反,请考虑一下图 1-1 所示的简单决策树。它是布里曼及其同事为了鉴别心脏病人的危险程度,仅使用三个变量设计的一

让我们更精明

简捷启发式

种简单决策模型(Breiman et al., 1993)。收缩压低于 91 的病人立即被判断为高危病人,不再需要任何其他信息,否则就要看第二条线索——年龄。62.5 岁以下患者被判断为低危病人。如果他或她的年龄超过了 62.5 岁,那么就需要更多线索(如,窦状心动过速)来鉴别高危和低危患者。然而,该决策树仅要求医生回答三个简单的“是/否”问题来作出最后决策,而不需要检测和考虑 19 项指标,这样可以保证治病救人的活动尽早展开。

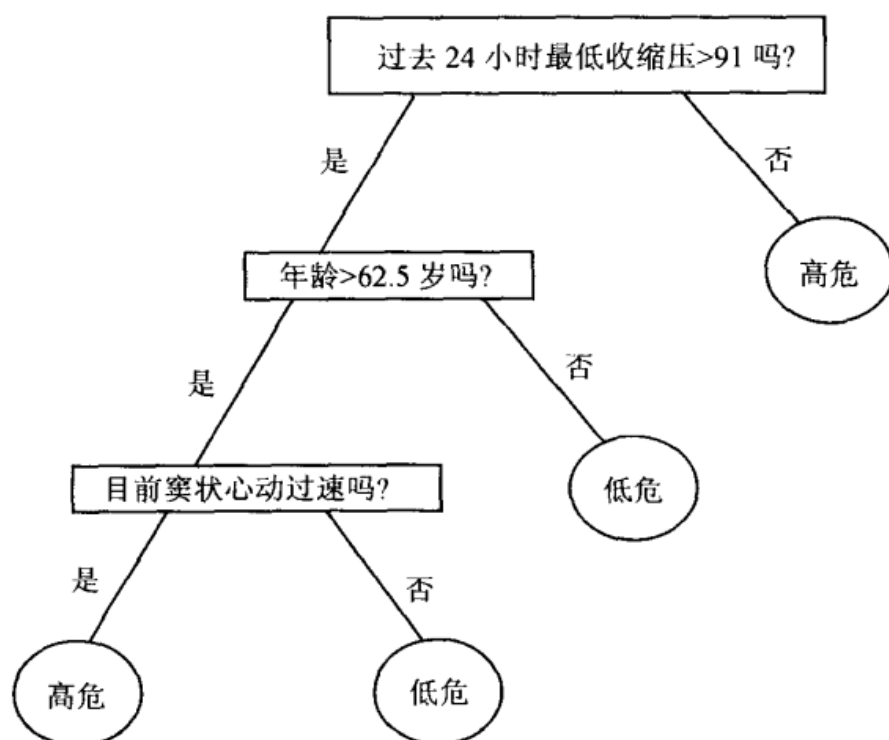


图 1-1 将入院心脏病患者归为高危或低危病人的简单决策树
(引自 Breiman et al., 1993)

这种决策制定策略在几个方面是简单的。首先,它忽略了众多可能的测量指标。其次,由于对三个问题只作简单“是/否”回答,它忽略了大量信息。例如,它不必关心病人比 62.5 岁这个分界线大或小几岁。第三,这种策略是一个逐步展开的过程。它也许在回答了第一个问题后就终止了,而不必将三种指标的值结合起来(如,加权或求和)加以考察。至多问三个“是/否”问题对于作出一种决策来说是快速而节俭的策略。它是快速的,因为它不涉及

大量计算；它是节俭的，因为它只搜索部分有用信息。它的节俭性也许会引起这样的怀疑：相对于把所有可用信息结合起来加以处理的标准统计分类方法来说，它也许是极不准确的。其实不然，实际上它在鉴别心脏病危险程度方面比复杂统计分类方法具有更高的准确性(Breiman et al., 1993)。这种与我们直觉相反的发现的更一般的形式——快速节俭决策能够像使用全部可用信息和精心计算的策略一样准确——构成了我们研究方案的基础之一。

本书是探讨决策制定过程中快速节俭启发式问题的——它们是如何运作的？什么时候和为什么它们会获得成功？这些启发式既可以被视为生命有机体的行为模式，又可以被视为人工系统的运作方式。从描述观点看，它们试图说明在时间和知识有限条件下人脑是如何作出决策的；从工程学观点看，这些启发式表明了建构人工智能系统的一些方法——这种决策制定系统不会由于缺乏大量知识和巨大运算能力而瘫痪。快速节俭启发式在上述两方面的应用并不相互排斥——的确，图 1-1 中的决策树还可以被用来描述处于无援状态的人脑的活动，或者被用来制造一种辅助的应急机器。*

理性概览：从全能神灵到有限理性

人类和动物利用有限的时间、知识和运算能力作出关于周围世界的各种推断。相反，许多理性推断模型把人脑视为仿佛一个超自然的神灵，拥有魔鬼般的推理能力、无限的知识及无穷无尽的其他用于作出决策的能量。这种关于理性的看法常常与现实相矛盾，但是我们不妨以其为参照，来帮助我们阐明自己关于生态理性的观点，这种理性是根据其与现实的拟合程度来界定的。我们通过两次观念上的革命来阐述自己的观点：第一次革命是必然性梦

* 本例中的决策树是简明易懂、易于使用的，但它的建构却建立在大量的运算基础之上。在本书中我们将会看到快速节俭启发式是如何通过规避代价高昂的运算过程被建构起来的。

让我们更精明

想的破灭和偶然性微积分原理的诞生,即概率理论的产生。这次革命在科学领域和日常生活中都以“概率革命”而著称(Gigerenzer et al., 1989; Kruger et al., 1987);第二次革命是本书首次提出并意欲着重探讨的问题,它主要关心人脑适应或应付瞬息万变的未知世界的方式。我们建议摒弃无所不知的全能人脑的想法,而代之以能量有限的人脑的观点。人不是靠计算复杂难懂的概率和反复比较外界物体的效用来适应环境的,而是利用多种快速节俭启发式装备起来的适应工具箱来适应环境的。

让我们简要回顾一下第一次革命,因为它涉及到了我们关于人脑和理性的观点。在亚里士多德之后的两千余年时间里,西方学术界在两类知识间作出了区分:一类是实证性知识,另一类是或然推断知识。前者提供了确定性知识,后者仅仅产生不确定性知识。在 16 世纪的变革与反变革过程中,确定性观点的源泉——尤其是宗教和哲学——受到了众口一致的非难。结果是,实证性思想衰落了,而或然推断的思想得到了发展(Daston, 1988)。到了 17 世纪中叶,一种放弃了传统必然性思想的新实用主义理性观产生了。它是一种“谦卑”的观点,借用了当时已经存在的概率计算方法来表明其关于人类理性的态度。这种思想的“谦卑”之处表现在它接受这样的看法:卑微的人类只能获得关于他们自身和周围世界的不确定性知识。拥有理性就意味着驾驭生活的不确定性。布莱斯·帕斯考(Blaise Pascal)著名的赌问说明了这种新理性观的一些道德后果。在一种坚信上帝必然存在的浓烈的宗教氛围中,帕斯考问道:相信上帝合乎理性吗?尽管如此,他建议一个人应该牺牲尘世的欢乐以增强来世得救的期盼,因为不管上帝存在的可能性多么小,过一种基督徒式的生活换来的都将是永恒的来生,而且这种所期盼的奖赏——其概率(也许很小)乘以其无限的价值——亦将是无限的。另外一种选择——永远罚入地狱——是无比可怕的,不管其可能性如何。由概率计算所表达的新理性观不仅仅是思想上的一场革命,而且也是道德和宗教态度上的一场革命。

概率革命形成了我们关于人脑功能的基本看法,其影响渗透到了从认知科学到经济学、再到动物行为等诸多研究领域。心理机能被假定为依据事物产生的概率及其效用而进行的运算(Gigerenzer & Murray, 1987)。照这种观点看来,概率规则为推理、判断和决策制定等过程提供了一种准确可靠的手段。心理概率的观念产生了许多独到而精致的理论,但也引起了许多棘手难办的问题。最重要的一点是,人们并非生活在心理学家和计算机科学家为他们设定的情境中,而是置身于他们实际生活于其中的现实情境中。在这种情境中概率模型所要求的时间、知识和运算能力是难以达到的。结果是,当这些理想化的推理模型遇到严峻的现实时,它们就只能使心理学更适合于解释超自然神灵的行为,而不是真正人类的行为。

在本书中,我们强调第二次革命,它提出了关于理性的一种大胆设想。我们的基本假设是:人类多数的推理和决策制定都可以借助于快速节俭启发式而建立模型,这些启发式仅使用有限时间和知识进行推断。它们不涉及大量运算,也不需要计算概率和反复权衡事物的效用。它们是一种有限理性模型。像第一次革命一样,第二次理论上的革命同样强调世界的不确定性,但无论是作为对人类行为的描述,还是作为一种切实可行的人类行为准则,它都不像第一次革命那样看重概率论。在第二次革命中彼此处于持续斗争中的四种主要理性观点见图 1-2。

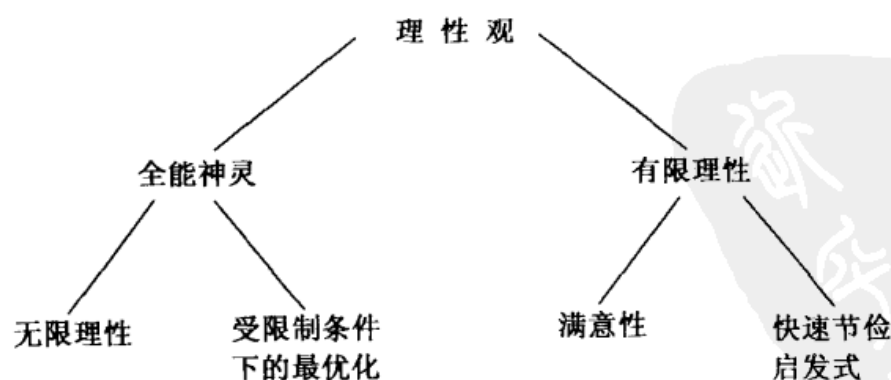


图 1-2 理性概览

理性可以通过许多方式表现出来。图 1-2 的左半部分代表这样的理性模型,它假定人类头脑从本质上说拥有不受限制的魔鬼般的或超自然的推理能量,由此假定我们实际上仅使用了很有限的理性在进行推断。该模型包含两种神灵:“无限理性”(unbounded rationality)和“受限制条件下的最优化”(optimize under constraints)。无限理性是一种忽略甚或完全无视人们所面临的时间、知识和计算能力限制的决策制定策略,传统上它是借助于概率论来建构其模型的,其最著名的表现是“预期效益最大化”(the maximization of expected utility)和“贝叶斯模型”(Bayesian models)。有限理性也包含两种形式:系列搜索可用选项的“满意性启发式”(satisficing heuristics)和利用很少信息和运算作出多种决策的“快速节俭启发式”(fast and frugal heuristics)。我们将结合现实生活中人们通常面对的决策问题,来说明这四种理性观,以表明神灵究竟潜伏于何处。

在完成了关于英国小猎犬的具有历史意义的科学探索两年之后,29 岁的查尔斯·达尔文(Charles Darwin)将他的注意力转向了个人家庭问题。他用铅笔在一张废纸上潦草地写下了如下便条,该便条分为两栏,像对称的表格一样,标题为“这是一个问题”:

结 婚

孩子——(如果幸运的话)——稳定的伴侣,(年老时的朋友)她将对人感兴趣,而不是热衷于被一只小狗或别的什么宠物所喜欢,满足于和它们嬉戏——家和一个照顾房屋的人——音乐和与一位女性聊天的惬意和满足。这些都有利于一个人的健康。被迫拜访和接

不结婚

没有孩子,(没有第二次生命)年老时无人照料……自由自在,想到哪去就到哪去——自由选择社团,不必多费思量。在俱乐部与聪明的男人聊天。——不必拜访亲属和在琐碎小事上点头哈腰——不必为孩子花费金钱和忧心忡忡——也许还有争吵。

待亲属朋友而损失大量宝贵的时间。上帝,想一想要消耗一个人整个生命,像一只无性的工蜂一样,忙碌,忙碌,到头来一无所获,实在是令人难以忍受。——试想,一个人住在伦敦的一幢乌烟瘴气、脏乱不堪的房子里,孤独无助地度过一天又一天。——反过来试想,娇好温柔的妻子坐在沙发上,炉火烧得旺旺的,也许还有书籍和音乐——请比较一下这种美景和加特·马波罗教堂里圣徒们惨淡的现实吧!

时间损失——不能够晚上读书——发胖和游手好闲——焦虑和责任——缺乏购书的钱——如果许多孩子被迫争吃一个人的面包。——(但是工作过多对于一个人的健康是十分不利的)

也许我的妻子不喜欢住在伦敦,这对我来说无异于一种流放和贬黜的判决,我一下子就会成为一个游手好闲、无所事事的傻瓜——

(Darwin, 1887/1969, pp. 232—233)

达尔文最后得出了应该结婚的结论,在第一栏下面决然写道“结婚,结婚,结婚——证毕”。达尔文还详细考虑了这个决定对他的自由可能造成的后果,最后在这张废纸背面恍然大悟地写道:“有许多幸福的奴隶!”第二年,他便娶了他的表妹埃玛·卫奇伍德(Emma Wedgwood),后来他们生了10个孩子。达尔文究竟是怎样依据他所预想到的可能后果(孩子、时间损失、稳定的伴侣)作出结婚的决定的?他并未告诉我们。但是我们可以用他的“问题”进行一次思想上的试验,以说明多种多样的理性观点。

无限理性

当达尔文还是一个刚刚5岁的小男孩时,天真、文静,没有显露出日后的才华和天赋。那时,天文学家 and 哲学家皮埃尔·西蒙·德·拉普拉斯(Pierre Simon de Laplace, 1814/1951)就已经在思考“终极天赋”(the ultimate genius)问题,他将无所不知的



简捷启发式

超级神灵描述如下：

假定有一种神灵，它包容了使大千世界变得栩栩如生、生机勃勃的所有力量，通晓其所由生成的人类的各种不同境况——它博大精深，足以使来自自然界和人类社会的所有信息得到分析……（对于它来说）没有任何事情是不确定的，未来像过去一样，都在它的掌握之中。（Laplace, 1814/1951, p. 4）

拉普拉斯是一位忠实的宗教信徒，他所设想的超级神灵是上帝的一种现世化身。约翰·洛克(John Locke, 1690/1959)曾经将无所不知的上帝和生活在“概率的微弱光线”(twilight of probability)下卑微的人类作了鲜明对照，拉普拉斯则用他虚构的超级神灵将这种对立世俗化了。从上帝和拉普拉斯的超级神灵的角度来看，自然界是一成不变的、确定的，但是对于人类来说自然界是无常的、不确定的。凡人不可能通晓这个世界，必须依靠不确定的推断和打赌，而不是依靠确实的证据。尽管对于任何真实的系统而言，无所不知和确定无疑都是难以企及的，但是拉普拉斯超级神灵的阴魂始终飘荡在无限理性王国的上空！

我们不妨设想一下，达尔文在解决他的个人问题的时候，试图达到一种最佳的主观预期效益。为了算出结婚的个人预期效益，他将不得不确定婚姻可能带来的所有后果（如，孩子、稳定伴侣以及未包含在他的便条中的其他一系列的可能后果），把量化的概率附加到每一种后果上，估计出每种后果的主观效益，用每一种后果出现的概率乘以它的主观效益，再将所有计算出来的值加在一起。同样地，为了算出不结婚的主观预期效益，他不得不重复上述步骤。最后，他还必须根据两种选择（结婚或不结婚）各自的主观预期效益总量，在它们之间作出取舍。为了获得有关各种后果及其出现概率和预期效益的可靠信息，达尔文也许不得不投入多年时

间进行研究——这些时间本来是可以用来研究甲壳动物和撰写《物种起源》的。

无限理性是一头不可思议、贪得无厌的怪兽。一方面,用这种方法看待理性的研究者们承认上帝或拉普拉斯的超级神灵与凡夫俗子之间的差异。人类必须隔着不确定性面纱作出推断,而上帝则清楚地看到了一切;人类思想的传播是或然的,而上帝所说的一切都是必然的。另一方面,当被问及他们认为不确定的推断是如何进行的时候,相信无限理性的人们又总是用上帝的偶像来描绘人类。上帝和拉普拉斯的超级神灵并不担心时间、知识和运算能力的限制——它唯一的问题是缺乏超凡的确定性。在图 1-2 中,无限理性出现在“全能神灵”(demons)标题下的推理模型之列。我们是遵照其希腊文的原义(即一种神圣的超自然的神灵)来使用“demons”这个术语的,不包含任何“邪恶”的意思。

无限理性的拥护者们通常也承认他们的模型假定了一种非现实的心理能量,尽管如此,他们仍然竭力为其辩护。他们指出,人类的行动仿佛表明他们拥有无限理性似的。依照这种理解,概率规则说明的不是推理过程,而只是推理结果。对这种观点的另外一种辩解是,它充分展现了数学的优越性和便捷性。最后,一些辩护者索性说:“不要挑剔成功!”(参见 Conlisk, 1996; Selten, 1991)。

无限理性的最大弱点是,它未能说明现实生活中人们思考问题的方式。正像下面的故事表明的那样,它甚至并未说明哲学家们思考问题的方式。有一位哲学家在为作出一项决定苦苦思索:究竟是留在哥伦比亚大学,还是接受另外一所颇具竞争力的大学提供的工作职位?其他人劝他:“算出你的最佳预期效益就是了——你不总是这样做的吗?”这位哲学家略带不满地回应道:“拜托了,这可是件严肃的事情!”

由于它的非自然性,无限理性在 20 世纪下半叶遭到了多方面的抨击。然而,当一颗(无限理性的)头颅被砍掉的时候,另一颗与其相仿的头颅通常就会在原地生根发芽:这颗头颅即无限理性的

受限制条件下的最优化

思想是冒险,意味着进入未知世界。由于不可避免地植根于不确定性,借用杰罗姆·布鲁纳(Jerome Bruner)的警句来说,我们的推断活动被迫“超出给定信息”。但是情况也许比这还要富有挑战性,因为许多情况下几乎没有给定信息。相反,我们必须搜索信息——将心脏病病人归为高危病人的线索、结婚的理由、股票市场涨跌的指标,如此等等。信息搜索通常被认为是一种内部过程,依赖于个体记忆的内容。然而,需要强调的是,多数信息搜索是外在的,是对暗藏于周围环境中的信息的细心检索。这种外部搜索包括对朋友、专家等无处不在的社会性信息载体的查询,还包括对图书馆、互联网等人造信息资源的访问。

无限理性与图 1-2 中所示的其他三种理性的关键区别在于,后三种理性都涉及到了有限的信息搜索,而无限理性模型则假定搜索可以无限制地进行下去。在合乎情理的决策模型中,搜索必须受到限制,因为真实的决策制定者只能将有限的时间、知识、注意力,甚或金钱花费在特定决策活动上。受限制的搜索要求找到一种方法,即一种“终止规则”(stopping rule),以便决定何时停止对信息的搜索。我们所谓的“受限制条件下的最优化”模型假定,终止规则能够保证所投入的有限时间、运算能力、金钱以及其他资源得到最有效的利用,以获得最佳的搜索效果。更具体地说,该种理性观点坚持认为,大脑应该计算出搜索每一种信息的收益和代价。一旦代价超过了收益,立即终止搜索(例如,Anderson & Milson, 1989; Sargent, 1993; Stigler, 1961)。乍一看,这种“代价超过收益即终止”规则似乎是合情合理的,但是仔细分析起来,这种受限制条件下的最优化甚至比无限理性需要的知识和运算步骤还要多(Vriend, 1996; Winter, 1975)。

为了说明这一点,我们不妨设想有两个达尔文:一个是拥有

无限理性的达尔文,另一个是受限制条件下寻求最优化的达尔文。他们两个都试图决定是否结婚。在根据较高的主观预期效益在两种选择间作出决定之前,拥有无限理性的达尔文必须搜索结婚和不结婚可能带来的所有可能后果。现在我们假定,受限制条件下寻求最优化的达尔文已经列出了婚姻可能带来的两种后果——稳定伴侣和孩子,并且已经估计出了这两种后果各自出现的概率和效益。在他搜索第三种后果之前,他必须计算出继续搜索信息所带来的收益是否超过了它的代价——如果不是,那么他可能会就此终止搜索。为了算出进一步搜索的收益,可怜的达尔文将不得不重新考虑所有可能的第三种后果是什么,估计它们的收益和概率,计算每一种后果将会在多大程度上改变他最后的决定,将它们加以平均,以确定继续搜索的预期效益。每考虑一种有待继续搜索的后果,都不得不进行同样的运算。因此,与拥有无限理性的达尔文作出决策时所使用的知识相比,对继续搜索的效益进行计算所需要的知识并无两样。然而,这对寻求最优化的达尔文来说还不算完——他还必须确定继续搜索可能付出的代价。举例来说,继续搜索信息可能要付出失去良机的代价:寻求最优化的达尔文必须确定在他用于考虑他的决策的这段时间内,他可以做哪些其他事情(如,解剖一只动物),以及专注于当前问题而不去做其他事情可能付出的代价是什么。到头来,这样做反而会导致巨大的退步。寻求最优化的达尔文不仅需要考虑会失去哪些良机,而且需要考虑进行代价—收益权衡本身的成本,如此等等(Conlisk, 1996)。结果是,拥有无限理性的达尔文反而能比追求最优化的达尔文更快地作出决策!

我们不妨回过头来看一看:用受限制条件下的最优化来代替无限理性,其本意是要建立一种充分考虑到人类头脑局限性的更加现实的决策模型。它的自相矛盾之处在于,在假定人类头脑具有进行代价—报偿权衡的无限时间和知识的前提下,又试图建构人类决策的有限信息搜索模型。由于受到了无限理性之光的照

简捷启发式

射,最优化梦想只有当该种模型屈尊就范于其竞争对手的情况下,才有可能实现。当解决达尔文的问题或其他现实问题的时候,受限制的最优化模型都不得不允许无限理性从后门悄悄潜入它的神圣领地。

当然,几乎没有人认为现实的人类会拥有完成大规模运算所需要的时间和知识,上面所设想的两个达尔文中的任何一个都是普通人难以企及的。相反,这些理性境界通常是作为人类推理应该追求的最高理想而提出来的。但是这种追求会使得真实的人类推理活动看起来是有缺陷的和不合理的。在我们看来,这些不切实际的追求才是有缺陷的——我们将证明不需要无限的时间和知识也能作出有力和准确的推理。

与无限理性相比,受限制条件下的最优化模型的贡献在于强调对环境结构的分析。例如,在安德森等(Anderson, 1990; Oaksford & Chater, 1994)的“理性分析构架”(rational analysis framework)中,来自环境的条件限制被用来调整人们对特定场合最优化行为的理解。这种分析并不直接表明哪些心理机制可能会产生最优化的行为模式,但是至少它允许我们设定一种对于评价这种心理机制更加现实的标准。 12

作为对上述神话般推理模型的替代,下面我们将注意力转向有限理性观。在认知科学、经济学和动物行为等研究领域,即便不是多数研究者,至少也有许多研究者将“有限理性”这个术语理解为受限制条件下最优化的同义语。这是我们强烈反对的一种错误用法。也许正是这种理解导致人们经常无视有限理性而倾心风行已久的神化模型。在作此种理解时,经济学家托马斯·萨金特(Thomas Sargent, 1993)指出,当一个人以人们理性“有限”的观点为指导建构推理模型时,他的模型中必然要使用大量参数,从数学运算的角度看这是更加繁琐的。他相信,研究者们(尤其是经济学家)之所以坚持无限理性模型,是因为他们对只需要使用少量参数的模型的渴求未能被有限理性观点所满足:“简约化并非有限理性

所能保证的。”(p. 4)然而,这是对有限理性的一种误解——理性不一定是最优的,有限不一定是受限。

有限理性: 满意性

有限理性之父赫伯特·西蒙(Herbert Simon)强烈反对将有限理性曲解为受限制的最优化模型:“有限理性并非关于环境任务最优化的研究。”(Simon, 1991, p. 35)在有关的个人谈话中,他曾经幽默而又不无愤怒地谈到,他考虑过要起诉那些误用他的有限理性概念来建立更加复杂和非现实人类决策模型的人。

西蒙的有限理性观点包含两种相互联系的成分:人类头脑的局限性和头脑在其中发挥作用的环境结构。第一种成分意味着,人类判断和决策制定模型应该建立在我们对人脑能量实际了解的基础之上,而不是建立在虚构的能量基础之上。在许多现实情境下,最优化策略都是未知的或不可知的(Simon, 1987)。即便是在下国际象棋这类确实存在最佳走法的简单游戏项目上,也没有任何策略可以在允许的时间内计算出最佳走法(不论是人脑还是计算机),尽管在此类活动上有待搜索的选项已经得到了很好的限定。在不受限定的自然情境中,我们确认一种可用的最优化策略的希望将会变得更加渺茫。由于头脑的局限性,人类“必须采用近似的方法去对待多数任务”(Simon, 1990, p. 6)。这些方法包括可最大限度地避免进一步信息搜索的再认过程、指导搜索并决定何时终止搜索的启发式规则,以及充分利用既得信息的简单决策规则。在本书中我们将详细探讨这些不同类别的方法。

西蒙有限理性观点的第二种成分(即环境结构)具有决定性意义,因为它能够解释何时和为什么简捷启发式规则会很好地起作用:只有当启发式规则的结构拟合于环境结构的时候。西蒙(1956a)的一个经典例子涉及到了两个正在觅食的有机体,它们只有一个简单的目标——食物。一个有机体生活在食物随意摆放的环境中,它可以侥幸地实施一种启发式规则,即随意地四处游荡,

13

让我们更精明

简捷启发式

直到发现食物为止。为此,该有机体需要一些观察和运动能力,但并不需要学习能力。第二个有机体生活在另外一种环境中,在那里食物不是随意摆放的,而是被有意放置于隐蔽的地方,它们的位置可以从一些线索中推断出来。该有机体能够使用一些更加复杂的策略,诸如了解线索和食物之间的关系、记住正在储存的信息等。最基本的要点是:要了解一个有机体使用哪种启发式规则,以及何时和为什么这种规则会起作用,就需要看一看该环境的信息结构。西蒙(1956a)并非唯一一个证明这一论点的人。在他的著作出版之前(例如, Brunswik, 1943)和之后(例如, Anderson, 1990; Shepard, 1990),人们在多种情况下论证了这一论点。甚至有人极端地声称只有环境需要研究,而不是脑的机制(例如, Gibson, 1979)。但是总体上看,西蒙论文的第二部分——“理性选择和环境结构”——在主流认知科学中一直未受到重视(有时甚至被西蒙本人所忽视——参见 Simon, 1987)。

我们用“生态理性”这一术语将环境结构迎回有限理性的怀抱。当一种启发式规则适应于环境的结构时,从生态学的角度看它就是合乎理性的(参见下文)。因此,简捷启发式和环境结构双双携手为最优化理想提供了一种现实选择,不论这种最优化是不受限制的还是受限制的。

有限理性的表现形式之一是西蒙提出的“满意性”(satisficing)原则——该词发源于位于苏格兰边境的英国的北阿姆布里尔(Northumbria)地区,在那里它的意思是“满足”(to satisfy)。满意性原则是指当人们对随后可能出现的目标知之不多的情况下,从相继遇到的一系列候选目标中作出选择的一种方法。在此种情况下,不存在关于何时终止进一步搜索的最佳方案——例如,一旦达尔文决定要结婚,将没有一种最佳方法可以帮助他决定什么时候可以停止寻找未来的婚姻伴侣,安心和其中的一个过日子(参见第13章有关寻找配偶的满意性启发式规则的更详细的论述)。而满意性原则走了一条捷径,它设定了一种可调节的抱负水准,一旦遇

到超过抱负水准的目标,立即终止对候选对象的搜索(Simon, 1956a, 1990)。

满意性原则在时间十分紧迫的情况下必须作出生死抉择的消防指挥员身上得到了充分体现。比方说,为了扑灭一座四层公寓楼地下室的火焰,他们不是反复比较和审视所有可能的行动措施,而似乎是随意捡取一种行动方案,迅速地在头脑中预演一遍。如果足够有效——也就是说,如果它的结果超过了预先为成功设定的抱负水准——他们就会毫不犹豫地“上演它”(act it out)。如果心理预演的结果未达到他们的抱负水准,他们就继续下一种方案,重复上述预演过程,直至找到一种满意的行动措施(Klein, 1998)。

有限理性:快速节俭启发式

满意性原则是从一套候选目标中作出抉择的方法之一,它考虑到了人类时间和知识的有限性:不像受限制条件下的最优化模型所做的那样,它无需发现和推测未来可能出现的所有选项和结果。然而,满意性原则的某些方面仍然需要决策者进行审慎细心的思考,如首先要设定一种适当的抱负水准,还要计算当前的选择怎样才能与抱负水准相匹配(Simon, 1956a)。为了杜绝非现实的心理演算潜入我们所描绘的人类理性画面,我们进一步聚焦焦点,将注意力集中于决策制定的快速节俭启发式上。

快速节俭启发式利用最低限度的时间、知识和运算能力作出现实环境中的适应性选择。它们能够通过设定目标或选项来解决系列搜索问题,就像满意性原则所做的那样。它们也能够同时在出现的适用目标之间作出选择。在此种情况下,必须受到限制的是对与目标有关的信息(如,线索、特征、后果等)的搜索,而不是对目标本身的搜索。快速节俭启发式使用易于操作的终止规则限制它们对目标或信息的搜索,并使用容易操作的决策规则作出它们的选择。因此,我们将满意性规则和快速节俭启发式规则视为有限理性的两种相互交叉重叠但又互不相同的类别。有些满意性规

简捷启发式

则是快速节俭的,有些从运算上看是不合理的;有些快速节俭启发式规则在相继出现的选项中作出决策,有些则在同时出现的选项中作出抉择。我们认为快速节俭启发式以最完美的形式代表了有限理性的含义。

达尔文怎样采用快速节俭启发式规则解决他的婚姻问题呢?他可以用少量时间和知识在两种选择——结婚或不结婚——之间作出决定的一种方法是“单一理由决策制定”(one-reason decision making)。在这种方法中他只需要找到一则信息帮助他作出抉择。实际上,就在达尔文“证毕”前的那一段里,我们可以看到,他作出结婚的决定只有一个理由,即有一个稳定的伴侣:“试想,一个人住在伦敦的一幢乌烟瘴气、脏乱不堪的房屋里,孤独无助地度过一天又一天。——反过来试想,娇好温柔的妻子坐在沙发上……”一个人所以将决策建立在单一理由而不是多种理由结合的基础之上,有一个非常重要的原因:将来自不同线索的信息结合起来需要把它们转换成一种“通用货币”(common currency),这种转换即便不是不可能的,也是代价昂贵的。例如,为了在几条线索结合起来的单一评估基础上作出决策,达尔文将不得不考虑与聪明的朋友进行多少次谈话等同于有一个孩子,以及在乌烟瘴气的住所里呆上多少个小时可以被沙发上温柔瞬间的益寿效果所抵消。标准的最优化模型,不管是受限制的还是不受限制的,均假定对所有信仰和愿望而言有一种“通用货币”,即量化的概率和效益。尽管从数学运算的角度看这是一种方便易行的假设,但我们看待世界的方式并不总是符合于它。一些东西未明码标价,不可能被转化成任何“通用货币”,并据此与其他东西进行兑换。例如,爱、真诚的友谊、军人的荣誉以及博士学位等都被视为无法估价的,因此无法与商场货架上出售的商品用共同的尺度来衡量。当许多理由不能够用单一货币进行兑换时,除了采用仅依据一个好的理由作出决定的快速节俭策略之外,别无选择。然而,正如我们将要看到的那样,不可比性并非采用“单一理由决策制定”策略的唯一原因(参见

第4—6章)。

在我们进一步深入探讨快速节俭启发式之前,让我们先来总结一下迄今所做的讨论。有限理性在许多领域都成了一个流行术语,过分庞杂的内容被包罗在该术语的范畴内,包括受限制条件下的最优化。图1-2有助于在有限理性和神化的理性观之间作出明确区分。无限理性不关心搜索的成本,而有限理性通过终止规则明确限制了搜索活动。受限制条件下的最优化模型也限制了搜索活动,但它是通过计算最佳终止点(即进一步搜索的代价超过其收益)做到这一点的。相反,有限理性敢于在不寻求最优化搜索和终止规则的简捷启发式上“下赌注”。最后,我们在快速节俭启发式规则上找到了有限理性的最完美形式,它在目标(本着满意性原则)或线索间进行有限搜索,并利用环境结构产生适应性决策。

快速节俭启发式的“ABCs”:一项新的研究方案

本书按照人们作出多数推断和决策的方式提出有限理性观。它还可以为发展人工系统的决策制定启发式提供有用的参考。这项关于有限理性的研究方案涉及到了以下几个方面:(a)设计候选简捷启发式规则的“计算模型”(computational models);(b)分析这些简捷启发式规则在其中发挥作用的环境结构;(c)测试它们在现实环境中的效果和作用;(d)确定人们是否使用这些简捷启发式规则,以及何时使用这些规则。其中(b)、(c)、(d)三个阶段的调查结果被用来充实阶段(a)提出的理论。这项研究方案的各个阶段依赖于多种方法,包括建立启发式规则的理论模型、对这些模型的运作机制进行计算机模拟、对启发式规则 and 具体环境之间的拟合程度进行数学分析以及实验室实验等。下面我们依次讨论每一个研究阶段。

计算模型

启发式规则的计算模型把产生决策所涉及的信息搜集和加工

简捷启发式

过程具体化为精确的步骤,以便它们能够被编制成计算机程序,在计算机上加以演示。对快速节俭启发式而言,这意味着计算模型必须将指导目标和信息搜索、终止搜索和作出决策的规则具体化。

指导搜索的启发式规则 决策必须在多种选项之间作出抉择,依赖于有关这些选项的信息。在许多情境下,这些选项及信息需要通过积极活动去发现。指导搜索的启发式规则,不管是对选项还是对信息的搜索,都是为了给出搜索的方向(假如有方向的话)。例如,对线索的搜索可能仅仅是随意的,或者是按照一些事先计算出来的有关它们有用性的标准依次进行的(参见第6章),还可能依赖于对先前作出同样决策时生效的线索的回忆(参见第4章)。类似地,对选项的搜索也可能是随意的或有顺序的。指导搜索的快速节俭启发式并不利用繁琐运算和广博知识去算计下一步到哪里去。

终止搜索的启发式规则 在我们的有限理性概念中,人类头脑(或任何其他现实的决策制定系统)的时限性必须像其他限制一样受到足够重视。这意味着对选项或信息的搜索必须在某一点上被终止。而且决定何时终止搜索的方法不应该是过于复杂的。例如,一种简单的终止规则是:一旦找到支持某种选项的第一条线索或理由,即停止对信息的搜索,立即作出决策(参见第4章)。这种基于线索的终止规则不需要像受限制条件下最优化模型那样算出最佳的代价—收益比率。事实上,它们根本不需要计算任何代价或收益。对于在选项间进行的搜索来说,可以采用简单的抱负水准终止规则,就像西蒙的满意性原则所做的那样(Simon, 1956a, 1990; 亦可参见第13章)。 17

作出决策的启发式规则 一旦搜索受到指导去发现某种适当的选项或信息,然后被终止了,最后一套启发式规则就会被唤起,以便根据搜索结果作出决策和推断。这些规则也可能是十分简单的和计算上非常有限的。例如,一种决策或推断可能依据唯一的线索和理由,不管搜索中找到的线索有多少(参见第2章至第6

章)。这种单一理由决策制定不需要对线索进行加权或结合,所以无需决定线索间的“通用货币”。还可以通过简单排除过程作出决策,这样做时一个个选项被后继线索所拒绝,直至仅留下一一种最终的选择(参见第 8 章至第 11 章)。

这些启发式规则是快速节俭启发式的“构筑块”(the building blocks)或“ABCs”。假定头脑是生物而不是逻辑实体,经过长期的自然生长、环境造化过程,由多种成分精制而成,那么我们似乎有理由设想新的启发式规则应该建立在老的启发式规则基础上,而不是从零做起(Pinker, 1997; Wimsatt, 出版中)。据此,我们采用两种主要方法建立快速节俭启发式的计算模型:联结构筑块和将其嵌入现有启发式。启发式规则可以用多种方法联结起来,尽管不是随意的,就像单一理由决策规则在本书中将以各种面貌出现一样。例如,对于简单的双项选择问题来说,一种快速节俭启发式规则是在造成选项差异的第一条线索出现时即终止搜索,所采用的实际上就是单一理由决策规则。所有快速节俭启发式规则均可以通过将一个嵌入另一个而联结起来。举例来说,再认启发式(参见第 2 章和第 3 章)依靠一种简单认知能力(再认)而生效,但它也能够作为吸收了其他认知能力(如,回忆)的更为复杂的启发式规则的第一步。无论从个体发生或是种族进化的角度来看,再认能力都比回忆能力发展得更早,类似地,启发式规则的嵌入可以被视为将一种新的适应性附加到一种已有适应性上。

本书中我们将快速节俭启发式作为精确的可逐步分解的模型加以详细阐述。这种模型是高度清晰和精细具体的:很容易理解和辨认它们是如何运作以作出决策的。由于他们涉及到了不多的参数和很少的计算,这种规则系统的每一个步骤都是一目了然的,容易得到审查和检验。这些简捷启发式与更加复杂和涉及更多计算的模型形成了鲜明对照。尽管后一种模型也许能够产生人类行为的“良好近似物”(good approximations),但它的运作机制常常是模糊不明的。例如,联结主义(connectionism)在 20 世纪 80 年

代的复兴产生了一大批神经网络模型,它们对说明许多心理现象来说都是相当不错的模型,但即便对这些模型的创建者来说,其潜在运作机制也是神秘莫测的。只有通过持续的努力,才能摆脱对这些“黑箱”(black-box)联结模型的迷恋,发展新方法以窥视暗藏于内的真相(Regier, 1996; Rumelhart & Todd, 1993)。与黑箱方法紧密相关的“黑色魔法”(the black magic)的诱惑力在许多领域都存在,诸如用复杂难懂的统计软件包去分析人类行为等(Boyd & Richerson, 1985; Gigerenzer, 1993)。公开透明的快速节俭启发式模型避免了人们对其所涉及过程的误解和迷惑,虽然它们的确因此而失去了作为未知事物的些许诱惑力。

生态理性

正像下一部分我们将会看到的那样,传统的理性定义主要关心内部信仰和推理的有序性。但是现实的有机体花费大量时间应付外部环境的无序性,试图作出使它们得以生存和繁衍的决策(Tooby & Cosmides, 1998)。面对环境的挑战,为了很好地适应它,有机体必须能够作出快速、节俭而准确的推断。现实世界的要求导致了一个合乎时宜的推理概念的诞生——生态理性。与特定环境结构相匹配的快速节俭启发式使得有机体从生态学角度看是有理性的。因此,生态理性研究涉及到了分析环境结构、启发式结构以及二者之间的适配性,正如本书通篇所要阐述的那样。

生态理性怎样才可能生效?也就是说,快速节俭启发式怎样才能如其所愿地起作用,而避免在诸如速度和准确性等不同现实标准间进行权衡?它们所以能够成功,主要原因在于它们仅在一个维度上进行权衡,即通用性—特殊性维度。尽管连贯性标准是普遍适用的——如,逻辑一致性适用于任何一个领域——但测量启发式规则在现实环境中成效的对应性标准却因具体领域不同而不同。在一个领域能够很好地作出快速准确推断的启发式规则,在另一个领域也许不起作用。因此,不同的环境可以有不同的快

速节俭启发式,它们利用该环境特定的信息结构作出适应性决策。但是特殊性也可能是危险的:如果每一种些微不同的决策制定环境都需要不同的启发式,那么我们将需要难以数计的启发式规则,而且无法将任何一种规则推广到先前未曾遇到过的环境中去。快速节俭启发式绕开了这个陷阱,其简明性允许它们在面对环境变化时仍然是有效的,并使它们能够很好地被推广到新的情境中去。

环境结构的利用 快速节俭启发式能够从环境中信息构成的方式上受益。例如,第10章我们将会介绍一种用于数量估计的快速节俭启发式规则,它建立在诸如城市人口规模等许多现实变量的偏态分布基础之上。这些变量构成了环境结构的一部分。传统统计估计技术通过对数据的标准化,要么忽略这些环境变量,要么有意排除这些变量。标准统计模型以及标准理性理论的宗旨是尽可能地泛化或普遍化,因此对它们将要进行分析的资料总是作出尽可能少而普遍的假设。但是现实环境中信息构成的方式常常并不成全这种方便简单的假设。例如,尽管多数统计模型都是在假定平均数和方差相互独立的前提下,对成批数据资料加以处理的,但是卡尔·皮尔逊(Karl Pearson,1897)注意到在自然情境中这两种统计指标常常是相关的,因此其中的一个可以被用作推断另一个的线索(Einhorn & Hogarth,1981,p.66)。虽然一般的统计方法力图忽略诸如此类的限制它们发挥作用的因素,但进化论却非常重视此类颇具信息量的环境从属关系,并借助于具体的启发式规则充分利用它们,如果它们赋予作出决策的有机体适应上的优势的话。

耐用性 特定领域的简捷启发式怎么会像利用了许多不确定参数的复杂通用策略一样准确呢?答案是它不过于具体。简捷启发式固然要应用于具体环境,但它们并不包含与任何一种具体环境都密切匹配的细节。另一方面,适用于广泛领域的通用策略过于精细和具体,以至于难以付诸实际应用——不得不应付大量不确定参数可能妨碍了它的使用。试想有那么一位细致入微的天气

观察者,他有大量空闲时间,所以决定用一年时间详细记录每天的气温和雨量变化情况。假如他利用这种确切的日志资料来决定下一年如何选择着装,而且仅仅由于去年的4月15日是炎热和干燥的,所以今年的4月15日就早早地穿上了短衫,那么他未免要常常挨冻和淋雨。尽管他的详细的天气资料对于描述他所建立的特殊模型而言是准确无误的,但在其他情境——其他时间和地点——其预言价值是极其有限的。这种推论活动的失败被称为“匹配过当”(overfitting)现象(例如,Geman et al., 1992; Massawo, 1988b),它导源于这样一种假设——每种细节都是至关重要的。相反,如果我们的天气观察者在他的模型中使用了较少参数,例如,只记录每周的平均气温和雨量,并利用这种记录决定随后一年怎样着装,他可能会作出更加准确的预测,因为这种资料符合年复一年天气变化的一般趋势。正像我们在本书许多章节中将会看到的那样,拥有许多不确定参数的推断模型——从多元线性回归到神经网络模型——都会遇到像这位天气观察者一样的问题。

20

因此,在同一种策略的两种常用方法之间存在着重要差异:一种用法称为“拟合”(fitting),即让一种策略适合于一套给定资料。在此种情况下,一种模型拥有的参数越多,使用的信息(线索)越多,它将越符合给定资料。另一种用法称为“推论”(generalization),即采用一种策略来预测新的资料。在这种情况下,拥有的参数越多并不一定越好。相反,仅利用少数可用信息、在计算上非常简单的策略反而可能比需要大量信息、计算上相当复杂的匹配过当策略更加耐用,并能对新资料作出更加准确的预测。

耐用性与速度、准确性,尤其是信息的节俭性密切相关。快速节俭启发式忽略许多线索固有的“噪音”(noise),而寻找最重要线索中隐藏的“潜力”(swamping force)。因此只利用一种或少数几种最有用的线索能够自然而然地产生耐用性。而且,重要线索终究是重要的,当环境发生变化时它所包含的信息量的关系保持不

变——例如，年复一年北方地区的四月份总是伴随着雷阵雨天气。相反，“噪音”的随机波动，甚或一些非系统因素的作用常常是不稳定的——例如，五月份花卉的开放依赖于许多变化因素，诸如气温、雨量、种子的疏密程度以及病虫害的情况等，这些因素每年都有很大不同。由于如此，将注意力放在系统的、有信息量的线索上而忽视更多变化无常线索的快速节俭启发式将能够摆脱环境变化的影响，始终保持良好预测效果。拉普拉斯的超级神灵从来不会匹配过当，因为他压根就不需要作出不确定预测。然而，试图像拉普拉斯的超级神灵那样建立人类推理模型注定是要匹配过当的，因为该种模型“吞下”了比它能够消化的数据多得多的资料。

研究生态理性使我们能够超越一种广为流传的虚幻设想，它认为将决策制定建立在更多信息和运算基础上将会导致更加准确的推断。这种设想主宰了许多研究，从而导致了认知过程计算模型的产生，该种模型以信息需求量相当大的统计方法为基础（Gigerenzer & Murray, 1987）。更一般地说，上述设想被普遍看成是评价认知活动优劣的标准。例如，在认知心理学和社会心理学中存在着许多“双过程”（dual process）理论，它们采用非好即坏的两分法，区分出了诸如分析性—启发式、爱争论的一易受暗示的、讲原则的一随意结交的，以及深思熟虑的一做事轻率的等双极特征（例如，Evans, 1989; Sloman, 1996）。隐藏在这些理论背后的一种不言而喻的假设是：制定一种策略付出的劳动愈多，计算起来愈复杂，包含的启发式成分愈少，由它产生的判断将会愈加准确。这种“愈多愈好”（More-is-better）的想法忽视了认知策略的生态理性问题。结果是，当人们违背这些双极论者心目中的理想标准而做得更好时，这些理论家们常常感到惊疑和困惑。例如，有时候人们依据自己的直觉能够作出比基于逻辑推理更好的判断（Wilson & Schooler, 1991），有时候增加了更多新信息反而降低了人们对股市涨落预测的准确性（Davis et al., 1994），有时候在贝叶斯推理中简单的直觉策略也能做得很好（Mckenzie, 1994；另见

Ambady & Rosenthal, 1992)。似乎有那么一个临界点,超越了它,过多的信息和信息加工可能是有害而无益的。认知活动可以被理解为集中注意相关信息而忽略其他信息的一门艺术。我们将采用这种观点来建立认知模型。 21

在现实环境中的成效

正如前面提到过的那样,有限理性常常被看成解释人类认知活动局限性的一种观点。这种理解是不全面的和包含着错误的。如果我们想要了解现实人类的头脑是如何发挥作用的,那么我们不仅要看一看相对于超自然的神灵而言我们的推理活动是如何受到限制的,而且还要看一看我们的头脑是如何适应于现实环境的。有限理性这两个方面的含义有助于我们确定一种更加现实的标准来评价启发式规则的效用。

人们通常用来评价判断和决策活动的一种标准是逻辑规则和概率理论。举例来说,如果一个人作出的判断是始终如一的(如,我始终认为事件 A 比事件 B 更可能发生)和可传递的(如,我认为事件 A 比事件 B 更可能发生,事件 B 比事件 C 更可能发生,所以事件 A 比事件 C 更可能发生),那么人们通常就会认为其潜在的决策策略是合理的。如果一个决策制定者违背了这种标准,人们就会认为他在判断上出了问题。这些逻辑和概率规则被称为“连贯性”(coherence)标准,因为它们主要关心判断的内在逻辑一致性,而不关心这种判断是如何帮助我们在现实环境中作出有效决策的。如果你相信艾尔韦斯(Elvis)有 90% 的可能性生还,有 10% 的可能性已经死去,那么你的这种信念至少是连贯的,因为你给出了两种不同概率,而且它们加在一起刚好是 100%,这正是概率论要求你去做。但是,如果你由此决定花费几个小时在玉米地里等待艾尔韦斯驾驶的宇宙飞碟着陆,那么这种信念对你适应现实环境就是有害而无益的。由于未考虑决策的实际内容和现实的适应问题,以往多数旨在说明人类和动物理性或(常常是)非理性活

动的研究均使用了抽象的连贯性标准。例如,许多人声称人类推理活动中存在着系统的非理性错误,作出这种判断的主要依据就在于人们的许多推理活动违背了逻辑和概率规则(例如,Tversky & Kaheman,1983;Wason,1983)。类似地,有人声称猴子也具有理性(Mcgonigle & Chalmers,1992),因为他们观察到松猴也能够作出符合可传递性标准的选择。

本书采取了一种关于人类理性行为的全然不同的适应性观点。我们不把人类判断过程与逻辑和概率规则进行比较,而是考察它在现实环境中的效用。启发式的作用不在于连贯性,相反它们的效用在于利用有限时间和知识作出关于现实社会和物理世界的合理的、具有适应价值的推断。因此,我们应该采取一种能够反映这种作用的新标准来评价启发式规则的效用。将决策制定策略与外部世界而不是内在连贯性联系起来的评价标准被称为“对应性”(correspondenc)标准(Hammond,1996),它主要考察决策的准确性、节俭性和速度。正如艾冈·布伦斯维克(Egon Brunswik,1957)注意到的那样,头脑和外部环境就像一对必须相互磨合以达成和谐的夫妇一样。然而,由于多数关于推理和决策制定的研究过于关心连贯性,使得这对夫妇彼此疏远了。我们的目标是让这对夫妇重归于好,哪怕他们彼此都是不连贯的。

的确,上述两种评价标准——连贯性和对应性——有时候可能是彼此相左的。举例来说,在诸如竞争性游戏项目和食肉动物与其捕捉对象相互争斗等社会情境中,为了达到最佳适应效果,避免蒙受损失和遭到捕获,一种非常有用的策略就是表现出并非始终如一的行为(Driver & Humphries,1998)。在本书的第4章和第5章中,我们还将介绍一种类似的非逻辑化的启发式规则——最少化启发式,它违背了连贯性原则的可传递性标准,但在特定环境中能够作出相当有用和准确的推断。因此,不可传递性不一定意味着不准确性,可传递性也不能确保高准确性——从逻辑学的角度看,合逻辑性和行为适应性是彼此不同的两个概念。

总之,启发式不是最优化策略的蹩脚翻版。事实上,在许多现实环境中原本就不存在最优化策略。当然这并不是说现实世界中不存在判断成绩好坏的客观标准。作为对启发式规则的评定尺度,我们将其成效与现实环境的实际要求进行对照,这些要求包括作出准确的决策、使用少量时间和信息等。因此我们用与现实世界决策成效有关的多种对应性标准代替了源于逻辑和概率规则的多种连贯性标准。然而,在这两套标准之间还存在着进一步的差异:就连贯性而言,只有当所有标准全都得到满足时,决策过程才能被视为合理的,而就对应性而言,可以将多种标准综合起来加以考虑。例如,在一些环境中,尽快作出决策而不是过多考虑准确性或许是更加重要的。然而,本书所报告的一种令人惊奇的实验结果是:简捷启发式不必总是进行这样的权衡。我们将证明,与一些标准策略相比,快速节俭启发式是更加迅速、更加节俭,同时也是更加准确的。它无需考虑任何复杂的权衡问题。

人们使用快速节俭启发式吗?

迄今所描述的研究方案涉及到了三个大的问题:(1)用来指 23
导和终止信息搜索以及利用搜索结果作出决策的启发式规则是什么?(2)什么时候和为什么这些启发式规则能够很好地生效,即从生态学的角度看为什么它们是合理的?(3)在现实环境中快速节俭启发式的实际效果如何?如果我们只对查明一些应用领域(如,人工智能和辅助决策系统)中新的启发式规则感兴趣,那么弄清楚上述三个问题已经足够了。但是,如果我们还想找到自然界中人类和动物行为的规则,那么我们还必须在我们的研究方案中加入第(4)个问题:何以证明人类和动物使用了快速节俭启发式?

关于人类和动物所使用的有限搜索和终止搜索的启发式规则,我们还知之甚少。一个重要原因在于通常的实验任务从一开始就排除了搜索活动(但可参见 Connolly & Gilani, 1982; Payne et al., 1993; Saad & Russo, 1996)。研究者们通常采用将所需要

的信息——常常是两到三种信息——非常现成地提供给被试的实验任务,从而回避了被试的信息搜索问题。我们将这类任务称为“基于给定信息的推断”(inference from givens),以区别于“基于记忆的推断”(inference from memory)或“基于外部环境的推断”(inference from the external environment)。后两种推断都需要信息搜索活动的参与(Gigerenzer & Goldstein, 1996a)。例如,关于分类问题的多数心理学研究都使用了仅在很少几个属性上变化的人造物体(如,面部和各种不同鱼类的画像)。为了对一个新对象作出归类,被试不需要在记忆或环境中搜索信息,而只需要利用刺激物中已经提供的少数即刻可用的线索(参见第 11 章有关该问题以及一种快速节俭分类启发式的讨论)。许多认知理论和用来检验它们的实验任务都忽视了有限的信息搜索和终止规则,但是排除了信息搜索活动的实验任务对于检验生态和有限理性模型来说是不适宜的,因为这种模型将有限信息搜索作为不可缺少的核心成分。

具有讽刺意味的是,人们所以忽视有限信息搜索的启发式规则,原因恰恰在于他们使用了快速节俭启发式。科学探索中通常采用的“从工具到理论”(tools-to-theories)的做法就是一种典型的启发式。它假定在认知科学家的常规研究中确立下来的实验方法所包含的逻辑实际上也是普通人进行推断时遵循的逻辑。20 世纪 50 年代和 60 年代,推断统计方法在实验心理学中被作为一种通用工具确立下来,它是费舍(Fisher)的虚无假设检验和奈曼(Neyman)及皮尔逊(Pearson)的决策理论的混杂物。与其他著名统计学工具(如, Wald 的序列分析方法, 1947)相比,这两种常用工具均未涉及信息的搜索和终止搜索规则。与这种“从工具到理论”的做法相应,许多研究者完全按照他们的理论赖以形成的统计工具和方法来建立认知过程模型,因此也忽视了为信息搜索和终止活动建立理论模型的必要性。例如,在凯利(Kelley, 1967)的归因理论中提出了一种模型,认为普通人的头脑将采用类似于认知科

简捷启发式

学家通常采用的方法对结果进行归因,即借助于费舍的方差分析方法进行计算。再如,在非常有名的信号检测论中,唐纳和斯怀兹(Tanner & Swets, 1954)指出,人类头脑能够在两种刺激信号之间作出区分,就像奈曼-皮尔逊学派的统计学家检验两种统计假设时所做的那样。如上所述,费舍和奈曼-皮尔逊的统计方法并没有包括搜索的概念,因此按照从工具到理论的观点,由它们所产生的理论也不可能考虑信息搜索过程。这就是为什么源于这些统计方法的认知理论和用来检验这些理论的实验任务全都不考虑信息搜索过程的原因所在。这就是我们的先人馈赠给我们的科学遗产(Gigerenzer & Goldstein, 1996b; Gigerenzer & Murray, 1987)!

与关于指导和终止信息搜索的启发式规则的研究相当贫乏的现象形成了鲜明对照,决策制定规则一直是大量研究和理论关注的中心问题。例如,心理学家们一直在探索这样一些问题:作出决策时儿童是否采用一种累积的、渐进的和其他与众不同的方法来对信息进行整合(例如, Gigerenzer & Richter, 1990)? 成人是否采用贝叶斯规则或计算平均数之类的方法来综合信息(例如, Birnbaum & Mellers, 1983)? 但是,即便对于这些问题,研究者们仍然采取惯用的实验设计方法进行研究,只向被试提供 2—3 种现成线索,不考虑被试本人对信息的自发搜索活动。结果是,这些研究结果也许并不能告诉我们在更加现实的情境中人们所采用的启发式决策规则的更多信息,因为在这种情境中人们必须积极地搜索信息和有关选项。

研究方案的总结

上面所描述的研究方案用来阐明理性的三个相互区别但又具有内在联系的方面(参见 Chase et al., 1998)。

有限理性 现实世界中的决策制定者必须使用有限的时间、信息和计算能力作出他们的判断。为了寻找有限理性的推断机制,我们设计并检验了快速节俭启发式的计算模型,以及它们的心

理构筑块,即指导信息搜索的启发式规则、终止信息搜索的启发式规则和作出决策的启发式规则。

生态理性 决策制定机制能够利用环境信息结构获得更富有适应性的结果。为了理解不同启发式规则怎样才能具有生态学合理性,我们详细描述了不同决策环境中信息结构化的方式,以及启发式规则怎样才能与这种结构化方式协调起来,成为快速的、节俭的、准确的,同时又是适应于现实的。

社会理性 一个决策者所处环境的重要方面常常是由与之相互作用的其他决策者构成的。因此,食肉动物必须作出关于其捕捉对象行为的关键性推断(参见第12章),雄性和雌性动物必须决定它们有兴趣与哪些其他动物进行交配(第13章),而为人父母者必须找到帮助他们孩子的有效方法(第14章)。社会理性是生态理性的一种特殊形式,为了研究它,我们设计了利用社会环境信息结构与其他对象进行适应性交往的快速节俭启发式,并检验了它们。这些启发式包括具有社会适应性的构筑块,诸如愤怒情绪、父母之爱以及社会规范等,它们可以作为指导和终止信息搜索的启发式规则的补充。关于这一点我们很快将会作进一步阐述。

上述理性的三个方面趋向于一个共同目标:按照其适应具体环境的本来面目来了解人类行为和认知活动,并发现指导适应性行为的启发式规则。

ABC 研究方案如何与先前的启发式思想联系起来

“启发式”这个术语起源于古希腊,意指“积极发现或探索”。18世纪早期它被引入到了英语中来,从那时起一直到20世纪70年代,“启发式”都是指解决难以用逻辑和概率理论加以处理的问题的一种有用的、甚至不可缺少的认知过程(例如,Groner et al., 1983; Ploya, 1954)。70年代以后,在心理学领域和决策制定研究

中,它的第二种意思产生了:人们常常滥用于本来应该使用逻辑和概率理论的情境中去的并非必不可少的认知过程(例如,Tversky & Kahneman,1974)。下面我们将探讨启发式概念在20世纪所发生的更加细微的变化,并说明我们对这个术语的使用是怎样与历史上的有关用法相一致的。

1905年,26岁的爱因斯坦(Albert Einstein)发表了他在量子物理学方面的第一篇重要论文,题目为“论有关光的产生和转换的一种启发式观点”。在这篇曾经获得诺贝尔奖的论文中,爱因斯坦用“启发式”这个术语来表明这样一种意思:他认为他所提出的观点是不完善的,甚至是错误的,但却是非常有用的。在该文中,爱因斯坦还不能够毫无保留地接受他开始发展的量子光学观点,但他相信它在方法上对于建立一种更加正确的理论具有重要的暂时用处(Holton,1988,pp. 360—361)。正像爱因斯坦所用的那样,启发式就是对待就目前知识来看还不一定完善的问题的一种观点,因此它不可避免地包含着错误,尽管如此,它对于人们沿着正确方向进行思考却是有益的(亦可参见 Wimsatt,1987)。

几十年以后,马克斯·威特海默(Max Wertheimer,爱因斯坦的亲密朋友之一)、卡尔·唐克(Karl Duncker)和其他格式塔学派的心理学家论述了启发式推理问题,但采用了与爱因斯坦略有不同的用法。格式塔心理学家将思维活动描述为一种内部心理活动和外部问题结构交互作用的过程。依据这种观点,人们首先用“查看周围环境”(looking around)和“审视问题”(inspecting the problem)之类的启发式方法搜索环境中的适当信息,然后借助于内部过程“重建”(restructured)和“改组”(reformulated)它们(例如,Duncker,1935/1945)。正是沿着这种传统思路,赫伯特·西蒙和艾伦·纽艾尔建立了信息搜索的启发式模型,但他们用更加精密的计算模型代替了格式塔学派某种程度上有些模糊不明的用法。随着认知心理学中信息加工观点的产生,人们开始用启发式这个术语来表示一种指导信息搜索的有用捷径、近似算法或简捷规则,

诸如象棋大师在对弈过程中每走一步棋用来缩小问题搜索空间的策略等。

像这些早期的理论家一样,我们在积极意义上使用启发式这个术语,一方面强调它在指导信息搜索方面的有益作用,另一方面又遵循纽艾尔和西蒙的做法,试图创建一种精密计算模型。然而,我们打破了过去仅用结构良好的人工问题研究启发式的传统,诸如数学问题(Polya,1954)或纽艾尔和西蒙(1972)使用的对弈问题和密码算术题问题等。相反,我们的研究强调快速节俭启发式怎样才能作出关于现实世界未知领域的推断。

与关于快速节俭启发式的ABC研究方案最紧密相关的研究是在机器学习领域有关适应性决策制定和简单分类规则的研究。在一项关于适应性决策制定的研究中,培恩、巴特曼和约翰逊(Payne,Bettman & Johnson,1993)探讨了当面临多种可供选择策略时,被试在准确性和可能付出的努力之间进行的权衡。他们使用的任务包括词典编纂规则和多线索排除问题等(Tversky,1972)。培恩及其同事强调指出,决策制定者有大量适用策略,必须依据给定限制(如,时间紧迫性)条件下它们各自的代价和准确性作出选择。正像贯穿全书我们将会看到的那样,这项重要研究与我们的研究方案有密切关系,但也存在明显差异。区别之一在于,培恩及其同事集中注意在诸如假想的工作候选人之间或只是偶尔才会遇到的必择其一的选项之间作出偏好选择,而不关心对非常现实问题的推断,诸如哪一个足球队将会赢得比赛,或两个城市中哪一个更大一些等。正由于如此,他们不能够依据一种策略对现实世界的预测能力来测定它的价值。相反,他们只能依据该策略和一种加权的人为附加规则之间的匹配程度来测定它的准确性,传统上这种人为附加规则对于合理偏好判断而言是一种“黄金”标准。结果是,在培恩、巴特曼和约翰逊的研究中,任何一种启发式规则在准确性上都不可能比上述人为附加规则做得更好(尽管他们进行的计算和付出的努力更少)。相反,当借助于外部标准

来评定所有策略的成绩时,我们发现无论在理论上(第4章)或是在实际上(第5章),快速节俭启发式都比人为附加规则更加准确。类似地,在有关机器学习的研究中,将注意力集中在关于现实世界的推断上也能使准确性得到客观测定。有关仅使用一条或少数几条线索的简单分类规则的研究已经证明,快速节俭方法可以是准确的和耐用的,因为它们限定了所使用参数的数量(例如,Holte, 1993; Rivest, 1987)。

20世纪70年代早期,心理学中出现了一种不同观点,强调使用启发式会导致系统错误和推理“失误”(lapses),这表明人类认知活动的非理性特征。这种由特威斯基和卡那门(Tversky & Kahneman, 1974)倡导的“启发式—偏差”(heuristics-and-biases)研究范式“污染”了有关简单心理机制的思想,因为它用一个单独的、不可分割的短语将这些简单心理机制“捆绑”到了一个具有价值偏向的术语中。在这项研究方案中,每当发现人类推理活动的错误(主要是违背了概率规则)时,就将这种错误归咎于使用了不适当的启发式规则。尽管特威斯基和卡那门(1974)一再强调启发式可能是成功的,也可能是失败的,但他们的实验结果通常被理解为揭示了人类推理活动的各种谬误,而这些谬误通常被归咎于使用了下面三种不适当的启发式规则之一:典型性规则(基于典型范例的推断)、可用性规则(基于容易得到信息的推断)和锚定规则(anchoring)或适应性规则(基于首先获得信息的判断)。

“启发式—偏差”方案暗示,普通人都是使用很少信息或心理资源进行推断的认知的“吝啬鬼”,因此不可能对概率或风险作出准确估计。有人甚至据此认为不应该让普通公众参与重要的社会和政治决策,如涉及核工业和其他具有危险性的高新技术领域的决策等。在《新周刊》杂志上曾经发表了一篇有关“启发式—偏差”研究的报道,借用这篇文章的话来说:“非常令人遗憾,多数人都是混沌不清的信息加工者,他们常常失足于错误选择的便捷途径,得出错误结论。”(McCormick, 1987, p. 24)然而,特威斯基和卡那门

又指出,将决策活动留待专家们去做也并非明智之举,因为他们也会犯同样错误。根据这些悲观论调,我们无从知道合理的决策究竟应该到哪里去寻找!

在“启发式—偏差”研究方案中所使用的狭义的“谬误”一词不仅仅是指不合理性或荒谬性,而且被视为人类有限理性的象征(Thaler,1991,p4)。就像将有限理性与受限制条件下的最优化等同起来一样,用这种方法将有限理性和不合理性等而视之也是一种严重的概念混淆。有限理性既不是受限制条件下的最优化,也不意味着不合理性。

我们关于快速节俭启发式的研究方案与“启发式—偏差”方案具有一些共同特征。二者都强调简单心理机制在人类思维活动中所起的重要作用,而且都想找到能够利用这些启发式规则的环境。但这些类似之处并不能掩盖它们在关于人类理性潜在本质的观点上的根本差异。这种差异导致了两种背道而驰的研究方案:在我们的研究中,将启发式看成是人类头脑充分利用环境信息结构达到合理决策的方法,因此我们更加重视人们能够利用简捷启发式作出准确、有用推断的方面和情境。相反,“启发式—偏差”观点将启发式看成虽然其决策成效差强人意、但有限的人类头脑不得不倚重的援手,因此其研究者总是寻找这样的情形,在那里启发式总是受到挑剔和责备,因为以它们为基础作出的推断是不准确的。在这种基本差异基础上,下面我们将简略讨论二者之间其他两种重要区别。更加详细的分析可参见吉戈伦尔的有关论文(Gigerenzer,1991b,1991,1997;Gigerenzer & Murray,1987)。有关两种观点正反两方面的意见,可参见吉戈伦尔(1996)与卡那门及特威斯基(1996)之间进行的辩论。

区别之一是,ABC 研究方案选择了为启发式建立详细计算模型的做法,而不是仅仅停留于使用笼统的启发式概念。在潜心研究了 30 年之后,“启发式—偏差”方案仅仅提出了关于(不准确的)推理过程简单机制的一些似乎合理、但却含混不清的设想,主要涉

简捷启发式

及前面提到过的三种启发式规则,即典型性规则、可用性规则和锚定规则。这些简单概念能够解释的东西太少,同时又太多。太少,是因为其潜在的过程没有被具体化;太多,是因为这些启发式规则几乎适合于解释任何一种实验结果。例如,“忽视基础比率”(base-rate neglect)倾向,即作出关于多种选项的决策时,忽略每种选项在现实环境中产生频率的现象,通常被归因于典型性启发式规则的作用。然而,相反的结果——过分重视基础比率(或称“保守主义倾向”,Edwards,1968)——又很容易用锚定规则(基于基础比率)加以解释(Gigerenzer & Murray,1987)。一种理论的失败可能有两种情况:(1)它是错误的;(2)它并没有错误,只是含混不定或不够精确。“启发式—偏差”方案很可能属于后一种情况。然而,我们却甘冒犯第一种错误的危险,因为含混不定的理论妨碍科学的进步,它根本就无意于证明、推翻,或改进一种理论。因此,在本书中我们提出了一种详尽的启发式计算模型,并将我们理论观点的“名片”摆到桌面上,以便其他人审视它们——甚至摆弄它们和玩味它们。

第二种区别更加典型。ABC 研究方案不再将连贯性标准(如,概率规则)作为理性的衡量尺度,而是采用了对应性标准,即根据启发式规则在现实环境中的成效来衡量它们,在这种环境中最优化的连贯性标准是未知的或不可行的。相反,“启发式—偏差”方案的拥护者们通常假定,每一种推理任务都有一种确切的标准答案,它来源于一种具体内容不明的概率规则的使用,而不管任务和环境的要求是什么。我们的观点将快速节俭启发式从“启发式—偏差”方案的桎梏中解放了出来,使它不再被视为违背连贯性标准的非理性的象征,而被看成是在现实环境中作出准确适应性决策的有用工具。

让我们简单总结一下我们的研究在启发式概念演变过程中所处的地位。ABC 研究方案采用了将启发式作为合理推断必要的适应性工具的传统观点。我们用指导和终止信息搜索和作出决策

的具体规则建立了一种计算模型,从而使快速节俭启发式的作用和功能比过去更加具体和明确。作为我们有关生态理性研究的一部分,我们借助于对启发式规则在现实环境中的准确性、速度和节俭性的详细分析,代替了连贯性标准内容不明的评定方法。最后,尽管“启发式—偏差”方案将启发式规则看成是对有效合理推断的一种妨害,使得“智者”(Homo sapiens)不那么明智了,我们却将启发式视为能够让我们在现实环境中作出合理推断和采取合理行动的重要方法——没有它们,“智者”也将迷失方向。

适应工具箱

哥特弗雷德·威尔哈姆·莱布尼兹(Gottfried Wilhelm Leibniz)曾经有一个美丽梦想。他梦想有一种通用的逻辑语言和特征系统,来代替所有推断(Leibniz, 1677/1951)。这种通用特征系统包括两个部分:(1) 人类思想的元素表,即一种原始的字母符号,它们代表了不能再加以分解的简单概念,并形成现实事物的明细表;(2) 将这种元素表中的要素结合起来进行推理的微积分规则。虽然罗伯特·鲍耶尔(Robert Boyle)早已研究了化学元素表,但莱布尼兹设定了一种更加雄心勃勃的目标:了解上帝描绘自然界奥秘的通用语言。他将上帝创造的这种知识看成是任何人都可以孜孜以求的最重要目标。莱布尼兹相信,这种通用特征系统将结束学者们的辩论和争吵——人们将停止对别人大喊大叫,而是手里拿着铅笔和纸张,坐下来平静地对别人说:“让我们算算看!”所有科学和道德问题都将得到解决,我们将生活在最为理想的世界里。年轻的莱布尼兹还对这种通用特征系统的发展作出了非常乐观的预言:“少数选择出来的精英也许能够在5年时间内完成它。”(Leibniz, 1677/1951)(多年来,在心理学和人工智能领域,我们经常可以听到类似的声音!)曾几何时,许多开明思想家们相信,有关概率的数学理论能够使莱布尼兹的梦想变为现实。但是,到了19

简捷启发式

世纪 40 年代,数学家们已经放弃了将理性推理活动还原为一种通用微积分规则的想法,认为这是一项出力不讨好、甚至与数学宗旨背道而驰的任务(Daston,1988)。然而,正如前面我们讨论超自然神灵时所指出的那样,迄今为止在许多领域仍然有许多理论家不愿放弃这种难以实现的梦想所暗含的逻辑。

构成莱布尼兹人类思想元素表的大量简单概念都是借助于概率论之类的单一通用工具起作用的。但是我们无从找到这样一种通用工具。犹如一个机械修理工在维修汽车发动机时所做的那样,时而拿出了老虎钳,时而拿出了扳手,时而拔出了火花塞,而不能仅用一把大锤子敲敲打打。不同的思想领域也需要不同的思维工具。这就是适应工具箱的基本思路:把自然进化赋予人类头脑来进行各种推断活动的具体认知机制(包括快速节俭启发式)集中起来,形成思维活动的工具箱(亦可参见 Bettman,1979;Cosmides & Tooby,1992;Payne et al.,1993)。这种把多种单功能工具混杂在一起的工具箱思想失去了莱布尼兹单一通用推理工具的美丽光环。相反,它祈求“粗鄙的机修工和旧货零售商”(像 Wimsatt 所描述的那样)所具有的那种平实无华、但却令人惊奇的能力,他们能够仅仅用手头的东西恰当地解决所遇到的任何问题。

这种适应工具箱容纳了心理学的(不同于形态学或生理学)适应机制(Tooby & Cosmides,1992)。这些适应机制既包括所谓的“低级”(lower order)心理机能,诸如深度知觉、听觉定位分析以及面部再认等知觉和记忆过程,它们都能够以相当自动化的方式起作用,又包括建立在低级心理机能基础之上、并至少部分地可被意识到的“高级”(higher order)心理机能,包括我们前面讨论过的一些例子,如将一个心脏病患者判断为高危或低危病人、决定是否结婚等。本书讨论的重点是建立在线索知觉和记忆等低级心理过程基础上的高级认知过程的快速节俭启发式。

诸如面孔和声音再认等低级知觉和记忆过程是复杂的和难以阐明的,部分原因在于它们使用了大量的平行计算过程。还没有

任何一个人试图制造一种机器,其对面孔的再认能力能像2岁儿童一样好。现在考虑一种依据这些低级过程作出推断的高级决策机制——在第2章将作详细介绍的再认启发式。快速节俭启发式能够利用再认过程迅速地作出关于现实世界未知方面的推断:例如,一个人也许觉得能够再认其味道的食物比不能再认的食物更安全,听说过其名字的大学比从未听说过的大学会提供更好的教育。尽管再认记忆的机制也许是错综复杂、难以阐明的,但再认启发式却能够仅用几个步骤来描述一种算法式规则系统。我们无需知道再认记忆是如何使基于它的启发式规则生效的。这个例子说明了一个表面上自相矛盾、但却毋庸置疑的命题:高级认知机能常常能够比低级认知机能更容易通过简单算法规则而模式化。

上述命题并非我们的新发现。在过去一个世纪里,它曾经被人们以不同方式提到过。例如,20世纪早期,符茨堡(Würzburg)心理学派的拥护者们(Kusch, 1999),以及较为近期的谢泼德(Shepard, 1967),都曾论及这个命题。当然,该命题也有它的局限性:一些高级过程,诸如达尔文发展自然选择理论所涉及的创造性过程,也许超出了快速节俭启发式所能解释的范围。但是,正如本书所表明的那样,我们相信简捷启发式规则既可以被单独使用,又可以结合起来加以使用,去说明表面上似乎需要更加复杂解释的多种高级心理过程。

情绪、社会规范和模仿

尽管在本书中我们主要探讨认知过程的启发式规则,但需要强调指出的是,情绪也可以作为指导和终止信息搜索的启发式规则起作用。例如,坠入情网可以被看成是一种强有力的终止规则,它结束了对新伴侣的进一步搜索(至少暂时如此),并使当事人对他所爱的人忠心不二。类似地,婴儿乖巧可爱的样子和甜甜的笑容勾起的父母怜爱之情,可以避免为人父母者在多种近似目标之间进行代价—收益计算,所以父母亲们通常不会为是否值得度过

简捷启发式

无数个不眠之夜和投入大量财力和精力来照料孩子这样的问题多费思量。在多种工具各司其职而非简单通用的适应工具箱中,情绪是一种非常典型的具有针对性的构筑块(Tooby & Cosmides, 1990),例如,父母的爱怜之情可以帮助他们无微不至地保护和照料婴儿,满足他的各种需要,而厌恶情绪又总是使他们尽可能避免孩子吃下变质食物而生病。不存在能够应付所有决策问题的通用情绪。

社会规范和社会模仿也能帮助我们使用有限时间和知识作出决策。例如,“不听老人言,吃亏在眼前”或“老婆总是别人的好”之类的启发式,可以帮助人们减少对直接经验的依赖,简化信息搜索过程,迅速作出决策。这些社会理性形式在动物界也广泛存在。例如,雌性古比鱼有一种模仿其他雌性古比鱼选择交配对象的倾向,这种倾向足以改变它们原有的择偶偏好(Dugatkin, 1996)。雌性鹌鹑采用了类似的同伴模仿机制(Galef & White, 1998)。对于人类来说,除了以个别化方式进行模仿或通过传媒相互影响之外,学术上的相互影响似乎也遵循类似的启发式规则。

因此,除了前面列出的认知构筑块以外,我们的适应工具箱中还容纳了充分利用情绪、社会规范及模仿的决策启发式。在社会理性领域这些附加的启发式规则是特别重要的。

32

怎样选择启发式规则

头脑是如何挑选适应工具箱中的启发式来解决具体问题的?关于这个问题并不存在我们想像的那么多选择,主要原因在于具体任务需要具体工具——这就是说,每一种启发式规则都只适用于特定种类的问题,这意味着在给定情境下适应工具箱中的多数工具是不适用的。在本书中我们主要讨论了两种相互交叉重叠的具体任务领域,它们能够确定在特定条件下选择何种启发式。这两种任务即具体的适应性任务(如,选择配偶和父母投资)和具体的推断任务(如,分类和估计)。很显然,用于在两个选项间作出选

择的启发式规则将不适合于分类任务,寻找配偶的启发式规则也将无助于对栖息地好坏的判断。适应工具箱中的每一种工具都有其特殊用途。

当面临两种或两种以上工具均可使用的任务时,决策制定者具有的知识可能被用来作出选择。例如,如果一个人知道一些线索可用在两种选项间作出抉择,但不知道哪一条线索比其他线索都好,他将没有足够(关于线索排列顺序的)信息应用我们提出的“采纳最佳”(take the best)启发式(第4章)。然而,他可以采用更简单的“采纳最近”(take the last)启发式,这种启发式只需要知道使用什么样的信息,而无需知道以什么样的顺序来使用它们。如果他连应该使用什么线索都不知道,但至少认出了一种选项(而不是其他选项),他还可以采用也许是最为简单的启发式:再认启发式(第2章)。通过这种方法,一个人的知识水平就可以帮助他在适应工具箱的多种启发式中作出选择。

其他一些外部因素,诸如时间紧迫性和所取得的成效,也能帮助决策者在多种启发式规则中作出选择(Johnson & Payne, 1985)。在一些情形下,交替使用多种工具也是一种重要适应手段。这样做时,要么会随机产生一些不可预测、变化多端的行为——这对于雇主在多个工作申请者之间作出选择或弱小动物免遭食肉动物捕捉都是非常有用的策略(Driver & Humphris, 1988);要么可能会使人们在作出决定时产生一些系统的个体差异,从而避免大家在同一时间去做同一件事情时发生冲突(例如,它可以避免人们在每周的同一时间拥入当地的同一酒吧。Arthur, 1994)。我们假定:在从适应工具箱中选择决策工具时,上述所有影响因素都将以快速节俭的方式帮助人们作出决定,而不必被移交给一个潜在的具有超凡运算能力的神灵,由它用最优化标准来判定快速节俭启发式的优劣。

适应工具箱中的每一种工具都是由更加原始的材料构成的,这些材料包括前面讨论过的搜索信息、终止搜索信息和作出决策

简捷启发式

的启发式规则。新工具可以通过对旧工具的加工而成型,这就好像将手柄安在一块劈东西用的石头上就成了斧头一样。例如,在本书第4章将要介绍的“采纳最佳”启发式是在第2章介绍的更加原始(从进化的角度看)的再认启发式和其他一些启发式规则基础上发展起来的。这种看待心理工具的构造观点将适应工具箱的形象比喻与把头脑看成是功能齐全的瑞士军刀的隐喻(Cosmides & Tooby, 1992)区别开来。这两种比喻都强调头脑使用了多种具体适应性策略而不是单一通用工具,但工具箱的隐喻更加重视对多种工具和构筑块灵活地进行重组,以形成新的启发式。

超越神灵与梦魇

在学术思想界,有一些很少有人愿与之为伍的恶魔,那就是非理性主义和神秘主义。但也有一些令许多研究者心向往之的神灵,一有机会就会被引入到心理学理论中来,那就是处于神性王国最高层的无限理性和受限制条件下的最优化。这些深受欢迎的神灵毫无疑问集中体现了人们渴望拥有的特性,诸如能够预卜先知、预测未来等,但对于生命有限的人类而言,它是难以企及的。为什么如此众多的社会科学家用一些超级神灵才具有的特性来美化人类的知识和能力,而不愿意正视他们所受到的限制呢?难道这些特性就是达尔文的物种进化论产生之前人们将人类与动物区别开来并引以为自豪的那些特性吗?

对思想和行为最优化的心驰神往在一定意义上反映了道德和美学的理想。正像拉普拉斯无所不能的超级神灵一样,莱布尼兹关于通用微积分规则的梦想充分体现了这种理想的道德和审美价值。当代认知科学家、经济学家和生物学家们追寻着同样的梦想,他们赋予有机体进行理解、记忆和计算的无限能量,在此基础上建立了复杂的推理模型。然而,这种天堂般的梦想在现实的物理和心理真实性面前破碎了。生命苦短的卑微人类不可能奢望享受到

这些梦想施与的恩泽,相反,当用这些最优化标准判定人类时,他们反而愈加显得是非理性的和机能不全的。到头来,天堂般的梦想变成了一场噩梦!

面对这种梦魇般的进退两难困境,许多研究者仍然抱着人类能够逐渐接近这些确定标准的幻想,而不肯屈服于人类理性有限和愚蠢无知的想法。然而,也许并非必须在非现实的梦想理性和现实的梦魇般理性之间进行选择,还存在着可以避免这种尖锐对立的第三条路线:朴素的理性和节俭的准确性。 34

本书讲述了我们如何取得成功和面对挑战的故事,它告诉人们头脑怎样才能借助于不确定线索,以快速、节俭,但同时又准确的方式,作出关于未知世界的适应性推断。它也是关于我们这个跨学科研究组的故事,我们结伴进行了一次探险,从神圣的理性大陆出发到达了生态理性的彼岸。在这个过程中,我们学会使用多种多样的调查方法,并彼此从对方的观点中吸取了有价值的东西。大约三年之前,当我们刚刚踏上旅途时,我们未曾想到这次旅行会是如此迷人,也未曾想到会经历如此多的新鲜事物!

让我们更精明

第二部分

基于无知的决策制定

DECISIONS
THAT MAKE US
SMART

蘇州大學
PDG

第2章

再认启发式——无知如何让我们精明

丹尼尔·G·古德斯腾(Daniel G. Goldstein)

哥德·吉戈伦尔(Gerd Gigerenzer)

人类的思维能力首先是由大量再认能力组成的,其次是由选择性搜索能力组成的。

——赫伯特·A·西蒙(Herbert A. Simon)

在

苏格兰的一条乡村小道上,马克格莱格(MacGreger)遇见了他的老同学马克阿利斯特(MacAlister),和他打招呼。马克阿利斯特愣了半天未说出一句话。他认出了马克格莱格的面孔,但怎么也想不起他的名字、在哪里见过他,以及其他任何事情。马克格莱格感到很奇怪,心想他的老同学一定是正在“犯晕”(tartle),也就是说,仅有一种似曾相识之感而不敢相认。正如“犯晕”这个苏格兰词汇所表示的那样,再认和回忆是两种互不相同的记忆能力,有时候它们之间的差异还是非常大的。举例来说,有一个54岁的警察,名字缩写为R. F. R。他患有严重的健忘症,以至于不能说出照片上他妻子和母亲的名字(Warrington & McCarthy, 1988)。看起来R. F. R似乎失去了再认能力,其实并非如此。在一项实验中,研究人员向他提供了一些名人和他以前从未见过的人的照片,要求他指出哪些是名人的照片。对于这项任务他完成得相当好,仿佛他的记忆力从未受到损害似的。尽管他缺乏对照片上人物(如,他的母亲)名字的回忆能力,但他仍然保持着正常的再认能力,所以他能够认出以前曾经看到过的著名人物的面孔。

让我们更精明

像 R. F. R 和那位“犯晕”的苏格兰人一样,当我们无法确定看到的物体、听到的声音、尝到的味道、闻到的气味以及触摸到的东西究竟是什么的时候,我们并非不知道那是我们以前曾经见到过的,即便我们不能回忆起更多信息。一些研究者指出,我们的“再认感”(sense of recognition)构成了一种特殊记忆系统,他可以独立于其他记忆能力而免受记忆缺损的侵扰。例如,记忆力衰退的老年人(Craik & McDowd, 1987; Schonfield & Robertson, 1966)和特定种类的脑损伤患者(Schacter & Tulving 1994; Squire et al., 1993)不能说出他们对于一种物体所知道的一切,甚至不知道在哪里看到过它,但却能表现出一定的行为,这些行为证明他们以前曾经接触过这种物体。类似地,有关实验室研究已经证明,在需要分配注意力的学习任务上,由于存在许多分心因素,所以更加有效的记忆活动难以进行。然而,即便在这种情况下,再认记忆仍然能够对信息进行编码(Jacoby et al., 1989)。作为对以前我们是否经历过某些事情的简单的“非此即彼”式的二元判断,简单再认记忆充其量仅仅表明我们对有关事物仍有那么一丁点了解。那么我们的头脑为什么要保留这样一种能力呢?简单再认记忆究竟有什么益处呢?

38

本章我们将介绍一种本书中最为简单的启发式——再认启发式(the recognition heuristic)。它充分利用再认记忆这种丰富有效的认知资源作出关于现实世界未知方面的推断。潜藏于面孔、声音和名字再认背后的认知过程远非那么简单,在认知科学中还远远未予澄清。然而这并非我们所关心的问题,我们仅关心利用这种复杂认知过程产生的后果——再认,来构建一种简捷启发式推断规则。再认启发式是如此节俭,甚至在缺乏有关知识时,它反而显得更加有效。在本章中我们采用简单规则形式来界定启发式,这使得我们能够借助于数学分析和计算机模拟方法来测定它的效果。我们表明,在一定条件下会产生一种与我们直觉相反的“少即多效应”(less-is-more effect),即缺乏再认对作出推断反而

是有利的。我们还将说明如何对再认加以测定,以便从实验上探明人们实际上是否采用了再认启发式。

在许多领域人们都使用“再认”这个术语,所以我们有必要首先澄清我们将怎样使用它。假如马克阿利斯特上了一辆公共汽车,里面的乘客可以分为三类,我们用图 2-1 所示的三个纵栏分别来表示他们。有一些乘客他根本不认识,也就是说他确信他以前从未见过他们,用左边的纵栏来表示;有一些乘客他只是认识(感到面熟),但不能确认或回忆出关于他们的任何事情(即让他“犯晕”的那些乘客),用中间的纵栏来表示;还有一些乘客他不仅认识,而且能够确认他们(如,知道他们的职业是什么等),用右边的纵栏来表示。

39

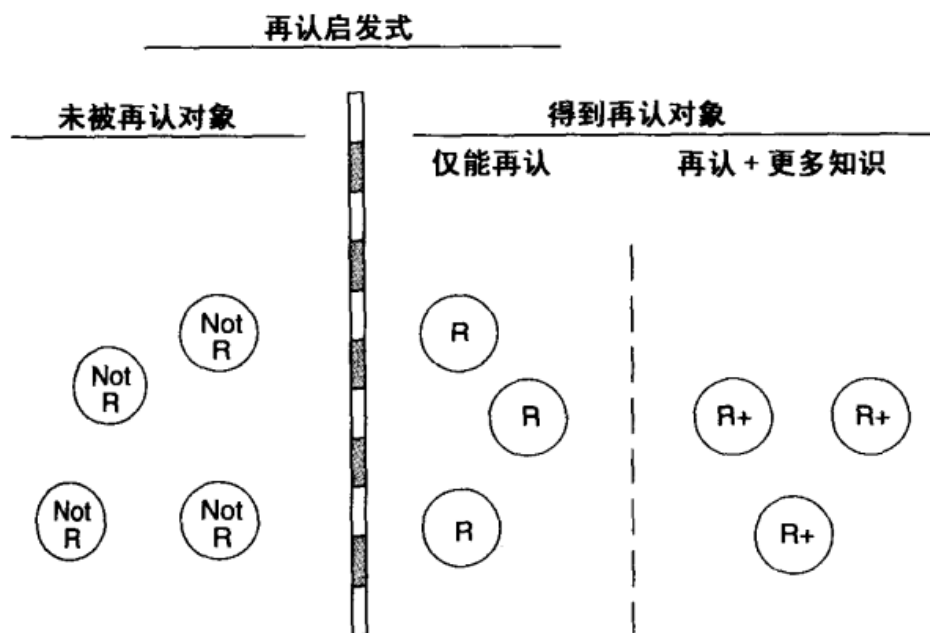


图 2-1 可应用于未被再认新对象(Not R=未被再认)、仅能再认对象(R)以及了解更多的对象(R+)的再认启发式。与再认启发式有关的区分是未被再认对象和其他两类对象间的区分。

正如图 2-1 中带斑纹的框条所表示的那样,我们用再认这个术语将世界划分为两个部分:陌生部分(左边纵栏)和先前经历过的部分(右边两个纵栏)。例如,对地界标志的再认就是以陌生的和经历过的部分之间的简单二元式区分为基础的,它具有帮助人

让我们更精明

简捷启发式

或动物返回家园的适应性功能。应该将简单再认与“了解程度”(degrees of knowledge)和“熟悉性”(familiarity)之类的概念区分开来。例如,有关理论假定,对物体的态度倾向于随着该种物体的重复出现而变得更加积极(例如,Zajonc,1968),对一种观点的信仰会随着它的重复而增强(例如,Gigerenzer,1984;Hasher et al.,1977)。从第4章起我们才开始介绍使用超出简单再认的更多知识的启发式规则。还应该将我们所使用的再认概念与一种非常流行的用法区分开来,这种流行用法即:将再认理解为一个确认一种物体在先前实验序列中是否出现过的能力。此类研究常常不能在陌生的和经历过的物体之间作出区分,因为这些研究中使用的刺激(多数情况下是数字或常用词)实验之前对被试而言就不是新奇的。例如,常用词“猫”(cat)对多数被试来说在实验之前就已经不是新奇的刺激,而非词“flink”则更可能是新奇的。相反,如同下面我们将要提供的例子那样,使用从未见过的照片进行实验更能体现我们使用“再认”一词所代表的意思。

再认记忆是一种广博的、自动化的和可靠的认知资源,尽管有时会产生似曾相识的错觉。谢泼德(1976)曾经要求被试按照他们自己喜欢的步调浏览612张图片,然后用成对图片(一个是前面呈现过的,一个是新的)立即对他们进行再认记忆测验,结果被试平均能够再认先前呈现过图片中的98.5%。斯丹丁(Standing,1973)将图片数量增加到1000张(根据它们的鲜艳程度分为“普通图片”和“醒目图片”两类),并将每张图片的呈现时间限定为5秒钟。48小时之后,用谢泼德所用的那种配对图片方法进行测验。结果被试在1000次测验中分别有885次(普通图片)和940次(醒目图片)选中了先前呈现的图片。在排除了猜测成分之后,分别还有770次和880次选中了先前呈现的图片。后来斯丹丁将其研究推进了一步,总共考虑到了可能影响图片再认的10种因素。在他所做的一项也许最为复杂的再认记忆测验中,向被试提供了10000对图片,结果被试有8300次都作出了正确选择(排除

40

猜测成分后仍有 6 600 次)。关于对“醒目”图片的再认成绩,斯丹丁推算道:“如果在这些条件下可以呈现 1 000 000 对项目的话,那么将会有 731 400 个项目被保持下来”(p. 210)。值得注意的是,尽管随着所呈现图片数目的增加,保持率在下降,但得到再认图片的绝对数量却具有一种增长趋势。我们推测,在实验室实验中,一个人不可能超出其再认记忆的能量限制,甚至一个人终其一生也难以超出其再认记忆的能量限制。

怎样从无知中受益?

高等有机体再认记忆的巨大能量很可能由于许多适应性机能而得到了进化。考虑一下野生老鼠的饮食习惯,它们表现出一种强烈的“恐新倾向”(neophobia)——不愿意吃它们不认识的食物(Barnett, 1963)。这种习性在避免食物中毒方面具有明显的适应价值:一只活鼠所吃的每一种食物都肯定不会置它们于死地(Revasky & Bedarf, 1967)。挪威鼠特别喜欢吃它们通过品尝或嗅其他老鼠呼出的气味而辨别出来的食物(Galef, 1987; Galef et al., 1990)。即便呼出气味的老鼠正在生病,这种选择食物的启发式规则仍然生效。这就是说,再认信息的作用超过了疾病信息的作用。在本章的后面我们将介绍用人类被试所做的有关实验,它们表明:当与其他信息发生矛盾时,再认信息仍能超越其他信息而起主导作用。野生鼠的食物选择行为分明与再认启发式是一致的!

下面我们将详细描述再认启发式,并探讨其在推断上的准确性。我们详述了一些具体条件,在这种条件下再认启发式能够使拥有较少知识的有机体比拥有较多知识的有机体作出更加准确的推断——一种我们称之为“少即多效应”的反直觉现象。我们将采取“解剖麻雀”的方法,从一个非常具体的问题领域——即一个很容易理解的例子——入手,来探讨再认启发式问题,这个领域就是对有关地理问题的推断。

在我们的认知结构中,对有关事物适当名称的再认构成了一个独特系统。它可能独立于其他语言技能而单独受到损伤(McKenna & Warrington,1980;Semenza & Zettin,1989;Semenza & Sgaramella,1993)。一个人的地理知识主要是由适当的名称(城市、国家、山脉,如此等等)以及它们在地球上的分布位置构成的。41 地理知识总是不完善的,这使它成为研究再认问题的一个理想领域。我们将借助于计算机模拟、数学分析以及实验室研究来分析再认问题。在一些研究中,我们使用了被试(芝加哥大学学生)不十分熟悉的地理问题:有关德国城市的问题。特别是,我们讨论了人口在10万以上的83个德国城市的排序问题。我们的美国被试只能够认出这些城市中的大约1/4。正如我们将会看到的那样,正是这种再认能力的欠缺帮助他们很好地完成了推断任务。

我们将要考查的是一种常用任务:从一套对象中选出一个子集。在本章中,我们着重探讨从两个对象中选出一个的问题。这种双项选择问题不仅在实验心理学中是一种经典任务,而且对于理解一些更加复杂的问题(如,多项选择问题)也可以起到抛砖引玉之效。该种双项选择任务的一个例子是:“哪一种货币更值钱:英镑或是马克?”或者在地理学领域:“哪一个城市有更多人口:慕尼黑或是多特蒙德?”

再认启发式

不妨考虑一下依据某些效标判断两个对象中哪个具有更高值(如,哪一个更快、更高、更强)的问题。在处理这类问题时,再认启发式可以被简单地描述为:如果两个对象中的一个得到了再认,另一个未得到再认,那么可以推断,得到再认的对象具有更高的值。

举例来说,一个人从未听说过多特蒙德,但却听说过慕尼黑,那么他将推断:慕尼黑有更多人口,而且碰巧是正确的。这种再

认启发式只有当两个对象中的一个未被再认时(即在部分无知情形下),才是适用的。请注意:当再认与一种效标呈负相关时,再认启发式定义中的“更高”就应该被置换为“更低”(此种情况下,我们可以推断:得到再认的对象具有更低的值——译者)。

再认与环境结构

再认启发式是针对特定领域的推断规则,即只有当再认与某种效标相关时,再认启发式才是有效的。那么,如何估计再认与效标之间的关系呢? 在一些领域,这种相关的方向是由遗传决定的(例如,老鼠推测“不认识的食物是值得怀疑的”似乎就属于这种情况)。在另一些领域,这种相关的方向必须通过后天的经验来习得。然而,在现实的推断或预测情形下,有机体常常难以直接接近这种效标。尽管如此,但正如图 2-2 所示的那样,现实环境具有双重性:一方面它反映了(而不是揭示了)效标,另一方面又是可以感觉得到的。在这种环境中存在着一些“信息媒介”(mediators),通过它们人们可以找到一种效标。例如,一个人也许没有

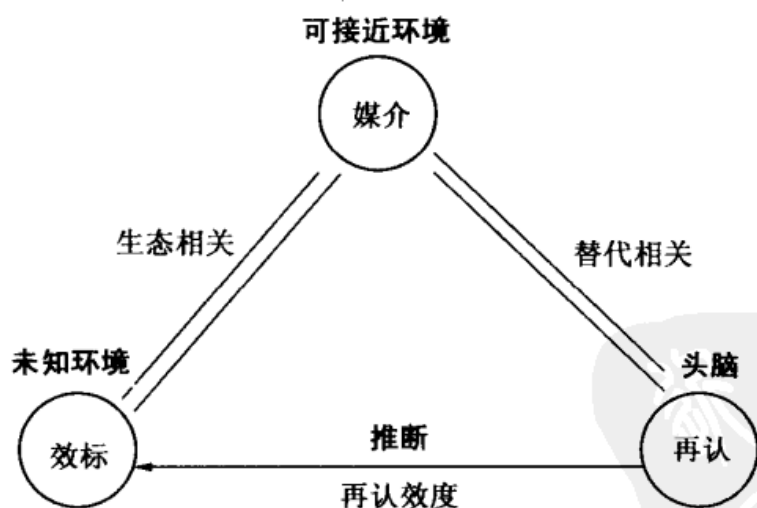


图 2-2 再认启发式的生态合理性。媒介变量反映了、但并未揭示不可接近的效标。媒介变量影响再认概率。头脑反过来利用再认对效标进行推断。

关于各个大学经费状况的直接信息,因为这种信息并非总是容易 42
得到的。然而,一所大学的经费状况也许会在该大学被新闻媒体
提及的次数多少上反映出来。因为新闻媒体是很容易接近的,所
以它就成为非常典型的信息媒介。一个名字在新闻媒体上出现得
越频繁,人们就越有可能再认这个名字。例如,斯坦福大学就比
“明尼斯库”(Miniscule)州立大学更经常地被新闻媒体报道。多亏
有了这种媒体,一个人才可以作出关于哪所大学有更多经费的推
断。用来描述效标、媒介和头脑三者之间关系的变量分别被称为
“再认效度”(recognition validity)、“生态相关”(ecological correla-
tion)和“替代相关”(surrogate correlation)。

生态相关描述了效标和媒介之间的关系。在上述关于大学经
费状况的例子中,效标是经费状况,媒介仅仅是大学在新闻媒体上
被报道的次数(而不是任何有关其经费状况的信息)。在挪威鼠的
例子中,效标是一种食物的毒性大小,媒介是吃了这种食物的老鼠
呼出气体的味道(而不是关于这些老鼠身体健康状况的任何其他
信息)。替代相关是媒介(它作为对不可接近效标的替代物而起作
用)和再认记忆内容之间的相关。在我们关于大学经费状况的例
子中,替代相关即大学的名字在媒体上被报道的次数与对这个名
字再认成绩之间的相关。替代相关可以根据对一个人的再认记忆 43
来测定(在此种情况下信息资料是双极的),或者依据对一个群体
的再认成绩来测定,对此我们在后文中将详加说明。

再认和效标之间相关程度的大小被称为再认效度,我们将其
界定为:在给定的参照评定等级中,得到再认对象比未得到再认
对象有更高效价次数所占的比例。因此,再认效度 α 为:

$$\alpha = R / (R + W)$$

在这里 R 是在一个对象被再认、另一个未被再认的所有二择
一问题中用再认启发式作出正确推断的次数,而 W 是在同样情况
下作出错误推断的次数。

凭借再认的推断是精明的吗？

老鼠的食物选择行为受到了再认的引导，但是有智慧的人类作出的推断又怎样呢？基于再认（或更恰当地说是基于无知）的推断会不会比仅仅依靠猜测做得更好呢？让我们看一看人们使用再认启发式的两个例子。

哪一个美国城市有更多居民：圣地亚哥或是圣安东尼奥？我们分别向芝加哥大学和慕尼黑大学学生提出了这个问题。结果芝加哥大学仅有 62% 学生选择了正确答案，尽管作为美国公民他们似乎应该做得更好。然而，100% 德国学生都作出了正确选择。德国学生是如何作出圣地亚哥有更多居民的判断的呢？所有德国学生都听说过圣地亚哥，但他们中的多数人不知道圣安东尼奥，所以他们能够运用再认启发式，从而作出正确推断。美国学生对这两个城市都很熟悉，他们还不够无知，所以不能够运用再认启发式。

哪一个足球队将赢得比赛？50 个土耳其大学生和 54 个英国大学生对参加英国足总杯第三轮比赛的所有 32 支球队的成绩进行预测（Ayton & Onkal, 1997）。土耳其学生对英国足球队知之甚少，而英国学生则对它们了如指掌。尽管如此，土耳其学生作出的预测几乎像英国学生一样准确（准确率分别为 63% 和 66%）。英国足球队通常是用它们所在城市的名字来命名的（如，曼彻斯特联队），因此对英国各个球队水平了解不多的人可以用对城市名字 44 的再认作为预测各个球队成绩的线索。拥有强劲实力球队的城市通常是大城市，而大城市恰好又是容易得到再认的。实验结果表明，土耳其学生确实使用了这种再认启发式：在一个球队得到再认（一定程度上熟悉）、另一个球队未得到再认的所有成对球队中，前者有接近 95% 的机会（662 次机会中有 627 次）被判断为赢。如同上一个例子一样，再认启发式能够将部分无知转化为合理推断。

这两个例子均说明再认启发式具有生态学的合理性。之所以会如此，是因为它充分利用了自然环境的信息结构：在这些环境

44

精

让我们更精明

中缺乏再认是系统的、经常性的,而非随机的、偶然的。如果无知与一个人希望推断的事物有关,那么它就是有益而无害的。这种启发式规则并非一种通用策略,因为在一些情形下上述相关是存在的,但并非在所有情形下这种相关都存在。在涉及二择一或多择一之类的竞选问题时(例如,两个大学中哪个排名更靠前,或者两个球队中哪个将赢得比赛),再认启发式是有效的。然而,在有些任务上再认启发式并不能起到很好的预测作用。下面让我们更加详细地作一分析,看一看再认启发式究竟何时有效。

再认启发式的准确性

在双项选择问题上,一个人借助于再认启发式可望得到的正确答案比例是多少?假定有一种参照效标,可以用它对 N 个对象加以评级,而且每次测验都是对随机抽出的两个对象加以评定的二择一问题。这样每次抽出的两个对象可能属于下列三种情况之一:一个能再认,另一个不能再认;两个都不能再认;两个都能再认。假定共有 n 个能够再认对象,所以也就有 $N-n$ 个不能再认对象。这意味着有 $n(N-n)$ 对一个能再认、另一个不能再认的情形。采取类似的数学方法,我们可以算出有 $(N-n)(N-n-1)/2$ 对两个都不能再认的情形和 $n(n-1)/2$ 对两个都能再认的情形。为了将上述每种情形出现的绝对次数换算成比率形式,我们只需要用其中的每一个除以 N 个对象可能配成的所有对数 $N(N-1)/2$ 。

要想算出在这种测验上得到正确答案的比例,有必要知道每一种配对情况下获得正确答案的概率。如前所述,再认效度 α 实际上就是一个能再认、另一个不能再认情形下获得正确答案的概率。而在两个对象都不能再认情形下,必须作出随机猜测,所以获得正确答案的概率为 0.5。最后,令 β 为“知识效度”(knowledge validity),即两个对象均能再认情形下获得正确答案的概率。将所有这些条件综合在一起,在所有配对测验中可望得到正确推断

的比例 $f(n)$ 为:

$$f(n) = 2\left(\frac{n}{N}\right)\left(\frac{N-n}{N-1}\right)\alpha + \left(\frac{N-n}{N}\right)\left(\frac{N-n-1}{N-1}\right)\frac{1}{2} + \left(\frac{n}{N}\right)\left(\frac{n-1}{N-1}\right)\beta \quad (1)$$

等式右边分成三个部分: 最左边的是由再认启发式作出正确推断的比例; 中间部分是完全凭猜测作出正确推断的比例; 最右边的是当一个人的知识超出了仅仅能够再认情况下作出正确推断的比例。对这个等式进行仔细分析, 我们可以看到: 如果得到再认城市数目(n)是 0, 那么所有问题都只能凭猜测来回答, 所以得出正确答案的比例将会是 0.5。如果 $n=N$, 即所有城市均得到了再认, 那么左边两部分均为 0, 得出正确答案的比例将为 β 。我们还可以看到, 当被试处于“半无知”(half ignorance)状态, 即刚好能再认一半对象时, 再认启发式最为有效, 因为在此条件下一个能再认、一个不能再认的配对数目 $n(N-n)$ 最多。

总之, 以再认效度 α 、知识效度 β 以及无知程度(即 $N-n$)为依据, 等式(1)具体描述了一个人使用再认启发式可能作出正确推断的比率。下面我们将看一看再认启发式可能导致的一种与我们直觉形成明显反差的现象: 少即多效应。

少即多效应

我们不妨设想一下: 马克阿利斯特的三个儿子在学校不得不参加一项有关德国城市的测验。这项测验采取随机抽题方式进行, 题目是涉及 50 个德国大城市人口规模的双项选择问题。关于这些问题, 小弟最为无知, 他以前甚至从未听说过德国(更不用说德国城市了)。老二还不错, 他道听途说, 能够认出 50 个城市中的 25 个。在所有配成对的城市中, 老二认出的城市恰好有 80% 的机会比他不认识的城市大, 也就是说他的再认效度 α 是 0.8。老大最有学问, 听说过所有 50 个德国大城市。在所有配成对的城市中, 他有 60% 的机

会都能作出正确选择,也就是说他的知识效率 β 是 0.6。

不管兄弟三人实际上做了什么,我们假定他们都使用了再认启发式,那么他们当中谁将在测验中获得最高分数呢?图 2-3 表明了兄弟三人的成绩,它是用公式(1)计算出来的。图中将各个点连接起来的平滑曲线说明公式(1)计算出来的变量应该是连续的。

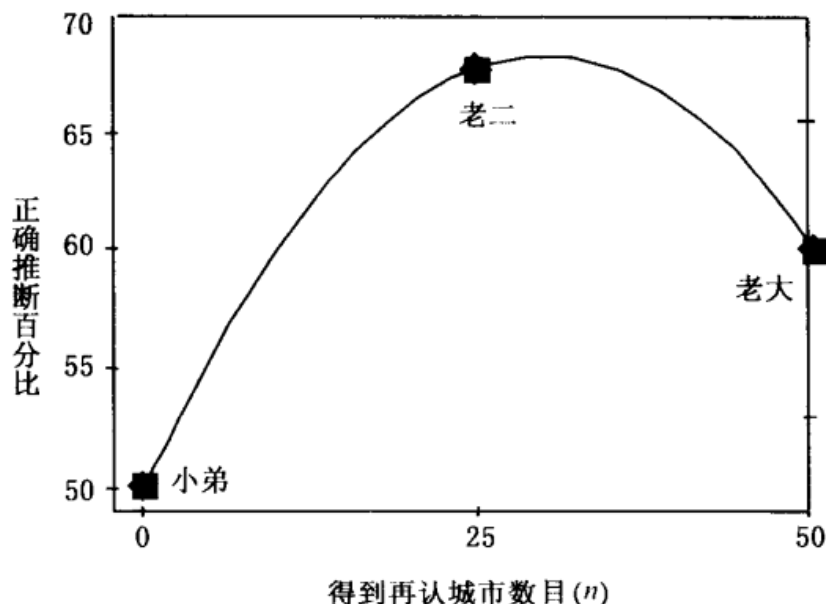


图 2-3 “少即多效应”的一个范例。小弟从未听说过任何德国城市,在机遇水平上作出推断。老二能够再认 50 个城市中的一半,因此在大约一半问题上可以应用再认启发式,这使得他可以作出大约 67.5% 的正确推断[根据公式(1)计算; $\alpha=0.8, \beta=0.6$]。老大听说过所有城市,因此比老二知道得更多,却只作出了 60% 的正确推断——一种被称为“少即多效应”的现象。连续的曲线表明了不同再认水平的中间状态[均可以用公式(1)计算出来]。请注意:曲线的顶峰并非出现在老二的成绩上,而是在他的成绩稍微偏右一点,其原因在于 β 值是 0.6 而不是 0.5。

小弟的成绩刚好在机遇水平,老大的成绩稍好一些,达到了 60% 的正确率。引人注意的是,虽然老二比老大知道的更少,但却作出了最为准确的推断。他是兄弟三人中唯一一个能够使用再认启发式的人。而且,他能够充分利用他的无知,因为他碰巧只能再认这些城市中的一半,这使得他能够最频繁地使用再认启发式。因此,再认启发式导致了一种似乎自相矛盾的情况:知道得更多的人反而不如知道得更少的人作出的推断准确。



何时产生少即多效应?

我们可以采用通用的专门术语来表述少即多效应产生的前提:在通常可以应用再认启发式的双项选择测验中,每当再认效度 α 大于知识效度 β 时,少即多效应就会产生。

如果这个条件不具备,那么当越来越多的对象得到再认时,推断准确性将随之增加。下面我们将通过数学推导来证明这一点(Goldstein & Gigerenzer, 1998)。

数学论证总是建立在简单假设基础之上的。例如,在上述例子中我们实际上已经假定:再认效度 α 在图2-3的X轴上保持恒定。也就是说,图2-3仅描绘了具有不同知识状态、但却有固定 α 的多个个体(兄弟三人)的推断成绩。与该图所描绘的情形不同,当某个人能够逐渐再认越来越多的对象时,再认效度通常是随之变化的。对于具有不同再认水平的许多不同个体而言,让每一个个体具有同样的再认效度是可能的。也就是说,不管他们各自能够再认的具体数目是多少,得到他们再认对象中具有更高效标值(即用某种效标对其所作的等级判断,如:大或小、高或矮、多或少等)的对象所占比例可能是一定的,这种比例被称为 α 。然而,当一个单独的个体逐渐能够再认越来越多对象时,再认效度将发生变化,因为每一个得到再认的新对象将会增加或降低再认效度,这取决于它自身效标值的大小。这就是说,每当再认了一个具有更低效标值(如,更小)的对象时,就会降低再认效度,而每当再认了一个具有更高效标值(如,更大)的对象时,就会增加再认效度。

因此,必须提出这样一个问题:当 α 并非随着 n 的变化保持恒定时,我们能够用现实的序列学习任务来论证少即多效应吗?换句话说,在双项选择任务上,当得到再认对象并非总是(按照确定的比例)具有更高效标值(如,更大)时,少即多效应还存在吗?

为了弄清楚这个问题,我们编制了一个计算机程序,让其按照德国城市的著名程度顺序逐个学习它们。为了估计出这种顺序,

我们对 66 名芝加哥大学学生进行了调查,要求他们从一个列表中选择他们能够再认城市的名字,然后根据每个城市被选中的人次数排列其顺序。通过这种资料,我们希望能够粗略反映德国各个城市在美国人心目中的知名度。计算机程序按照上述排列好的顺序逐个学习再认每个城市。例如,首先学习再认德国最著名城市慕尼黑,紧接着接受一次配对判断测验,即让其判断配成对的德国城市中哪个有更多人口。在熟悉了慕尼黑之后,接下来学习再认德国第二个最著名城市柏林,并接受同样测验。以此类推,直到能够再认所有城市。在一种条件下,计算机仅仅学会识别城市的名字,因此只能借助于再认启发式作出所有推断。其推断结果用图 2-4 最底部的曲线(“无线索”)来表示。所有城市都未得到再认时,其推断成绩处于机遇水平。随着学习的不断深入,出现了一个倒“U”形的曲线,就像图 2-3 一样。然而该曲线不像图 2-3 那样平滑,而是显得参差不齐,因为如上文所述,在这里再认效度并非恒常不变,而是随着得到再认城市数目的增多自由变化的。

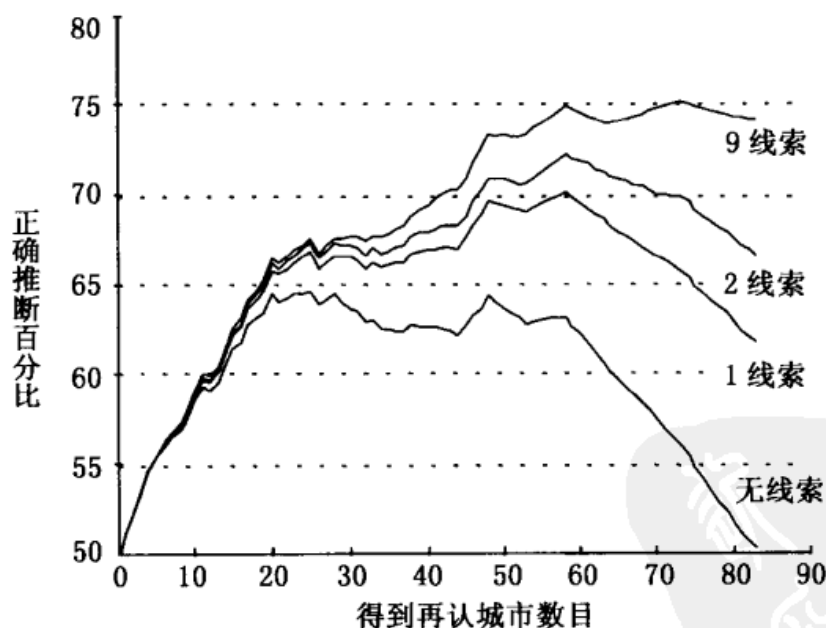


图 2-4 城市按照一定顺序逐个得到再认时的少即多效应。这个顺序来自实际再认数据,被试仅依靠再认(无线索)或辅之以 1 条、2 条或 9 条线索作出推断。

如果计算机程序并非仅仅学习再认城市名字,而且还学习了其他一些对于推断城市人口有用的信息,少即多效应会消失吗?为了回答这个问题,我们设置了一系列附加条件。在这些条件下,计算机除了要学习再认城市名字外,还要学习 1 条、2 条或 9 条与推断城市人口有关的线索(参见 Gigerenzer & Goldstein, 1996a)。在“单线索”条件下,在计算机学会再认一个城市名字后,还要了解这个城市是否有过一个展览会。被选为设展地点是一个城市人口多少的一项非常有力的预测指标,其生态效度高达 0.91(参见 Gigerenzer & Goldstein, 1996a)。^{*} 然后计算机采用一种被称为“采纳最佳”(take the best)的决策策略作出哪个城市更大的推断。在此我们尚无意详述“采纳最佳”策略,只需要知道它是一种依据多条线索作出推断的非常准确的(像多元回归一样准确)策略,而且将再认启发式作为其首要步骤。

有关展览会的附加信息“洗刷掉”(wash out)了少即多效应 49 吗? 没有。曲线的顶峰稍微向右移动了一点,但仍然保持了倒“U”形状。即便附加了关于展览会的信息,当得到计算机再认的城市超过了 58 个时,推断准确性也开始降低。在“双线索”条件下,计算机知道一个城市是否举行过展览会和是否有大的足球俱乐部——另外一条附加线索,其生态效度为 0.87。正像随着知识增多我们所预期的那样,这种条件下少即多效应有所降低,但仍然十分明显地表现出来。能够再认所有城市并了解两条线索包含的所有信息(该曲线上最靠右边的点)居然导致了比仅能再认 23 个城市更加不准确的推断。最后,在“9 线索”条件下,计算机拥有关于适用于它的所有 9 条线索的信息。对于预测德国城市人口来

^{*} 0.91 的生态效度意味着当一个城市举行过展览会而另一个没有时,有 91% 的可能性前一个城市更大。生态效度是一条线索与一种效标之间的关系,与特定的个人无关。它与知识效度不同,后者是对两个对象均能再认时一个人获得正确答案的比率,不管多条线索值的大小。生态理性是根据一个对象所具有的某种特性来界定的,例如两个城市均被再认,但其中举行过展览会的城市更加特殊。 α 和 β 均属于特定个人的特征。

简捷启发式

说,这实际上比大多数德国居民知道的还要多。如此多的知识应该足以超过无知的作用了吧?图2-4显示,少即多效应终于被“摆平”(flattening out)了。然而它并未完全销声匿迹:即便所有747(9×83)条线索都被掌握了,而且所有城市都得到了再认,该曲线最靠右边的点仍然比曲线上多于四分之一的部分要低。适当的无知能够比详尽的知识获得更高的准确性!

该项计算机模拟研究得到了两个主要结果。对于预测少即多效应来说,再认效度 α 必须保持恒定的简单假设并非必不可少。而且,即便提供了关于9种预测指标的完整知识,这种反直觉现象仍然存在。

至少在三种不同情形下我们可以观察到少即多效应。首先,在对两个群体的推断成绩进行比较时,可以发现这种现象。在给定领域中,拥有较多知识的群体比拥有较少知识的群体经常作出更加不准确的推断。一个典型的例子是前面介绍过的德、美两国学生关于圣地亚哥和圣安东尼奥哪个更大的推断。其次,当对两个不同领域进行比较时会产生这种效应,即与熟悉的领域相比,同样一组人对他们不甚了解的领域作出了更多准确推断。下面我们将提供这方面的实验范例。第三,在对不同时间作出的推断进行比较时也存在这种效应,即随着对某一个领域了解的增多,作出的推断反而越来越不准确。例如,图2-4的模拟实验结果说明了准确性是如何首先增加、继而随着知识增多而降低的。

迄今我们已经从数学上论证了何时会产生少即多效应,并证明它在违背数学模型假设的实际学习情境中也会出现。但是这种效应在现实的人类身上也能观察到吗?很可能进化已经掩盖或抹去了再认启发式具有的优势和准确性!在下面的部分中,我们考查了人类判断是否遵循再认启发式,以及少即多效应是否能够从实验上加以证明。

人们使用再认启发式吗？

我们通过一项简单实验来检验未受任何指导的人们是否会自发地使用再认启发式。从德国城市中分别挑选出 25 个($n=6$)或 30 个($n=16$)最大城市,将它们一一配对,分别组成 300 个或 435 个双项选择问题。在每对城市中都要求美国被试选出人口较多的城市。之后,我们让被试把自己在测验中选择的城市与他们在测验前或测验后再认了的城市(顺序是随机安排的)进行核对。根据这种信息,我们能够算出被试共有多少次机会按照再认启发式作出选择,并将其与他们实际上按照再认启发式进行选择的次数进行比较。图 2-5 是 22 个美国被试的实验结果。请注意:再认启发式预言将会存在一定的个体差异。由于人们再认出来的具体城市不同,他们关于各个城市人口数量的推断也将会有系统变化。

50

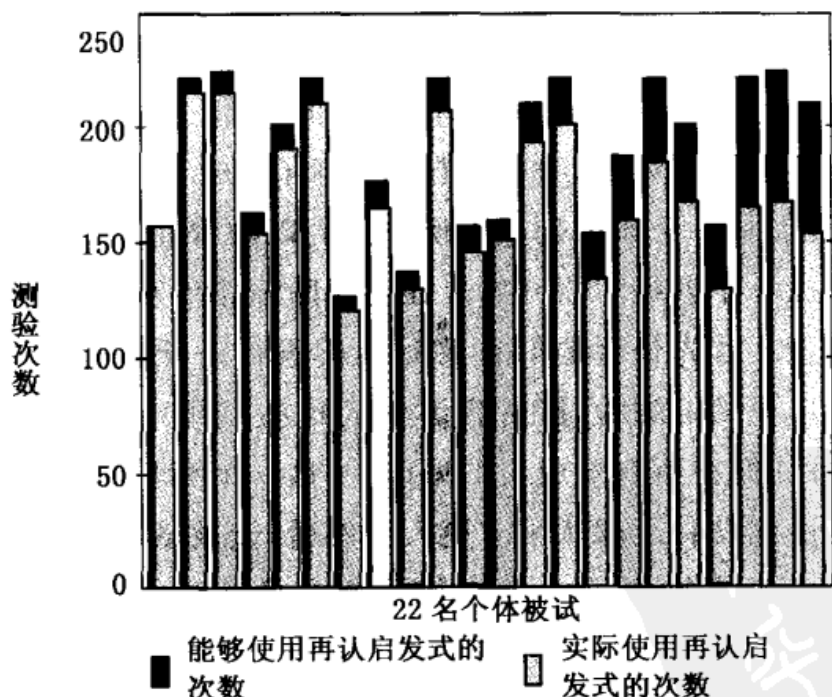


图 2-5 能够使用再认启发式的几率和 22 名被试实际使用它的几率。按照各个被试的判断与再认启发式相符的程度从左到右排列他们的成绩。暗条形框的高度各不相同,因为不同被试再认的城市数目不同。

每个被试的成绩都用两个条形框来表示。暗条形框表示一个人共有多少次机会使用再认启发式,亮条形框表示这个人实际上用再认启发式作出判断的次数。例如,最左边的一对条形框表明,这个人共有 156 次机会按照再认启发式进行选择,而且每一次他都这样做了;第二个人共有 221 次机会采用再认启发式作出选择,实际上他有 216 次这样做了,如此等等。被试按照再认启发式作出选择的几率在 73% 到 100% 之间,其中数为 93% (平均数是 90%)。

这项简单实验表明,大多数情况下人们是遵循再认启发式的。下面让我们对它进行更加严格的实验检验:当给出了其他与之相左的信息时,人们还会依赖于它吗?

尽管存在矛盾信息,人们仍然使用再认启发式吗?

在这项实验中,我们教给被试一种与再认启发式相互矛盾的信息,即关于城市是否有足球俱乐部的信息——在德国这是一项有关城市人口多少的非常有力的预测指标。我们想要看一看人们将会选择两个城市中的哪个作为较大城市:未被再认城市,或是得到再认、但却获悉其没有足球俱乐部的城市。为了了解我们的被试能够再认哪些城市,我们安排了一项有 26 个被试参加的预备测验,让他们从一个城市列表中挑出他们以前曾经听说过的城市。

实验从一个训练阶段开始。在这个阶段指导被试写下所有要使用的信息。首先告诉他们将要对他们进行一项有关 30 个德国大城市人口规模的测验,接着告诉他们这些城市中的 9 个有足球队,并告诉他们在所有配成对的城市中,这 9 个有足球队的城市有 78% 可能性比没有足球队城市更大。然后,让被试随机从 30 个城市中划去 8 个城市,并看一看它们是否有足球队。作为一种实验操纵,我们使每个被试都能划去同样的 4 个有足球队的著名城市和 4 个没有足球队的著名城市。我们对被试进行了测验,以确保他们能够确切地复述这些信息,否则实验将不继续下去。在此之前或之后,向被试出示一个德国城市列表,要求他们标出参加实验

之前曾经听说过的那些城市。

让被试将记录着上述信息的便条放在身旁,然后我们向他们呈现配对的德国城市,要求他们从每对中选出较大的一个。为了激发他们认真严肃地从事这项任务,告诉他们如果达到一定正确率的话,将有机会得到 15 美元报酬。需要重申的是,实验的关键之处在于考查被试将会选择哪个城市作为较大城市:他们以前从未听说过的城市,还是预先能够再认、但刚刚获悉没有足球队的城市?从训练阶段提供的信息(它们未涉及任何再认问题)来看,人们将会预期:与前面介绍的实验相比,本实验中将会有更多被试选择未被再认城市。为什么?这是因为未被再认城市要么有足球队,要么没有足球队,二者必居其一。如果它们有足球队(从前面提供的信息看其概率为 $5/22$),那么从有关足球队的信息来看,它们有 78% 的概率是更大的;如果它们没有足球队,那么有关足球队的信息将会成为无用的,必须作出盲目的猜测。训练阶段提供的信息对未被再认

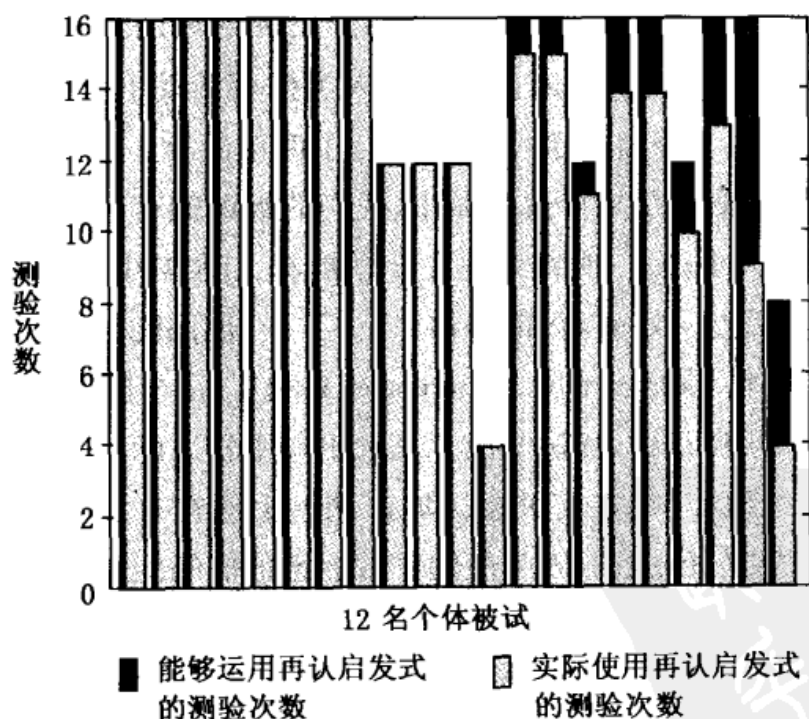


图 2-6 存在矛盾信息情况下按照再认启发式作出选择的几率。按照被试判断与再认启发式的相符程度从左到右排列其成绩。暗条框高度不同,因为不同被试再认的城市数目不同。

简捷启发式

城市是有利的,因为无论它们拥有足球队的几率有多大,都有利于将其判断为更大城市。该实验的结果示于图 2-6。

像图 2-5 一样,每个被试的成绩用明、暗两个条形框来表示,其含义也同图 2-5。暗条形框高度各不相同,因为不同被试实验前能够再认的城市不同,以至于应该能够应用再认启发式的次数也各不相同。21 个被试中有 12 个完全按照再认启发式作出了选择,而其他被试中的多数人仅仅在一二次测验中偏离了再认启发式。总的来说,在总共 296 次关键性配对测验中,被试有 273 次都遵循了再认启发式。按照启发式进行推断所占比例的中位数高达 100%(平均数为 92%),尽管存在与之矛盾的信息。这些数据几乎像上一项实验一样高。很显然,附加信息并没有被应用于推断过程,这是符合再认启发式的。

在人类推理活动中少即多效应会产生吗?

我们已经证明,再认启发式可以描述在一些任务上人类是如何进行推断的。这种结果对少即多效应将会出现的理论预测提供了实验支持。但是,迄今我们尚未在现实的人类推理活动上看到这种效应。我们曾经对 52 名芝加哥大学学生进行了两项测验:一项测验是有关美国 22 个大城市的,对于这些城市,我们的被试拥有充分信息来推断其人口规模大小;另一项测验是关于德国 22 个大城市的,对于这些城市他们知之甚少,或者仅仅能够再认它们——他们能够再认的不到这些城市的一半(Goldstein & Gigerenzer, 1998)。每个问题都由两个随机抽出的城市所组成,被试的任务是选出其中较大的一个。人们也许会期望这些美国学生在有关本国城市测验上比在有关外国城市测验上取得明显更好的成绩,因为他们非常熟悉他们的国家。我们则将这项研究作为对少即多效应耐用性的一种考验。相对于纸上谈兵式的理论论证而言,少即多效应这种奇特现象是难以用现实人类被试来证明的,因为我们前面提出的理论和提供的计算机模拟研究都涉及了不确定条件下的推断问题,但现实的人类常常具有关于特定效标的确切知识。例如,许多美国人和几乎所有芝加哥大学学生都能够按顺序说

出美国 3 个最大城市。仅凭这一点,他们就能够正确回答所有问题中的 26%(读者不妨用排列组合规则算算看)。如果一个人能够按顺序说出 5 个大城市的名称,那么他将能够正确回答 41%的问题。这种有关城市排列顺序的确切知识,加之美国人所具有的关于他们城市的一些日常知识,本应该能够使他们在有关本国城市测验上获得的成绩远远高于在有关外国城市测验上获得的成绩的。

实验结果是,这些美国被试在有关美国城市测验上所获得正确率的中位数是 71%(平均数是 71.1%),而在有关较不熟悉的德国城市测验上正确率的中位数为 73%(平均数为 71.4%)。尽管具有关于本国城市的可靠知识,包括一些关于哪个城市较大的非常具体的知识,再认启发式仍然表现出了些微的少即多效应。对于其中一半被试,像前面介绍的两个实验一样,我们记下了他们能够再认的那些城市。相对于他们总共应该能够采用再认启发式的次数而言,他们实际按照再认启发式进行选择次数所占百分比的中位数是 91%(平均数是 89%)。而且,当被试能够再认的德国城市平均达到 12 个(大约为总数的一半)时,他们采用再认启发式的几率达到了最高。在与这项研究恰恰相反的一项研究中,以熟悉德国城市的奥地利学生为被试,获得了类似的少即多效应:这些学生对美国城市比对德国城市作出了更加准确的推断(Hoffrage, 1995;亦可参见 Gigerenzer, 1993)。

再认起源于何处?

对一些重要适应性任务来说,诸如避免食物中毒和辨认血缘关系等,有机体似乎天生禀承了按照再认启发式行动的倾向。野生挪威鼠不需要学习就更喜欢能够再认的食物,而不是新异食物。如果一种选择有危及生命的后果,那么不得不学习再认启发式的有机体也许在其学会这种启发式之前就已经死于非命了。血缘关系辨认也是一种重要适应性任务,其机能似乎在于避免乱伦和增进亲和力(包括适配性)(Holmes & Sherman, 1983)。例如,雌性黄蜂利用气

简捷启发式

味(它们在巢穴中学会识别的气味)再认来推断另一只黄蜂是否属于同胞姐妹。有人曾将刚刚产生的蜂王转移到另一个巢穴中来捉弄它,以考验这种机制的通用性,结果发现在那里该蜂王仍然能够学会识别同巢(非亲属)黄蜂的气味(Pfenning et al., 1983)。另一方面,还有一些领域,在那里有机体通过经验懂得了再认的预测效力。下面让我们详细考查一下人们对地理名称的再认究竟起源于何处。

新闻媒体对我们再认适当地理名称究竟起多大作用?如果其作用是相当大的,那么一个城市在新闻媒体上被提到的次数就应该与能够再认该城市的读者所占比例有明显相关。仅在伊利诺斯州,《芝加哥论坛》周日的发行量就达到了 100 多万份。我们对 1985 年到 1997 年 7 月间《芝加哥论坛》上发表的有关“柏林”与“德国”二词联系起来的文章作了一个粗略统计,共有 3 484 篇。采用同样方法,我们对德国超过 10 万居民的城市都进行了统计。《芝加哥论坛》上对一些城市名称的拼法与世界上流行的拼法不太一致。我们发现“Nuremberg”曾经被拼写为“Nurnberg”、“Nurnburg”、“Nuernberg”和“Nuremburg”等(还不包括母音变化)。因此,在一些情况下,我们不得不努力从记忆中搜索,甚至凭借想像确定它指的究竟是哪一个城市。表 2-1 说明,对 12 个德国大城市来说,报纸上提及其名字的文章数目是其名字能否被再认的一项很好的预测指标。其替代相关(即报纸上提及一个城市名字的文章数目与该城市得到再认次数之间的斯皮尔曼相关)达到了 0.79。但是,各个城市的实际人口规模又怎样呢?其生态相关(即报纸上提及其名字的文章数目与其实际人口数量之间的相关)达到了 0.70。最后,人们再认城市名字的次数与实际人口数量的相关为 0.60。*

55

这些研究结果显示,个体再认与新闻媒体更加一致,而不是与实际环境更加一致,它表明对于城市名字的再认也许主要来源于

* 该相关反映了平均的再认效度。它是综合了许多人的成绩计算出来的,而再认效度应该是单个人的特征。效度与相关之间的关系将在第 6 章详加分析。

新闻媒体。对于大多数人来说,人口规模是未知的,但他们能够仅仅依靠再认作出相当准确的猜测。

然而,在不同文化背景中,这种结果仍然能够站得住脚吗? 我们曾经查阅了一份德国大报《时代周刊》(*Die Zeit*),记录了超过 10 万人口的每个美国城市在该报上被专文提及的次数。我们将其与能够再认每个城市的奥地利萨尔茨堡(Salzburg)大学学生的比例(Hoffrage,1995)加以比较。表 2-1 再一次显示,新闻媒体十分准确地预测了能够再认各个城市名称的人数比例。具体地说,报纸上文章的数量与各城市再认之间平均的替代相关是 0.86,文章数量与各城市实际人口之间的生态相关为 0.72,而再认与城市人口排列顺序之间的相关为 0.66。这和来自美国被试的实验结果是相当一致的,各项相关还略微高一些。在两种情况下,都是替代相关最强,生态相关次之,再认与效标之间的相关最弱。下一部分我们看一看有关机构或部门是如何借助于广告来充分利用这种关系的。

表 2-1 对德国和美国城市的再认

城 市	文章数量	再认(%)	城 市	文章数量	再认(%)
柏 林	3 484	99	纽 约	493	100
汉 堡	1 009	96	洛 杉 矶	300	100
慕 尼 黑	1 240	100	芝 加 哥	175	97
科 隆	461	82	休 斯 顿	73	80
法兰克福	1 804	96	费 城	67	63
埃 森	93	28	圣地亚哥	78	47
多特蒙德	84	19	凤 凰 城	56	53
斯图加特	632	63	达 拉 斯	39	100
杜塞尔多夫	381	81	圣·安东尼奥	4	23
不 来 梅	140	44	底 特 律	66	80
杜伊斯堡	53	7	圣 约 瑟	13	17
汉 诺 威	260	88	印第安纳波里	20	50

左边:12年间《芝加哥论坛》发表的有关德国12个大城市的文章数目和芝加哥大学67名学生能够再认每个城市的人所占比例。按照这些城市的实际人口排列其顺序。右边:2年中《时代周刊》发表的有关美国12个大城市的文章数目和萨尔茨堡大学30名学生能够再认每个城市的人所占比例。

充分利用再认启发式

奥里维罗·托斯卡尼(Oliviero Toscani)是一系列臭名昭著的班尼通(Benetton)广告活动的幕后策划者。他敢于拿自己的职业生涯下赌注——在他策划的一系列广告中从不提及有关产品的信息,而是用诸如躺在一滩血迹中的死尸或死于艾滋病的病人等令人惊骇的画面来诱使人们再认其所推销产品的名称。在他所写的一本书中,托斯卡尼(1997)报告,这种广告活动是极其成功的,使得班尼通的知名度一跃而超过了香奈儿(Chanel)公司的知名度,并跻身于全球五大名牌行列。那么,一个企业的知名度,无论如何取得的,都将有益于商业活动吗?在社会领域,知名度常常与财富、资源、品牌质量、权力等密切相关。广告商花费巨额资金,只是为了能够在普通公众的再认记忆中占有一席之地。我们已经逐渐习惯于观看仅提及产品名称而不涉及产品信息的广告(这在一个人出国访问时尤其明显,在那里他甚至不知道广告中的产品名称指的是什么)。无名的政客、大学、城市,甚至较小的国家常常四处活动,以求对他们名字的再认。他们全都信奉一条原则:如果我们能够再认他们,我们将会赞许或支持他们。

有证据表明,一个人能够不知不觉地、甚至无意识地产生对一个名字的再认。杰克比及其同事(Jacoby, Kelley, Brown & Jas-echko, 1989; Jacoby, Woloshyn & Kelley, 1989)所做的“一夜成名”(overnight-fame)实验证明:人们常常对一个名字是否在前面的实验系列中出现过或是他们在实验前就曾接触过它感到困惑。他们的研究表明:向人们呈现一些不著名的名字,过一个晚上之后,让他们对这些名字和其他一些真正著名的名字进行声望判断,结果他们将这些不著名名字和著名名字搞混淆了。这说明再认感是如何愚弄我们,致使我们将普通人误认为著名人士的。

仅能再认与了解程度

我们将再认视为一种“双极”(binary)现象:一个人要么能再认,要么不能再认。人们对某种东西的了解程度在主观上是难以评估的,同时对于快速节俭启发式而言也是无关紧要的。这两种特征,即再认的双极性和更多知识的无关紧要性,将再认启发式与诸如“可用性”(availability; Tversky & Kahneman, 1974)、“熟悉性”(familiarity; Griggs & Cox, 1982)或“知晓感”(the feeling of knowing; Koriat, 1993)等概念区别开来。“可用性”和“熟悉性”这两个术语常常被用作一种普通意义上的解释模型而不是过程模型。可用性适合于记忆中的项目,而且常常通过项目被回忆的顺序或速度,或一个类别能够产生的范例数量来测定(参见第10章)。相反,如同图2-1表示的那样,再认主要涉及到了记忆内、外项目的差异(Goldstein, 1997)。可用性是关于回忆的,而不是关于再认的。“熟悉性”通常被用来说明一个人关于某种任务或对象所拥有知识或经验的程度,它并未涉及对于再认启发式来说最重要的区分——即能够再认对象和不能再认对象之间的区分。尽管可用性和熟悉性之类的概念看起来是非常形象直观的,但仍然需要将其字面意义转换为精确的启发式模型(Gigerenzer, 1996)。如果这样做了,那么我们就可望对其作出更加深入、更加详细的理解,这有可能导致包括少即多效应在内的意想不到的后果。

用考雷特(Koriat)的话来说,“知晓感”就是一个人对未来能够从记忆中提取某些信息的可能性的评估。例如,“加拿大的总理是谁?”这样一个简单问题可能使许多非加拿大人处于一种“舌尖状态”(A tip-of-the-tongue state),即产生一种明明知道他是谁、却说不出的感觉。与再认启发式不同,知晓感预先给定了一种不仅仅能再认的知识,即问题中所包含的信息。另外一种关键性区

别是：再认启发式能够利用再认预测现实世界中的一些效标，而知晓感仅能预测未来的记忆成绩。

作为快速节俭启发式原型的再认启发式

本书中我们主要研究快速节俭启发式的结构原理和效用。再认启发式是所有这些适应工具中最为简单的工具。它使用了再认这种经数百万年进化造就的能力，使得有机体能够从其自身的无知中受益。这种启发式能够快速地生效，而且仅使用有限知识——甚至需要一定程度的无知。它用于搜索、终止搜索和作出决策的构筑块令人惊奇地简单。搜索仅限于再认记忆——不需要任何超出再认之外的知识回忆。由于搜索活动受到了限制，终止搜索的规则也是相当简单的——一旦两个对象中的一个得到了再认，立即终止搜索。结果是，决策也只建立在单一信息基础之上，那就是再认。由于缺乏再认对于作出一种决策来说是必不可少的条件，所以我们将这种启发式称为“基于无知的决策制定”(ignorence-based decision making)。这种启发式规则代表了一种规避矛盾策略，即毋需在指向不同方向的多条线索之间进行权衡(就像一个人再认了一个城市，但却获悉其没有足球队情况下所做的那样)。

58

快速节俭启发式，包括再认启发式，是建立在再认之类的心理能力和基于无知的决策制定及单一理由决策制定(仅依靠一种信息而不是多种信息的结合)之类的启发式规则基础之上的。人们常常力图规避权衡而集中注意一个好理由的现象已经得到了无数次证明(例如，Baron, 1990; Hogarth, 1987; Payne et al., 1993)。然而，许多学者(包括心理学家)仍然不相信这些启发式规则的效力，而将其视为一厢情愿的执迷不悟和非理性的。我们不这样认为。再认启发式不仅仅是一种合理的认知适应机制，因为在许多有限知识情形下人们别无选择，而且也的确是适应于环境的，因为

在一些情形下(包括本章论及的那些情境)缺乏信息反而能够比获得大量信息作出更加准确的推断。在这些情形下,我们可以说再认启发式从生态学上看是合理的,因为它能够用一种简单优雅的方式利用环境的信息结构。

第3章

无知能在股市赢利吗

本哈德·伯格斯特(Bernhard Borges)

丹尼尔·G·古德斯腾(Daniel G. Goldstein)

安德鲁斯·奥特曼(Andreas Ortman)

哥德·吉戈伦尔(Gerd Gigerenzer)

一个好名字胜过巨大的财富。

——塞万提斯(唐·吉珂德)[Cervantes(*Don Quixote*)]

为

了制定成功的股票投资方案,一个投资者需要多么复杂的决策工具?为了实现这个目标,一个人必须获得多少种特殊信息?职业投资公司用来进行投资决策的工具和信息非常人所能及,而且专家建议的价值一直受到怀疑。用亿万富豪华伦·巴夫特(Warren Buffet)的话来说:“股票预测者的唯一价值是让算命先生看上去也是好的。”(1987, p. 40)在本章中我们提出了一种快速节俭启发式——它充分利用人们知识的贫乏而不是特殊的市场信息和工具来制定投资方案。特别是,我们考察了被称之为再认启发式(第2章已详述)的基于无知的决策机制在股市上是否能够赚到钱。这种启发式仅依据一则信息——对上市公司名字的再认——来作出投资决策。不需要搜集关于上市公司的任何具体信息,也不需要利用任何复杂的分析和统计工具,唯一需要的是“适度的无知”(a beneficial degree of ignorance)。

我们将再认启发式与5种标准股票选择程序——互助资金、市场指数、随机化或“镖靶”投资组合、个别投资决策,以及陌生股

票投资组合——作了直接对比。本章比本书中其他任何一章都走得更远：我们在充满险情的金融领域进行了一次大胆尝试——将拟议中的尚不成熟的快速节俭启发式投进了瞬息万变、利益巨幅波动、无常规可循的声名狼藉的股票市场。它能够很好地起作用吗？我们投入的资金将会得到什么样的回报呢？ 60

投资理论与实践

股市的难以预测性是尽人皆知的，这对于再认启发式这样的简单投资策略来说是一项严峻考验。许多理论家和实践者抱着股票市场定能赚钱的想法，将无数时间和巨额资金投到了股市上。有人断定，股市上不可能有常胜将军。例如，虽然新经典主义经济学家们把投资者理解为拥有无限理性的投资群体，完全有能力形成合理的投资预期，但这种观点受到了“高效市场假说”（the efficient market hypothesis，缩写为 EMH）* 的反对，后者坚持认为，长远地看投资者不可能得到高于平均收益的赢利（例如，Lucas，1980；Muth，1961）。用库特纳（Cootner，1967）的话来说：

如果某个投资群体对股票价格的预测总是比多数人的平均预测更加准确，那么他们将使目前的股票价格更加接近其真实价值。相反，在预测准确性上不如多数人平均水平的投资群体所起的作用将会越来越小。果真如此的话，那么目前价格将是反映未来价格的最好信息。（p. 80）

尽管当初受到了一些挑战（例如，Rozef & Kinney，1976；Special Issue，*Journal of Financial Economics*，1978），但是高效市场

* 具体来说，EMH 不仅假定经纪人是足够理性的，而且假定每一个经纪人都知道其他经纪人也是按同样方式行动的。

简捷启发式

假说仍然被纳入到了主流的标准经济学模型之中。例如,它已经成为广为流传的“资本资产定价模型”(capital asset pricing model)的有机组成部分,该种模型形成了当代有价证券(管理)理论的基础。

许多专业投资分析家(例如, Soros, 1994)一直怀疑高效市场假说的真实性和适用性,并转向了充分利用投机机会的“技术交易模型”(technical trading models)。事实上,近期有证据表明技术交易模型在市场上取得了微小但却显著的收益(例如, Brock et al., 1991)。而且,以行为经济学家们(例如, Arthur et al., 1997; Debondt & Thaler, 1985; Shleifer & Summers, 1990; Thaler, 1992, 1993)的独到观察为基础的当代金融界也已经开始认识到高效市场假说的局限性,并开始关注理性受限条件下人们的心理期望是如何影响金融市场的。

不管学术研究领域发生了什么变化,许多投资基金专营机构的实际业绩却表明,长远地看要想在市场上保持长盛不衰是非常困难的。例如,美国几家主要投资和联营公司的记录表明,大多数经验丰富的投资专家的业绩都远远低于市场预期。^{*} 这样严峻的事实对目前金融界惯用的复杂运作模型无疑是莫大的嘲讽。的确,1996年美国几乎75%的股票专营机构的业绩都低于“标准普尔500”(“the Standard & Poor's 500”,美国一种通用的股票指数,常缩写为“S & P 500”)的业绩标准(Kadlec, 1997)。

虽然投资专家们努力搜寻尽可能多的信息,并采用最好的金融管理模型对这些信息进行分析和处理,但从理论上讲他们仅能获得平均收益,从实践上看他们获得的收益常常更低。在这种情形下,我们还敢奢求“目光短浅的普通人”将他们有限的认知和财

^{*} 当然,这种糟糕业绩至少应该部分地归咎于两种周而复始的“费用”因素:第一,资金需要流通,因此并非收到的用于投资的钱都被投进了股市,而是以现金账户形式回笼,这部分资金如果有收益的话,也只是很少一点。其次,专家建议是有代价的,被征去的许多管理费没有或很少被算到投资基金的收益上去。

力资源投放到金融市场上去吗？

基于无知的投资决策

由于专业技术和知识在预测股市方面尚不像人们通常推测和断定的那样有用，所以人们一定会对基于无知的投资启发式在股市上如何生效感到困惑。无知，只有当它系统地而非随机地分布时，才会有利于再认启发式发挥作用。最初，古德斯腾和吉戈伦尔是将再认启发式作为解决双项选择问题的一种机制提出来的，其含义可以用相当简单的语言来表述：当在两个对象中进行选择、且其中只有一个能够再认时，选择能够再认的那一个。推而广之，我们可以将这种规则应用于从一套对象中选出一个子集的情境，即：当从一套对象中选择一个子集时，选择能够再认的那些对象。这样一来，再认启发式就可以被用作进行投资决策的一种策略。

对于一个个体而言，再认启发式意味着只选择他或她能够再认的那些股票。对于一个群体的集体再认（如同我们将要介绍的研究所做的那样）来说，再认启发式意味着：选择该群体一定比例（如，90%）人数能够再认的所有股票。

要想使用再认启发式，需要一定程度的无知（即缺乏再认）。例如，能够再认所有股票的金融专家不可能使用再认启发式在多种股票中进行选择。另一方面，从未听说过任何股票的完全无知的人也不能使用这种启发式规则。在此二者之间，大多数人处于我们所谓的“适度的无知”状态。

敢于拿再认启发式这个“卵”来击股市这块“石”并非一定是愚蠢之举。有两个理由使我们深信再认启发式将不至于一败涂地。首先，消费者倾向于选择他们听说过的商品——被广告商们利用的一种行为——而从本质上说股票就是一种其价格由人们的选择行为决定的“商品”。其次，一些成功的投资专家（例如，Peter Lynch）指出，缺乏对某种股票名字的再认是从多种供选择股票中

62

让我们更精明

简捷启发式

排除它的重要依据(Lynch, 1994)。那么,再认启发式能够与“高深莫测”的理论家们和华尔街(Wall Street)上猎取股市利润的无所不知的投机商们所利用的工具和知识相媲美吗?更具体地说,利用再认启发式能够制定出至少能在市场水平上运作的投资方案吗?

公司再认

我们曾经要求德国人和美国人从一个公司列表中指出哪些是他们能够再认的公司。表中列出的公司均为在纽约股票市场和德国几家股票市场上上市的公司。共调查了 480 个人对 798 家公司的再认情况。这些公司中包括 500 家美国“标准普尔 500”公司[其中包括“道琼斯 30”(Dow 30)公司]和 298 家德国公司[其中包括“达克斯 30”(Dax 30)公司]。受调查者分为四组:美国普通人、美国专家、德国普通人、德国专家。普通人是 360 个芝加哥和慕尼黑商业区的普通行人,其中每个人都对所有公司中的 1/6 提供再认信息。专家是芝加哥大学和慕尼黑大学 120 位金融或经济专业的研究生,其中每个人都对所有公司中的一半提供再认信息。表 3-1 列出了得到各组受调查者一致再认公司的名字,而图 3-1 表明了得到一定比例人数再认公司的数目。例如,有 8 家德国公司得到了 100%德国普通人的再认,有 25 家德国公司得到了至少 90%德国普通人的再认。只有 1 家德国公司——汉莎——得到了 100%美国专家的再认,只有 5 家德国公司得到了至少 90%美国专家的再认。

哪一组再认了最多公司?哪一组最为无知?美国专家再认了最多公司,接下来依次是美国普通人、德国专家和德国普通人。有点出乎意料的是,德国专家比芝加哥商业区的普通行人再认的公司还少。造成这种结果的原因可能有二:一是美国公众更为积极地参与股票投资活动,二是美国比德国有更多的股票。跨国公司再认率最低,如受测的美国行人居然全都连一个德国公司也不能再认。

表 3-1 得到每组被试中所有人再认的公司(G=德国公司;US=美国公司) 63

德 国 普 通 人		美 国 普 通 人	
Allianz AG	G	Amoco	US
Bayerische Vereinsbank	G	Chrysler Corp.	US
Commerzbank AG	G	Coca-Cola Co.	US
Daimler Benz	G	Ford Motor Co.	US
Dresdner Bank AG	G	General Mills	US
Lufthansa	G	Hilton Hotels	US
Porsche AG	G	Maytag Corp.	US
Siemens	G	Sears Roebuck & Co.	US
德 国 专 家		美 国 专 家	
Adidas AG	G	Allstate Corp.	US
American Express	US	American Express	US
BASF AG	G	Ameritech	US
Bayer AG	G	Avon Products	US
HypoBank	G	Bell Atlantic	US
Bayerische Vereinsbank	G	Black & Decker Corp.	US
Daimler Benz	G	Citicorp	US
Dresdner Bank Ag	G	Coca-Cola Co.	US
Escada AG	G	Dow Jones & Co.	US
Karstadt AG	G	Eastman Kodak	US
Kaufhof AG	G	Ford Motor Co.	US
Lufthansa	G	Intel Corp.	US
Microsoft Corp.	US	J. P. Morgan & Co.	US
Munchener Ruck.	G	Kmart	US
Telekom AG	G	Lufthansa	G
Volkswagen AG(VW)	G	Merck & Co.	US
		Merrill Lynch	US
		Morgan Stanley	US
		Procter & Gamble	US
		Southwest Airlines	US
		Whirlpool Corp.	US

注：鉴于对公司名字的翻译可能因人而异，且保留英文不影响理解文章内容，所以这里未对表中公司名进行翻译——译者

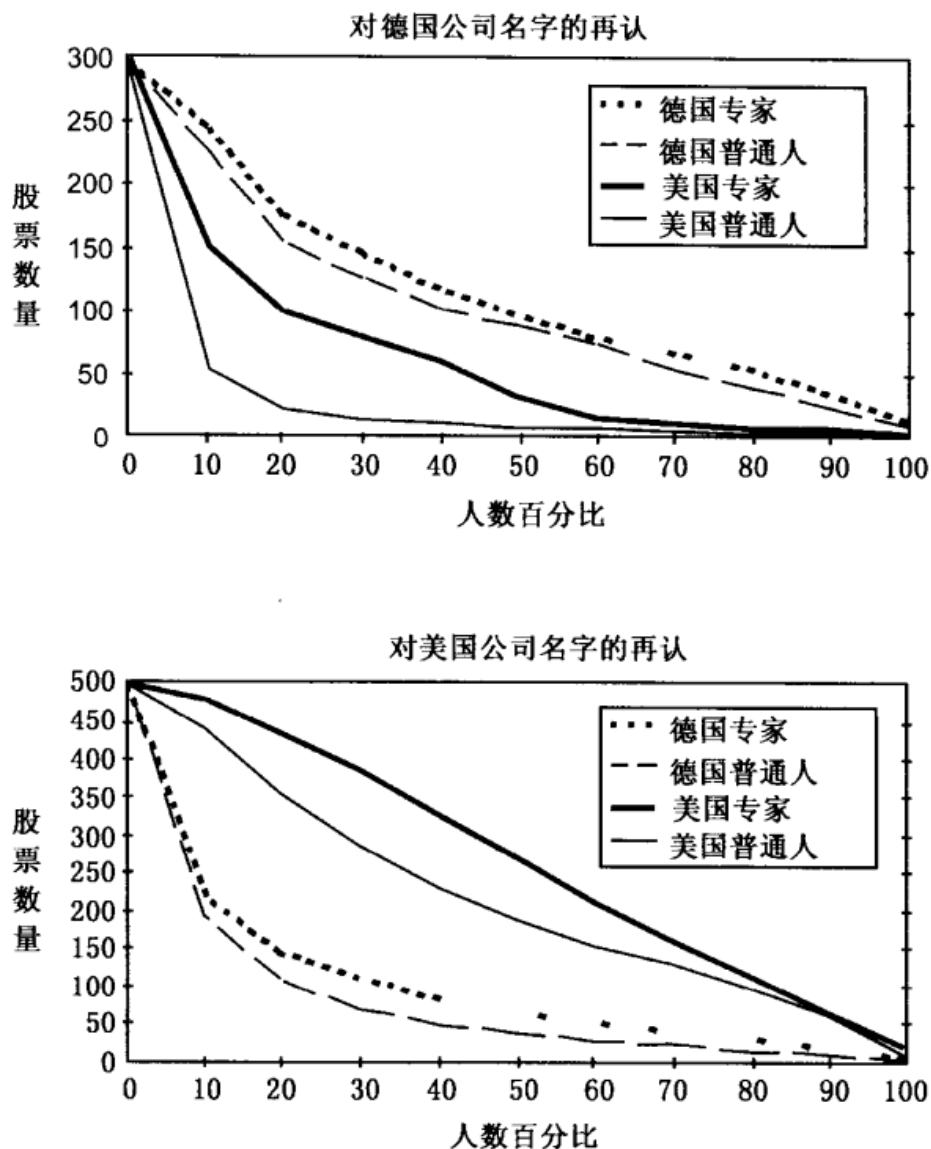


图 3-1 得到不同比例人数再认公司的数量。如, 14 个德国公司得到了 100% 德国专家的再认, 33 个德国公司得到了至少 90% 德国专家的再认, 等等。

64

再认启发式能赚钱吗?

为了测定再认启发式在股市上的业绩, 我们用再认率较高的公司为每组受测者制定了两套投资方案: 一套方案是由再认率较高的本国公司组成的(“本国再认”), 在这里再认率较高是指一组

受测者中有 90% 以上的人都能再认；另一套方案是由各组受测者能够最经常地再认的 10 家外国公司组成的（“跨国再认”）。这样总共有八种基于再认的投资方案，如图 3-2 所示。需要重申的是：再认启发式意味着只把资金投放到再认率较高的股票上去。 65

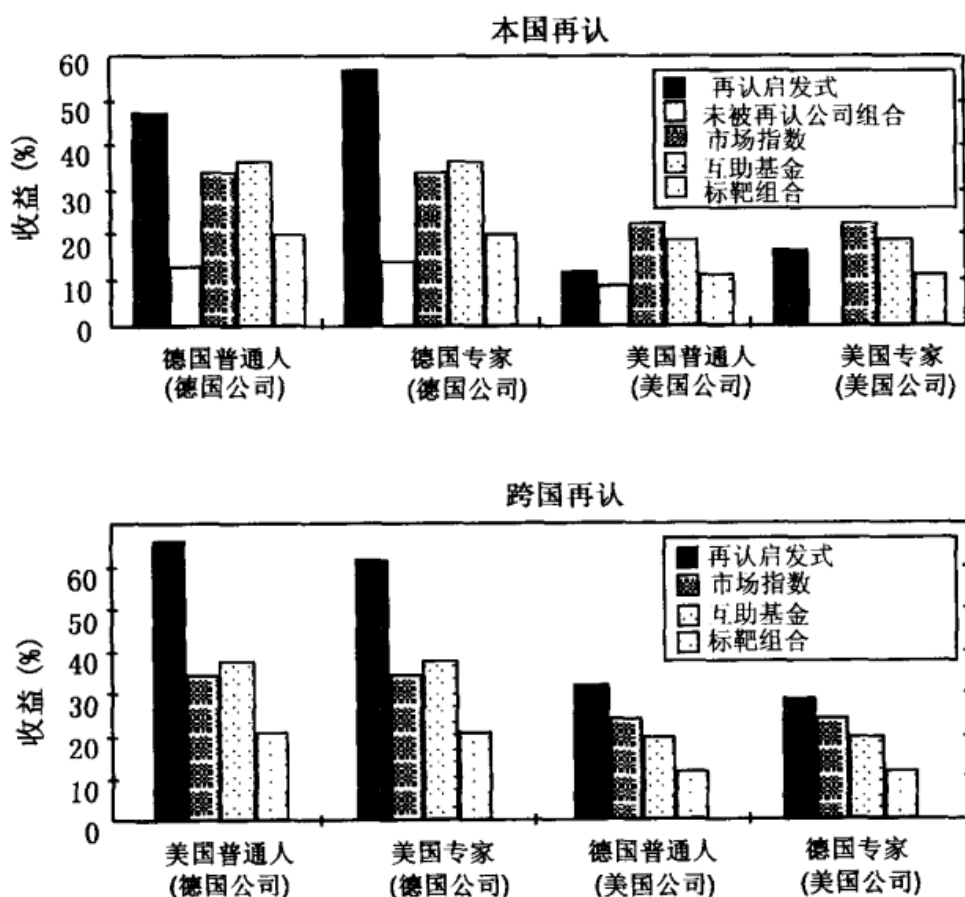


图 3-2 在本国或跨国再认条件下再认启发式方案的投资业绩。此为自从再认测验之日起 6 个月内（即 1996 年 12 月到 1997 年 1 月）的投资结果。为比较起见，同时列出了未被再认股票（0% 到 10% 再认率）、市场指数（道琼斯 30 或达克斯 30）、互助资金（保值增长基金或 Hypo 银行公正基金）和机遇投资（随机抽取的 5 000 股股票的平均收益）等方案的业绩。

从 1996 年 12 月 13 日完成再认测验之日起，我们用 6 个月时间考察了上述八种基于再认投资方案的市场业绩，并将其与以下五种投资策略的收益作了比较：(a) 选择未被再认的公司，即选择低于 10% 的受测者能够再认的公司；(b) 市场指数；(c) 互助基金；(d) 随机化策略；(e) 个别投资选择。在比较结果尚未明了之

简捷启发式

前,我们中的两个人决定将我们个人为数不少的钱投放到慕尼黑行人能够再认的股票上。我们值得为普通人的无知下如此大的赌注吗?在随后的日子里我们会为把辛苦挣来的钱如此随意地花掉而后悔吗?

再认启发式与选择未被再认公司业绩的比较

以德国 180 名普通人集体再认为依据的德国股票投资方案 6 个月内获得了 47% 的收益。^{*} 而投资到未被再认公司上的股票仅产生了 13% 的收益。图 3-2 显示,在本国再认条件下的 4 种投资方案中,再认启发式的这种投资优势都表现了出来,无论是专家还是普通人,也无论是德国人还是美国人。在两个最为无知的群体(德国普通人和德国专家)中,再认启发式的优势尤为明显。将资金投入德国专家再认的公司上的收益达到了 57%,而投资到未被再认公司上仅得到 14% 的收益。在本国公司再认条件下,4 种再认启发式投资方案的平均收益比投资到未被再认的公司上的平均收益高出 3 倍还多。

再认启发式与市场指数的比较

得到再认股票远远超过了未被再认股票的业绩,但这也许不会引起股票投资者的多大兴趣,因为他们的主要兴趣是如何在市场上立于不败之地。股票市场的总体业绩通常是由股票指数来测定的,诸如美国股票市场上的“道琼斯 30”指数和德国股票市场上的“达克斯 30”指数等。在我们进行研究的那 6 个月中,达克斯指数增长了 34%,然而同一时期产生达克斯指数的 30 种股票的价格总量却上扬了 41%。因此,构成“达克斯 30 指数”的 30 种股票的实际市场业绩要好于达克斯指数本身(这是可能的,因为为了能

^{*} 该收益的计算方法如下:原始分数=(股值 t_1 /股值 t_0)-1,标准分数=[(股值 t_1 /股值 t_0)/(市场指数 t_1 /市场指数 t_0)]-1。少数公司的价格变动在研究期间难以记录,所以分析时被剔除了。

够反映总体市场走向,股票指数通常要被加权计算)。298家德国公司的股票价格上涨了24%。与此同时,美国的道琼斯30指数增加了23%,而产生该指数的30家美国公司的股票价格仅上扬了8%,所有美国500种股票的价格上扬了10%。很显然,道琼斯指数是难以企及的!

再认启发式的业绩能够逼近道琼斯和达克斯市场指数吗?为了回答这个问题,我们分别对本国和跨国再认启发式的业绩进行了分析,这是因为如同图3-1所示,本国公司再认率远远高于跨国公司再认率。

本国再认 就原始收益来看,基于德国普通人再认基础上的股票投资收益(47%)比达克斯市场指数高出10%(图3-2),而基于德国专家再认基础上的股票投资收益(57%)比达克斯市场指数高出17%。 67

以美国两组受测者(他们并不像德国受测者那样无知)的再认为基础的投资方案的业绩又怎样呢?基于美国普通人再认基础上的投资方案也赚了钱,但其收益却比道琼斯市场指数低10%,而未被再认的股票收益比道琼斯市场指数低12%。类似地,基于专家再认的投资方案产生了16%的收益,仍比道琼斯指数低6%,而未被再认的股票是毫无价值的投资,没有任何收益。

跨国再认 由于在跨国再认中不同受测者能够再认的外国公司的名字五花八门,我们用10个最经常得到再认公司的股票制定了一种投资方案,即只将资金投放到这10种固定股票上。我们所以采用10个最经常被再认的公司,而不是以被90%的人再认为标准,还因为对跨国公司的再认率非常低(如同图3-1所示),以至于在一些情形下没有任何一个公司得到了一组受测者中90%人的再认。

德国人能够最经常地再认的10种美国股票的业绩如何呢?事实上,此种以跨国再认为基础的启发式投资方案获得的收益超过了道琼斯市场指数——这是其他投资方案均未能做到的。对于被德国专家再认的美国股票来说,也获得了同样的结果。相比较

简捷启发式

而言,基于知识较少普通人的再认而制定的投资方案比基于专家再认的投资方案反而获得了更高收益。

这种对跨国股票的无知所具有的益处在大洋彼岸的美国人身上也能表现出来吗?最经常得到美国普通人再认的 10 种德国股票的收益比市场指数高出 23%,类似地,基于美国专家再认的 10 种股票的收益也远远超过了市场指数。

在 4 种跨国再认条件下,再认启发式投资方案的收益均超过了有关市场指数。而且,在 4 种情况下,基于跨国再认的投资方案都比基于本国再认的投资方案得到的收益更高,基于普通人再认的投资方案得到的收益也都略高于基于专家再认的投资方案。总之,情况似乎是:无知的程度越高,所选择股票的收益越大。

作为一种控制,我们还计算了 4 种跨国再认条件下最经常得到再认的 20 种和 30 种股票的业绩。图 3-3 表明,在 4 组受测者中再认率较高的股票都得到了更高收益。

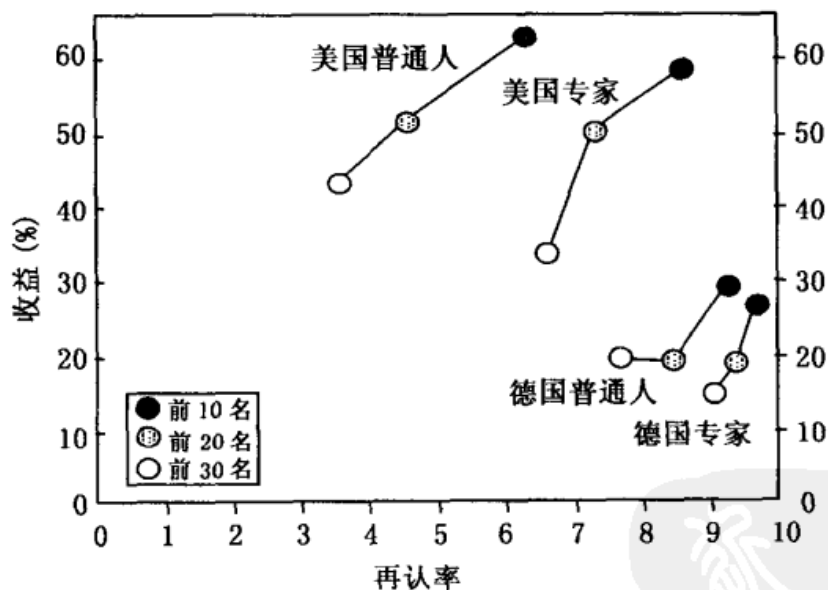


图 3-3 跨国再认率与收益。在每组受测者中再认率与收益之间都存在直接的关系。前 10 名股票即最经常得到跨国再认的股票(参见图 3-2)。当将范围扩大到前 20 名或前 30 名最经常得到跨国再认的股票时,再认率降低了,收益率也随之降低。

再认启发式与经营资金的比较

我们已经通过与市场指数和未被再认股票的比较,考查了再认启发式的成效。到现在为止,它运作得相当好,在所有4种比较中都打贏了未被再认股票,而且在与市场指数进行的8场较量中贏得了6场。那么,与专业股票经纪商所使用的知识和工具相比,基于再认的启发式投资策略又怎样呢?我们将两种主要互助基金——美国的“保值蓝筹股增长基金”(The Fidelity Blue Chip Growth Fund)和德国的“Hypo 银行投资基金”(Hypobank Investment Capital Fund)作为再认启发式的基准。在我们进行投资评价的6个月内,保值基金增长了19%,Hypo 银行基金增长了36%。

图3-2表明,在所有8种比较中,再认启发式贏得了6种。例如,慕尼黑商业区180位普通行人的集体无知比美国和德国专业经纪商的知识和经验更具有预测力。我们再一次看到,跨国再认比本国再认显示出了更大优势。此外,两个最为无知群体——德国普通人和德国专家——从他们的无知中获得的益处最多。

再认启发式与随机股票组合的比较

按照前述的高效市场假说,一个人不会总是比随机选择股票的投资策略做得更好。例如,《华尔街杂志》著名的投资专栏一再指出,通过非常复杂的“镖靶”操盘机制随机地选择股票常常能比专家亲自操盘取得更好的业绩。我们分别用美国和德国的10种股票随机地组合成5000种投资方案,并对它们的市场业绩进行了评价。图3-2显示,这些随机组合方案的平均收益是22%(德国股票)和11%(美国股票)。再认启发式在7种比较中都战胜了随机组合方案,在剩下的1种比较中也与其持平。很显然,在股票选择上再认启发式远远好于随机组合策略。

再认启发式与个别投资选择的比较

专家和普通人在选择哪一种股票进行投资时做得如何呢?谁

简捷启发式

将制定出更好的投资方案呢？我们分别要求德国专家和普通人从一个上市公司列表(均为再认测验中使用过的公司)中挑出他们愿意投资的 10 种股票,然后根据最经常被他们选择的 10 种股票制定了一种投资方案。

德国普通人倾向于挑选得到较高再认的股票进行投资,他们最经常选择的 10 种股票的平均再认率达到了 80%。然而,专家更喜欢选择再认率较低的股票,他们选出的 10 种股票的平均再认率为 48%。这里再认启发式作出如下明确预测:选择再认率较高股票的那一组将获得更大收益。情况的确如此,普通人的选择获得了巨额收益,而专家的选择实际上赔了钱(图 3-4)。

就德国人所选择的 10 种美国股票的情况来看,平均再认率更低(0.27),而且普通人和专家之间没有明显差异。结果是,这两种跨国选择的投资组合均不如图 3-2 中得到再认股票的收益高。在此,普通人的选择再一次全面战胜了专家的选择(图 3-4)。

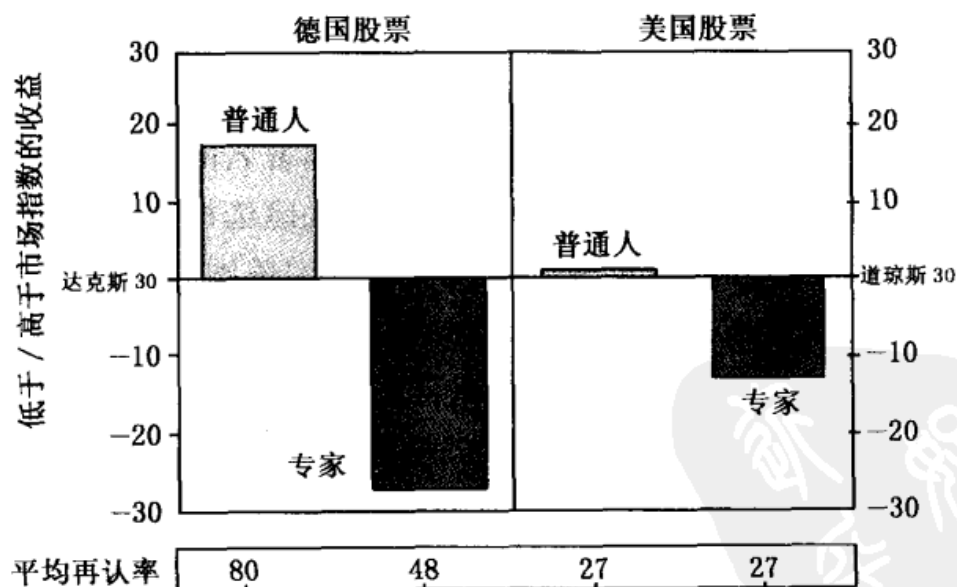


图 3-4 在两种股市上德国普通人和专家选择股票的业绩。用公式 $(1 + \text{投资方案收益的原始变化}) / (1 + \text{市场指数的原始变化}) - 1$ 来计算收益高于或是低于相应市场指数。

从再认到富有

充分利用人们的无知而不是丰富知识的快速节俭启发式能够赚到钱吗？经过一个时期的研究，我们获得了如下结果：

1. 再认率高的股票的市场业绩超过了未被再认股票的市场业绩。该结果在所有 4 种本国再认和所有 4 种跨国再认条件下均得到了证实，而且股票收益率随着其平均再认率的增高而增加(图 3-3)。
2. 在跨国再认的 4 种条件下，利用再认启发式得到的收益全都超过了市场股票指数，而且受测者能够再认的公司愈少，从再认启发式得到的益处愈多。在本国再认的 4 种条件下，再认德国公司的收益超过了达克斯指数，但再认美国公司的收益未达到道琼斯指数。因此，在再认启发式与市场指数的 8 场较量中，再认启发式赢了 6 场。
3. 与两种主要经营基金(美国的保值蓝筹股增值基金和德国的 Hypo 银行投资基金)相比，再认启发式在 8 种条件中的 6 种条件下都取得了更好业绩。
4. 随机化投资方案的平均收益始终低于再认启发式投资方案的收益。这种结果在 8 种条件中的 7 种条件下都保持 71 不变(在剩下的 1 种条件下二者势均力敌)。
5. 当人们按照再认启发式进行投资选择时，他们制定的投资方案获得的收益都远远高于达克斯指数。在其他 3 种情况下，当所选择股票的平均再认率较低时，其收益也更低。专家倾向于选择再认率较低股票，该种股票的市场业绩较差。

再认启发式强大的预测力证实了这样一种观点：缺乏再认可能隐含着像显性知识一样有用的隐性知识。而跨国再认相对于本国再认以及普通人相对于专家表现出来的优势支持这样一种观点：一定程度的无知是有益的。

再认启发式何以会在股票市场上产生如此好的业绩呢？有关战略管理和市场经营的研究表明，市场占有率与公司收益之间存在正相关（例如，Buzzell et al., 1975; Ramanujam & Venkatraman, 1984）。因此，拥有主导性市场股份的公司最有可能既是可再认的、又是高收益的公司。与公司收益有关的另外一种因素是“核心竞争力”（core competence），即“组织内的集体学习，尤其是学会如何使多种产品的生产技术协调一致，并将多种技术力量凝聚在一起”（Hamel & Prahalad, 1994, p. 82）。例如，本田（Honda）通常被认为是一个在机械动力行业具有核心竞争力的企业，它以生产轿车、除草机、轮船发动机及其他动力产品而著称。有研究似乎证明，企业的核心竞争力与其超出平均水平的股票业绩有关（例如，Hamel & Prahalad, 1994; Prahalad & Hamel, 1990）。如果能够证明在企业核心竞争力与对其股票的再认率之间也存在明确的关系（这是一个值得进一步研究的问题），那么就可以解释为什么得到再认公司会有更多的收益了。

再认和股票市场业绩之间也许还存在另外一种关系。专业股票经纪人似乎也在利用人们对企业名字的再认进行交易活动，仿佛企业名字是股票价值的一种基本标志似的。例如，“Hugo Boss”公司投资关系经理尤格·维格·穆勒（Jörg Viggo Müller）经常被引用的一句话是：“Boss 股份不是蓝筹股（blue chips，即来自于最著名公司的股票），但分析家们宁愿把我们作为蓝筹股”（Deutsche Börse, 1997, p. 35）。这就是说，投资界未作充分分析就将好的品质赋予了“Hugo Boss”公司。类似地，“保时捷”（Porsche）公司信息关系经理曼弗雷德·阿亚瑟（Manfred Ayasse）指出：“像消费者一样，投资者很难抗拒拥有一部保时捷轿车带来的那种‘酷感’。”（Deutsche Börse, 1997, p. 35）因此，尽管投资专家们也许不能直接应用再认启发式本身（因为他们实际上能够再认所有股票），但他们会把公众对股票名称的再认作为股票价值的一部分。名字是有价值的，这不仅反映在企业良好声誉所

带来的外显价值上,而且被无数有关公司冠名权的诉讼案件所证明。

基于再认的投资方案在股票指数不断攀升的“牛市”(bull market)上表现出了不俗业绩,但我们尚不知道这种结果在股票指数下跌的“熊市”(bear market)上能否保持下来。对再认启发式良好业绩的一种解释是,它总是选择“大”公司,这些公司的股票通常在股指上扬的市场上运行得很好。对于此种假设,我们可以在股指下跌的疲软市场上加以检验,因为在这种情况下大股的业绩通常低于市场指数。如果得到再认股票的业绩在股指回升时超过了“大股”,而在股指跌落时没有蒙受那么大的损失,那么我们就有了区分再认效应和大股效应的依据。如果这两种效应存在某些区别的话,我们认为至少股指回升时期基于再认投资方案的业绩将会超过达克斯和道琼斯等指数所涉及的大公司的总体收益。

72

当一个企业由于一些“负面报道”(如,媒体曝光了某石油公司的严重漏油事件)而被公众再认时,它能够从再认启发式受益吗?人们或许会预期该公司的股票价格将会暴跌。然而,金融市场是变化莫测的,股票价格将根据预期的亏损而得到调整。一旦有关某企业的坏消息通过媒体泄露出去,其股票价格甚至还会上扬,因为消息灵通的内部人士会对这种消息作出各种解释和说明,以至于关于未来能否赚钱的疑问反而减小了,甚至被排除了。因此,通过不利渠道再认了一个企业的“无知”投资者也有可能意外地赢利。

无知中隐含着智慧

在本章中,我们采取了一项大胆举措,将以无知为基础的再认启发式派上了金融贸易的“战场”。基于再认投资方案所产生的惊人收益为上一章提出的观点提供了重要证据:在现实环境中再认启发式可以作出准确推断。就本章所关心的具体领域来看,我们

简捷启发式

至少可以说：普通行人的再认知识战胜了深思熟虑的金融专家们的智慧。股票市场是一个纷繁复杂的现实环境，在这里缺乏再认并非随机偶然的现象，而是经常性的。在投资活动中，我们也许需要一种基于无知的智慧！



第三部分

单一理由决策制定

THAT MAKE US
SMART

新学网
PDG

第4章

为一个好理由下赌注—— “采纳最佳”启发式

哥德·吉戈伦尔(Gerd Gigerenzer)

丹尼尔·G·古德斯腾(Daniel G. Goldstein)

认知心理学研究的全部问题即有限理性。有限理性研究不是对环境任务最优化的研究。

——赫伯特·A·西蒙(Herbert A. Simon)

正如约翰·洛克(John Locke, 1690/1959)声称的那样,上帝“仅供给我们概率的微弱光线;我想他也许乐意置我们于一种平庸的、接受考验的状态,就此而言他的做法还是蛮适宜的……”在前面的两章中,我们指出人类能够很好地利用这种平庸的不确定性。幸运的是,对于现实环境的无知常常是系统的、经常性的,而不是随机的、偶然的,这使得有机体能够利用再认启发式航行于概率的微弱光线之下。在本章中我们将对依据再认之外的其他信息作出推断的启发式加以详细介绍。这些信息的来源可能是直接的观察、从记忆中回忆、亲身经历或传说等。例如,达尔文(1872/1965)曾经观察到,人们能够利用闪烁不定的眼神和低垂的眼睑等面部线索来推断一个人的犯罪感。夜晚漫行于沼泽地中的蟾蜍能够利用对手鸣叫声的高低来推断其大小,以决定是否与之争斗(Krebs & Davies, 1991)。对于现实世界的推断常常都是建立在这些非常不确定的线索基础之上的:我们的眼睛可能会欺骗我们。同样地,

让我们更精明

黑夜里“中等大小的人种学家”(“medium-sized ethologist”——一种幽默的说法)也能够模仿大青蛙发出低沉的叫声来欺骗我们。正像本杰明·富兰克林(Benjamin Franklin)在1789年的一封信中所评论的那样:“这个世界上没有任何东西是确定的,除了死亡和不堪忍受的重负。”(Smyth, 1907, p. 69)

人们如何依据成堆的不完善线索和信号作出推断、预言和决策呢?关于不确定条件下理性判断的传统观点可以用富兰克林的“道德决策代数学”(moral algebra)来加以说明。在经常被引用的写给英国科学家约瑟夫·普里斯特莱(Joseph Priestley)的一封信中,富兰克林(1772/1987)详细解释了如何依据不确定线索(他称之为“理由”)在两种行动方案中作出选择: 76

“我的方法是,用一条线将一张纸分成两栏,一栏的上方写上‘正方’(Pro),另一栏的上方写上‘反方’(Con)。在随后三至四天的考虑过程中,我经常会产生一些赞成或反对某一方的想法,我随时以简短提示的方式将其记录在不同栏目下面。当我将它们放在一起加以考虑时,我力图估计出它们各自的权重。如果发现分属不同栏目的2条理由具有同样权重,我同时将它们划去;如果发现‘正方’栏目下的1条理由抵得上‘反方’栏目下的2条理由,我同时划去这3条理由;如果我断定‘反方’栏目下的2条理由抵得上‘正方’栏目下的3条理由,我同时划去这5条理由。如此进行下去,最终就能找到两个栏目的平衡点所在。如果在随后一两天时间的进一步考虑中,双方都没有发生新的变化,我最终就能作出决定。虽然这些理由的权重无法进行精确的代数运算,但当分别考虑每一条理由,并仔细比较它们时,完整的结论离我已经不远了。我认为我能够作出更好的判断,不大可能作出轻率鲁莽的选择。事实上,我从这种等式运算中已经发现了巨大的益处,这种等式运算也许可以被称之为道德决

策的或审慎的代数学。”(p. 878)

富兰克林的道德决策代数学(或者我们可以称之为“富兰克林规则”)是:搜寻所有理由,不管是正面的或是负面的;仔细权衡每一条理由;将它们放在一起,看一看平衡点在哪里。这种将多种理由结合起来的线性模式实际上包含着关于理性行为的一种基本观点,即:仔细地搜索点点滴滴的信息,审慎地权衡手头的每一条信息,并将它们联系起来形成一种判断。富兰克林方法是启蒙运动时期产生的经典理性观(参见第1章)的一种变式,只不过经典理性观不强调多种理由的线性联结,而是假定概率规则是人类头脑遵循的基本规则,至少是“受过教育的开明人士”(hommes éclairés,参见 Daston, 1988)遵循的规则。正如皮艾尔·西蒙·拉普拉斯(1914/1951, p. 196)指出的那样:“归根结底,概率论只是普通意义上的一种计算方法。”

然而,在足够丰富复杂的现实情境中,实现这种经典无限理性理想所必需的知识、时间和运算能力是不可企及的——它们对于笨拙的人类而言要求过高了,而且对于最为有力的计算机而言也常常是要求过高了。例如,如果一个人想要用一种经过改进的现代方法——多元线性回归,来解决富兰克林式的多种理由的线性结合问题,那么在将这些理由线性地结合起来以前,他必须能够充分估计每种理由的权重,而且这种估计应该在“最小平方”意义上将误差降至最低——这是一种大多数人没有计算机帮助难以完成的任务。如果一个人还想更进一步,用非线性的贝叶斯网络模型来解决富兰克林式的问题,那么这项任务计算起来就太过复杂了,以至于计算机也难以完成它。

尽管从心理学上看是难以置信的,但是自从20世纪60年代认知革命以来,心理学中备受青睐的认知过程模型仍然一直是下列这些假想的神灵模型:选择活动的预期效益最大化模型、判断过程的多元回归模型、问题解决过程的贝叶斯模型,以及无处不在

77

让我们更精明

简捷启发式

的神经网络模型。具有惊人运算能力的神灵不仅席卷了整个认知心理学领域,而且蔓延到了经济学、最优化搜索理论、人工智能及其他许多领域。对此,赫伯特·西蒙抗议道:“没有充分的证据表明,在任何复杂的人类选择情形下,这些运算能够或者实际上在起作用。”(1955a, p. 104)

西蒙主张建立一种有限理性模型,而不是最优化模型。但是实际情形怎样呢?如果不采用最新的统计技术来分析心理过程,我们又如何来分析心理过程呢?

简单终止规则

在本章中,我们将讨论第2章使用过的那种任务:依据一种效标来评定两个对象中哪一个得分更高。这种任务可以作为更为一般的问题——从一类对象中选出一个子集(如同第2章所做的那样)——的一种特例。我们使用的具体例子有:制定治疗方案(如,以治疗后的预期寿命为效标,急诊室中的两个病人中哪个应首先得到治疗?)、金融投资(如,以收益率为效标,两种证券中买哪一种?),以及人口统计学方面的预测(如,两个城市中哪个有较严重的环境污染、较高的犯罪率或死亡率等)。

为了说明本章将要介绍的启发式,不妨首先考虑下面的双项选择问题:

两个城市中哪个有较多的人口?

(a) 汉诺威

(b) 比勒费尔德

假如一个人听说过这两个城市,因此不能够使用再认启发式。他必须搜索一些有助于表明哪个城市有更多人口的信息。这种搜索可能是内部的(如,从记忆中搜索),也可能是外部的(如,到图书

馆查资料)。在此,快速节俭启发式最重要的特征是搜索的有限性:并非所有适用信息都得到搜索,结果是只有很少一部分信息影响判断。(相反,在许多实验室实验中非常现成地将信息呈现和展示在被试面前,这就排除了被试搜索信息的可能性。与之相应,许多有关认知过程的理论甚至从不涉及信息搜索问题。)

有限搜索暗含着一种终止规则。快速节俭启发式使用了简单的终止规则。它们不采取传统的做法,即:只要觉得获得更多信息的收益稍微超过了其代价,就不停顿地搜索下去(Stigler, 1961)。头脑能够而且将会不厌其烦地进行这种最优化代价—收益权衡的假设在信息搜索模型中一直处于主导地位,但却是难以置信的(参见本章开头引用的警句)。

我们用图 4-1 来说明一种简单终止规则。该图代表了一个人关于 a、b、c、d 四个对象(如,城市)的知识,包括 5 条线索(如,这个城市有无大的足球俱乐部、是不是州府等)和再认的情况(如,这个人以前是否听说过这个城市)。举例来说,如果一个城市有足球俱乐部,而另一个没有,那么有足球俱乐部的城市可能(但不一定)有更多人口。假定我们希望确定城市 a 和 b 哪个更大。a 和 b 均得到了再认,因此不能使用再认启发式。对记忆中的信息进行搜索使我们注意到线索 1,即足球队线索。城市 a 有足球俱乐部,而城市 b 没有(图 4-1 中其线索值分别用“1”和“0”表示)。因此这条线索就可以将两个城市区分开来,搜索就此终止,我们得出结论:城市 a 是更大的。更一般地说,对于这种全或无式的“双极”(binary)线索而言,简单终止规则是:如果一个对象有正线索值(“1”),而另一个没有(即“0”或未知),那么终止搜索。

79

为了方便起见,我们使用“1”作为正线索值,表示有较高效标值(如,较多人口),而使用“0”作为负线索值,表示较低效标值。如果终止搜索的条件未被满足,那么继续搜索另一条线索,如此进行下去。例如,当在图 4-1 中的 b 和 c 之间作出选择时,线索 1 无法将二者区分开来,但线索 2 可以。依据这条线索,对象 b 被判断

简捷启发式

为更大的。有限搜索是逐步展开的,线索逐个被搜索,直到终止规则被满足为止(类似于米勒等人采用的“测试—运行—测试—退出”程序)。如果得到尝试的线索中无一能满足终止规则,就作出一种随机猜测。为了终止搜索,无需进行任何代价—收益计算。下面将要介绍的启发式——最少化启发式、采纳最近启发式和采纳最佳启发式——全都使用了这种简单终止规则。它们还共同使用了一种作出决策的规则,即单一理由决策规则——它仅将推断和决策建立在一种理由或一条线索基础之上。这些启发式的区别仅在于搜索线索的方式不同。

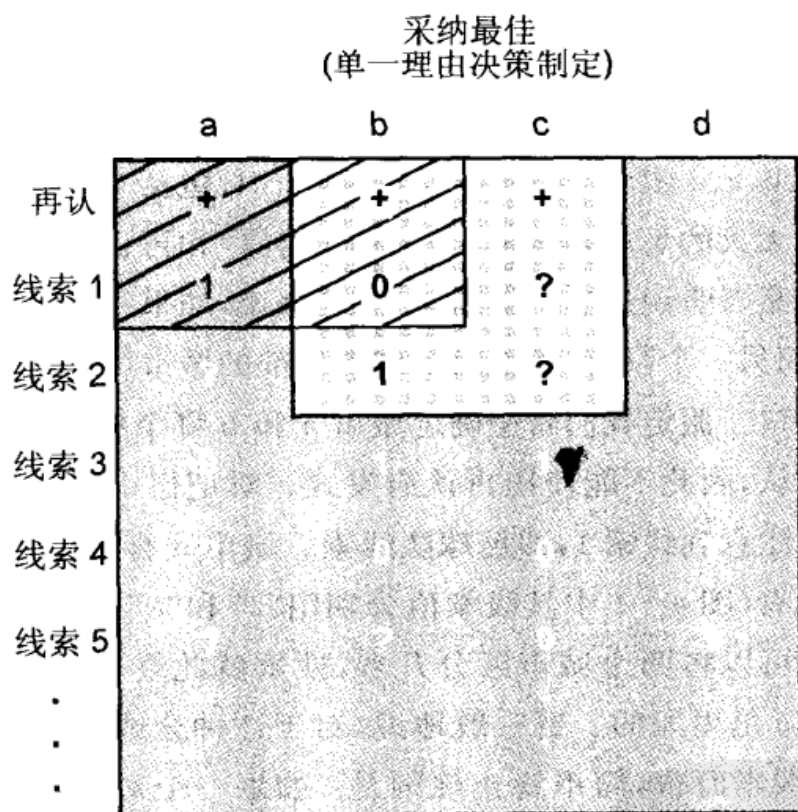


图 4-1 借助有限知识进行有限搜索的例证。对象 a、b、c 得到了再认(+),d 没有被再认(-)。线索值是双极的(0 或 1);未知线索值用“?”号表示。例如,为了推断 a 是否大于 b,采纳最佳启发式将仅搜索以线条为界的长方形中的线索值。为了推断 b 是否大于 c,搜索仅限于以圆点为背影的白色方框。其他线索值不被搜索,所以用阴影背景来表示。

启发式规则

最少化启发式

基于线索的推断所需要的最起码的直觉知识是了解一条线索所指明的方向,如是否拥有足球俱乐部意味着较多人口或是较少人口。这种方向能够从所了解的少数样例中估计出来(所估计的方向有时可能是错误的,参见下文)。最少化启发式仅需要这种起码的直觉知识,除此之外不需要知道任何信息(如,哪一条线索比其他线索更有预测力)。结果是,最少化启发式采用的搜索规则是按照随机顺序检测线索的。只要有可能,它将充分利用再认启发式(参见第2章)。然而,在一些情形下不能使用再认启发式。例如,当两个对象都得到了再认或再认与效标无关时,将无法使用再认启发式。

最少化启发式可以用下面的具体步骤来表述:

步骤0. 若适用,则利用再认启发式。这就是说,如果只有一个对象得到再认,就预言它具有较高效标值。如果无一得到再认,就猜测。如果两个均得到再认,就到步骤1。

步骤1. 随机搜索:随机地挑出一条线索(不再放回原位),审视两个对象的线索值。 80

步骤2. 终止规则:如果一个对象有正线索值(“1”),而另一个对象没有(“0”或未知),那么终止搜索,到步骤3。否则,返回步骤1,搜索另一条线索。如果未发现任何线索,就猜测。

步骤3. 决策规则:有正线索值的对象有较高效标值。

采纳最近启发式

像最少化启发式一样,采纳最近启发式仅有关于线索方向的直觉知识,而无关于哪些线索更加有效的知识。它与最少化启发

简捷启发式

式的不同之处仅在于步骤 1,在那里它采用了一种类似于著名的“态度定势”(Einstellung set)策略的启发式搜索规则。卡尔·唐克(Karl Duncker)和其他格式塔心理学家早已证明,当人们解决系列问题时,每遇到一个新的、看起来类似的问题,他们都倾向于用解决上一个问题时生效的策略来解决它(Dunker, 1935/1945; Luchings & Luchings, 1994),借以建立一种可供以后尝试的态度定势。对于第一个问题,采纳最近启发式像最少化启发式一样随机地尝试一条线索,但是从第二个问题开始,它将从上一次成功地终止了搜索的线索开始。如果该线索未能终止搜索,它将尝试更上一次终止了搜索的线索,依次类推。由于最近一次成功地终止了搜索的线索常常比其他线索更可能终止当前搜索(即它们有较高的区分率),所以采纳最近启发式可能会比最少化启发式搜索的线索更少。例如,如果上一次决策是根据足球队线索作出来的,在解决下一个问题时采纳最近启发式将会首先尝试足球队线索。与最少化启发式不同,采纳最近启发式需要记住过去成功地将两个对象区分开来的线索。其步骤 1 是:

步骤 1. 态度定势搜索: 如果存在有关过去哪些线索终止了搜索的记录,那么选择在最近的问题解决中成功地终止了搜索的线索,不再尝试其他线索。审查两个对象的线索值。否则,尝试随机线索,形成这样一种记录。

采纳最佳启发式

在一些环境下,人类和动物不仅知道(正确地或错误地)线索所代表的意义,而且知道哪些线索优于其他线索。对线索顺序的了解可能是生就的(如,许多动物选择配偶时所依据的线索),也可能是通过观察习得的。在后一种情况下,线索排列可以从其预测效标的相对频率上估计出来。例如,足球队线索的效度就是拥有足球队城市大于没有足球队城市的相对频率。这种效度很容易依据有无足球队城市的所有配对测验估算出来。如果人们能够按照

其知觉到的效度来排列线索——不论这种主观排列顺序是否对应于线索的生态顺序——那么对线索的搜索便可以按照这种顺序来进行。采纳最佳启发式首先尝试有最高效度的线索,如果该线索不能区分两个对象,就尝试效度居于第二位的线索,如此进行下去。该种启发式遵循的基本准则是“采纳最佳,忽略其余”。采纳最佳启发式仅在第1个步骤上与最少化规则不同。它搜索线索的第1步是:

步骤 1. 顺序搜索: 选择在该种任务上尚未尝试过的有最高效度的线索,审查两个对象的线索值。

值得注意的是,采纳最佳启发式所使用的顺序不是“最优化”顺序——毋宁说它是一种简约的顺序。它并不试图掌握线索之间的依从关系,也就是说它无意根据条件概率或部分相关来建立线索之间的顺序(参见第6章)。这种简约的顺序能够从少数对象或线索样例中估计出来(参见第5章)。

总之,上述三种快速节俭启发式充分体现了下面这些特征:利用逐步展开程序的有限搜索、简单的终止规则和单一理由决策制定。单一理由决策制定(仅依据一条线索作出推断)是通过所使用的某种特定终止规则来体现的,而不是通过所有三种终止规则来体现的。而且,单一理由决策制定并非一定依据上述三种启发式所使用的终止规则。例如,一个人可能搜索许多有助于区分两个选项的线索(就像一个人必须为自己的决定进行辩护时所做的那样),但仍然将决策建立在一条线索基础之上。

不妨将这些简捷启发式所代表的观点与富兰克林规则作个比较。一个非常明显的区别是:所有三种启发式规则都实施单一理由决策制定。相反,富兰克林的道德决策代数学却总是奉劝我们搜索所有理由——至少在几天的考虑过程中是如此——仔细权衡每一种理由,并将它们放在一起看一看平衡点在哪里。上述三种启发式回避了指向不同方向的多条线索之间的矛盾,这使得它们成为不可代偿的:任何来自随后线索的

反面证据都无法取代或抵消基于前面线索作出的决策。图 4-1 中关于 a 大于 b 的推断就是一个典型例子：无论是 b 随后所具有两种正线索值，还是 a 随后所具有的“0”线索值，都不能改变上述推断。把决策完全建立在单一理由基础之上毫无疑问是大胆的，但它也是精明的吗？

从心理学上看似乎合理，但却笨拙吗？

首先考虑一种动物，它们遵循一种非常类似于采纳最佳启发式的单一理由决策规则进行推断。在群居的古比鱼中，选择配偶这种重要的适应性任务通常是由雌性来完成的。影响它们作出选择的一般有两类线索：一类是身体线索，另一类是社会线索(Dugakin, 1996)。在身体线索中，它们最重视的是雄性古比鱼的身体大小和鲜艳的橘黄色外表。它们所依据的主要社会线索是正在考察的雄性古比鱼以前是否和其他雌性古比鱼交配过。这些线索似乎是按层次等级排列的，橘黄色线索居于支配地位。如果一个雌性要在两个雄性之间作出选择，而且两个雄性中的一个比另一个具有更加鲜艳的橘黄色外表，它将选择更鲜艳的这一个。如果两个雄性在颜色上接近，它将更喜欢与其他雌性古比鱼交配过的那一个，即便它在色泽上稍微逊色一点。在这里颜色线索的终止规则是：一个雄性的橘黄色外表必须比其他雄性更加鲜艳(大约超出 40%)。雌性古比鱼的配偶选择行为充分体现了有限的搜索规则、简单的终止规则和单一理由决策策略。

不仅仅是低等动物，人类也常常仅搜索 1 或 2 条相关线索，力求规避对相互矛盾证据的搜索，并使用“不可代偿策略”(noncompensatory strategies; 例如, Einhorn, 1970; Einhorn & Hogarth, 1981, p. 70; Fishburn, 1988; Hogarth, 1987; Payne et al., 1993; Shepard, 1967a)。例如，采纳最佳启发式(不同于最少化启发式和



采纳最近启发式)与其他研究者曾经使用过的“词典编纂策略”(lexicographic strategies)有关。所谓“词典编纂策略”是指,以固定的效度顺序搜寻每一条线索,就像词典中按照字母顺序排列词汇一样。阿拉伯数字(共有 10 个基本数字)和巴比伦数字(共有 12 基本数字)都是符合词典编纂规则的。要想判断由同样多数字组成的两个数哪个更大,一个人必须首先看一看左起第一个数字。如果这个数字是更大的,那么整个数就是更大的。如果它们是相同的,必须看左起第二个数字,如此等等。这种简单方法对于罗马数字是不适合的,因为这种数字系统不符合词典编纂规则。有关实验研究似乎表明,在时间紧迫情形下,采用词典编纂策略搜索信息是很有效的(Payne et al., 1993; 亦可参见第 7 章)。此外,采纳最佳启发式和更加通用的概率心理模型(Gigerenzer et al., 1991)在综合多种实验现象方面一直都是很成功的(DiFonzo, 1994; Gigerenzer et al., 1991; Juslin, 1993; McClelland & Bolger, 1994)。

然而,利用单一理由作出决策、规避矛盾冲突并采用不可代偿原则的简捷启发式常常被怀疑为不合理的,因为与传统的重视连贯性而忽略在现实环境中成效的理性标准相比,它们看起来是愚蠢的。例如,当苛尼和莱法(Keeney & Raiffa, 1993)讨论词典编纂策略时,他们一再警告人们这种策略“在实际中使用得过于广泛,这是不值得的”,因为“它几乎简单到了天真的程度”,因此“很难通过理性的检验”(pp. 77—78)。他们实际上根本就没有进行这样一种检验,我们将进行!

快速节俭启发式可能是准确的吗?

人们常常用内部连贯性准则来评定启发式,而不是用它们在外部的成效来评定它们。这些外部标准应该包括准确性、节俭性、速度以及其他一些指标。在判断和决策制定研究中,培恩等

简捷启发式

人(Payne et al., 1993)的研究也许是最为与众不同的。他们将简单策略的“努力—准确性”权衡与加权累积规则(富兰克林规则)作了系统比较——这种规则常常被作为偏好判断的标准(亦可参见 Beach & Mitchell, 1978; Beach et al., 1986)。与我们的研究不同,培恩及其同事研究的是对人工问题的偏好判断,而不是对现实世界的推断。这样做的结果是难以找到评定准确性的外部效标(如,一个城市的实际人口),因此必须人为构建一种效标。在他们的研究中,加权累积规则被作为“黄金”标准,准确性被界定为一种策略与这种标准的接近程度。因此,没有任何一种策略会比这种标准更加准确。^{*}然而,当作出关于现实世界的推断时,将这种加权累积规则作为一个人可能采取的最佳规则并不一定可行。

下面两条规则常常被作为理性判断者必须遵守的戒律。那么,违背了这两条戒律的启发式规则的准确性如何呢?

全部搜索 应该寻找所有适用信息。如果由于时间或计算能力受限不能找到这些信息,那么就要算出一个终止点——在那里继续搜索的代价超过了其收益。一直搜索下去,直到达到这个终止点。

可代偿性 应该将所有适用信息结合在一起。不应该仅仅依靠一种信息。

尽管富兰克林规则非常重视这两条戒律,但最少化、采纳最近和采纳最佳启发式却背离了它们。这些启发式规则不寻找所有线索(有限搜索),而且使用了一种简单终止规则。它们也不把线索值结合起来(不可代偿性)。此外,最少化规则可能还违背了可传递性标准——内部连贯性的神圣戒律。^{**}

^{*} 当这种加权累积规则仅使用有限信息进行判断时,可能会有一种意想不到的结果。

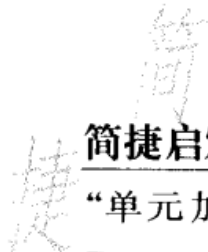
^{**} 最少化规则的不可传递性可能导源于这样一个事实:它按照随机的顺序采样信息,如同图 4-1 表明的那样。例如,恰好线索 1 适用于 a 和 b,线索 2 适用于 b 和 c,线索 3 适用于 a 和 c,我们就得到了一种不可传递的推断,即 $a > b$, $b > c$, 而 $c > a$ 。

为了回答快速节俭启发式是否准确的问题,我们将它们放在一种竞争情形下考察其成效,即让上述三种快速节俭启发式与三种标准统计策略一决高下。我们的目的在于考察:在寻找较少线索值的情况下,究竟哪一种策略产生了最为准确的推断。

竞争对手

为了提供比较标准,我们引入了三个不违背上述理性判断戒律的竞争对手。第一个竞争对手是线索的加权线性联结方法,我们将其称为富兰克林规则,因为它应用了富兰克林解决双项选择任务时所遵循的思路。与富兰克林原来的道德决策代数学不同的是,这种方法的实证性更强,因为它的加权不是主观的,而是依据资料计算出来的。在我们的模拟研究中,将线索生态效度作为它的权重,下面我们将对其加以简单界定。富兰克林规则用每一条线索值乘以它的权重,将得到的数加在一起,然后推断:拥有较大数值的对象是更大的对象。而在我们的模拟中,正或负线索值分别被编码为1和0。 84

像富兰克林规则一样,另外两个竞争对手也是线索的线性联结方法。其中的一个需要相当多的知识和计算,另一个对知识和计算的要求较低。前者即多元线性回归方法,它主要关心线索之间的相互依从性,而且要通过复杂计算将每条线索权重的误差降至最低。在心理学研究中人们常常将加权线性模型的各种变式作为描述认知过程的规范模型。例如,在安德森(N. H. Anderson, 1981)的“信息整合理论”(information integration theory)和有关社会判断的研究(Brehmer, 1994; Brunswik, 1955)中,都是这样做的。作为对心理过程的描述,加权线性模型,特别是多元线性回归模型,是很成问题的,因为他们都需要事先假定人们具有超凡的计算能力(Brehmer & Brehmer, 1988; Einhorn & Hogarth, 1975; Hogarth, 1987)。一种从心理学上看更加现实的线性策略利用了 85



简捷启发式

“单元加权”(unit weights)方法,就像鲁宾·达维斯(Robyn Dawes,1979)所做的那样。这种方法仅仅求出具有正值的线索数目,然后减去具有负值的线索数目。因此它是快速的(不涉及大量复杂计算),但并不节俭(需要搜寻所有线索)。为了简便起见,我们将这种策略称为“达维斯规则”。

在我们所报告的模拟研究中,将上述三种线性模型作为评价快速节俭启发式成效的基准。需要注意的是,富兰克林规则和多元线性回归方法利用了三种启发式规则所使用的所有信息,甚至更多。它们还对这些信息进行了更加复杂的计算。

环境

1990年重新统一之后,德国共有83个人口超过10万的城市。这些城市以及表明每个城市人口规模的9条线索形成了我们模拟研究的环境。这9条线索是依据预备试验中人们的报告挑选出来的(Gigerenzer et al.,1991;Gigerenzer & Goldstein,1996a)。任务是推断两个城市中的哪一个有较多人口。每一条线索都有两个重要特征:生态效度和区分率(discrimination rate)。线索的生态效度是指该线索正确预测效标(根据一种参照等级来界定,在这里参照等级即人口在10万以上的所有德国城市)的相对频率。举例来说,如果一个人检查了所有的有无足球队的城市配对,发现在87%的情况下有足球队的城市也有更多人口,那么足球队线索的生态效度即为0.87。一般地说,第*i*条线索的生态效度 v_i 为:

$$v_i = \text{正确预测数} / \text{预测数}$$

在这里预测数即一个有正线索值、另一个有负线索值的两个预测对象配成的对数。各条线索的生态效度在一定的取值范围内各不相同(参见表4-1)。

表 4-1 线索、生态效度和区分率

线 索	生态效度	区分率
国家首都(该城市是首都吗?)	1.0	0.02
展览馆(该城市曾经举办过展览会吗?)	0.91	0.25
足球队(该城市有大足球俱乐部吗?)	0.87	0.30
城间火车(该城市在城间火车线上吗?)	0.78	0.38
州府(该城市是州府吗?)	0.77	0.30
牌照(其缩写只有一个字母吗?)	0.75	0.34
大学(该城市是大学所在地吗?)	0.71	0.51
工业带(该城市位于工业带吗?)	0.56	0.30
东德(该城市位于前东德吗?)	0.51	0.27

然而,如果一条线索的区分率较低的话,它的较高的生态效度也不一定很有用。区分率是指一条线索依据特定参照等级在所有成对对象间作出有效区分的相对频率。区分率既是线索值分布的函数,又是参照等级中所有预测对象数目的函数。分别令正、负线索值的相对频率为 x 和 y 。那么第 i 条线索的区分率 d_i 为:

$$d_i = \frac{2x_i y_i}{1 - 1/N}$$

正像简单的计算告诉我们的那样,如果 N 是非常大的,那么区分率将接近于 $2x_i y_i$ 。

一条线索的生态效度越高,根据它作出的推断越准确;一条线索的区分率越高,它就越有可能被用来作出推断。9 条线索配对

86

相关系数的范围在 -0.25 和 0.54 之间,其绝对值的平均数为 0.19 。

不同策略从环境中抽取不同的信息。例如,最少化策略不关心哪条线索比其他线索包含更好的信息,它只需要估计出一条线索所指明的方向。采纳最佳启发式则抽取有关线索排列顺序的信息,以确定按照什么样的顺序尝试每条线索。而所有竞争对手都充分利用了来自环境的各条线索值,以便计算生态效度或回归系数之类的参数。

让我们更精明

我们模拟出了对环境具有不同知识水平的被试。知识的有限性可以采取两种形式：一种是对有关对象再认的有限性；另一种是关于被再认对象各条线索值知识的有限性。为了将关于再认的有限知识具体化，我们模拟出了能够再认 0 个到所有（83 个）德国城市的被试（即共有 84 种不同再认水平）。为了将关于线索值的知识具体化，我们模拟出了 6 类被试，他们分别了解 0、10、20、50、75 和全部（100%）与得到再认对象有关的线索值。将这两类有限知识结合起来就产生了 6×84 类被试，每一类被试都有不同程度或不同种类的知识。对于每一类被试，我们都模拟出了 500 个个体，他们所了解的特定对象和具体线索值各不相同。

模拟需要具有真实性，即模拟出来的被试应该能够像真正人类被试那样借助于再认启发式完成任务。因此，必须对模拟被试能够再认的那些城市加以仔细挑选，以便在一定比例上得到再认城市比未被再认城市更大。为了对城市再认与其人口规模之间的真实关系作出粗略估计，我们用人类被试进行了一项测验，所使用的被试是芝加哥大学学生。我们发现在 80% 的情况（即配对测验的数目）下，他们能够再认（从 83 个德国大城市中）的城市都比不能再认的城市更大。我们在进行模拟研究时充分利用这种信息为每一类被试挑选再认城市（即为每一种知识状态的被试挑选适合于它们的再认城市数目），以确保得到再认的城市在大约 80% 的配对测验中大于未被再认的城市。因此，就城市再认和其人口规模的关系来看，模拟被试保持了人类被试的水平。关于模拟的细节，请参见吉戈伦尔和古德斯腾的有关研究（1996a）。

每一个模拟被试都要采用六种策略中的每一种作出关于两个城市中哪个更大的推断。这六种策略即三种快速节俭启发式策略（采纳最佳、采纳最近、最少化）和三种线性方法（回归、富兰克林规则、戴维斯规则）。在归纳推理中，我们还不曾考查过快速节俭启

发式在现实环境中的成效问题。如果简捷启发式是适应于环境结构的,那么他们应该不至于一败涂地。

这些启发式有多么节俭?

我们用一种启发式需要寻找线索的数量来测定它的节俭性。三种线性模型总是寻找并整合所有 10 条线索(9 条生态线索和 1 条再认线索)。而在所有有限知识状态下,采纳最近启发式平均只需要寻找 2.6 条线索,最少化启发式平均寻找 2.8 条线索,采纳最佳启发式平均寻找 3.0 条线索(表 4-2)。采纳最近启发式之所以最为节俭,也许是由于利用了态度定势,它倾向于采集区分率最高的线索。最少化启发式寻找的线索少于采纳最佳启发式,可能是由于线索的生态效度和区分率是负相关的(参见表 4-1),所以随机选择的线索倾向于比根据生态效度选择的线索有更高区分率。总之,三种启发式平均利用的线索数量还不到线性模型利用线索数量的 1/3。

表 4-2 三种快速节俭启发式(最少化、采纳最近、采纳最佳)与三种线性策略(达维斯规则、富兰克林规则和多元回归)的对比

策 略	关于线索的知识	节俭性(寻找线索的数目)	准确性(%)
采 纳 最 近	方 向	2.6	64.5
最 少 化	方 向	2.8	64.7
采 纳 最 佳	排列顺序	3.0	65.8
达 维 斯 规 则	方 向	10.0	62.1
富 兰 克 林 规 则	效 度	10.0	62.3
多 元 回 归	beta 加权	10.0	65.7

注:这是所有有限知识水平(即:有限再认和所知道的有限的线索数)条件下的平均结果。例如,最少化策略平均只寻找 2.8 条线索,而作出 64.7%的正确推断。

这些启发式有多么准确?

在只寻找很少一部分适用信息情况下,三种启发式还可能较高

的准确性吗? 譬如最少化规则, 它平均只寻找 2.8 种信息, 采用了单一理由决策规则, 还不了解哪条线索比其他线索更好, 甚至可能违背了可传递性标准, 那么它注定是要失败的了。然而, 表 4-2 中的数据显示最少化策略平均获得了 64.7% 的准确性, 略微高于采纳最近启发式(64.5%), 而低于采纳最佳启发式(65.8%)。但是使用所有适用线索、并将它们综合起来加以计算的三种推断策略的准确性如何呢? 它们获得了更高的准确性吗? 图 4-2 中的结果是令人惊奇的。尽管达

88

维斯规则充分利用了最少化和采纳最近启发式拥有的所有信息(只有采纳最佳启发式了解线索的排列顺序, 这种信息对达维斯规则是不适用的), 但其推断准确性比三种启发式中的任何一种都低。富兰克林规则利用了三种启发式拥有的所有信息, 甚至比这更多的信息, 但其推断准确性仍然连三种启发式策略中最为节俭的启发式都不如。

与更为强劲的竞争对手相比, 简捷启发式的准确性如何呢? 多元回归方法算出了对于线性预测来说最优化的一套权重, 要做到这一点, 需要相当大的计算量。尽管它作出了比最少化和采纳最近策略更为准确的推断, 但与快速节俭的采纳最佳启发式相比, 至多只能算是势均力敌。

图 4-2 表明了六个竞争对手的推断准确性随着得到再认城市数量增加而变化的情况。请注意: 这里所显示的是所有竞争者推断准确性的最佳水平, 即当关于各条线索的知识处于 100% 状态下的成绩。该图表明: 当得到再认城市数量有限时, 最少化和采纳最近启发式的准确性与其他策略的准确性持平, 但当所有城市都被再认(即全部信息均适用)时, 它们的准确性低于其他策略。在无任何城市或所有城市得到再认两种极端情况下, 富兰克林规则和达维斯规则的准确性与采纳最佳启发式的准确性持平, 但当再认城市数量处于中等程度时, 它们的准确性低于其他策略。为什么会这样呢? 原因在于这两种策略均背离了再认启发式这种简单的“智慧”。它们过于倚重每个对象所携带的具体线索值, 所以有时会

89

会将未被再认城市判断为大于得到再认城市。在我们模拟的

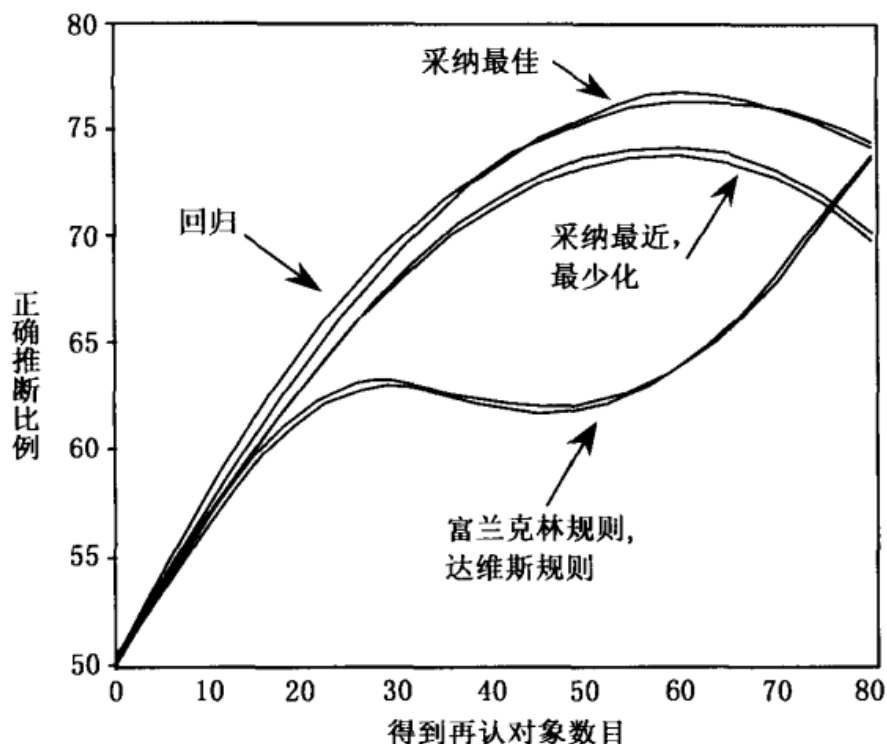


图 4-2 有关线索的知识达到 100%，但再认率变化时，各种策略推断准确性之比较

环境下，大多数城市所携带的负线索值多于正线索值。例如，多数城市都不是州府所在地，而且没有大足球俱乐部，等等。这样，当戴维斯规则用一个对象所拥有的正线索数减去其负线索数时，常常会得到一个负值，即便对于得到再认城市来说也是如此，有时候得到再认城市算出的负值还会大于未被再认城市，以此为基础作出判断时，就会将其判断为小于未被再认城市。富兰克林规则同样存在这个问题，它要对所有理由（线索）进行加权计算（Gigerenzer & Goldstein, 1996），并依据加权计算结果来决定哪个城市更大，因此未被再认城市常常会被判断为大于得到再认城市（即它们的线索加权值常常大于得到再认城市），这在 80% 情况下得到再认城市大于未被再认城市的环境中显然是不适合的一种策略。只有当无法使用再认启发式时（无一再认或全部再认），这两种线性策略的准确性才会与采纳最佳启发式持平。换句话说，如果人们能够用再认启发式这种简单机制来辅助线性策略的话，那么即使

在部分再认情况下,它们也应该达到更高的准确性。

从图 4-2 我们还可以看到,六种策略中的四种都表现出了少即多效应(参见第 2 章)。与图 2-4 不同的是,那里的曲线是以参差不齐的形式出现的,而这里的曲线是以平滑、精确的形式出现的。这是因为在图 2-4 中再认效度是随着得到再认城市的增多不断变化的,而这里的再认效度是恒常稳定的(0.8)。

准确性与节俭性的权衡

三种启发式策略存在同样的权衡问题:越节俭(寻找的线索值越少),可能越不准确。然而,当我们将三种启发式与三种线性策略加以比较时,产生了一些非常有趣的结果。与多元回归方法相比,采纳最佳规则并没有因为它的节俭性而牺牲准确性——它做到了两全其美。与达维斯和富兰克林规则相比,三种启发式都更加准确,同时也更加节俭。

当我们刚刚获得这些结果时,我们也难以相信它们。为了排除我们的主观愿望或想法可能对结果产生的影响,我们分别在美国和德国雇用了一些独立的程序员来运行我们的模拟程序。在我们发表这些结果时,我们还将有关模拟环境的详细资料包括在了论文中,以便大家都能够自己亲自来重复它,而且许多人真的这样做了(Gigerenzer & Goldstein, 1996a)。快速节俭启发式的确未必以牺牲准确性为代价来换取它们的节俭性!

节俭性和准确性可以兼而得之吗?

快速节俭启发式能够作出关于世界未知特征的准确推断,即等同于甚至好于三种线性策略的准确性。在设计这项模拟研究时,我们曾经怀疑这些启发式是否会一败涂地。在我们报告这些结果之前,判断和决策制定领域的三位知名专家曾经预言:采纳

最佳启发式的准确性可能会比线性策略的准确性低 10 到 5 个百分点。然而,三种启发式中的每一种都比这些预言做得好,甚至超过了某些线性方法的准确性。采纳最佳启发式甚至等同于或超过了所有三种线性方法的成绩。当时我们并不知道为什么会产生这种令人费解的结果。这个问题的答案——我们称之为生态理性的概念和观点——在我们经过了进一步的艰难探索后才找到,并将在第 6 章详加分析。这里我们仅将我们的几点看法总结如下。

线性模型的“极限平铺”(flat maximum)效应是我们观察到的现象之一。在线性模型中,如果采用许多套权重获得了像仅采用一种最优化权重一样好的成绩,就被称为极限平铺。这种现象是通过达维斯及其同事(Dawes & Corrigan, 1974)的研究工作被引入到决策研究领域的,但实际上很久以前就受到了有关研究者的注意。自从威尔克斯(Wilks, 1938)关于“等同加权”(equal weights)通用性的著作出版以来,许多研究者已经指出:无论对于人工系统(如,某种 IQ 测验)作出预测而言,还是对于描述人类实际推断活动来说,加权都是无关紧要的。在心理测验中,对一套测验的各种成分进行加权是很少见的做法,因为不同加权方案常常导致惊人相似的合成份数,即极限平铺(例如, Gulliksen, 1950)。极限平铺似乎在多条线索间呈高度正相关时较容易产生。快速节俭启发式取得的成绩表明,极限平铺效应可能超出加权问题而被推广到决策规则本身:基于唯一一条最佳线索作出的推断能够像基于所有线索的加权线性联结作出的推断一样准确。

此外,还有一些零星的早期证据表明,简单的、不可代偿的启发式能够取得很好的成绩。然而,由于多数早期研究主要关心偏好判断(而不是推断)和人工问题(而不是现实环境),所以常常难以获得评定成绩的外部标准。正像前面提到过的那样,与采纳最佳启发式最为接近的是词典编纂策略。培恩等人(Payne et al., 1993)的研究表明,词典编纂策略有时能够取得相当于加权线性模型的成绩,但他们未曾使用评定准确性的外部标准。近似于采纳

简捷启发式

最佳启发式的第二种方法是机器学习领域中的简单算法规则,它能够非常准确地完成分类任务(Holte, 1993; Rivest, 1987)。采纳最佳启发式较为疏远的“亲属”是“逐项排除程序”(elimination by aspects; Tversky, 1972),它也采用了有限搜索和终止规则,但主要关心偏好判断问题,而不是推断问题,而且并不利用线索的排列顺序进行判断,而是利用概率标准进行搜索,这种标准需要了解每条线索的多种效度指标。这样它就无法使用再认启发式,也不能利用单一理由决策规则。采纳最佳启发式的另外一个“远亲”是“分类和回归树”(classification and regression trees, 缩写为CARTs),它使用了一种简单决策树和单一理由决策规则,但建立这种简单决策树时所使用的知识和计算方法不同于采纳最佳启发式。例如,布里曼等人(Breiman et al., 1993)仅用三条双极的、有序的线索建立了一种分类和回归树,用来将心脏病患者鉴别为高危或低危病人。这种遵循可代偿原则的简单决策树比使用多达 19 项变量的标准统计分类方法还要准确(参见第 1 章)。其实际意义是十分明显的:在急诊室中,医生可以迅速地完成对 1 种、2 种或 3 种变量的测定,而不必进行任何复杂计算,因为毋需对信息进行整合。所有这些理论均假定人类推理活动使用了类似于采纳最佳启发式的机制。关于这些理论的更多论述,请参见“适当说”(relevance theory; Sperber et al., 1995)和“优化说”(optimality theory; Legendre et al., 1993; Prince & Smolensky, 1991)。

总之,极限平铺现象、简单机器学习规则和分类回归树的良好成效以及培恩等人的研究工作给了我们信心,使我们相信在初步的令人惊奇的发现背后还有更多东西有待我们去探索。

匹配终止规则与环境

为了取得像它们已经做到的那样准确的成绩,快速节俭启发式能够利用现实环境中的何种信息结构呢?在何种情况下它们将会

失败呢？第5、6章我们将着重讨论这些问题。这里我们仅结合终止规则的“正值偏向”(positive bias)来说明生态理性——即头脑与环境的匹配——思想。如前所述，一种正线索值和一种未知线索值相结合能够终止搜索，但一种负线索值和一种未知线索值相结合却不能终止搜索。这种不对称性即我们所谓的“正值偏向”。在人类推断和决策研究中，人们已经观察到了多种多样的正值偏向(例如，Klayman & Ha, 1987)，而且该种偏向能够使人们作出比无偏向终止规则更加节俭和准确的推断。不妨考虑一种无偏向终止规则，它既需要正线索值，也需要负线索值(就像吉戈伦尔等人所做的那样，1991)。这种终止规则是不够节俭的，因为在知识(关于未知线索值)有限的情况下，它将不得不比利用正值偏向策略花费更长时间搜索线索。现在我们再考虑另外一种快速的无偏向终止规则，它总是在正线索值适用时终止搜索，但只有当负线索值和未知线索值不存在时才这样做。与这种无偏向终止规则相比，在负线索值比正线索值出现频率更高的环境下，可以证明正值偏向策略将会作出更加准确的推断。一个非常直观简单的原因在于此种环境下未知线索值更有可能是负线索值。然而，如果未知线索值是负线索值，那么当一个人依据一种负线索值和未知线索值终止了搜索时，将会导致更不准确的推断，因为这实际上常常意味着他是依据两种负线索值终止搜索的。因此，在负线索值数量超过正线索值数量的环境下，采取正值偏向策略终止搜索从生态学上看是合理的。一个典型的例子就是本章所研究的环境，在这里只有相当少的城市有足球俱乐部，而且只有很少的城市是州府(亦可参见 Oaksford & Chater 的“稀少”假设，1994)。更一般地说，在正值指标稀少而分散的环境——如，表明疾病的罕见症状、暗示某种能力的不寻常特征——下，采用正值偏向策略的终止规则将被证明具有生态学的合理性。

92

可推论性

采纳最佳启发式如何估计线索的排列顺序呢？采纳最近和最

简捷启发式

简化启发式又如何了解线索所代表的方向呢？有几种方法可以帮助人们掌握线索及其排列顺序。首先，线索（或者了解线索的心向或准备状态）可能通过进化得到与生俱来的编码。在这方面，距离知觉、配偶选择以及食物规避等适应性行为所利用的线索可以作为例证（例如，Buss, 1992）。其次，线索也可以通过知识或技能传递方式来掌握。例如，学习某种专长所需要的线索可以通过学徒期或交流秘诀而习得。最后，线索也可以通过直接观察来掌握。例如，仅了解 10 个德国城市部分线索值和其中一些配对中哪一个较大的人，可以利用这种知识估计出所有德国城市具有的线索值及其方向。相反，在本章所报告的模拟研究中，每一种推断策略都依据全套资料算出了其需要的参数（如，线索方向、线索顺序、线索效度、回归系数等）。

那么，如果不得不从少数样例推断出全套线索的话，采纳最佳启发式做得如何呢？如前所述，采纳最佳规则仅从了解的样例中抽取线索顺序及其含义，这相对于更加复杂的统计方法所利用的真值加权、回归系数或条件概率等信息而言只是很少量的一部分信息。因此，在学习情景下，采纳最佳启发式仅从一套有限样例中抽取出了少量信息。相反，回归方法从有限样例中抽取出了相当多的信息。那么，哪一种策略更好一些呢？

图 4-3 表明了随着学习样例数的增多采纳最佳、采纳最近、最少化以及多元回归四种策略准确性的变化情况。每一种策略都从不同比例（10% 到 90% 之间）德国城市的线索中估计出各自所需要的参数，并作出关于剩余城市人口多少的预测。将所有城市区分为训练项和测验项，从训练项了解有关参数以及作出关于测验项预测的过程均重复 500 次。在这项模拟研究中，再认不被作为一种变量，即我们假定所有城市均得到了再认。我们首先看一看训练项和测验项的全部线索值都得到了了解的情形（图 4-3a）。例如，训练项达到所有城市的 50% 时，采纳最佳启发式达到了 72% 的正确预测率，而多元回归达到了 71% 的正确预测率。更一

般地说,在整套测试中采纳最佳启发式的成绩都超过了多元线性回归的成绩,特别在训练项较少的情况下,更是如此。图 4-3b 表明了一种更加困难的情境,即区分训练项和测验项之前先排除了一半可利用的线索,在这种情况下采纳最佳启发式的优势更加明

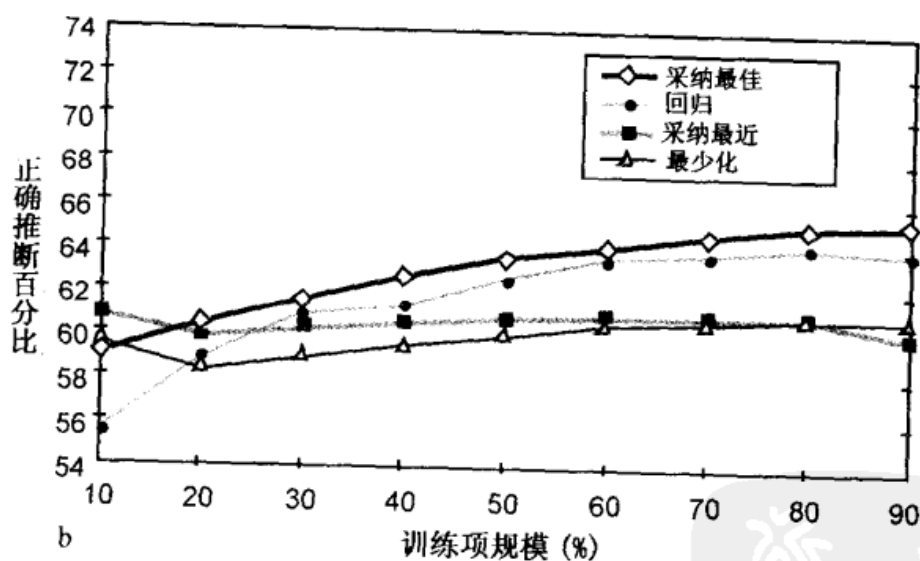
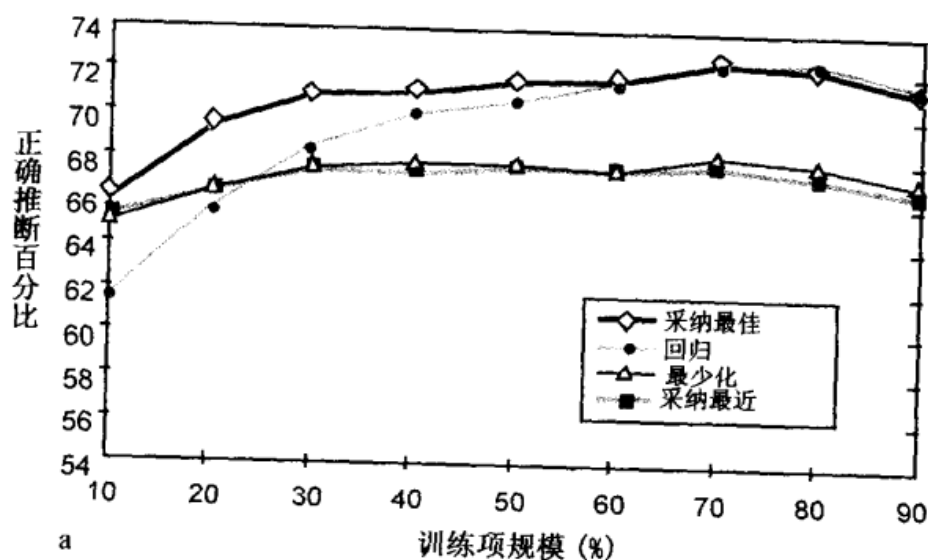


图 4-3 从训练项到测验项的可推论性。图中表明了训练项为全部项目的 10% 到 90% 时的推断成绩。被试仅作出关于未训练项目的推断,因此 10% 的条件意味着测验项目为剩余的 90%。我们还改变了关于环境的知识数量。图 4-3a 显示的是训练项和测验项的全部线索值均得到了解情况下的成绩;图 4-3b 显示的是区分训练项和测验项之前先随机排除 50% 线索值情况下的成绩。

简捷启发式

显。而且当训练项较少时,两种最为简单的启发式——采纳最近和最少化——取得了等同于或更好于其他策略的成绩。这种结果表明,在给定的这套资料中采纳最佳启发式比多元回归方法更加耐用,而且其成绩较不容易受训练项目数的影响而变化。在知识有限情境下,更简捷的策略也是更耐用的。

94

那么计算起来比多元回归还要复杂的策略的可推论性如何呢?以德国城市为对象,查特等人(Chater et al., 1997)将采纳最佳启发式与包括神经网络模型和概念范例模型在内的一些复杂策略进行了比较。像多元回归方法一样,这些策略中没有一个是使用终止规则,而是全部使用了所有适用信息。当训练项目数少于测验项目数的40%时,采纳最佳启发式的成绩超过了所有竞争对手;当训练项目数达到最少时,采纳最佳启发式的优势也最明显(超出10个百分点)。只有当训练项目数超过所有德国城市的40%(这实际上比大多数德国居民所具有的知识还多,包括德国本地人)时,竞争对手的成绩才超过了采纳最佳启发式的成绩,但最多也只超过了5个百分点左右。然而,值得指出的是,查特等人的模拟研究使用了没有任何未知线索值(如图4-1中的问号所表示的那样)的情境。

这些让我们惊奇的研究结果表明,在知识有限而必须从少数样例推断整体的情境下,简捷启发式规则可以做得相当出色。第5、6章将更加详细地阐述快速节俭启发式的耐用性。

适应工具箱

最少化、采纳最近和采纳最佳启发式是我们适应工具箱中“成批”启发式规则中的一些候选工具。我们所以将强调的重点放在“成批”一词上,是因为没有任何一种工具能够解决不确定条件下的所有推断问题。例如,上述三种策略都是为了估计两个对象中哪一个更大、更有效、更危险,等等。它们不能够对一个对象所具

有的量化值作出估计。然而,有些构筑块——简单终止规则、单一理由决策规则——可以被结合起来形成新的启发式规则,来完成数量估计、分类和其他任务,正像我们在本书后面的有关章节将会看到的那样。

人们也许可以将适应工具箱中成批的启发式规则看成是由进化造就的“元件”组成的,而不是一项宏伟的研究计划设计出来的。因此,也许可以通过把一些特征附加到现有工具上,来充实或发展适应工具箱,而不是用全新的一套工具来代替现有工具。例如,本章提出的三种启发式全都是围绕再认启发式构建起来的。如果能够使用再认启发式,就毋需搜索进一步的知识。如果不能使用再认启发式,就借助于附加工具作出推断。在这里这两类启发式规则的先后顺序是按照其可能的进化和发展顺序排列的:相对于采纳最佳之类的启发式所利用的回忆能力而言,再认和再认记忆是更为基本的适应机能,不太容易受到衰老和脑损伤的破坏(参见第2章)。

本章提供的最重要的结果是:充分利用简单心理机制的快速节俭启发式至少能够产生像复杂标准线性统计策略一样准确的推断。这种结果将我们从一种广为流传的观点中解放出来——这种观点认为只有“理性的”算法式规则(从富兰克林规则到多元回归方法)才可能是准确的。人类推断活动未必因为节俭性而抛弃准确性。我们的头脑完全能够做到两全其美。

当我们首次在有关这些研究结果的报告中作出上述论断时(Gigerenzer & Goldstein, 1996a),在我们内心深处始终萦绕着挥之不去的疑惑。启发式真的是快速的、节俭的,同时又是准确的吗?或许我们所使用的城市人口问题或德国城市问题有一些特别之处,以至于得到了上述的结果。这种集快速、节俭和准确于一身的启发式规则的效力能够推广到其他领域吗?它们利用了自然环境中的哪些信息结构呢?什么情况下它们将不能生效呢?下面的章节将告诉读者迄今我们对这些问题所找到的答案。更令人惊奇的发现还在后面呢!

第5章

简捷启发式有多好

金·塞勒林斯基(Jean Czerlinski)

哥德·吉戈伦尔(Gerd Gigerenzer)

丹尼尔·G·古德斯腾(Daniel G. Goldstein)

心理学已经忘记了它是关于有机体与环境之间关系的科学，而变成了关于有机体的科学……这在某种程度上让人联想起了中世纪那些夸夸其谈的男性神学家们所采取的立场：灵魂仅属于男人，而不属于女人。

——艾冈·布伦斯维克(Egon Brunswik)

斯

迪威·鲍尔(Steve Bauer)赢得了1991年“环法自行车拉力赛”(Tour de France)第一天的比赛，但在三周比赛结束后他仅排在了200位参赛者的第97名(Abt, 1991)。这种消耗体力的自行车拉力赛要穿越各种各样的地形，因此赢得了第一天的比赛并不能保证其他时间也取得好成绩。同样地，采纳最佳启发式在吉戈伦尔和古德斯腾(第4章)用德国城市人口问题所设置的竞赛情境中的成功并不能保证它在其他竞赛中也能获得成功。因此，在出售你的复杂的多元回归软件或皈依快速节俭方法之前，请你先留意阅读一下这一章。所有策略将在跨越20种环境的两轮比赛中一决高低，其预测对象涵盖从鱼类生殖力到燃油消耗量等领域。第一轮比赛是资料拟合：所有策略都要事先在比赛将要经过的道路上进行赛前训练。第二轮比赛更加困难，因为比赛开始之前不允许参赛者(一些策略)了解实际的比赛路

线。它们必须自己辨别方位。这两轮比赛的结果将有助于我们澄清何时采用快速节俭启发式较为有利、何时使用多元线性回归之类的复杂策略较为合适。

认识环境

环法自行车拉力赛的魅力就在于它要穿越各种不同地形，从平原到丘陵，再到山地。斯迪威·鲍尔赢得了第一天的比赛，因为他擅长于在平原地带比赛，但他最终输掉了比赛，因为他不擅长山地和丘陵地带的比赛。获胜者必须是一个全能选手。我们的竞赛没有什么不同，它将由 20 种不同环境所组成。每一种环境都包括三部分：一些评价对象、与每一种评价对象有关的预测效标和可能有助于作出预测的一些线索。比赛的内容是：依据一种效标推断两个对象中的哪一个将获得更高分数，如两个高级中学中哪一个有较高失学率等。对作出这种推断可能有用的线索包括一所学校低收入学生所占比例、平均标准入学分数以及父母亲对他们孩子学业的关心程度等。我们设置的环境涵盖了很不相同的领域，从绵长高速公路上发生交通事故次数的客观记录，到对公众人物个人魅力的主观评价（表 5-1）。环境的复杂程度也各不相同，其包括的对象从 11 个（在 11 种场合测到的旧金山市空气的臭氧水平）到 395 个（395 种鱼类的生殖能力）不等，其拥有的线索也从 3 条（在策略之间进行区分所需要的最低限度的线索数量）到 18 条不等。要想赢得比赛，一种推断策略必须在多种不同环境下都做得很好。我们使用的多数环境选自统计学教科书，在那里它们通常是作为讲授多元线性回归原理的案例来使用的。这将使得采纳最佳启发式策略战胜回归方法更加不易。

98

表 5-1 本次竞赛所使用的 20 种环境资料的说明

心理学

男性魅力：依据一些被试对 32 位著名男性可爱程度的等级评定、能够再认（仅仅见过名字，而未见过照片）每个男性的人数比例以及该男性是否是美国人等线索预测这些男性的一般魅力水平。[资料来源于 Henss(1996)用 115 名男性和 131 名女性德国被试（年龄范围是 17—66 岁）所做的研究]

女性魅力：依据同上的线索预测 30 位著名女性的一般魅力水平。（资料来源同上）

社会学

高中失学率：根据低收入学生所占比例、有色人种学生所占比例以及平均标准入学分数等线索预测芝加哥市 57 所公立高级中学的失学率。[资料来源于 Morton (1995)和 Rodkin(1995)的有关研究]

无家可归者：依据当地气温、失业率、贫困线以下低收入居民比例、房屋空闲量、是否有租赁限制以及公房比例等线索预测 50 个美国城市无家可归者的比率。（资料来源于 Tucker, 1987）

人口统计学

死亡率：依据 1 月份的平均气温、空气污染程度、有色人种比例等线索预测 20 个美国城市居民的死亡率。（资料来源于 McDonald & Schwing, 1973）

城市人口：依据是否有足球队、大学、城间铁路、展览馆等线索预测 83 个德国 10 万以上人口城市的人口数量。（资料来源于 Fischer Welt Almanach, 1993）

经济学

房价：依据当前资产税率、卫生间和卧室数量、占地面积、总居住面积、车库大小、房龄等线索预测伊利湖上 22 所房屋的售价。（资料来源于 Narula & Wellington, 1977; Weisberg 报道, 1985）

地租：依据所有可耕地平均地价、奶牛稠密度、牧场所占比例以及是否需要石灰来种植紫苜蓿（常被用来喂养奶牛）等线索预测明尼苏达州（1977 年为种植紫苜蓿的农业州）58 个县每亩地的地租。（资料由 Douglas Tiffany 提供；Weisberg 报道, 1985）

教授工资：依据性别、级别、得到目前级别的年数、最高学位以及获得该学位的年数等线索预测中西部学院 51 位教授的工资收入。（Weisberg 报道, 1985）

交通运输

交通事故：依据每条道路的长度、平均车流量、货车比例、速度限制、车道数目和宽度、侧道宽度以及岔路口数目等线索预测 1973 年明尼苏达州 37 段高速公路上车辆每行驶 100 万英里的事故率。（资料来源于 Carl Hoffstedt 未发表的硕士论文；Weisberg 报道, 1985）

燃油消耗：依据各州人口数量、获驾照者人数、燃油税、车辆流通量、主要高速公路长度等线索预测美国 48 个毗邻州机动车的人均燃油量。（根据 Christopher Bingham 为 1974 年美国年鉴搜集的资料，它还被收入了 1974 年的世界年鉴；Weisberg 报道, 1985）

健 康

18岁时的胖瘦情况：依据从2岁到18岁进行的跟踪身体测量预测46个儿童18岁时的身体胖瘦情况。测量包括身高、体重、腿围和腿长等指标。（根据伯克利指导中心的跟踪监控资料，Tuddenham & Snyder, 1954; Weisberg 报道, 1985）

身体发胖：依据对218位男性身体各种围长的测定预测其在水下称得的身体重量（对身体肥胖的一种更加准确的测量）。（来源于A. Garth Fisher根据Penrose等人的研究编辑的资料, 1985）

生 物 学

鱼类生殖力：依据每条鱼的重量、年龄以及所产卵的平均重量预测395条北极嘉鱼的产卵量。（资料由Christian Gillet提供, 1996）

哺乳动物睡眠：依据脑重量、体重、寿命、妊娠时间、掠食习惯和危险性预测35种哺乳动物的平均睡眠时间。（资料来源于Allison & Cicchetti, 1976）

奶牛耗氧量：根据对奶牛每日粪便的分析（分析指标包括生物氧含量、化学氧含量、含氮量、固体物含量以及挥发性固体物含量）预测其日耗氧量。（Moore, 1975; Weisberg 报道, 1985）

环境科学

物种差异：根据岛屿面积、海拔高度、与最近岛屿的距离、最近岛屿的面积及与海岸的距离等线索预测加拉帕哥斯群岛26个岛屿上的物种数量。（资料来源于Johnson & Raven, 1973; Weisberg 报道, 1985）

雨量：根据云彩的类型、云彩遮盖的比例、云彩是否由飞机喷射以及实验开始后的天数等线索预测Coral Gables地区24天的雨量。（资料来源于Woodley et al., 1977; Weisberg 报道, 1985）

洛杉矶的氧化水平：依据每天的风速、气温、湿度和日晒量（对日光的一种测定指标）预测洛杉矶17天的氧化水平。（资料由洛杉矶污染控制中心提供; Rice 报道, 1995）

旧金山的臭氧水平：依据过去2年冬季的平均降水量以及位于海湾南端的San Jose市的臭氧水平预测旧金山市一年中11种场合的臭氧量。（资料来源于Sandberg et al., 1978; Weisberg 报道, 1985）

注意：对于每一种环境，我们都详述了它的效标、用来预测效标的主要线索以及资料的来源。如前面有关章节所述，线索或者是双极的，或者以中数为界一分为二。任务均为预测两个对象中哪一个在效标上得分更高。

认识竞争对手

在吉戈伦尔和古德斯腾设置的竞赛情境（第4章）中，让多种

简捷启发式

推断策略相互较量。下面我们简单地描述一下这些策略中的 4 种,更加详细的内容请参见上一章。

采纳最佳启发式

试想有那么一辆自行车,它由几个不同参赛者各自所喜欢的零件组装而成:一个人提供了车架,一个人提供了轮胎,还有一个人提供了车把。与自行车零件不同的是,采纳最佳启发式是由认知构筑块组成的,这些构筑块即指导搜索的规则、终止搜索的规则和作出决策的规则(参见第 1 章和第 4 章有关的界定)。

采纳最佳启发式的第一步是再认启发式。在我们设置的两种竞赛情境中,所有对象均得到了再认,而且所有线索均为已知的,因此不能够利用再认启发式(参见前面的有关章节)。如前所述,采纳最佳启发式将依次尝试每一条线索,寻找能够将两个对象区别开来的那条线索。例如,当推断两个教授中谁的工资更高时,可能会首先尝试级别线索。如果两个教授处于同一级别(如,均为副教授),那么也许尝试性别线索。如果其中一个是男性,另一个是女性,那么我们就可以说性别线索是有鉴别力的。一旦找到了一种有鉴别力的线索,就将其作为推断的基础,其他线索就被忽略了。例如,如果两个教授性别不同,那么就可以据此推断男教授将挣更多工资,而不必再考虑工作年限、获得的最高学位之类的其他信息。这种单一理由决策规则可能是准确的吗?本章将回答这个问题。 101

由于采纳最佳启发式未对信息加以整合,也无需进行复杂计算,所以它是快速的;由于它采取了一种终止规则限制信息搜索,所以它是节俭的。在我们设置的竞赛情境中,采纳最佳启发式按照线索的效度大小顺序搜寻线索,这种顺序必须从一套训练项目中估计出来。如前所述,一条线索的效度是指使用这条线索作出正确推断的次数除以推断的总次数(参见第 4 章)。

最少化启发式

快速节俭的最少化启发式按照随机顺序搜索线索,当找到了一条能够将两个对象区别开来的线索时,即终止搜索。除此之外它与采纳最佳启发式是一样的。由于同样理由,在我们的模拟研究中,它将无法利用再认启发式。

多元线性回归

在我们挑选的选手中多元线性回归策略是最为娴熟和准备得最为充分的选手。它依靠复杂计算而不是快速节俭的构筑块来参与竞争。回归方法假定可以借助于一条最佳曲线和一个平均值为0的独立的、恒定的误差分布对数据作出近似估计。然后它试图找到这样一条最佳曲线,使得所有数据点与它的垂直距离的平方和为最小。寻找这样一个最佳拟合平面并非仅用铅笔、纸张或标准袖珍计算器就能很容易完成的任务——必须拥有一台功能齐全的计算机。当使用回归方法进行预测时,需要搜集所有可用线索,并将每一条线索并入到回归模型中,因此它是不够节俭的。而且,由于多元线性回归需要复杂计算,它也是不够快速的。

达维斯规则

102

达维斯规则是回归方法的一种简化形式。这种模型仍然是线性的,但放弃了最优化加权的做法,而代之以单元加权(+1或-1)方法(Dawes,1979)。这就是说,它仅将积极证据的数目加在一起,然后减去消极证据的数目。在本模拟研究中,为了使单元加权法成为可操作的,每当一条线索的效度超过了机遇值(0.5)时,我们即赋予其+1的权重,而每当一条线索的效度低于机遇值时,我们即赋予其-1的权重。由于使用达维斯规则作出预测仍然需要所有线索,所以它仍不够节俭,但不像回归方法那样,它是快速的,因为其加权方案是简单的。

让我们更精明

简捷启发式

我们决定让上述四种策略作为参赛选手。谁将赢得跨越 20 种不同环境的拉力赛呢？

第一轮比赛：拟合已知环境

在我们的第一轮比赛中，允许选手们在比赛之前详细查看比赛将要经过的路线和适应比赛环境。与第 2、3、4 章介绍的研究不同，在该项竞赛中没有遗漏任何线索值，也没有任何未被再认的对象。所有线索和效标值均可以被用来计算线索效度和进行线性加权。在此基础上所有策略都要对效标值作出预测（它们实际上已经见到过这些策略）。这种类型的竞赛被称为“资料拟合”（data fitting）。我们将详细描述在拟合上做得最好的两种环境——高中失学率和教授工资，以便了解各种策略是如何相互竞争的。然后我们将直接到比赛终点，去看一看谁将赢得整个比赛以及赢了多少。

失学率

第一轮比赛对于美国社会是重要的，因为它涉及到了在美国社会中相当普遍的一种现象——失学率问题。我们以芝加哥市 57 所公立中学为例来预测美国中学的失学率。该市 1995 年的失学率被披露在《芝加哥杂志》上（Morton, 1995; Rodkin, 1995），同时还列出了与失学率有关的一些线索，如学生的社会经济和种族成分、班级大小、学生出勤率、家长参与程度以及在多种标准考试中学生的分数等。

为了使其适合于我们所采用的四种策略，我们对杂志上的原始资料进行了适当改变，即将原本为具体数字的所有线索值均以中数为界限转换成双极线索值（0 或 1），0 表示较低效标值，1 表示较高效标值。

将资料转换成适当格式之后，就可以用我们的方法对它们的

统计特性进行测定了。总起来看,由失学率环境得到的结果是相当富有挑战性的。与在德国城市人口资料上得到的高达 0.76 的平均线索效度相比,失学率资料的平均线索效度仅为 0.63,其最高的线索效度也仅有 0.72。这种统计特性为实施采纳最佳启发式造成了相当大的困难,因为该种策略需要依靠一条最佳(线索效度很高的)线索来区分两所学校。而且,失学率环境包含的线索总量多达 18 条,为城市人口环境线索数量的两倍。这对于达维斯规则来说是较为有利的,因为它随着线索数目的增多而愈加准确(参见第 6 章)。但对于采纳最佳和最少化策略而言,线索数目的增多将会增加搜索线索的难度。

在说明我们采用的四种策略预测失学率的准确性之前,有必要回顾一下上一章有关德国城市人口的研究结果(图 4-2 最靠右边的值,即所有城市均得到再认情况下的推断准确率)。如前所述,多元回归作出了 74% 的正确推断。相对而言,当所有对象均得到再认时,达维斯规则做得相当好,其准确率也达到了 74%。令人惊奇的发现是采纳最佳启发式这种快速节俭策略取得了与上述两种线性策略匹等的准确性。就连最为简单的最少化策略也取得了接近 70% 的准确率。

那么,在我们所使用的更加困难的中学失学率环境下将会获得什么样的结果呢?尽管线索效度较低,回归方法仍然获得了 72% 的推断准确率(表 5-2,拟合度)。也许较多的线索数量弥补了较低线索效度的不足。达维斯规则似乎并未从更多线索数量上获得多少好处,仅达到了 64% 的准确率。采纳最佳启发式作出了 65% 的正确推断,略微好于达维斯规则,但却比回归方法低了 7 个百分点。最少化策略仍然为成绩最差的策略,但也取得了 61% 的准确率,与采纳最佳策略相比落后得并不太多。采纳最佳和最少化策略平均只寻找很少线索(表 5-2,节俭性)。在更加困难的中学失学率任务上,启发式规则为它们的快速性和节俭性付出了大约 7 个百分点的准确性代价。

表 5-2 预测高中失学率

策 略	节 俭 性	准确性(%正确)	
		拟 合 度	可 推 论 性
最 少 化	2.7	61	58
采 纳 最 佳	3.4	65	60
达 维 斯 规 则	18	64	62
多 元 回 归	18	72	54

注：预测成对芝加哥中学中哪一个有较高失学率时两种快速节俭启发式(最少化、采纳最佳)和两种线性策略(达维斯规则、多元回归)的成绩。共有 57 所公立中学和 18 条线索(表 5-1)。依据预测的节俭性(寻找线索的数目)和准确性(%正确)测定成绩。准确性包括拟合资料(测试项目数=训练项目数)和可推论性(测试项目数≠训练项目数)。可寻找的线索数对于两类竞争者来说是一样的。

对制定公众政策的意义 发现哪种策略最能够拟合于哪些资料对于制定有关政策具有重要意义。例如,采纳最佳启发式依次将学生出勤率、写作课分数及社会科学分数作为预测失学率最有效的线索。相反,线性回归方法的前三项预测指标分别是西班牙籍学生、英语水平有限学生和黑人学生所占比例。因此,每一种策略都将导致我们采取不同措施来帮助学校降低失学率。采纳最佳启发式的使用者可能会建议学校设法增加学生出勤率,并采取更加有力的教学措施来教他们,而回归方法的使用者可能会建议学校设法帮助少数民族学生,使其尽快融入主流文化,并支持将英语作为第二语言的计划(ESL 计划)。由于在本研究中回归方法拟合率最高,所以回归方法的使用者似乎能够对降低失学率问题提出更好的建议。

104

教授收入

现在让我们根据下面 5 条线索看一看四种策略在预测个别教授收入方面做得如何,这 5 条线索是:性别、级别(助教、副教授、正教授)、获得目前级别年限、最高学位和获得最高学位年限。具体资料来源于中西部学院的档案材料,教授们的真实姓氏被隐去

精

让我们更精明

了。很明显,这种环境中已经有一条双极线索(性别),其他四条线索以中数为界限双极化。

该环境最高的线索效度为 0.98,平均的线索效度为 0.79,接近于城市人口环境的平均线索效度。它只有 5 条线索,大体上为城市人口环境线索数的一半。如此高的线索效度和如此少的线索数将会对采纳最佳和最少化启发式产生什么影响呢?一种非常直观的回答是它将会使两种启发式的预测准确性赶上或超过整合所有线索的算法规则。让我们看一看是否如此。

第一个到达终点的仍然是驾御最为精美和昂贵赛车的选手——多元回归方法,它令人惊奇地达到了 83%的正确率(表 5-3,拟合度)。这是令人惊奇的,因为除了线索数目将近少了一半之外,这种环境似乎等同于上一章所用的城市人口环境(但却没有得到同样的结果)。仅采用一条线索,采纳最佳启发式以 80%的预测准确性排在第 2 位。达维斯规则——高级赛车(线性模型)的初学者——达到了 75%的正确率。它被线性回归策略拉下的距离像在预测失学率时一样远。最少化策略仍然以 73%的准确率落在了最后,但其成绩与达维斯规则的差别并不大。

表 5-3 预测教授工资

策 略	节 俭 性	准确性(%正确)	
		拟 合 度	可 推 论 性
最 少 化	2.1	73	72
采 纳 最 佳	2.3	80	80
达 维 斯 规 则	5	75	75
多 元 回 归	5	83	80

注:在预测两个教授中哪一个有较多工资时两种快速节俭启发式(最少化、采纳最佳)和两种线性策略(达维斯规则、多元回归)的成绩。共有 51 位教授和 5 条线索(表 5-1)。依据节俭性(平均寻找线索的数量)和准确性(%正确)测定成绩。准确性包括拟合资料(测试项目数=训练项目数)和可推论性(测试项目数≠训练项目数)。对于两类竞争者来说可寻找的项目数一样。

该项竞赛表明,对于预测教授工资来说最好的线索是级别(其

简捷启发式

线索效度达到了 0.98), 其次是性别(线索效度为 0.88)。在该竞争环境中, 回归也相当一致, 将最高权重赋予了教授级别, 接下来依次是最高学位和性别。

第一轮比赛的最后获胜者

现在我们已经知道参赛者之间是如何相互竞争的, 并了解环境特征是如何影响每个参赛者的。下面让我们看一看哪种策略赢得了跨越所有 20 种环境的比赛, 即不仅适应于平原地带的比赛, 而且也适应于山地和丘陵地带的比赛。

简捷启发式有多么节俭呢? 为了作出一种决策, 最少化启发式平均仅搜索 2.2 条线索。采纳最佳策略搜索的线索略多一些, 平均为 2.4 条。而两种线性策略总是利用所有适用信息, 平均搜索 7.7 条线索(它们不使用指导和终止搜索的启发式)。因此, 两种启发式搜索的线索数不到所有线索的 $1/3$ 。它们是如此节俭, 那么在准确性上如何呢?

毫不奇怪的是, 第一个到达终点者是使用所有信息、并进行复杂计算的多元回归策略(表 5-4, 拟合度)。在跨越所有 20 种环境的比赛中, 回归方法获得了 77% 的正确率。然而, 第二个完成比赛者却有点令人费解。快速节俭的采纳最佳启发式仅以落后于回归方法 2 个百分点的劣势完成了比赛, 其正确率达 75%。这与吉戈伦尔和古德斯腾的发现(第 4 章)是十分接近的, 说明我们选用的环境中更多情形类似于城市人口资料, 而不是类似于中学失学率资料。快速而不节俭的达维斯规则以 73% 的正确率落后于采纳最佳启发式 2 个百分点。在采纳最佳启发式更加节俭、且未对搜集到的线索加以综合情况下, 达维斯规则仍然比它做得更糟, 这是颇令人费解的。最少化启发式以 69% 的准确率落在了最后。考虑到它的极端节俭性, 这是一个可以接受的成绩。采用快速节俭启发式付出的代价是很小的: 对于采纳最佳启发式而言, 仅为大约 2 个百分点, 而对于最少化策略而言, 仅为大约 8 个百分点。

106

而且,更多的信息并未保证更加准确的成绩,因为采纳最佳启发式使用较少信息反而比达维斯规则取得了更高的准确性。

表 5-4 跨越 20 种环境的总成绩

策 略	节 俭 性	准确性(%正确)	
		拟 合 度	可 推 论 性
最 少 化	2.2	69	65
采 纳 最 佳	2.4	75	71
达 维 斯 规 则	7.7	73	69
多 元 回 归	7.7	77	68

注:跨越 20 种环境条件下两种快速节俭启发式(最少化、采纳最佳)和两种线性策略(达维斯规则、多元回归)的成绩。20 种环境平均的线索数目是 7.7 个。依据节俭性(平均寻找线索的数量)和准确性(%正确)测定成绩。准确性包括拟合资料(测验项目数=训练项目数)和可推论性(测试项目数 $\neq$ 训练项目数)。对于两类竞争者来说可寻找的项目数一样。关于使用略微不同的一套环境获得的结果,可参见 Gigerenzer et al. (1999);而关于其他不同策略在上述 20 种环境下得到的结果,可参见表 8-1。

我们早知如此

面对一种新异论断,人们的第一反应通常是:“那是不可能的。”然而,吉戈伦尔和古德斯腾(1996a)早已证明他们的论断——快速节俭启发式也可能是准确的——是成立的。本章进一步证明它不仅有时是可能的,而且事实上经常都是如此。节俭性需要付出昂贵代价的情形(如,预测中学失学率的环境)反而似乎是例外的和不合常规的。

当一种“不可能”的新异论断最终得到了证明时,人们的第二种反应通常是:“我们早知如此。”(参见第 9 章,在那里这种记忆错失现象被作为“事后通偏向”进行了详细论述)这里我们仅对有关的心理学文献加以回顾,看一看关于快速节俭启发式相对于更加复杂策略表现出来的优势,人们早已知道了什么。

我们不可能将我们的发现与早期的研究进行直接比较,因为早期研究在许多方面不同于我们的研究。首先,早期研究中参与

竞争的各种策略主要局限于线性模型的不同加权方案；其次，早期研究所选用的竞赛环境通常都是在假定多条线索值和效标值均呈正态分布基础上人为设定的资料；第三，竞赛的类型也不同。早期研究通常不涉及拟合给定一套资料的问题，而只关心如何将有关发现推论到新资料中去，即仅用一部分资料训练一种算法规则，然后用这种规则作出关于另一部分资料的预测。如果两部分资料的容量同样大，那么这种方法就被称为“交叉验证”(cross-validation)。下一部分我们将返回起点，用交叉验证方法看一看各个参赛者的成绩。在此之前我们首先回顾一下先前的有关文献。

简单策略研究真正的开端是 20 世纪 70 年代中期。借助于计算机模拟和数学分析等手段，施密德(Schmidt, 1971)、达维斯和科雷根(Dawes & Corrigan, 1974)以及爱因霍恩和霍格斯(Einhorn & Hogarth, 1975)等研究者发现：总体上看，“单元加权”(unit-weighted)线性模型(我们所谓的“达维斯规则”)能够作出几乎像多元线性回归模型一样准确的推断——而且就其效果来看更加耐用(一种“耐用”的策略或模型意味着当将其推论到一套新资料时仍然能够保持预测准确性，即经得起交叉验证)。例如，在预测不同年级的分数时，单元加权线性模型作出了与真值相关为 0.6 的预测，而交叉验证回归模型作出的预测与真值的相关仅为 0.57。值得注意的是，由于回归模型采用的是交叉验证方法——对一套不同于训练项目的资料作出预测——所以它的预测成绩可能会低于单元加权模型(它不采用交叉验证)。在其他 3 种任务中的 2 种任务上，单元加权方法的准确性也都超过了交叉验证回归方法(Dawes & Corrigan, 1974)。正像鲍·米尔(Paul Meehl)指出的那样，“多数情况下，未加权的少数‘重要’变量的简单相加比复杂回归算式做得更好”(引自 Dawes & Corrigan, 1974, p. 105)。

在较为近期的有关研究中，恩伦伯格(Ehrenberg, 1982)采用复杂分析方法对回归加权和其他加权方案进行了系统比较。他证明：对于典型的单线索预测问题(如，在效标和线索之间有 0.7 的

相关)来说,使用一种不同于最佳拟合曲线的斜率(从最佳拟合斜率中加或减 30%)仅使未被解释的误差增加了 4%。达维斯和科雷根[1974,引用 Winterfeldt & Edwards(1973)未发表的手稿]将这种现象称为“极限平铺”:大略近似于最佳拟合的权重导致了几乎像最优化权重一样好的预测结果。

这些研究似乎表明达维斯规则几乎总是作出像多元回归方法一样准确的推断。但是生活远非如此简单。首先,在能够利用现实环境进行预测的情形下,很少有人利用这种环境进行预测。其次,交叉验证仅用于多元回归方法,而未被用于达维斯规则,因为“作出预测的是了解预测变量与所关心效标之间向量关系的人类判断者”(Dawes,1979,p. 573)。然而,即便是专家也必须采取一些方法来估计线索的方向,因此在下一部分我们所做的交叉验证模拟研究中,将要考察这样一个问题:当达维斯规则也必须估计线索方向时,它的预测成绩如何?我们暂且不考虑米尔提出的仅使用少数重要变量的建议,这个问题留待下一部分加以探讨,现在我们将继续使用所有线索。我们的研究超越了以往的研究,因为它涉及到了达维斯规则的所有方面——将交叉验证的达维斯规则与交叉验证的回归方法进行对比——看一看先前的发现能否保持下来。

我们的研究超越以往研究之处还在于,它更加强调这样一个权衡问题:在不损失太多准确性前提下,推断策略怎样才能做到尽可能节俭?我们的模拟研究不仅涉及到了达维斯规则与多元回归方法之间的比较,还涉及到了与采纳最佳和最少化策略的比较。曾经有一些分散的研究也尝试过非常简单的启发式规则(例如, Hogarth & Makridakis, 1981; Kleinmuntz & Kleinmuntz, 1981),但只有培恩等人(Payne, Bettman & Johnson, 1988, 1990)的工作形成了连贯的研究方案。他们的方案主要关心偏好判断(如,两种赌注中更喜欢哪一个),而不是基于线索的推断,而且他们是通过一种选项与某种加权附加规则(预期结果)的相关来评定它的成绩的,而不是借助于外部标准来测定成绩的(因为对于主观选择而言

简捷启发式

不存在这样的外部标准)。例如,在他们所做的比较中,更加简单的启发式规则取得的成绩通常都能达到加权附加规则基准成绩的60%到70%,但用这种测定方法,简捷启发式规则不可能比加权附加规则“理性的”(rational)答案做得更加准确。只有借助于外部标准来评定推断准确性,才有可能证明简捷启发式能够比更加复杂的策略做得更好。由于先前的研究主要关心偏好判断而不是基于线索的推断,关心人工环境而不是现实环境,所以未能使简捷启发式在作出关于现实世界推断方面的优势显示出来。

在本章中,我们利用比先前研究更加广泛的经验资料构成的环境来检验启发式规则。我们让这些启发式参加一项环法自行车拉力赛!

第二轮比赛:推论到新环境

设想有那么一个选手,他花费大量时间在美国中西部地区的平原地带进行刻苦训练,然后想要参加穿越多种地形的环法自行车拉力赛,那么将会发生什么事情呢?他很可能会在山地和丘陵地带的比赛中一败涂地。这并不是说他必须到法国去进行训练,只要他能够到科罗拉多山区、中西部平原和蜿蜒的新英格兰街道等多种不同道路上进行训练,他就能够为参加比赛作好充分准备。在一种道路上进行训练、而到另一种道路上进行比赛(测试)是我们所谓的推论,而不仅仅是拟合。

更严格地说,推论意味着一种策略依据全部对象中的一部分样例(即训练项)建构其模型(即计算回归权重、确定线索排列顺序或方向等),然后用所建立的模型对剩余对象(测验项)作出预测。相对于用同样对象进行训练和测验而言,推论是对一种策略的更加困难和现实的考验。在我们的模拟研究中,随机地将所有对象一分为二:一半为训练项,另一半为测验项。这就是所谓的交叉验证。一种策略的成绩就是用它预测测验项的正确率。为了使区

分资料造成的有利或不利影响得到平衡,每一种环境的 1 000 次测试都被对半地分成训练项和测验项(各 500 次)。

在训练阶段达维斯规则和最少化策略似乎不需要进行任何估计,但实际上它们需要利用前一半测试(训练项)对线索方向作出估计。在我们的模拟中,它们通过计算线索效度高于或低于机遇水平来做到这一点(这相当于计算线索与效标之间有正的或是负的“Goodman-Kruskal 等级相关”,正像第 6 章所证明的那样)。采纳最佳启发式需要估计线索顺序,并按照从最好到最差的顺序排列它们。多元回归则需要估计出最优的 β 权重,并用它解释变量之间的关系。

现在我们将开始跨越每一种环境的比赛,看一看每种策略在更加困难的推论任务上的表现如何。在第 4 章中,我们已经看到过一种推论情境,即预测城市人口。当用一半城市对各种策略进行训练时(像我们的比赛中所做的一样),采纳最佳启发式的推论准确性(72%)略微高于多元回归和达维斯规则的成绩(71%)。这种结果可以推广吗?单一理由决策制定在跨越 20 种环境的比赛中仍然可能是更加准确的吗?

再看失学率

我们首先看一看预测中学失学率时各种策略的可推论性。最为简单的最少化策略的准确性从原先的 61%降低到了 58%;类似地,达维斯规则的准确性从 64%降低到了 62%。训练时既估计线索方向、又估计线索排列顺序的采纳最佳启发式蒙受了略多的损失,其预测准确性从原来的 65%降低到了 60%(表 5-2,可推论性)。最后,多元回归方法蒙受了 18 个百分点的巨大损失,其预测准确性从 72%陡降到了 54%,这也使它成为四种策略中可推论性最差的方法。情况似乎是:越是简单的策略,在可推论性上越耐用。如何解释回归方法预测成绩的陡然下降现象呢?

我们相信原因在于它的“匹配过当”。试想有那么一个自行车

简捷启发式

选手,每天训练时都按照比赛的要求,首先爬一段陡峭的上坡路,然后是一段长长的平路,最后是下坡路。日复一日,这个选手已经习惯于首先艰难地攀登,继而迅速地前进,最后轻松地下坡。这样做的危险在于该选手如此习惯于这种路程搭配模式,以至于难以应付其他坡路与平地的搭配方式。如果比赛时的路线首先是下坡,继而是平路,最后是陡峭的上坡,该选手也许难以适应它。同样地,对于各种推断策略来说也可能发生这种匹配过当现象:它们也许过分熟悉训练资料的特定模式,诸如具体的线索顺序、线索间的内部关系等,以至于难以摆脱它们的限制。一种策略越是想要拟合训练资料,其匹配过当的危险性就越大。

失学率环境包含了 18 条线索。如此丰富的线索为或然相关提供了足够机会。如果回归策略将这些或然相关也吸收进了所建立的回归模型,那么它对另一半资料(它们未必具有像上一半资料一样的或然相关)的预测将是不准确的。

对制定公众政策的意义 我们不妨再次考虑一下上述发现对制定公众政策的意义。在拟合比赛中,回归策略的使用者非常自信地建议学校采用将英语作为第二语言的教学计划(ESL 计划),以降低失学率。由于当时它的预测准确性是最好的,所以这种建议似乎也是最可取的。然而,在推论比赛中,采纳最佳启发式比回归方法预测的更加准确,而且回归方法似乎过于匹配于训练资料了。尽管依据拟合资料回归方法将最大的权重赋予了外语教学计划对失学率的影响上,但这也许是一种难以推广的侥幸成功。另一方面,采纳最佳启发式依据拟合资料提出的增加出勤率和加强基础课教学等建议则具有更好的可推论性。达维斯规则和最少化策略并未提出具体建议——只是建议在所有方面进行改进——因为它们同等重视所有预测指标。

本章后面我们将指出,在拥有更多预测对象(或更少线索)的容量较大的环境中,回归方法匹配过当的可能性将会降低。尽管如此,但芝加哥市实际上只有 57 所高级中学,而且如果为了适合

于回归方法而不得不增加学校数量的话,那我们宁愿放弃回归方法。没有必要为了它而搜集更多资料。有人也许愿意用其他城市或芝加哥市以前的失学率资料来训练回归方法,但这样做时又会产生另外一种匹配过当的危险:找到的是与其他时间和地方有关的因素,而不是与今天的芝加哥市公立中学有关的因素。

教授收入

让我们简单地看一看依据 5 条线索预测教授工资的可推论性。回归方法的预测准确性只有些微降低,即从 83%降低到了 80%。达维斯规则和采纳最佳策略分别保持了它们原先的 75%和 80%的预测准确性(表 5-3,可推论性)。最少化策略的准确性降低了 1 个百分点,从原来的 73%降低到了 72%。与对中学失学率的预测相比,对教授收入的预测依据的线索较少,但每条线索的效度较高。在这种条件下,其从拟合到推论变化过程中预测准确性的降低似乎比预测城市人口时要少。

我们已经有了一些关于推论如何影响各种策略的资料。那么,在跨越 20 种环境的比赛中究竟哪种策略将作出更加耐用的推论呢?

第二轮比赛的最后获胜者

平均起来看,在预测准确性上回归方法蒙受了惊人的 9 个百分点的损失,从拟合条件下的 77%降低到了推论条件下的 68%(表 5-4,可推论性)。同时,达维斯规则从 73%到 69%平均降低了 4 个百分点,而最少化策略也从 69%到 65%平均降低了 4 个百分点。这两种策略之所以蒙受的损失较小,或许是由于这样一个事实:它们在训练阶段作出的估计较少——仅仅估计线索的方向。

最后的获胜者是采纳最佳启发式。在所有四种策略中它以 71%的推论准确率居于榜首,尽管它具有快速、节俭的特点。在跨越所有 20 种环境的比赛中,回归方法和达维斯规则每作出一种推

断平均使用了 7.7 条线索,而采纳最佳启发式仅平均使用了 2.4 条线索,像在拟合任务上使用的线索一样少。启发式规则能够背离搜集所有适用信息的理性戒律而成为最为准确的策略,这当然是值得深入思考的问题。

111

当作出关于新资料的推论时,采纳最佳启发式平均超出多元回归方法 3 个百分点。这种结果是令人吃惊的,但却与先前的有关文献不谋而合,这些文献表明在几种类型的环境中回归方法的可推论性均不如更加简单的达维斯规则(未进行交叉检验)。然而,我们的第 2 轮比赛表明达维斯规则也有匹配过当的嫌疑。而且,如果过分简单和节俭的话,也是需要付出代价的:最少化策略的预测成绩始终落在了最后。但这种代价并不很高,就最少化策略的平均成绩来看,仅落后于多元回归方法 3 个百分点。

比赛规则的修改

我们的一些同事对快速、节俭,同时又准确的单一理由决策规则表示怀疑。他们建议我们修改一下比赛规则,并预测采纳最佳启发式异常的准确性很快将会销声匿迹。早先人们对吉戈伦尔和古德斯腾报告的研究结果(1996a;参见第 4 章)提出质疑,认为在他们的研究中城市再认对于预测城市人口规模的效度过高(0.8),这可能导致再认启发式在单一理由决策活动中扮演重要角色,从而使采纳最佳规则取得异常的准确性。本章我们格外注意这个问题:在两轮比赛中,所有对象均得到了再认,因此再认启发式不可能在单一理由决策过程中起作用。下面我们考虑四种可能的修改方案和建议。

使用确切的而不是双极的线索值 在我们报告的模拟研究中,以中数为界对所有线索值(除了本身具有双极值的线索外,如性别)都进行了双极化处理,而不是利用确切的线索值。有人假定这种方法仅保留了关于线索值的有限信息,因此对于模拟精确的线索值来说是不可靠的。每一个参赛者,无论是线性策略还是启

发式策略,均依据这种二元的或双极的线索值作出预测。一种似乎合理的推测是,当被应用于这种双极资料时,多元回归模型就失去了它的部分效力。一些同事建议我们让每种策略都拥有确切的线索值,然后重新进行比赛。有两种主要方法可供采纳最佳启发式之类的词典编纂策略利用精确线索值作出推断(Payne et al., 1988, 1990):一种方法是一直搜索下去,直到找到两个对象具有不同线索值的那条线索为止;另一种方法是一直搜索下去,直到找到两个对象线索值之间的差异超过了某个阈限值[“刚刚可觉察的差异”(just noticeable difference)]的那条线索为止。为了避免确定阈限值的主观随意性,我们决定采用第一种方案。

我们让采纳最佳启发式和多元回归方法在第二轮比赛条件下重新进行比赛。当采用确切线索值时,采纳最佳启发式甚至比原先做得更加节俭。对信息的搜索常常在找到第一条或第二条线索时就终止了,因为即便非常微小的线索值差异就足以终止搜索,立即作出决策。但是基于量化指标的预测准确性如何呢?我们的同事是正确的:当给出真实线索值时,多元回归方法的准确性的确得到了很大提高——但采纳最佳策略的准确性也提高了。与多元回归方法相比,在跨越 20 种环境的比赛中,采纳最佳启发式平均作出了 76% 的正确预测,而多元回归方法恰好也获得了 76% 的正确率。因此,即便使用确切的量化线索值,采纳最佳启发式在预测准确性上仍然能够与多元回归方法持平。这种反常的结果再一次使我们感到惊奇,但这时我们已经习惯于这种惊奇了!

仅使用重要线索,再给达维斯规则一次机会 如前所述,鲍·米尔假定:“在多数情况下,未加权的少数‘重要’变量的简单相加比复杂回归算式做得更好。”在我们的两轮比赛中,达维斯规则均使用了所有线索。然而,鲍·米尔却建议仅使用重要线索,即最为有效的线索。在达维斯规则利用训练项估计线索方向时,效度较低线索常常导致对其方向的错误估计。因此,我们排除了效度低于 0.17 的线索,并用剩下的“重要”线索来测试每种策略的准确性。我们将这

种简化的达维斯规则重新放到第二轮比赛的环境中去。在此种情况下,其预测准确性的确从原来的 69%(参见表 5-4)增长到了 71%,高出了多元回归方法 3 个百分点。跨越 20 种环境的比赛证明米尔的推测是正确的:仅使用少数重要线索、且不对它们进行加权的情况下,达维斯规则的预测准确性通常高于需要对所有线索进行加权的回归方法。但是米尔的推测还可以更进一步:仅使用把两个对象区分开来的“最佳”线索也能够做得同样好。

如果没有现成的线索排列顺序(就像第一轮比赛那样),需要依据少量样本估计其顺序时,采纳最佳策略做得怎样呢?如前所述,采纳最佳启发式并不试图估计最优化的线索顺序(例如,像分类树方法试图做到的那样。参见第 6 章和第 8 章)。相反,它利用简单、快捷的方法创建一种线索顺序(就双极线索而言,线索顺序可以通过预测对象用一种简单方法估计出来;参见 Czerlinski, 1998)。根据线索效度排列其顺序(如同采纳最佳规则所做的那样)并非所谓的“最优化”,因为这种方法忽略了线索之间所有的依从关系。我们已经看到,对于估计城市人口规模来说,当训练项目很少时,采纳最佳启发式损失了一些准确性,但多元回归方法受到的损失更大(图 4-3)。这种结果表明,当根据对少量样本的观察作出预测时,采纳最佳启发式是相当耐用的。但是这种结果能够推论到其他资料吗?我们考察了当不得不从随机选择的 10 个对象(而不是前一半对象)估计线索顺序时,采纳最佳启发式的准确性。然后像在正规的第二轮比赛中所做的那样,让它作出关于另一半环境的预测。

113

这种比赛有助于查明采纳最佳启发式估计线索顺序时在多大程度上依赖于信息的丰富性。跨越 20 种环境的比赛结果是:采纳最佳启发式平均作出了大约 66% 的正确预测,比它用固定的一半对象进行训练时低了 5 个百分点。这个数字基本上与第 4 章用城市人口进行研究时获得的结果持平。这种“10 个样本”的比赛就好像只留给自行车选手很少训练机会一样,也就是说仅提前几

天将比赛路线告诉他们。采纳最佳启发式在只有很少机会进行观察的情况下,似乎仍然能够作出相当准确的预测。

采纳最佳启发式不能估计数量,而多元回归方法却能够做到这一点。这种质疑涉及到了采纳最佳、采纳最近和最少化三种启发式的可推论性问题。他们均能够依据一种效标作出关于哪个对象有较高效标值的预测(如,哪一条高速公路是更加危险的),但却无法作出数量估计(如,一条高速公路上的机动车事故率有多高)。因此,这些启发式规则只适合于特定种类的任务,而多元回归方法则可以应用于更加广泛的领域。在第10章中,我们将介绍一种简捷启发式,它能够作出数量估计,但却使用了类似于采纳最佳策略的单一理由决策规则。我们所谓的适应工具箱是由成批的不同启发式规则构成的,但这些不同规则却是根据同样的构筑块设计出来的。这些构筑块(而不是由它们设计出来的特殊启发式规则)是通用的。每种启发式规则的特殊性使得它们能够成为快速的、节俭的和准确的——毋需作出过多权衡。

采纳最佳启发式何以做得如此好?

什么环境下采纳最佳启发式能够取得好成绩?什么环境下它将做得很糟?这两种环境之间的差异何在?这些问题涉及到了采纳最佳启发式的生态理性问题,即启发式结构与环境结构之间的拟合问题(第1章)。这个问题是下一章将要探讨的中心问题——这里我们仅提出这个问题,简单回顾一下先前的研究,并用实验资料考证一下生态理性的三种一般预言。

对环境特征的描述:文献回顾

关于回归方法(需要进行交叉验证)和达维斯规则(无需进行交叉验证)谁将赢得我们的比赛,有关的文献已经发现了一些在它们之间作出判断的变量。然而,我们的目标是理解快速节俭启发式的

简捷启发式

成效,因此这些早期发现是否与我们的目标有关尚不清楚。而且,先前的研究要么对成百上千种随意生成的人为环境进行模拟研究,要么对有关资料形态的无数简单假设进行数学分析。有理由怀疑这种发现是否能够推论到我们的实验环境。尽管如此,还是让我们看一看早期的研究究竟发现了什么,但请记住上面的警示。 114

施密德(1971)用随机化资料(多元正态分布)所做的模拟研究表明,只有当环境包含了大量预测对象时,交叉验证回归方法的预测成绩才会超过达维斯规则。例如,若只有4条线索,一个人需要容量至少为50个对象的样本,才能保证回归方法打败达维斯规则。若适用线索为6条,一个人就需要容量至少为75个对象的样本,才能做到这一点。若线索增加到了10条,则样本容量至少应该增加到100个。一般而言,当每种对象包含的线索少于10条时,就不应该采用回归方法。否则,单元加权的平均预测准确性将会超过回归加权。对于给定线索数目而言,如果样本容量过小,回归方法极容易匹配过当,也就是说解释了当前样本实际上所不包含的过多内部相关。类似地,一个自行车选手进行训练时经过的路线越少,他就越有可能过分适应于他所经过的道路。

爱因霍恩和霍格斯(1975;亦可参见 Hogarth, 1981)证实了施密德的发现,并增加了两种其他因素:(a) 回归模型的决断系数(coefficient of determination, 即 R^2)落在中等或更低(0.5 或更低)的范围之内;(b) 线索内部相关。在此两种情况下,达维斯规则都可能大致上像多元回归方法预测的同样准确。决断系数测定的是效标与线索之间的线性拟合程度。

综合上述三种因素,我们看到,有关文献已经表明:如果每条线索有很多预测对象、对效标有很高的预测力,且线索间的内部相关不是很高,那么回归模型的预测准确性要略高于达维斯规则。在相反的条件下,达维斯规则的成绩要略好于回归方法。

下面让我们看一看上述三种因素(每条线索的预测对象数目、线性预测力及线索内部相关)是否能够解释为什么采纳最佳启发

式在我们两轮比赛所用的现实环境中会做得如此出色。

是由于每条线索预测的对象数目吗？

图 5-1 表示的是随着预测对象数与线索数之间比率(即每条线索平均预测对象数)的变化,采纳最佳启发式超出多元回归方法的优势(预测准确率之差)。当每条线索预测的对象较少时,采纳最佳启发式赢得较多。然而,当每条线索预测的对象很多,甚至超过 10 个时,也不能确保多元回归方法的成绩赶上或超过采纳最佳启发式。例如,图 5-1 中对象与线索间的最大比率是在预测身体肥胖时获得的,即用 14 条线索来预测 218 位男性的肥胖程度,这样每条线索平均大约预测 16 个对象。即便在如此高的对象/线索比条件下,采纳最佳启发式仍然比多元线性回归方法作出了更加准确的预测。

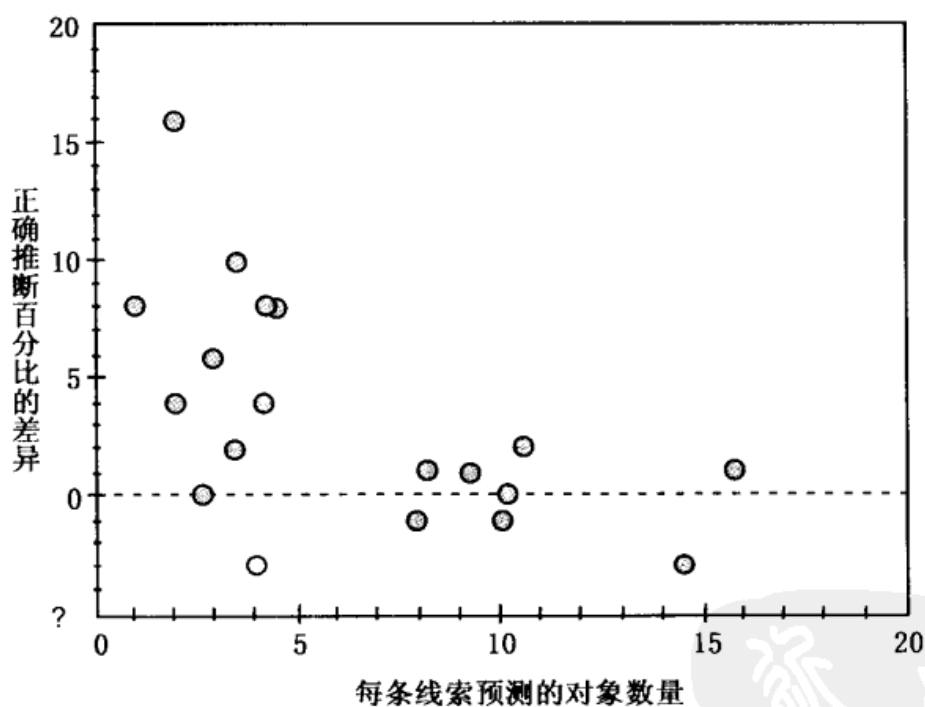


图 5-1 随着预测对象数与线索数比率的变化采纳最佳启发式相对于多元回归方法的优势(正确率之差)。资料依据于 20 种环境中的 19 种(第二轮比赛)。漏掉的一种环境为鱼类生殖力环境,它与其他环境差异过大(每条线索平均要预测 395/3 个对象),在该种环境下采纳最佳启发式的正确率落后于回归方法 1.8 个百分点。

简捷启发式

每条线索平均预测对象的数目在一定程度上可以解释采纳最佳启发式何以具有预测上的优势,这种发现支持如下假设:如果每条线索只预测很少对象,回归方法极容易匹配过当。对此种假设还可以采取更为直接的方法加以检验,即:当每条线索平均预测的对象数目变化时,将交叉验证的回归与仅仅拟合现有资料的回归进行比较(图 5-2)。结果产生了一种类似于图 5-1 的趋向:每条线索预测的对象数目越少,拟合条件下的回归超出推论条件下回归的优势也越大。上述两图均有一些不规则变化,因而显得散乱不齐。

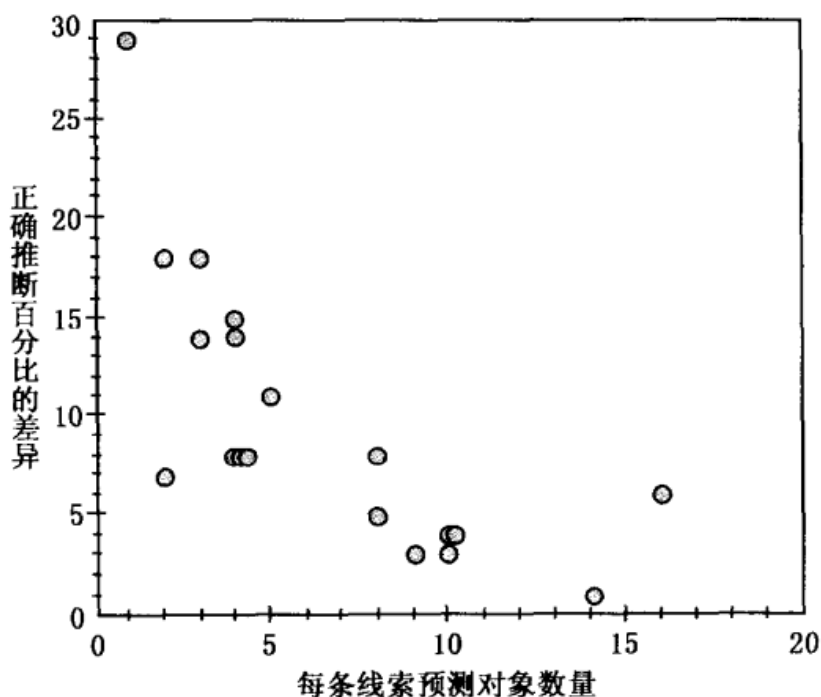


图 5-2 随着预测对象数与线索数比率的变化,多元回归方法在拟合任务上(第一轮比赛)的正确率减去其在推论任务上(第二轮比赛)的正确率。资料依据 20 种环境中的 19 种。由于同样原因,鱼类生殖力环境再一次被排除在外,该种环境下回归方法在拟合任务上的正确率超出在推论任务上的正确率 0.4 个百分点。

是由于线索的线性预测力吗?

环境的第二种特征是决断系数(R^2)。其基本想法是:在拥有

较高决断系数的环境中,多元回归方法将作出比达维斯规则更加准确的预测,尽管其过于精确的加权方案并不十分符合并非完全线性化的资料。我们通过对所有环境下获得的资料进行回归分析算出了 R^2 。然而,在 R^2 与采纳最佳和多元回归准确性的差异之间显然不存在任何相关。

是由于线索间的内部相关吗?

如果线索间是完全相关的($r = 1.0$),那么单一理由决策制定将像所有线索的线性联结一样准确。因此,我们的几个同事指出线索间的内部相关越高,采纳最佳启发式超出回归方法的优势将越多(也有人推测可能会出现相反的关系)。为了检验这种假设,我们算出了每一种环境中每一对线索之间的相关,取其绝对值,然后将它们加以平均。拥有较高平均绝对相关的环境将赋予采纳最佳启发式更大的优势(相对于回归方法)吗?我们并未发现这种趋势,因此该变量似乎不能解释采纳最佳启发式的成功。下面的几个变量也均未能成功地解释采纳最佳启发式的优势:最佳线索相关、最差线索相关和相关变异。

117

采纳最佳启发式究竟利用了现实环境的什么结构,以至于取得了如此好的成绩呢?从先前有关达维斯规则和回归方法的比较研究所发现的三种变量来看,只有一种变量(预测对象数目和线索数目的比率,即每条线索平均预测的对象数)与采纳最佳启发式超出回归方法的优势有关。因此,基于有关达维斯规则和回归方法的比较研究提出的假设只能部分地回答我们的问题,即:快速节俭启发式究竟利用了哪些环境结构?下一章依据精确的数学知识和证据对这个问题提供了更多的分析。

我们所知道的

为了抑制人们对新异结果作出“我们早知如此”的反应,我们

曾经要求判断和决策领域几位著名研究者对采纳最佳启发式在多大程度上能够接近多元回归的准确性进行预测。这些研究者均为不可代偿策略和多元回归方法方面的专家。他们的预测相当一致：多元回归方法的准确性将超出采纳最佳启发式 5 到 10 个百分点。正像让我们感到惊奇一样，我们的研究结果也让他们大吃一惊。

本章我们获得了以下几个令人惊奇的结果。

1. 由吉戈伦尔和古德斯腾(1996a)发现、并在第 4 章中加以总结的现象能够被推论到 20 种不同环境，而且适合于再认启发式不起作用的情形。这使得原来使用的德国城市人口问题或许包含着某些特别之处或错误的猜疑不攻自破。

2. 当用推论任务代替吉戈伦尔和古德斯腾的拟合任务时，快速节俭的采纳最佳启发式在跨越 20 种现实环境的比赛中甚至比多元回归方法更加准确。尽管使用的线索不到所有线索的 1/3，采纳最佳启发式仍然获得了其准确性。而且，最为节俭的最少化策略在准确性上也与多元回归方法相差无几。将早先的研究推进了一步，我们发现即便达维斯规则也采用交叉验证方法时，其准确性仍略微高于回归方法。

3. 有关变量并不能在很大程度上改变上述结果。例如，即便使用量化的而不是双极的线索值，采纳最佳启发式的准确性仍然匹等于或略高于多元回归方法。 118

本章我们未能解决的一个重要问题是“怎样”的问题：单一理由决策制定怎样才能像线性策略一样准确，有时甚至比它们更加准确——尽管后者使用了所有的线索，有些情况下还进行了复杂矩阵运算。图 5-1 中的结果表明线索和其预测对象数量之间的关系在一定程度上有助于回答这个问题。但这无助于解释原因和提供证据。下一章将对这个问题给出几种分析性答案，并提供有关证据。采纳最佳启发式能够利用什么样的环境信息结构呢？也就是说，什么样的环境结构使得启发式从生态学上看是合理

的呢？

本章开头我们引用了艾刚·布伦斯维克达尔文式的警句：为了理解头脑的运作机制，我们需要了解它过去和现在作用于其中的环境结构。然而，布伦斯维克也曾试探性地指出，多元回归方法是关于头脑如何推断其周围环境的一种有用模型。从那以后，许多新布伦斯维克主义者毫无保留地倚重于这些线性模型。相反，我们的结果表明，为了作出关于环境的准确推断，头脑完全可以采用不同于多元回归方法的简单策略。在本章所研究的情境中，简捷启发式能够获得同样目标。单一理由决策制定最终赢得了整个比赛！

第6章

为什么单一理由决策制定能够生效 ——生态理性的个案研究

劳拉·马蒂冈(Laura Martignon)

奥瑞齐·霍夫瑞格(Ulrich Hoffrage)

不应该假定存在比绝对必要的东西更多的东西。

——威廉·奥卡姆(William of Occam)

奥

卡姆简化论(Occam's Razor; 主张一切理论应力求简化——译者)最初是作为反对急剧增长的有名无实的东西的格言提出来的,现在已成为理论建构的一种方法学准则,它反映了人们追求简单化的倾向。用今天的科学术语来说,奥卡姆简化论可以改变为:宁肯选择能够对资料作出解释的最简单模型。对此类格言的需要本身就表明科学理论常常表现出一种相反的倾向:过分追求最优化,以至于变得高深莫测。在没有外来帮助的自然状态下人类的推断活动果真如此复杂和拐弯抹角吗?在认知心理学中有一种流行做法,即用科学研究所采用的原理来建构人类现实心理活动的理论。正像吉戈伦尔(1991a)指出的那样,有关心理机能的模型常常反映了科学推断过程所使用的相当复杂的统计工具包含的思想。本书采取了一种不同的观点,并提出了一些人类头脑在不失去其推断准确性前提下可以采用的一些简捷启发式规则。

就对偶比较问题而言,一种简捷启发式规则能够是快速

的、节俭的和准确的吗？我们对采纳最佳启发式在第5章所介绍的20种不同环境中的成绩感到惊奇，同时也对为什么它并未因为节俭性而丧失准确性而感到困惑。在那一章中，借助于对相关资料的详细分析，我们发现预测对象数目与线索数目之间的关系影响了采纳最佳启发式的成绩。正像第1章提到过的那样，简捷启发式成功的原因也许是它符合生态理性，更确切地说这是由于这些启发式的结构特性适合于它们应用于其中的环境结构。

在本章中，我们首先对有利于采纳最佳启发式和有利于其竞争对手的环境结构作了明确区分。其次，我们考查了采纳最佳启发式的耐用性，尤其是其简单信息搜索和终止规则的耐用性。最后我们从三个不同角度对采纳最佳启发式进行了理论分析：首先将其视为线性模型的一种特例，其次将其作为一种词典编纂方法，第三将其作为一种分类树。

环境结构

在描述有利于采纳最佳启发式发挥作用的环境结构之前，有必要首先阐明我们所谓的“环境”究竟是指什么。一种环境是由一套对象构成的。每一个对象都用一种效标值和一种线索配置方案来描述（表6-1是按照其效标值大小顺序排列的8个对象构成的一种环境，每个对象都用3条线索的不同配置方案来描述）。这里我们仅关心所有线索值均为已知、且所有对象均被再认的情形。因此，当使用“环境结构”一词时，我们采用的是对已知环境信息结构的一种简便易行的描述方法。我们对采纳最佳启发式在三种信息结构中的成绩进行了分析，它们是：不可代偿信息（noncompensatory information）、贫乏信息（scarce information）和丰富信息（abundant information）。

表 6-1 效标值依次排列的八个对象(a 的得分最高)构成的环境

对 象	线 索 1	线 索 2	线 索 3
a	1	1	1
b	1	1	0
c	1	0	1
d	1	0	0
e	0	1	1
f	0	1	0
g	0	0	1
h	0	0	0

注：每一个对象均用三条线索的不同配置方案来描述。对于采纳最佳启发式而言，这是一种完备的环境，因为当使用这些线索推断所有成对对象中哪个得分更高时，它获得了 100% 的正确率。

不可代偿信息

搜索信息的采纳最佳启发式是以单一理由决策规则为基础的。如前所述，这种决策规则在依据特定效标决定两个对象中哪个得分更高时，仅依据能够将两个对象区分开来的最为有效的那条线索。依据该线索作出的决策或许是错误的，但没有任何其他线索或线索的联结能够改变它。换句话说，采纳最佳启发式是一种不可代偿策略。如果一种环境有类似的信息结构，即它所包含的每条线索均比效度低于它的线索的任何一种联结更加重要，那么采纳最佳启发式将最为有效。例如，按字母表排列的单词顺序显然是不可代偿的。两个单词的第一个字母不同是排列它们时遵循的唯一准则。正是这种准则使得我们从字典中查取一个单词变得相当容易。另外一个例子是数的大小比较。如果两个大数有同样多的数字，那么一个人就可以通过比较最靠左边两个数字的大小来确定哪个数更大。如果最靠左边的两个数同样大，那么接着比较左边的第二个数字，依次类推，直到两个对应数字不同为止。这种方法非常类似于采纳最佳启发式搜索线索时遵循的规则。不

可代偿数字系统(如,阿拉伯数字系统)使得计算和比较变得非常容易。这就是它们所以比罗马数字系统更受欢迎的原因所在。

不妨考虑一下图 6-1 所示的视觉领域的不可代偿性的一个例子。左边剑手的腿在界内还是界外? 裁判和剑哪一个离观众更远一些? 在这里效标是距离,我们面对的是类似于第 4、5 章所考查过的那种任务,即:依据一种效标比较两个对象。关于如何加工有关线索以形成深度知觉的问题已积累了大量文献。许多实验证据表明,人们常常使用多条线索来形成深度知觉(例如, Bruno & Cutting, 1988; Massaro, 1988a), 尽管在这些线索如何被联结起来的 122 问题上还存在巨大分歧。然而,一些现象表明深度知觉线索有时是不可代偿的。在图 6-1 中,许多人透过裁判“看到”了剑,也就是说把剑看成是在裁判的后面。而且左边剑手的腿被知觉为在护栏的前面。由于这是一个两维画面,所以一些重要的深度知觉线索(如,立体运动、双眼视差)没有被保留下来,不能被用来区分有关对象。但仍然有几条线索可以使用,如对象重叠(较近对象部分地遮挡了较远对象)、对象大小(较大对象比较小对象更近一些)、透视、在平面上的相对高度(裁判的脚比剑更接近地面)以及一般感觉等。综合这些线索我们本应该将剑判断为在裁判的前面。

121

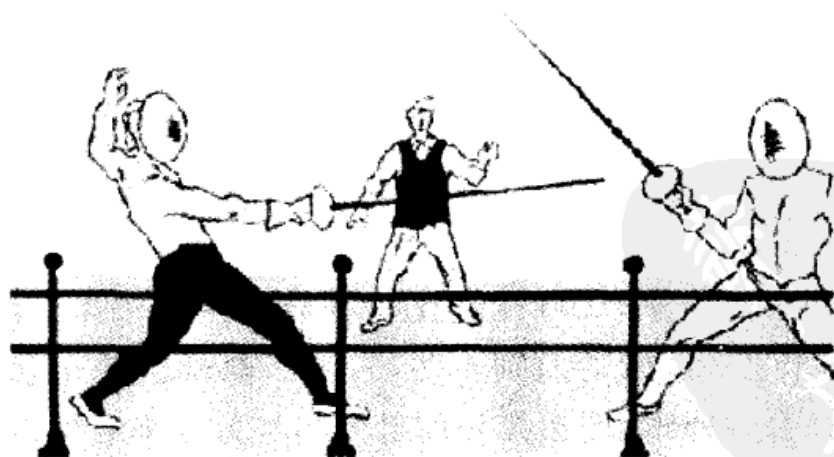


图 6-1 可以被归因于不可代偿性的一种视错觉——Petter 错觉(引自 Metzger, 1975; 得到了出版者的许可)

让我们更精明

上述错觉的潜在机制是对线索的不可代偿加工。如果对象重叠能够被用来判断剑和裁判的距离,那么仅依靠这条线索就足以作出决策。但由于画面中二者均为黑色的,该线索无法区分对象。第二条线索(对象大小)能够在两个对象间作出区分,据此作出的推断(裁判比剑更近一些)不能够被其他线索的联结所否决。

如果线索权重是不可代偿的,那么将线索线性地联结起来将会取得像按照词典编纂方法一次只加工一条线索一样的成绩(参见下面的推导 1)。在这种情形下,是词典编纂策略的成绩而不是加工过程等同于线性模型。通常的看法是此种情形下线索就是被线性地联结起来的(Bruno & Cutting, 1988; Jameson & Hurvich, 1959; 亦可参见 Julesz, 1971)。相反,我们的看法是:线索表面上看是被线性地联结起来的,但实际上它们是按照词典编纂策略得到加工的。这种看法与伦森(Runeson, 1977)及伦森和科斯塔(Costall, 1991)的观点是一致的,他们一再强调假定线索线性联结的模型也许并未领会深度知觉线索加工过程的精妙之处。线索的不可代偿排列顺序似乎是这种精妙过程的一部分。受有关知觉线索加工模型的启发,在布伦斯维克(1956)关于认知过程的观点中也包含了单一线索也许能够超过所有线索的思想,尽管他并未充分地证明它。

作为不可代偿加工过程的一个实际例子,我们不妨回想一下第 4 章提到过的雌性古比鱼的交配策略。在选择交配对象时,雌性古比鱼将橘黄色外表凌驾于其他所有线索之上(Dugatkin, 1996)。如果一个雄性古比鱼比另一个显然具有更加鲜艳的橘黄色外表,那么雌性古比鱼就会选择他作为交配对象。然而,如果这条线索的差别不明显,那么雌性古比鱼就会使用另一条社会性线索,即选择先前看到其与另一个雌性古比鱼交配过的雄性古比鱼。

上面我们列举了按照不可代偿方法使用线索的一些例子。下

面我们将用更为一般的术语对不可代偿策略加以界定。设想有一套(M 条)按顺序排列的双极线索, $C_1, C_2, \dots, C_M$ 。如果其中的一条线索 C_j 的权重超过了在它之后(即:从 C_{j+1} 到 C_M)线索的任何一种可能联结,那么这些线索对于给定的一种策略而言就是不可代偿的。在加权线性模型拥有一套完整权重 [$W = \{w_1, w_2, w_3, \dots, w_m\}$] 的特殊情况下,如果每当 $1 \leq j \leq M$ 时,我们都有 $w_j > \sum_{k>j} w_k$, 那么一种策略就是不可代偿的。换言之,就给定的权重排列顺序来说,如果每一种权重都大于随后权重的和,那么一种线性模型就是不可代偿的。例如,当一套线索的权重为 $\{1, 1/2, 1/4, 1/8, 1/16\}$ 时,情况便是如此。

拥有一套不可代偿权重的线性模型最终将作出像采纳最佳启发式完全一样的推断,反之亦然(Martignon & Hoffrage, 1999)。

推导 1: 不可代偿信息。采纳最佳启发式的成绩等同于拥有一套不可代偿权重(以类似于采纳最佳规则那样的顺序依次降低)的线性模型的成绩。如果一种环境由效度依次递减的不可代偿线索所构成,那么相应的加权线性模型的成绩不可能超过更加快速节俭的采纳最佳启发式的成绩。

简单地说,采纳最佳启发式体现了不可代偿的结构。如果一种环境有同样结构,那么拟合就产生了。这种拟合在一定程度上促成了采纳最佳启发式的生态理性。第 5 章所讨论的 20 种环境中有 3 种环境拥有不可代偿的回归权重,即其线索效度以同样次序递减。因此正像推导 1 所预言的那样,回归方法和采纳最佳启发式的成绩是一致的。如果拟合不太完善、但近似于完善,那么采纳最佳启发式仍然能够大致上像对应的线性模型一样准确。

采纳最佳启发式与多元回归方法的等价在表 6-1 所示的人工环境中明显表现出来。该环境是由 8 个对象(a, b, c, d, e, f, g, h)和它们在三条双极线索上的值构成的。假定 a 在效标上得分最

简捷启发式

高, b 次之, 依次类推。这是一种理想化的而非现实的环境: 线索配置方案刚好使得 8 个对象进行对偶比较时的效标值处于从 7 到 0 的 8 个等次上, 即在给定了位居前面的一条线索值之后, 该线索的效度和区分率(参见第 4 章)总是最优的。在这种环境中, 多元回归方法如何运作呢? 为了使用它, 我们必须为每个对象设定一种效标值, 如设想 a 的效标值是 8, b 的效标值是 7, c 的效标值是 6, 依次类推, 直到 h 的效标值是 1。在此种情形下, 第 1、2、3 条线索的“beta 权重”分别为 0.873、0.436 和 0.218(所有线索权重间的相关为 0)。这是一套不可代偿的线索权重, 因此推导 1 预言: 采纳最佳启发式与多元回归方法之间将取得完全对等的成绩。事实的确如此, 这两种方法取得了同样的成绩, 即均作出了 100% 的正确推断。然而, 作为在此种环境中最为简单的线性模型, 达维斯规则做得如何呢? 如前所述, 达维斯规则简单地将除了 0 以外的所有权重设定为 1。在这里所使用的具体任务上, 达维斯规则的分数的可以通过将每条线索配置方案中“1”的数目加在一起而计算出来, 而不是用“1”的数目减去“0”的数目。^{*} 因此在比较 d 和 e 这对对象时, 达维斯规则将会作出错误推断。在其他 6 种对偶比较(bc, be, ce, df, dg, fg)中, 它也不得不凭借投币式的方法进行随机猜测。这样它至多只能达到大约 86% 的正确推断。

124

采纳最佳启发式“赌定”环境是呈偏态分布的(图 6-2, 左), 而达维斯规则假定环境是均匀分布的(图 6-2, 右)。现实世界中很可能有更多的偏态分布, 而不是均匀分布(参见第 10 章), 因此“赌定”偏态分布也许能被证明是更好的策略。相反, 多元回归方法并不盲目地设赌, 而是精确地计算有关参数, 来拟合相应环境结构。这样做的代价是计算起来过于烦琐和模型的耐用性较差。

* 达维斯规则的分数的(参见第 4 章)通常是用线索配置中 1 的数目减去 0 的数目来表示的。就这里所用的任务来说, 达维斯规则等价于仅根据每条线索配置中 1 的数目所进行的比较。1 的数目的两倍减去线索的数目等同于 1 的数目减去 0 的数目。所以根据这两种记分方法为对象排列的等次是一致的。

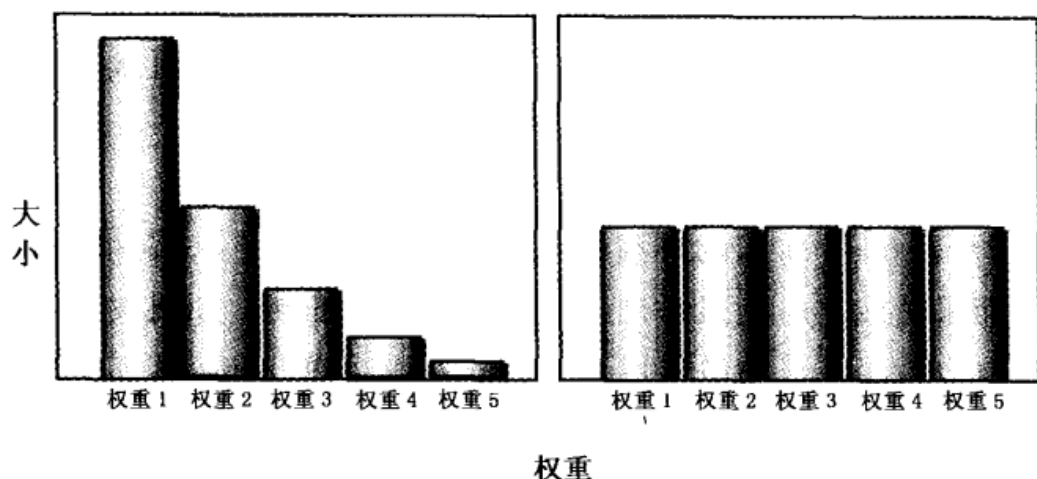


图 6-2 不可代偿(左)和可代偿(右)权重举例

贫乏信息

偏态分布并非唯一一种有利于采纳最佳启发式的环境特征。在许多现实情境中,我们不得不依据贫乏信息作出决策。下面我们首先对我们所谓的“贫乏信息”加以界定,然后看一看它是如何影响采纳最佳启发式的准确性的。

区分 N 个对象需要多少种信息呢? 为了回答这个问题,我们不妨考虑一种仅需要作出“是”或“否”回答的室内游戏。设想你和另外一个人在一个房间里,那里有 64 种不同玩具。你的任务是猜想另一个人心里最乐意选择哪一个玩具。你可以问一些仅需要对

方作出“是”或“否”回答的问题。要想准确猜出对方最愿意选择的玩具,你至少必须问几个问题呢? 答案是 $\log_2 64 = 6$ (个)。更一般地说,按照信息论(Shannon, 1948)的观点, N 个按等次排列的对象包含了 $\log_2 N$ 比特的信息。表 6-1 所示的环境包含了 8 个按等次排列的对象,它们恰好是由 3 种($\log_2 8$)双极线索来编码的。我们看到这些线索相互间非常协调一致: 线索配置方案独一无二地和明确无误地界定了各个对象。在多数拥有 8 个对象和 3 条线索的环境中,获得这种完善线索配置都是非常困难的。如果仅有 2 条线索,完善配置将是不可能的。这使得我们作出如下界定:

125

定义：如果 $M < \log_2 N$ ，那么拥有 M 种双极线索和 N 个对象的环境将仅提供贫乏信息。

在表 6-1 所示的环境中，达维斯规则作出了 86% 的正确推断，而采纳最佳启发式作出了 100% 的正确推断。如果仅考虑前两种信息，那么采纳最佳启发式将获得 93% 的正确率，而达维斯规则仍将获得 86% 的正确率。由前两条线索构成的环境提供了贫乏信息，但如我们看到的，此种情况下采纳最佳启发式的成绩仍超过了达维斯规则。那么，当将其推论到包含更多对象的环境中时，情况依然如此吗？通过详细计算，我们已经证明下面的推导对于“较小”环境——拥有 500 个对象以内的环境——来说是真实的 (Martignon & Hoffrage, 1999)：

推导 2：贫乏信息。在多数信息贫乏的小环境中，采纳最佳启发式比达维斯规则更加准确。

我们不妨分析一下在对象数目和线索数目固定、且现有线索仅提供贫乏信息的环境中，这两种方法的成绩。对上述推导的一种直观解释是：当使用大量线索时，达维斯规则能够补偿前几次推断可能造成的误差，而且很少需要进行猜测。而当只有贫乏信息时，达维斯规则的这种优势丧失了：它不能够利用补偿原则，而且不得不作出许多猜测。

将采纳最佳启发式与多元线性回归进行比较是更加困难的。用 30 个对象和 5 条线索进行的一项模拟研究表明，在信息贫乏环境中多元回归方法平均比快速节俭的采纳最佳启发式更准确一些（仅在拟合度上）。然而，如表 6-2 所示，也有一些例外。这里我们再次分别赋予 8 个对象以 8, 7, 6, ..., 1 的效标值。采纳最佳启发式的推断准确率为 91%，而多元线性回归的准确率是 87.5%。

我们尚无法用一个定律来描述采纳最佳启发式能够打败多元回归方法的环境,但我们的确知道:当回归权重不可代偿、且按照与线索效度同样的次序递减时,采纳最佳启发式与多元回归方法将总是作出同样的推断。

表 6-2 采纳最佳启发式的成绩超过多元回归方法成绩的一种环境

对 象	线 索 1	线 索 2	线 索 3
a	1	1	1
b	1	0	0
c	1	1	1
d	1	1	0
e	0	1	1
f	0	1	0
g	0	0	1
h	0	0	0

注:采纳最佳启发式的成绩为 91%,多元回归的成绩为 87.5%。

丰富信息

一般而言,增加贫乏环境中信息的数量将无助于采纳最佳启发式,尽管它能够纠正达维斯规则依据先前线索作出推断时所犯的错误。

当所有可能的不确定但却有效的线索存在时,一种环境就提供了丰富信息。在一种拥有 N 个对象和双极线索的环境中,可能的线索数量就是由 1 或 0 两个数字构成的长度为 N 的数字序列的数量。请注意:“所有可能的不确定但却有效的线索”并非指所有可能的真实线索,而是指不同的“1—0”数字序列。尽管真实的线索数量是不确定的(因为就每一个对象而言,不同的真实线索也许有同样的值),但不同的“1—0”数字序列的数量是确定的。举例来说,表 6-3 就是由 5 个对象(a,b,c,d,e)构成的一种丰富环境,在这里 a 在效标上得分最高,b 次之,依次类推。共有 8 条不确定但却有效的线索(即其效度大于 0.5 但小于 1 的线索)来描述 5 个

达维斯规则通过数出每个对象所包含的“1”的数目来计算其得分,因此作出了 100% 的正确推断。8 条线索对于确保在 5 个对象上作出完全正确的推断已经足够了(在仅由 7 条线索构成的环境中,达维斯规则将不得不作出一种猜测,即比较 b 和 c 时不得不凭借猜测,因此只能获得 95% 的准确率)。富兰克林规则(参见第 4 章)用每一个对象的效度乘以每一条线索的值,将得数加在一起,然后推断:有较高总得分的对象在效标上排列的也更靠前。在上述环境中它也能作出 100% 的正确推断。一般而言上述结果均为真实的吗?

表 6-3 拥有丰富信息的 5 个对象(效标值依次排列)构成的环境

127

对象	线索 1	线索 2	线索 3	线索 4	线索 5	线索 6	线索 7	线索 8	得分
a	1	1	1	0	1	1	0	1	6
b	0	1	1	1	0	0	1	1	5
c	1	0	1	0	1	0	1	0	4
d	0	1	0	0	1	1	0	0	3
e	0	0	1	0	0	0	0	1	2
$v =$	5/6	5/6	3/4	3/4	4/6	4/6	4/6	4/6	

注:丰富信息是指效度 v 大于 0.5、小于 1 的所有非多余双极线索。此种环境中达维斯规则和富兰克林规则均能作出完全正确的推断。

就拥有 5 个或更多对象的环境而言,下面的推导是真实的(有关的分析可参见 Martignon & Hoffrage, 1999):

推导 3: 丰富信息。当环境提供了丰富信息时,达维斯规则在所有对偶比较中均能作出正确推断,富兰克林规则亦如此。

相反,在此种环境中更加节俭的采纳最佳启发式难以做到十全十美,因为它所犯的错误不能够被随后的附加线索所弥补。例

如,在表 6-3 所示的环境中,采纳最佳启发式的成绩完全由前 3 条线索所决定,而其他的线索值不起任何作用。依据这 3 条最为有效的线索(不管线索 1 和线索 2 谁先得到检测,或线索 3 和线索 4 谁第 3 个得到检测),采纳最佳启发式获得的正确率为 90%。

关于采纳最佳启发式,表 6-3 中的资料产生了一种不寻常的现象。人们也许会以为采纳最佳启发式的成绩不可能比最为有效线索的效度更高,但这种想法是错误的。例如,在表 6-3 中,采纳最佳启发式的准确率($9/10 = 90\%$)就超过了最为有效线索的效度($5/6 = 83\%$)。事实上,共有 4 对对象是最为有效线索难以区分的,剩余的线索也只能正确地区分其中的一对(对于理解这种现象来说,条件效度的概念是很有用的,下面我们将对其加以介绍)。

总之,我们分析了三种信息结构——不可代偿信息结构、贫乏信息结构和丰富信息结构——并考察了每种条件下采纳最佳启发式相对于标准线性模型的成绩。但这只是有助于说明采纳最佳启发式生态理性的三个方面,也许还有更多的方面有待去探索。快速节俭启发式与环境结构的拟合能够解释为什么不需要总是在节俭性和准确性之间进行权衡——即何以简捷启发式能够兼具这两种似乎对立的特性。

采纳最佳启发式的耐用性

迄今我们已经考察了启发式规则与已知环境结构的拟合问题。下面我们将进入未知领域:当依据在一套训练项上获得的信息作出关于新对象的推断时,快速节俭启发式做得如何呢?节俭性也为超出训练项的推论奠定了基础吗?按照有关决策理论,毫无价值的多余信息对于决策制定者来说丝毫无损(如, Clemen, 1996)。在实践中,我们面对的问题通常是:如何从“噪音”中鉴别出真实信息。这不同于拟合已知资料时面对的问题,即:在哪种情况下可以实施最优化策略?在本章所分析的对偶比较任务中,

“配置记忆法”(the profile memorization method, 参见第 8 章)捕捉到了资料所包含的所有信息。*

然而,在依据已知信息推断未知资料时,并不存在可证实的最优化模型。试图从一套训练项“榨取”出尽量多信息的策略将会在测验项上一败涂地,其原因在于这种策略过度匹配于训练项。匹配过当现象很容易得到解释。例如,假如我们从一大套对象中抽取出了一个子集作为训练项。我们用抽出来的这些项目来训练有关模型,然后看一看它们在作出关于全部对象推断时的可推论性如何。如果存在这样一种可供选择的模型,它在训练项上的推断准确性不如正在使用的模型,但当将其推论到整套资料时,它取得了比正在使用的模型更加准确的成绩,那么我们就可以说当前正在使用的模型过度匹配于训练资料了。那么,为什么会产生匹配过当现象呢?我们认为原因在于训练项本身既包含有环境固有的内在结构,也包含有“噪音”。超出训练项的推论需要的正是训练项所包含的环境的内在结构,因此一种模型只有“捕捉”到这种内在结构,才容易推论到其他环境中去。如果一种模型过于“灵便”,以至于能够照应到训练项的所有细节(就像包含大量不确定参数的复杂推断模型那样),那么它将会超出内在结构而拟合于训练项的“噪音”。当这种模型被应用于训练项之外的情境(即推论到新环境)时,就有可能导致成绩的急剧下降。另一方面,简单模型——遵循奥卡姆的格言建立的模型——不大容易匹配过当,因为它们

* 按照这种方法,一个人首先要记住所有的线索配置以及它们如何获得效标分,然后依据这种信息推断一个对象是否比另一个得分高。如果所有对偶比较所依据的线索配置均不重复,一个人记住哪一种配置得分最高,并作出推断。如果一些线索配置不止出现一次,以至于在一些对偶比较中出现了两种线索配置一样的情况,解决的办法是数出一种对象得分高于另一个的次数。配置记忆法仅采用有最多机会得分较高的配置。拟合训练资料时,这种方法与皮尔逊(Persson, 1996)的基于范例的模型密切相关。在每种线索配置仅出现一次的环境中(如图 6-1),配置记忆法总能获得 100% 的正确推断。如果有重复的线索配置——因此也有重复的线索配对——这种方法将获得低于 100% 的正确推断,其他算法规则亦如此。

“十分吝啬”，使用的参数极少，因而避免了对“噪音”的拟合（例如，简单的达维斯规则就曾因为它的耐用性而名噪一时）。但是，节俭性也有一定限度，否则就会走向另一个极端，即“匹配不足”（underfit）。因此，一种模型可能由于两个原因而失败：拥有太多不确定参数，以至于匹配过当，或过于简单，以至于匹配不足。

在上述两个极端之间取得平衡是一门艺术。有两种截然不同的方法均可以做到这一点：一种方法是从所有可能的——简单的或复杂的——模型中选取一种“接近最优”（near-optimal）的策略来最大限度地降低匹配过当。最近十年人们设计了许多方法，试图控制匹配过当现象及与推论到新情境时成绩的突然变化相关的因素（例如，Geman et al., 1992）。其中最为著名的是贝叶斯范式（Bayesian paradigms），它已经成为在不同类别模型——决策树模型、回归模型，或神经网络模型——之间进行搜索的通用方法（Chipman et al., 1996; Friedman & Goldszmit, 1996; Mackay, 1995）。贝叶斯范式为估计环境结构和参数提供了一种通用一致的方法（参见第8章），但却背离了有限理性的观点。

第二种方法是本书极力主张采用的方法。它在环境结构上“下赌注”，强调采用快速节俭启发式规则，而不是通过复杂计算寻找最优化模型。例如，采纳最佳启发式“赌定”：环境信息结构以不可代偿的方式呈偏态分布，因此它的成败依赖于其假定的条件在多大程度上被环境所满足。

正像第1章所指出的那样，一种启发式可以用三种规则来表述：搜索规则、终止规则和决策规则。搜索规则指导人们如何搜索信息；终止规则决定人们何时停止搜索信息；决策规则指导人们如何利用搜索到的信息作出决策。其中搜索和终止规则决定一种启发式能否在匹配过当和匹配不足之间取得平衡。正像我们将要看到的那样，采纳最佳启发式的搜索过程既不过于复杂，又不过于简单，而它的终止规则也具有恰如其分的灵活性。

搜索规则的耐用性

我们已经将采纳最佳启发式与线性模型作了许多比较。值得注意的是,线性模型不需要指导搜索和终止搜索的启发式规则(尽管它们也需要作出决策的启发式规则,如达维斯规则就将其决策建立在所有线索最为简单的联结基础之上)。采纳最佳启发式是一种词典编纂策略,因此需要把搜索信息的启发式规则具体化。搜索信息的启发式将确定首先搜索哪一条信息,其次搜索哪一条信息,第三搜索哪一条信息,如此等等。词典编纂策略可以采用多种不同方法来决定这些线索的顺序。采纳最佳启发式采用了一种节俭的方法,但还有许多其他方法,如寻找最优化的顺序。在这一部分中,我们首先将采纳最佳启发式采用的节俭方法与其他已知方法联系起来加以考虑,然后对节俭策略的耐用性加以分析。

采纳最佳启发式按照线索的效度顺序来排列其等次。尽管人类也许能够直接估计线索顺序,但计算机必须首先算出线索效度,然后利用它们来对线索进行分类。因此,这里有必要首先对线索效度概念加以较为详细的分析。按照吉戈伦尔和古德斯腾(1996)的用法,我们将线索生态效度界定为基于它作出正确推断的比例:

$$v = \frac{R}{R+W} \quad (1a)$$

上式中 R 代表正确推断数量, W 代表错误推断数量。如果 $v=0.5$, 我们将一条线索称为中性的或无信息量的; 如果 $v>0.5$, 我们将一条线索称为有效的。值得注意的是, 只有当线索有效时, 采纳最佳启发式才能生效: 如果一条线索的效度小于 0.5, 该条线索的值就发生了急剧逆转, 即从 1 改变为 0, 反之则为 1。如果像本章所有表格所示的那样, 一种环境包含了效标值依次递减的对象, 那么公式(1a)能够迅速地算出每条线索的效度。一个人只需要数出获得正确答案 1 的次数超出 0 多少次和获得错误答案 0 的次数超出 1 多少次。为了算出需要进行比较的成对对象的数量

(即上式中的分母),一个人甚至可以采用比上述记数方法更加简单的方法:即用1的数目乘以0的数目。因此,另外一种十分不同但却实用的计算生态效度的方法是:

$$v = \frac{R_0 - \left(\frac{N_0(N_0 + 1)}{2} \right)}{N_0 N_1} \quad (1b)$$

在这里 R_0 代表所有0条目排列等次的总和, N_0 代表0条目的数量, N_1 代表1条目的数量。公式(1b)的优点在于线索效度能够在不知道所有配对数量的情况下计算出来。请注意上式中的分子相当于著名的“Mann-Whitney 测验”的 U 值,而分母相当于 $R+W$ (关于 1a 和 1b 的等价问题,参见 Martignon & Hoffrage 的有关推导,1999)。

生态效度 v 是用下面公式表达的著名的“Goodman-Kruskal 等级相关” γ 的一种变化形式:

$$\gamma = \frac{R - W}{R + W} \quad (2)$$

通过简单计算便可以看到 $\gamma = 2v - 1$ 。因此,当依据它们来排列线索顺序时,效度的这两种指标(v 和 γ)排出了同样的线索等次,二者均可以被用作排列线索等次的依据。^{*} 选择“Goodman-Kruskal 等级相关”作为效度指标有如下好处:中性或无效线索刚好满足于 $\gamma = 0$, 而有效线索满足于 $0 < \gamma \leq 1$ 。

确定一条线索效度的另一种方法是看一看当该条线索为可以使用的唯一一条线索时,采纳最佳启发式的成绩。每当该条线索

^{*} 线索效度也可以被用作线性模型的权重,就像富兰克林规则所做的那样。在表 6-1 所示的理想环境中,若用等级相关 γ , 富兰克林规则就有不可代偿的权重,即 1、0.5、0.25。相反,若用 v , 富兰克林规则就有可代偿的权重,即 1、0.75、0.625。这表明:就线性模型而言,线索是否可代偿不仅取决于环境,而且取决于测定线索效度的方法(Gigrenzer, 1981)。在表 6-3 中 γ 效度提供了一套不可代偿的线索。多元回归也有一套不可代偿的权重,但却依照不同的顺序。因此推导 1 不适用于该套资料。

不能够区分两个对象时,我们就作出一种猜测。我们将采纳最佳启发式之类的方法仅依据这样一条线索取得的成绩称为线索的“成效”(s=success),其计算公式如下:

$$s = \frac{R + 0.5(P - R - W)}{P} \quad (3)$$

式中 P 为对偶比较的总数量。该指标代表的是预期正确推断数量,已被一些启发式规则用作判断成绩好坏的标准。例如,第 10 章将要介绍的排除归类启发式就用它来界定效度。生态效度 v 与成效 s 也有关系:它是一条线索以辨别力为条件获得成功的概率。

如果没有肯德尔(Kendall)系数 τ ,上面列出的候选效度定义均是不完善的。该系数采用略微不同于求 γ 的方法来计算:

$$\tau = \frac{R - W}{\sqrt{P(R + W)}} \quad (4)$$

这里 P 仍然代表对偶比较的总数量, R 仍为正确推断的数量, W 为错误推断的数量(有关 γ 和 τ 异同点的详细讨论参见 Gigerenzer, 1981)。

一种经过精心设计的复杂启发式是借助于“条件效度”(conditional validity,简称 cv)来搜索信息的。条件效度是针对不能够被已使用线索区分开来的一套对偶比较的每一条线索而计算出来的。此类搜索首先使用最有效的线索,就像采纳最佳启发式所做的那样。其次使用对第一条线索不能区分的那些对偶比较而言最有效的线索,依次类推。如果就剩下的对偶比较来说,一条线索的效度被证明低于 0.5,那么这条线索的值就从 1 直接逆转为 0,反之亦然。下面是直接的推导:

推导 4: 当训练项与测验项一致时,利用条件效度进行推断的词典编纂策略的准确性大于或等于采纳最佳启发式的准确性。

词典编纂策略既可以利用节俭的启发式(如,使用生态效度)搜索信息,也可以利用复杂的启发式(如,使用条件效度)搜索信息,甚至可以尝试最优化搜索。表6-4是证明利用生态效度(v)搜索信息并非最佳方法的一个简单例子。我们不妨计算一下线索1的生态效度。正像前面已经提到的那样,一种简单的方法是数出获得正确答案1的次数超过0多少次,它实际上就是作出正确推断的次数。用这个数字除以1的数目和0的数目的积。这样计算,线索1比线索2有更高效率(分别为 $5/7=15/21$ 和 $10/15=14/21$)。实际上除了肯德尔系数 τ 外,其他效率系数均将线索1排在线索2之前。然而,首先检测线索1只获得了13个正确推断,而首先检测线索2却获得了14个正确推断。

表6-4 一种令人困惑的环境

对 象	线 索 1	线 索 2
a	1	0
b	1	1
c	1	1
d	1	1
e	1	1
f	0	1
g	1	0
h	1	0

注:尽管线索1比线索2有更高效率,但如果先检测线索2的话,一种词典编纂策略将获得更高成绩。

因此,采纳最佳启发式并不一定能够确保获得最优化的线索顺序。那么,怎样才能获得最优化线索顺序呢?施密特和马蒂格农(Schmitt & Martignon, 1999)证明,为了做到这一点,除了算出线索的所有可能顺序,并比较一种词典编纂策略在每种线索顺序上的成绩外,不存在更为简单的方法。这种方法当然很好,只是不够快速和节俭,因为在一种拥有 M 条线索的环境中,不得不对 $M!$ 种排列进行比较。寻找最优化顺序的问题在于它需要一种“非多项式可完

成的”(NP—complete)算法规则($M!$ 超过了 $2^{(M-1)}$; 关于“NP—completeness”的解释, 参见第 8 章)!

不同信息搜索规则如何影响词典编纂策略的成绩呢? 我们用比较德国城市人口的任务(参见第 4 章)来回答这个问题。图 6-3 表明了所有可能排列的推断准确率分布。最优化线索排列获得 75.8% 的成绩(该分布的右端)。此分布的平均值(70.0%)相当于最少化启发式的预期成绩, 该种启发式采用随机选择线索的最为简单的搜索规则(参见第 4 章)。所有其他搜索规则均获得了介于上述二者之间的成绩: 条件效度(cv)的成绩是 75.6%; 生态效度(v)的成绩是 74.2%, 这实际上就是采纳最佳启发式的成绩; 肯德尔系数(τ)的成绩是 73.5%; 线索成效(s)的成绩是 72.7%。只有 1.8% 的线索排列获得了高于采纳最佳启发式线索排列的成绩。考虑到采纳最佳启发式的搜索规则并未付出复杂计算的代价, 它取得的成绩还是令人满意的。就德国城市人口问题而言, 为了确定最优化的线索顺序, 即使最快速的计算机, 也需要大约 4 天时间进行计算!

133

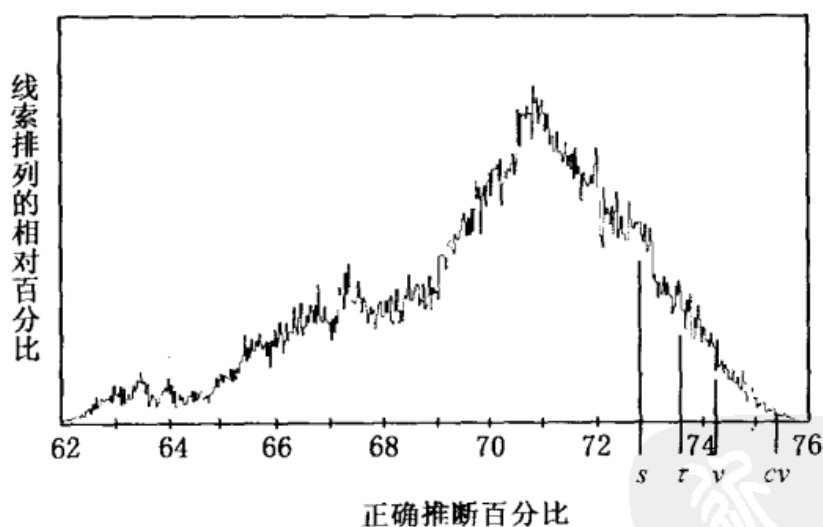


图 6-3 在德国城市人口环境下所有可能线索排列的成绩分布。该分布的平均值(70%)相当于最少化规则(随机排序)的预期成绩。按照生态效度(即采纳最佳)、成效、肯德尔系数和条件效度搜索信息的词典编纂策略的成绩分别用 v 、 s 、 τ 和 cv 表示。

相对于其他策略来说,采纳最佳启发式搜索规则的耐用性如何呢?为了回答这个问题,我们需要将拟合已知资料时的准确性与推论到未知资料时的准确性加以比较。通常,推论是在一个同质总体中进行的——从一个子集推论到总体或另外一个子集。*第5章所做的交叉验证属于后一种情况。这里我们采用同样方法考察搜索规则的耐用性:随机地选出一个包含一半德国城市的子集,按照图6-3所示顺序决定线索排列等次。首先考查它们在子集内(训练项)的成绩,然后考察它们在另外一半对象(测验项)上的成绩。用这种方法我们随机选择不同子集100次。结果示于图6-4。正像所预期的那样,对所有线索顺序来说,推论到测验项上的成绩均低于在训练项上的成绩。令人惊奇的是,采纳最佳启发式使用的搜索规则最为耐用。复杂搜索规则(最优化顺序和 cv)推

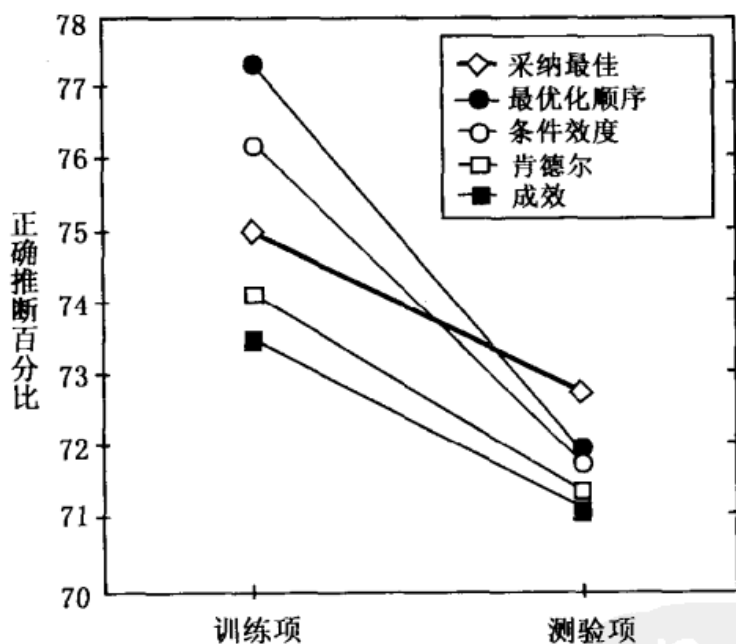


图6-4 在德国城市人口环境中拥有不同线索搜索顺序的词典编纂策略的耐用性。训练项由随机地从该环境中抽取出的半城市所组成,测验项是剩下的一半城市。成绩是100次试验的平均值。

* 在另一种情况下,没有包含训练项和测验项的总体,或者有这样的一个总体,但它是含糊不明的和异质的,如全世界教授的工资。达维斯(1979)举了一个医疗诊断资料的例子,在那里即使预测对象随时间而变化,达维斯规则仍然相当通用。

简捷

简捷启发式

论到新资料时成绩降低最多,说明它们拟合了训练项中的“噪音”。

终止规则的耐用性

采纳最佳启发式的终止规则允许它在必要时搜索所有提取出来的线索。当然,还有另外一些可能性。例如,一个人在检测了第一或第二条最有效线索之后,可能就会考虑终止搜索,而采用掷硬币式的方法对尚未被区分的对象作出猜测。另外一个极端是不采用任何终止规则,那就是线性模型所使用的方法,它们检测并使用所有适用信息。采纳最佳启发式介于二者之间,其终止规则的一个重要特征是它的灵活性:对于一对特定的对象而言,采纳最佳启发式也许在找到第一条线索后就终止了搜索(因为这条线索在两个对象间作出了区分),而对于另一对对象而言,它也许要搜索所有信息(因为没有发现能够区分两个对象的线索)。为了考查这种灵活终止规则的耐用性,我们用第5章介绍的20种资料对不同策略进行了交叉验证。图6-5显示了在跨越所有20种环境条件

135

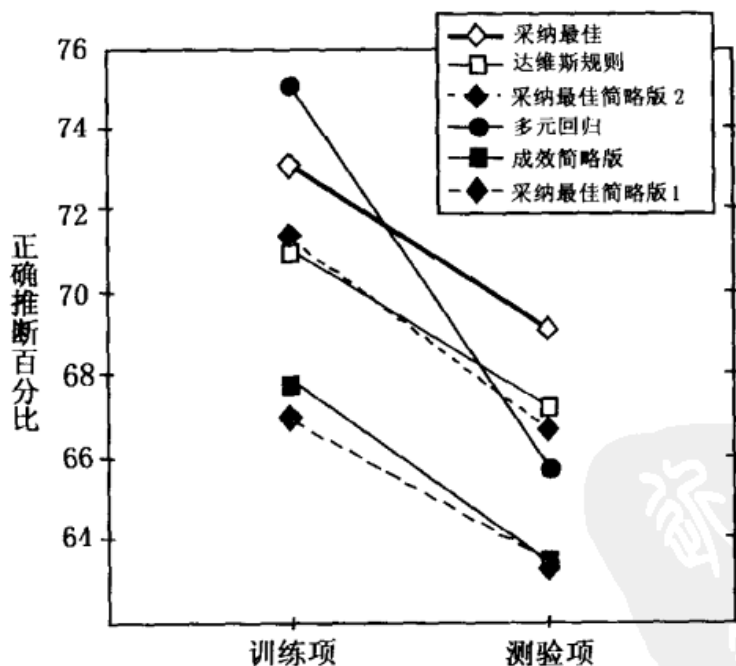


图6-5 拥有不同终止规则的多种策略在跨越20种环境(每种环境都作了100次试验)条件下的耐用性。多元回归匹配过当了,而采纳最佳启发式是相当耐用的。

下,不同策略拟合已知资料和推论到新资料时的平均成绩。

再一次,当从训练项推论到测验项时,所有策略的成绩均明显地降低了。拟合已知资料时多元回归方法做得最好,而就耐用性来看采纳最佳启发式做得最好,因为它在测验项目上获得了最高成绩。多元回归方法显然匹配过当了,达维斯规则也比它更加耐用。采纳最佳启发式的两种简略形式(一种限制了对第一条线索的搜索,称为“采纳最佳简略版 1”;另一种限制了对前两条线索的搜索,称为“采纳最佳简略版 2”)有些匹配不足。当限制了对一条线索后面的线索进行搜索时,生态效度就不再是最好的选择标准,因为此种情况下它不能够充分地解释线索的区分率。但是当单独使用时,甚至最成功的线索(成功率版本 1)也有些匹配不足。

采纳最佳启发式的终止规则从生态学上看是合理的:当终止并非恒常和固定,而是依赖于适用信息时,它是合理的。在我们所考察的 20 种环境中,采纳最佳启发式平均使用了 2.4 条线索(参见第 5 章)。它在耐用性上的生态合理性在于:在推论到未知资料时,这些为数不多的重要线索仍然是有效的。这也能够解释何以简化的达维斯规则(如同第 5 章所介绍的那样)会比正式的(未简化的)达维斯规则更加耐用。

关于采纳最佳启发式的三种观点

我们认为采纳最佳启发式是一种心理过程,可以用数学方法将其表征为一种算法规则。这种算法规则能够用一种以上的数学结构来体现。由于采取不同角度看待这种算法规则有助于丰富和深化对它的理解,下面让我们探讨其中三种不同的观点。

线性模型观点

线性模型是所有推断策略中最广为人们采用的工具。正像第 4 章提到过的那样,作出决策时将线索累积起来或进行加权的做

简捷启发式

法反映了启蒙运动以来流行已久的理性选择方法的道德理想。一般而言,对外部世界进行线性分析已经成为统计推断领域最基本的工具。从历史发展角度看,统计推断中线性回归方法的使用达到了该种模型发展的顶峰,它解决了一个非常重要的最优化问题:将预期效标值和观察到的效标值之间的离差平方和降到最低。线性回归方法的迅速扩散和在几乎所有科学研究领域的广为接受形成了统计学发展史上一道独特的风景线(Stigler,1986)!

从准确性的观点来看,线性模型远远不如其他一些非线性方法,如第8章将要讨论的贝叶斯网络模型等。但是,如果不得不同时考虑几种证据,那么贝叶斯网络之类的更加复杂的模型很快将会无法处理有关数据。事实上,尽管线性方法在结构上是简单的,但仍然能够取得相当好的推断成绩。对外部世界进行线性分析不失为一种可靠、明晰和简单的决策制定方法。

线性计算的一种局限性在于:如果只涉及很少条件的话,它们只能够在心理上被完成。富兰克林只需要纸和笔来平衡正面和负面理由。达维斯规则数出正面线索数量,然后减去负面线索数量。如果线索数量不是太多,它也可以在心理上被完成。对于现实条件下的决策来说,很少有机会依次加或减每一条信息,更不用说用纸和笔进行计算了。如果所有适用信息同时出现,可以平行地得到加工——即同时对它们进行估价——那么时间限制问题也就不复存在。然而,在我们所关心的任务中,线索必须被搜索,并相继地、逐步地或逐个地得到加工。线性模型既可以对信息进行系列加工,也可以进行平行加工:可以借助于系列机制逐个对每条线索进行加权,也可以同时对所有线索进行加权。就拥有不可代偿权重的线性模型来说,遇到能够区分两个对象的第一条线索即终止搜索的系列加工将导致完全等同于对所有线索进行平行加工的成绩。正是这些线性模型等价于采纳最佳启发式(推导1)。

137

用于比较的词典编纂规则

自然环境在层次结构、排列顺序和等次上是复杂多样的。然

而,正像“布尔巴基研究组”(Bourbaki,1963)一再指出的那样,在数学发展的历史上,数论被忽略了。将数论置于像代数学和拓扑学一样重要的地位来考查其广泛深远意义的功绩应归于19世纪的数学家们(主要是 Dedekind,1887/1987)。布尔巴基构建的数学王国所包含的三个领域实际上是三个不同学科——代数学、拓扑学和数论,其他的一切均可以从它们推导出来(Bourbaki,1963)。从数论观点来看,当(且仅当)将每一种线索配置分派给每一个对象的方案能够保证依据效标排列的一套对象和用双极线索表示的自然成员之间具有序列同质性时,采纳最佳启发式才会获得100%的正确推断。这意味着,如果按照效标对象 a 大于对象 b ,那么 a 的线索配置所代表的自然成员也应该比 b 的线索配置所代表的自然成员更大。

分类树

对于解决本章所讨论的对偶比较问题来说,另外一种候选方法是分类树。毫不夸张地说,分类树已经被人们使用了几千年。早在公元前3世纪,古希腊哲学家波菲里斯·马修斯(Porphyrus Malchus)就采用所谓的“种差法”(per genus et differentiam,即搜集同一物种的多种信息,再按照细微差别来区分它们)来对物种进行分类,这实际上就相当于简单的分类树。

解决对偶比较问题的分类树是由三个部分构成的。结点是线索,而结点之间的弧线长短代表弧线末端的线索值。从根到叶的每一条路径都由线索值的一种可能联结构成,而叶结点表明了拟作出的决策(如, $a > b$ 或 $b > a$)。为了对一个对象作出分类,从根结点开始,考查其线索,沿着对应于线索值的弧线进行搜索,首先检查每一条弧线开端的线索值,依次进行下去,一直到分类树的叶结点,作出一种决策。如图6-6所示,采纳最佳启发式相当于一种简单分类树。根结点是拥有最高效度的线索。如果一对对象在这条线索上不同,那么就可以作出决策,否则搜索第二条线索,直

138

到找到能将两个对象区分开来的那条线索为止。

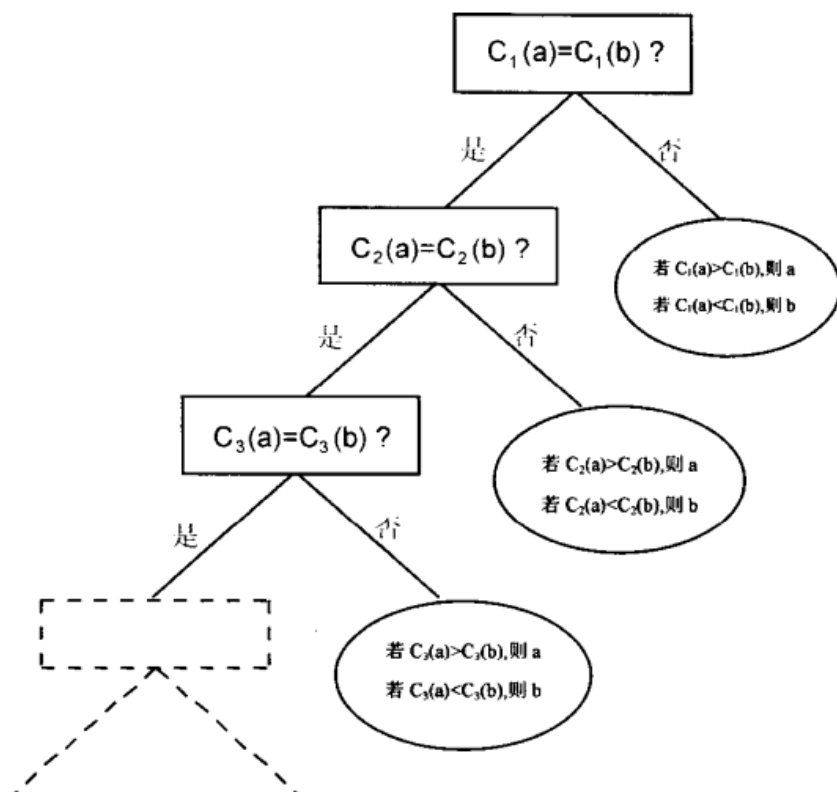


图 6-6 作为一种分类树的采纳最佳规则,依据线索 C_i 比较对象 a 和 b 。

为一套给定资料建立一个充分有效的分类树是极其困难的。在过去 20 年中,在统计学和机器学习领域依据给定资料建立分类树已经成了一个重要研究课题。人们的兴趣主要集中在如何发展对于建立能够很好地运行并推论到新资料上的分类树充分有效的启发式规则,并评价其效用。分类和回归树(classification and regression tree,常缩写为 CARTs;Breiman et al.,1993)作为一种能够保证准确性的分类树是非常成功的。贝叶斯范式也为建立好的分类树提供了有效搜索方法(参见第 10 章)。

建立分类树的快速节俭策略只是最近才在机器学习领域受到人们重视。与采纳最佳启发式相类似,由郝特(Holte,1993)提出的“1-R”分类树规则也以单一理由决策制定为基础。当面临在两个对象间进行比较的特殊任务时,“1-R”分类规则选择具有最高

效度的线索,并像采纳最佳启发式一样,依据这条单一线索作出决策。也就是说,如果该条线索在对象之间作出了区分,它将选择有较高线索值的对象。然而,与采纳最佳启发式不同的是,如果这条线索不能够区分对象,“1-R”规则将不再尝试第二条效度最高的线索,而是随机选择后面的线索。因此,在第一条线索之后“1-R”方法类似于我们前面所说的采纳最佳启发式简略版。就双极线索来说,“1-R”方法与采纳最佳启发式的标准版本相比是有缺陷的(无论是拟合已知资料,还是推论到新资料,参见图6-5),因为它必然要涉及过多的随机决策。

“1-R”规则实际上是针对基于真实线索值(而不只是1或0)的分类和多种不同范畴而设计出来的。郝特曾经用大量不同资料(特别是来源于“UC Irvine”信息库的几种不同资料——参见 Merz & Murphy, 1996)对“1-R”方法进行了测试。他获得了非常优异的成绩(Holte, 1993),仅比昆兰(Quinlan, 1986)提出的“C-4”规则的成绩略低一些(关于更多的细节,请参见第11章)。对于未被最佳线索(效度最高线索)区别开来的那些对偶比较而言,如果剩下的线索能够取得高于机遇值的成绩,那么最佳线索的成绩将会被超越,而事实上情况也的确如此。

简短评论

尽管旨在通用的信息整合方法不厌其烦地搜集点滴信息以便作出推断,但精明的启发式却巧妙地利用了对于具体任务来说非常关键的信息的结构特点。进化已经赋予我们“捕捉”环境信息有关特性的能力,使我们能够按照这些特性灵活变通地使用不同工具。

基于双极线索的词典编纂模型与拥有可代偿权重的线性模型获得了同样好的成绩。那么,作出决策时一个人究竟使用哪一种模型呢?如何区分呢?我们的推测是:词典编纂模型比线性模型

简捷启发式

更加自然一些。线性化模型常常是人为设计的策略,其目的在于以我们头脑能够理解的方式将外部世界模式化。然而,秩序不仅仅存在于我们的头脑之中,也存在于我们的头脑之外。外部世界充斥着各种各样的秩序结构。由于考虑到了环境中明显存在的信息结构,词典编纂模型提供了一种确保推断准确性的强有力的方法。它们充分利用了偏态分布。有时候它们做得甚至有些过分:将外部世界的信息结构理解得比其本来面目更加偏离常态。但是环境的确经常地偏离常态,以至于其信息结构常常是不可代偿的。在此种情况下,采纳最佳启发式与更加复杂的线性模型在拟合一套已知资料时取得了同样的推断成绩。

然而,也有一些极端环境,在那里两类不同模型取得了不同成绩:140
贫乏环境有利于采纳最佳启发式,而丰富环境有利于达维斯规则。这些理论上的推导与培恩等人(Payne, Bettman & Johnson, 1988, 1993)的实验结果是一致的。他们发现,人们按照环境提供给他们不同条件选择不同策略。例如,人们所选择的算法规则依赖于他们有多少时间作出一种选择和推断,以及作出这种决策的代价是什么。如果时间十分紧迫,他们将借助于词典编纂模型作出推断(参见第7章)。如果有一条便于使用的好线索,他们也会这样做。相反,如果作出一种错误选择的后果非常严重,或者有足够时间进行考虑,那么人们将会尽可能多地搜集信息,并将它们整合起来进行推断。“有时使用采纳最佳启发式,有时使用达维斯规则”,这应该像所罗门诗文中的格言一样,告诉我们如何在不同规则之间加以选择。经过数千年的演变,进化早已将判定多种策略基本差异的良好直觉能力赋予我们,因此也教给了我们何时使用哪种策略的方法。下一章我们将详细讨论在不同任务和环境条件下如何确认被试所采用的策略。

第7章

人们何时采用简捷启发式以及如何识别

尤格·雷斯卡姆普(Jorg Rieskamp)

奥瑞齐·霍夫瑞格(Ulrich Hoffrage)

我们很可能觉得能够同时顾及多种不同因素,因为尽管某些时候我们仅能专心于某一种因素,但我们却难以明确地意识到这一点。

——罗杰·N·谢泼德(Roger N. Shepard)

K先生刚刚继承了一大笔钱,正在考虑如何利用它们进行投资,以获取最大利润。他的两个朋友一直投资于股票市场,非常了解股市行情。其中的一个是“快”女士,她是一位股票经纪人,常常不得不在几秒钟之内决定购买哪种股票。另一个是“慢”先生,对他来说买股票只是一种业余爱好,尽管如此,他仍然非常认真地对待他的投资活动,常常花费整个周末来考虑一项投资决策。在本章中我们将把自己置于K先生的位置——他试图弄清楚他的朋友们是如何作出投资决策的。像K先生一样,我们想要弄清楚当人们不得不作出选择时,他们使用了什么样的策略。人们使用简捷启发式吗?我们如何才能识别呢?这就是我们强调的两个问题。

首先,我们问:人们是否使用简捷启发式?本书前面的章节已经证明快速节俭启发式可能是非常准确的。这些简捷启发式不仅是准确的,而且仅需要很少认知资源,这使得人们在实际中使用它们似乎也是合理的。我们假定人们有一种由不同策略组成的适应性机能系统——适应工具箱,在不同条件下他们从该工具箱中

让我们更精明

选用不同策略。哪些条件对策略选择有影响呢？一种可能性是时间紧迫性。“快”女士需要在时间紧迫条件下作出决策，所以她也许可采用不同于“慢”先生所用的那种策略。本书提出的快速节俭启发式高度地适应于时间和知识有限的环境。它们是快速的、易于使用的和需要很少信息的，但同时也是能够很好地完成任务的。第4章主要涉及了有限知识问题，本章将主要探讨有限时间问题。我们的问题是：时间紧迫性促进了简捷启发式的使用吗？

142

其次，我们将考虑另一个重要问题：如何才能识别一个人正在使用何种策略？为了调查和判别人们使用的策略，我们发展了两种主要研究方法：一种是“过程取向”(the process-oriented)的方法，它主要关心作出决策前的信息搜索过程，例如，可以通过考察获得信息的顺序做到这一点；另一种是“结果取向”(the outcome oriented)的实验方法，主要关心决策过程的结果，并建立预测这些结果的模型。

本章的结构如下：首先我们对多种决策策略作了描述，然后提供了一项实验，被试不得不在时间充裕或紧迫两种条件下从多种选项中作出选择。接下来我们描述了过程取向和结果取向的方法，并通过实验来说明识别有关策略时的方法学问题。最后，我们回顾了一些研究，这些研究探讨了包括时间紧迫性在内的可能对决策策略产生影响的条件。

作出决策的简捷启发式

启发式描述了决策过程的一系列连续阶段。(但请记住：个别人的实际决策过程或许不同于启发式，如果他采用了启发式未包含的附加操作的话)。在文本框7-1中，我们对文献中曾经提到过的多种决策策略作了简短总结。它们均适合于在几个选项中作出一种选择的任务。每个选项均用一条线索(属性)来刻画，而每条线索的重要性用它的权重(或效度)来体现。文本框7-1中

列出的决策策略远非完整,部分原因在于我们研究的是推断而非偏好判断(它常常涉及到了在几个赌项之间进行抉择)。主要用于偏好判断的策略未被包括在文本框 7-1 中。而且本章的目的并非对所有候选策略进行全面分析。我们相信我们的主要结论不依赖于文本框 7-1 中列出的策略是什么,而适用于广泛的策略。

文本框 7-1 决策策略

富兰克林规则 算出每一个选项每条线索的值与其对应权重(效度)的积,求其得分之和,选择有最高得分的选项。

达维斯规则 算出每一个选项所有线索值(乘以 1 以内的单元权重)的和,选择有最高得分的选项。

好特征规则(good features; Alba & Marmorstein, 1987) 选择好特征数目最多的选项。好特征是指超过某种具体切割分数的线索值。

加权正面线索规则(weighted pros; Huber, 1979) 选择有最多加权正面线索的选项。一条线索在一个选项上比在其他选项上有更高的值,那么对该选项而言它就是正面线索。每条正面线索的权重是由特定线索的效度来界定的。

词典编纂规则(LEX 或 lexicographic; Fishburn, 1974) 选择有最高线索值(最高效度)的选项。如果多于一个选项有同样的最高线索值,那么比较这些选项的第二个最高线索值,依次类推。词典编纂规则是采纳最佳启发式的延伸。

半词典编纂规则(LEX-Semi 或 lexicographic semiorder; Luce, 1956) 像词典编纂规则一样,只是附加了不显著差异假设:若两个选项线索值的差异不显著的话,这两个对象不被区分。

逐项排除规则(EBA 或 elimination by aspects; Tversky, 1972) 在检测第一条线索时,排除未达到特定线索值的所有选项。如果多于一个选项被保留下来,选取另一条线索。该过程持续下去,直到仅留下一个选项。每一条线索都依据与其权重相对应的概率加以选择。与此不同的是,在本章中逐项排除规则是依据线索效度来确定线索顺序的,因此在每种情况下,有最高效度的线索都将首先得到检测。

词典编纂附加规则(LEX-ADD 或 lexicographic additive) 它是两种策略的结合。首先使用半词典编纂规则选出两个中意的选项,然后用达维斯规则评价它们,选择有最高加权分总和的选项。

一个一般的问题是:一种策略怎样处理不同线索(属性)值的不可比性?举例来说,如果一条线索的取值范围是 0—100,而另一条线索的取值范围是 3 000—100 000,那么一种策略将如何处理它们呢?在第 5 章中,连续线索值依据中位数被双极化了,从而产生了双极线索值。在下面所报告的实验中,连续线索值依据一

让我们更精明

简捷启发式

种分级量表被分成不同等级,因此只有 1、2、3、4、5 五种线索值。从最低的原始线索值到第 20 个百分点的原始线索值均被转变为线索值 1,从第 21 个百分点到第 40 个百分点之间的原始线索值均被转变为线索值 2,依次类推。文本框 7-1 所列出的所有策略都能够利用这种等级线索值作为输入信息。为了简便起见,迄今我们尚未考察知识有限(线索值未知)时,这些策略是如何运作的。 143

本书第 1 章在搜索规则、终止规则和决策规则之间作了区分。搜索规则指导如何搜索信息,终止规则确定何时停止搜索,而决策规则具体决定如何使用搜索到的信息作出一种决策。这些规则可以用来描述任何一种策略。对于每种规则,我们都引入了一些区分。首先,一种策略可能以“基于选项”(alternative-wise)的方式搜索信息,也可能以“基于线索”(cue-wise)的方式搜索信息。基于选项的搜索在考虑下一个选项之前寻找一个选项的所有线索值,而基于线索的搜索在检测另一条线索之前寻找一条线索的所有线索值。例如,K 先生可以逐个地使用有关线索来比较不同股票(基于线索的搜索),也可以依据所有线索逐个地比较每一种股票(基于选项的搜索)。其次,一种策略也许有何时停止信息搜索、作出一种选择的终止规则,也许没有这种规则。尽管许多实验本身限制了大量线索(所以无需终止规则),但是在这些实验之外的现实情境中,终止规则是十分重要的。第三,一种策略可能是可代偿的,也可能是不可代偿的。如果至少有一条线索可能被其他线索所超过,那么一种策略就是可代偿策略;相反,如果有一条线索不能够被较不重要的其他线索的任何一种联结所代替,那么一种策略就是不可代偿策略(参见第 6 章)。可代偿策略对(至少一些)适用信息加以整合,在有关线索间进行权衡,以便对每一个选项形成总体评价。 144

搜索、终止和决策规则是相互联系的。例如,当一种启发式仅搜索唯一一种(有鉴别力的)线索时,就把可能采用的决策规则限制在无需对信息进行整合的规则范围之内。因此,如果一个个体

进行了很少搜索,那么过程取向的方法对启发式决策规则具有的特征将有决定性意义。另一方面,如果搜索扩展到了许多线索,将不会对所采用的决策规则产生任何限制。要么对搜索到的线索进行加权或整合,要么仅依据其中一条最好的线索作出决策。在后一种情况下,过程取向的方法暴露出了它的局限性。它仅关心信息搜索过程,但即便一个人搜索到了大量信息,也并不一定意味着他就会将决策建立在所有这些信息基础之上。尽管如此,过程取向的方法坚持认为,信息搜索与决策策略密切相关,信息搜索过程的特点对一个人实际采用的决策策略有重要影响。

下面我们根据上述三种规则对文本框 7-1 列出的决策策略进行归类。富兰克林规则、达维斯规则、好特征规则、加权正面线索规则以及词典编纂附加规则均为可代偿策略,它们均没有终止规则。对它们(除了加权正面线索规则和词典编纂附加规则)而言,基于选项的搜索是最为合适的。词典编纂规则、半词典编纂规则以及逐项排除规则均为不可代偿策略,它们均有终止规则,并采用基于线索的信息搜索方式。

可代偿策略并非一定要进行代偿。是否需要进行代偿取决于线索数量、线索值和线索效度的变异性,以及信息是如何被具体加工的。富兰克林规则加工所有适用信息,而达维斯规则用(正的或负的)单元加权代替了具体线索效度。(在本章中我们仅考虑正的单元加权,因为在我们的实验中较高线索值总是代表高效标值。)好特征规则类似于达维斯规则,但不需要线索值的累加。这三种策略的优点是:进行决策时不需要每遇一个选项都对所有选项进行一次总体评价。并非所有可代偿策略都是如此。例如,加权正面线索规则为每一个选项算出的分数都依赖于其他选项的线索值,这是因为“正面线索”本身就是通过比较所有选项的线索值来界定的。这个特点也能够说明加权正面线索规则的另外一个独特之处:它是唯一一种基于线索搜索的可代偿策略。词典编纂附加规则兼有可代偿和不可代偿以及基于选项搜索和基于线索搜索

145

简捷启发式

简捷启发式

两种特点。因此,加权正面线索规则和词典编纂附加规则在可代偿与不可代偿策略以及基于选项搜索与基于线索搜索之间架起了桥梁。

词典编纂规则、半词典编纂规则和逐项排除规则均为不可代偿策略,这意味着支持某个选项的一条线索不可能被较不重要的其他线索的任何联结所超过,即便它们全都支持不同的选项。这是因为这三种启发式不一定使用所有适用信息。词典编纂策略类似于采纳最佳启发式(除了不需要再认启发式作为起点之外,参见第4章)。事实上,它是采纳最佳启发式的推广,即从双项选择问题推广到多项选择问题。如果两个或更多选项打成了平手,也就是说,它们在具有最高效度的那条线索上有同样的值,而且没有其他选项有更高的值,那么有两种方法可以将采纳最佳启发式加以推广。第一种方法是继续检测对于具有最高线索值的那些选项来说效度居于第二位的线索,这也是本章所选用的方法。它强调选项的词典编纂顺序,从而导致了一种选择性的信息搜索过程,因为在决策过程的前几步就借助于启发式规则排除了一些选项。推广采纳最佳启发式的第二种方法需要对所有选项的第二条(也可能是第三、第四条,如此等等)效度最高线索进行重新考虑。这种方法强调单一理由决策制定概念,即在决策过程的每一步所有选项都要在一条特定线索上进行比较,而且仅依据这条线索就可以决定选择哪个选项。^{*} 由于词典编纂规则的高度选择性和基于线索

146

^{*} 举例来说,一个人可以使用上述两种方法中的每一种来决定 a、b、c 三个公司中哪一个有最高的收益。假如有两条不同的线索(资产总量和雇员数量)均与公司的收益呈正相关。公司 a 和 b 有较多的资产,公司 c 有较少的资产。然而,公司 a 有最少的雇员,公司 b 次之,公司 c 最多。如果最重要的线索是资产,那么一个人将难以在公司 a 和 b 之间作出区分,于是就会求助于第二条线索——雇员数量。在此种情况下,采用词典编纂顺序将会导致一个人选择公司 b,而采用另一种方法将会导致一个人选择公司 c。

索值或线索(亦可参见 Payne et al., 1988)。逐项排除规则的信息搜索过程类似于词典编纂和半词典编纂规则,但有一个重要差异:使用逐项排除规则时,项目的排除依赖于一个具体切割分数,而使用词典编纂和半词典编纂规则时,不需要任何具体切割分数,项目的排除完全是由它们的线索值决定的。

作出选择时需要做的全部事情就是确认在效标上得分最高的那个选项。既没有必要知道确切效标值,也没有必要知道次级选项的排列等次。在这方面,基于线索的搜索具有明显优势,因为在搜索过程的每个阶段它都能够使个体辨认次级选项,然后集中注意优等选项。相反,在基于选项搜索中,依次评价每个选项是必要的。只有当所有选项均得到考虑的情况下,才能明确判断哪个是最优的。因此,如果 K 先生想要在两种股票间作出选择,一种较为简单的方法是:依次使用不同线索来比较两种股票,并集中注意最有希望赢利的那一个。

如何依据执行策略时需要的信息加工量对策略进行比较呢?由于可代偿策略通常没有终止规则,需要搜寻所有线索,所以相对于不可代偿策略而言它们是不够节俭的。例如,富兰克林规则使用所有线索值和它们的权重,所以既不够节俭,计算上也不够简单(亦可参见 Payne et al., 1988, 1993)。可代偿策略通过计算得分单独评价每一个选项,然后不得不将每个选项的得分储存在记忆中,以便随后与其他选项进行比较。有限的人类工作记忆容量(Miller, 1956)对需要短期储存信息的策略提出了一个问题,因为这种策略对记忆容量的要求往往过高。对于一项同时提供许多信息的选择任务来说,基于线索的信息搜索有可能比基于选项的信息搜索需要较小的记忆容量和较少的运算步骤。它使得依次逐项使用每条线索对有关选项进行比较成为可能。这种分步比较程序可以被用来建立选项的先后顺序。个体只需要记住那些被认为重要的选项,而不必记忆任何线索值。只有这些重要选项才需要依据新信息加以进一步考查。相反,基于选项搜索的优势在于能够

对一个具体选项形成精确认识,但它却使得发现选项间的精细差异变得更加困难。

使用不同策略需要付出的“努力”程度不同,这是否有助于我们预测其中哪种策略将提供关于人们实际决策过程的更好模型呢?培恩等人(Payne, Bettman & Johnson, 1988, 1993)相信如此,并提出了所谓的“努力—准确性”权衡观点,其依据的基本假设如下:人们有一种能够依据代价和收益加以评定的不同决策策略的明细表。代价用加工有关策略需要付出的努力程度来衡量,而收益则用准确性来衡量。培恩等人(1993)指出:“决策制定者对适用的不同策略的多种代价和收益加以评定,然后选择对解决目前问题来说最好的策略。”(p. 91)按照这种观点,决策制定者能够预先估计一种策略需要付出的努力程度及其准确性,并以其为标准选择最好策略。培恩等人承认在策略之间进行选择并不一定是有意识的,在努力程度和准确性之间的权衡也许可以借助于经验而习得。在一些情况下,人们也许根本不依据其需要付出的努力和准确性对策略进行权衡,因为它们显然已经超出了个体能够承受的努力阈值。在培恩等人的模型中,对需要付出的努力程度加以评定主要有两种不同作用:要么作为除准确性之外评价一种策略的附加标准,要么作为从拟议中的策略中排除一种策略的限制条件。

147

与培恩等人(1993)相一致,我们假定人们作出决策时的条件限制(如,有限时间和知识)是判断和决策活动的重要决定因素。培恩等人的研究提出了这样一种可能性,即:这些条件限制是特定决策过程中必须得到评价的附加标准。与这种观点不同的是,我们将有限的时间和知识作为对决策过程的条件限制,而且人们已经发展和习得了一些在此种受限制条件下可以使用的精明启发式。这意味着个体的策略明细表中已经包括了一些充分考虑到这些条件限制的策略。我们并未假定努力—准确性权衡和对策略的评定需要在决策过程中通过精细计算来完成。鉴于个体先前的决

策制定经验,特定环境足以暗示或提醒他或她使用哪种决策策略。

培恩等人的模型和我们的观点导致了类似的假设:在特定条件下,尤其是在时间紧迫条件下,个体使用简单策略。我们假定人们已经拥有应付时间紧迫性的成套简单策略,使用这些策略几乎不需要付出任何努力。因此,使用较少信息、无需对信息加以整合,并采取基于线索搜索的简捷启发式将会有利于时间紧迫条件下人类的决策活动。

一种似乎合理的假设是:相对于时间充裕条件而言,时间紧迫条件下人们将获得更少信息。如果他们转而求助于不可代偿策略来作出决策,那么信息搜索顺序将由线索的重要性来决定,这样做时他们将会把注意力集中在更加重要的线索上。这导致我们作出如下两个具体假设:(1)相对于时间充裕条件来说,时间紧迫条件下使用简单的不可代偿策略作出决策的被试比例将会更大; 148
(2)相对于时间充裕条件而言,时间紧迫条件下被试将会表现出更有选择性的信息搜索,并获得更多关于重要线索的信息。

为了考证上述假设,我们进行了一项实验。在该项实验中,我们调查了在时间充裕和紧迫两种条件下人们使用的决策策略(参见文本框7-2)。在下面两部分中,我们将以该实验为例来说明用于考查人们所使用的决策策略的两种主要方法。

文本框 7-2 实验任务

被试必须从4个不同国有公司中选出年利润最高的公司(参照年份为1994年)。每一组(4个)公司均是从70个真实公司样本中随机挑选出来的,由它们构成一种推断环境。6条有关公司的变量作为线索。每条线索的重要性由它的线索效度来确定。这些线索(和它们的效度)是:投资总量(0.78)、公司股份总量(0.68)、雇员数量(0.64)、再认率(0.50)、1994年末在股市上的股票价格(0.48)、股票红利(0.37)。请注意:在这里0.25的效度代表机遇成绩。为了确定每条线索的效度,我们考察了70个公司能够配成的所有4个一组的推断环境。线索效度就是拥有最高线索值的公司同时也具有最高年利润的组在全部配成的组中所占的比例。(再认率是有关测验中能够再认某个公司的被试所占比例。)信息在一个计算机控制的信息板上呈现。

实验开始于一个由 7 种选择构成的训练阶段,每次选择后被试都得到一个有关正确答案的反馈信息。(主要目的是让被试熟悉任务,而不是让他们了解线索效度。实验之前向被试解释线索效度的概念,实验时将线索效度呈现在靠近监视器的一张卡片上。)训练阶段之后,被试首先在时间充裕(每种选择 50 秒钟)条件下作出 15 种选择,然后在时间紧迫(每种选择 20 秒钟)条件下作出 15 种选择。信息板左上角的钟表提醒被试已经用去的时间。点击一个方框打开一个信息箱。在打开另一个信息箱之前,必须再点击一次方框关闭上一个信息箱。时间一到,被试不能再打开剩下的信息箱,必须马上作出选择。每作出一次正确选择,被试将得到 1 马克(大约 60 美分)的报酬。

分析决策过程

149

为了充分了解他的朋友作出决策时采用的策略,K 先生可以采取的一种方法是观察他们作出决策前的一切活动。他们寻找什么信息,以及按照什么顺序搜集这些信息呢?更一般地说,作出决策前发生了什么事情呢?导致最终决定的潜在过程是什么?许多研究者倾向于采用这种过程取向的方法,并用这种方法获得了一些重要发现。例如,培恩等人(1988,1993)、比奇和米切尔(Beach & Mitchell,1978)以及斯文森(Svenson,1979)等的研究均为该领域作出了突出贡献。他们的主要兴趣在于观察被试是以什么顺序获取信息的、这些信息是如何在认知上得到表征的,以及得到表征的信息是如何得到转换,从而帮助个体作出最后决策的。他们采用的研究方法主要有两种:一种是要求被试描述他们作出决策的过程;另一种是对被试的信息搜索过程进行监控。

第一种方法涉及到了被试对自己决策过程的口头报告——或者在进行决策时借助于出声思维(think-aloud)技术来进行(例如,Timmermans,1993),或者在作出决策后通过反省或回忆的方式来实施。这种口语报告能够提供有关认知操作和决策策略的信

息。反省技术常常受到批评,因为人们并不总能接近或意识到他们的认知操作,也难以准确地描述他们的决策过程(Nisbett & Wilson,1977)。相反,一些研究者支持采用口语分析(protocol analysis)技术(Ericsson & Simon,1993;Montgomery & Svenson,1989;Payne et al.,1978)。出声思维的一种变化形式是吉戈伦尔和霍夫瑞格(1995)研究贝叶斯推理问题时提出的书面报告(write-aloud)技术。在我们的研究中也采用了这种技术,即要求被试在解决推理问题的同时作书面记录,之后研究者使用这种记录就有关推理过程问题与被试进行谈话。

第二种方法对决策制定者的信息搜索活动进行监控。以“选项×属性(线索)”矩阵的方式将信息呈现在一个计算机控制的信息板上。通常被试不能够直接看到信息,必须通过询问或一定的操作来获得属性值。一种类似的方法是眼动记录(eye-movement recording)程序,它根据被试的眼动记录来推断其信息搜索过程(Russo & Doshier,1983)。用这种方法获得的资料依据下面的假设加以整合:信息搜索顺序依赖于被试采用的决策策略。关于口语记录方法和监控技术孰优孰劣的比较,请参见比格、罗斯曼和塞根生(Bigg,Rosman & Sergenxian,1993)的有关研究。

广为人们采用的监控信息搜索的方法是被称为探测实验(mouselab)的计算机程序(Johnson et al.,1991)。在我们的实验中,该程序被用来将4个公司和6条信息呈现在一个信息板上(图7-1)。线索值隐蔽在每一个方框下面,点击方框可以看见它。一次只能打开一个方框。最后的公司选择通过点击屏幕底端的4个方框中的一个来完成。

对信息搜索过程的监控怎样才能被用来揭示被试采用的策略呢?探测实验程序给出了两类信息:被试打开方框的顺序和打开每一个方框的持续时间。有了这些信息,我们就可以对被试搜索信息过程涉及到的几个变量加以分析。下面所讨论的几个变量也是培恩等人(1988)曾经使用过的变量。

	公司 1	公司 2	公司 3	公司 4
股份价格				
再认率				
红利		4		
资本总量				
投资总量				
雇员数量				
	公司 1	公司 2	公司 3	公司 4

图 7-1 呈现 4 个公司和 6 条(隐蔽的)线索的计算机控制的信息板(由探测实验程序生成)。每条线索的值可以通过点击方框揭开。图中公司 2 的“股票红利”已经被揭开了。

搜索活动的一个非常明显的特征是寻找的信息量。例如,在我们的实验中,被试在时间充裕条件下平均打开 33 个方框。由于信息板共有 24 个方框组成,所以这意味着一些被试不止一次打开了一些方框。就表 7-1 列出的决策策略本身而言,没有任何一种策略有必要对信息进行重复检测。被试这样做的原因或许只是由于工作记忆的容量限制(Miller, 1956)导致他们忘记了先前看到过的线索值。另外一种可能性是:被试重复打开方框,是为了作出选择时更有信心。他们首先利用一种特定启发式规则选定一个选项,然后重复获得信息以便支持将该选项作为最终选择(类似的想法请参见 Svensson, 1992)。尽管如此,但从被试打开方框的数量来看,拟议中的策略中的任何一种都不能够准确描述信息搜索的全过程。在时间紧迫条件下,被试平均打开的方框数量是 19 个。打开方框数量的减少并不令人惊奇,它表明:至少对一些被试来说,时间紧迫条件下再想采用所有适用信息是不可能的。从更加广义的角度看,被试打开方框的平均数量触及到了这样一个问题:个体的信息搜索是怎样与所采用的决策策略相关的? 即便一个人获得了大量信息,也不一定意味着所有这些信

151

立在他采用的特定策略所需要的信息基础上。

信息搜索过程还可以通过打开每一个方框持续的时间来加以考查。我们也许会预期：时间紧迫性将会提高信息获取的速率，即：缩短打开每一个方框所用的平均时间将会导致在同样单位时间内获得更多信息。我们的实验结果是：时间充裕条件下打开方框的平均持续时间是 430 毫秒，而时间紧迫条件下打开方框的平均持续时间是 325 毫秒。如果我们计算一下打开不同线索或选项方框的持续时间的話，我们将会得到一些更加具体的信息。被试花费在最重要线索上的时间在全局用于搜索的时间中占有多大比例呢？时间充裕条件下此项比例平均为 21%，而时间紧迫条件下此项比例平均为 25%。（由于本实验共用了 6 条线索，所以每条线索平均占有所有适用信息的 16.6%。）从 21% 增加到 25%，说明时间紧迫条件下被试更为注意效度最高的线索。相反，花费在效度最低线索上的时间从 13.5% 降低到了 11%。更一般地说，相对于时间充裕条件而言，时间紧迫条件下花费在每条线索上的时间与该条线索的效度有更加紧密的关系。这可以用线索效度与花费在每条线索上时间的相关来表示：在时间充裕和紧迫两种条件下这种相关分别为 0.69 和 0.82。这表明，将注意力集中于哪条线索在很大程度上决定于线索的效度，一定程度上也决定于时间紧迫性。对这种结果的一种理解是：在时间紧迫条件下人们改变了他们的决策策略——更加倾向于不可代偿策略。

花费在每条线索上的时间的标准差随着时间紧迫性的改变而变化吗？时间充裕条件下，平均的标准差为 80 毫秒，时间紧迫条件下它增加到了 100 毫秒。因此，被试并未花费同样多的时间在各条线索上，时间紧迫条件下被试的注意力更加不均匀地分配到不同线索上。一种类似的变量是花费在每个选项上时间的标准差。时间充裕和紧迫两种条件下该标准差分别为 130 毫秒和 120 毫秒。时间紧迫条件下对选项的注意力分配的不均匀性并未增加。这种结果不支持如下假设：时间紧迫条件下，由于要更加频繁地使用不可代偿策略，对选项的

注意力分配的不均匀性将会增加。对这种结果的一种可能的解释是：时间充裕条件下被试利用多余时间对他们偏爱的选项作了更多考虑，以至于造成了选项之间较大的标准差。

另一种变量——“范型”(pattern)——是由培恩提出来的(1976)，用来刻画所使用的信息搜索类型。在关闭了一个特定方框之后，一个决策制定者要么打开同一个选项的另一个方框(即纵向搜索，被称为“选项内转换”)，要么打开同一条线索的另一个方框(即横向搜索，被称为“线索内转换”)。信息搜索范型就是用选项内转换的次数减去线索内转换的次数，再除以两种类型转换的总次数。该变量的取值范围从-1到+1。正值表示更多基于选项的信息搜索，负值表示更多基于线索的信息搜索。在我们的实验中，时间充裕条件下的平均“范型”值为-0.13，时间紧迫条件下的平均范型值为-0.28。这表明信息搜索更经常地是基于线索进行的，时间紧迫条件下更是如此。基于线索的搜索与任何一种不可代偿启发式的使用均是一致的，与加权正面线索规则也一致。

更重要的是，范型变量值受到了不对称原始信息的影响，这意味着如果线索数量多于选项数量(如同我们实验中所做的这样)，那么即便当信息搜索随机进行时，也应该获得一种正的范型变量值。因此，如果我们对这种不对称原始信息加以矫正的话(即增加选项数目或减少线索数目)，那么实验中将会有更多的证据支持基于线索的搜索。^{*}

* 而且，即便加权的正面线索策略或词典编纂策略这种典型的基于线索的搜索规则也可能包含选项内的转换。如果一个人使用加权的正面线索规则，获得了关于一条线索的信息，其线索值超过了其他所有选项的线索值，那么从该条线索向另一条线索的移动就只能在同一项目内进行。这种信息搜索对于使用加权正面线索规则的人而言是可能的，但是对于我们实验中所使用的信息板来说，这种搜索将只能产生-0.57的范型值(它远高于可能的最低值-1)。例如，如果另一个人使用词典编纂规则，在检测了具有最高效度的线索之后，仍然难以对四个选项中的两个作出区分，那么这个人将利用效度居于第二位的线索来比较剩下的两个对象。这种信息搜索只能导致-0.6的范型值。因此，即便对于以基于线索的信息搜索为特征的启发式策略来说，也难以达到最高的负范型值，相形之下，我们实验中所获得的范型值已经足以证明被试采用了基于线索的信息搜索方式(参见 Bockenholt & Hynan, 1994)。

虽然时间充裕和紧迫两种条件下有关变量的绝对值差异不大,但其中多数已经表现出了显著的主效应。这两种条件下被试打开方框的数量、平均持续时间、在最重要和最不重要线索上花费时间的比例以及在各条线索上花费时间的比例的标准差均有非常显著的主效应,而范型值的差异以及花费在各条线索上的时间与线索效度相关的差异只有中等程度的主效应[按照 Cohen(1988)的方法进行计算和分类]。

上述变量对于揭示和分析不同决策过程是有用的,但是它们也能够被用来鉴别和确认具体策略吗?一些研究者指出,这些变量的结合可以使人们做到这一点(Billings & Marcus, 1983; Westenberg & Koele, 1992)。韦斯腾伯格和库尔推测,通过“范型”变量负值(表明基于线索的信息搜索)和不同选项获得的不同信息量(表明一些选项比另一些选项得到了更多注意)的联结,可以鉴别和确认逐项排除启发式策略。这种方法的缺陷之一在于难以将逐项排除规则与词典编纂策略、半词典编纂策略等其他不可代偿策略(在韦斯腾伯格和库尔的研究中未考虑这些策略)区分开来。它的另外一种缺陷是,即便一个人为不同选项搜集的信息量只有些微不同,他所采用的决策策略也会被确认为不可代偿的。而且,这种情况经常会发生,尤其在拥有两个以上适用选项的任务上,更是如此。似乎有理由假定,即便人们使用了可代偿策略,他们有时候也会不为某些特殊选项搜集信息,或者忽略有关某些选项的信息。因此,我们认为上述变量还不足以使我们确认个别被试所采用的独特策略。

总之,对我们实验过程的分析表明,被试主要采用了基于线索的信息搜索方式。这一点最明显地体现在范型变量上。这种类型的信息搜索与词典编纂规则、半词典编纂规则、逐项排除规则和词典编纂附加规则的要求是一致的。然而,在这些启发式规则之间作出进一步区分是困难的。作为对过程取向方法的评价,我们认为,尽管它对于揭示不同条件下的不同决策过程是有用的,但是将其用

简捷启发式

作识别和确认特定决策策略的方法是成问题的。而且,尽管我们可以假定,相对于使用不可代偿策略的人来说,使用可代偿策略的人获得了更多信息,但是我们难以确定将这两种策略区别开来的具体临界值。尽管如此,我们仍有理由假定有关信息搜索过程的资料有助于减少似是而非的策略数量。表 7-1 表明了可代偿与不可代偿策略在有关信息搜索过程变量上的关系。加权正面线索规则——它并未出现在表 7-1 中——可以被安排在可代偿策略的行列中,但有可能在范型变量上出现负值。词典编纂附加规则也是一种可代偿策略,它也未出现在表 7-1 中,因为它来说各种预期变量值是词典编纂规则和达维斯规则预期变量值的混合。

表 7-1 区分可代偿与不可代偿策略的过程变量

154

过 程 变 量	策 略	
	可代偿	不可代偿
	富兰克林规则 达维斯规则 好特征规则	词典编纂规则 半词典编纂规则 逐项排除规则
打开方框的数量		>
打开方框的平均持续时间		<
花费在最重要线索上的时间所占比例		<
花费在不同选项上的时间比例的标准差		<
花费在线索范型上的时间比例的标准差		<
范型值	正值	>
花费在每条线索上的时间与线索效度的相关		<
		负值

注:策略的含义见文本框 7-1。

虽然许多研究者建议既要考查被试最终作出的选择,又要考查其决策前的过程(例如, Einhorn & Hogarth, 1981; Einhorn et al., 1979),但多数过程取向的研究并未考查最终的选择(例如, Payne et al., 1988)。既然有关搜索活动的资料表明被试采用的是基于线索的信息加工方式,但却不能在不同的具体策略间作出区分,那么对最终的实际选择加以分析也许有助于弥补过程取向

方法的不足。

分析决策结果

K 先生还可以通过分析各种股票的信息(输入)与其朋友最终作出的股票选择(输出)之间的关系,来模仿他的朋友作出选择。实际上,许多研究者早已采用了此种研究策略。其中最具有代表性的是肯·哈蒙德[Ken Hammond, 1955; 亦可参见 Brehmer & Joyce(1988), Cooksey(1996), Doherty(1996)等人的有关评论]提出的带有新布伦斯维克主义倾向的社会判断理论(social judgment theory)。作为一种描述人们判断策略的理论构架,社会判断理论主要使用了线性模型,尤其是多元回归模型。在一种被称为判断分析(judgment analysis)或策略探取(Policy capturing)的技术中,这种统计模型被用来描述被试是如何利用适用线索作出判断的。一个人采用的“策略”被描述为回归方程中权重的向量,它使得判断成为线索的线性函数。尽管新布伦斯维克主义者声称仅限于将多元回归作为描述判断过程的一种工具,但这种分析原则上也适合于检验其他候选模型。事实上,在艾刚·布伦斯维克的著作中,一个人很难找到有关限制使用回归方法的论述。最近,肯·哈蒙德(1996b)指出,过于倚重回归方法是一种可悲的错误!

155

在下面报告的结果分析中,文本框 7-1 列出的每种策略都被用来预测被试的选择。^{*}标准的策略探取技术从来自被试的实际资料推导出多元回归的权重,因此建立起来的回归模型与被试实际选择之间的拟合度较高。与这种方法不同,我们用线索效度(参见文

^{*} 然后将这些预测与被试实际的选择相比较。每一种策略的预测都是以所有 24 个方框中的信息为基础的。这也许被认为是成问题的,因为——尤其在时间紧迫条件下——被试并不一定搜寻所有这些信息。然而,将预测仅仅建立在实际寻找的信息基础之上将是更成问题的:当一个被试没有寻找一种策略所需要的信息时,我们将不知道如何确定该种策略将作出什么预测。

简捷启发式

本框 7-2)来表示权重,先人为地对有关策略的运作机制加以描述。我们的方法无法取得像标准策略采取技术那样高的(与被试实际选择的)拟合度,因为后者为不同被试推导出了不同权重。

然而,在进一步讨论之前,我们有必要指出标准采取技术存在的一种局限性。不同的策略常常导致同样的预测,因此一个个体的决策很难被明确地归因于某一种策略。为了说明这个问题,我们不妨以第 4 章讨论过的城市人口问题为例,来看一看采纳最佳启发式(类似于词典编纂策略)与多元回归方法的预测是如何相互重叠的。采纳最佳启发式和多元回归方法均被用来预测两个城市中哪个有较多人口。共有 83 个城市和 9 条双极线索构成了推断环境。我们通过 83 个城市组成的所有对偶比较来考查两种策略预测的重叠情况。如果在特定对偶比较中,采纳最佳启发式和多元回归作出了同样预测,就记为“1”;如果它们作出了不同预测,就记为“0”;如果其中一个或两个都无法作出预测,不得不凭猜测,就记为“0.5”。结果表明,多数情况下采纳最佳启发式和多元回归方法作出的预测是一致的:在所有对偶比较中 92%的情况下两种策略都作出了同样预测,排除了其中一种策略必须凭借猜测的情况后,这个比例增加到了 96%。

如此高的重叠对于重新理解策略采取技术有重要意义。假如我们从上述城市人口环境中随机抽取出 100 对城市。一个被试始终用采纳最佳启发式进行推断,另一个被试始终用多元回归方法进行推断。那么,一个采用策略采取技术的新布伦斯维克主义者能够探测到这两个被试之间的差异吗?答案是否定的。这两个假想被试所采用的策略是用多重相关和回归方程中的权重来表示的,而他们彼此间在这些方面并无不同(Hoffrage et al., 1997)。因此,我们应该认识到线性模型并不比霍夫曼(Hoffman, 1960)所谓的“同质异形”(paramorphical)表征——也许并非对实际推理过程作出有效描述的一种数学表征(亦可参见第 6 章有关不可代偿权重的论述)——做得更好。当然,这个问题不限于线性模型,在描述人类判

156

断和决策活动的其他模型中也经常产生。而且,线性模型对于预测结果和探测个体间差异仍然不失为一种很有用的模型。

总之,像分析决策过程一样,对决策结果进行分析也面临着同样的问题——策略的可分离性,即难以在不同策略间作出鉴别或区分。解决这个问题的一种方法是,对呈现给被试的选项加以精心挑选,迫使不同策略作出不同预测。然而,人为地挑选选项以最大限度地减少不同策略所作出的预测间的重叠常常会使所用选项缺乏代表性,以至于研究结果难以推广。这里似乎存在着一种两难困境:要么选项具有代表性,研究结果可以推广,但却无法鉴别和区分有关策略;要么对选项加以挑选,以提高对被试采用策略的鉴别力,但研究结果很可能难以推广。

在本实验中,我们以 70 个随机选择的国有公司为推断环境。每个问题都由四个一组的公司所形成,被试的任务是从四个公司中挑选出利润最高的公司。我们构建了两套不同任务。首先是“有代表性”任务,它包括 30 个问题,每个问题都由随机挑选出的四个公司组成。此种情况下,文本框 7-1 中列出的策略作出预测的重叠率相当高:所有策略在 30 个问题中的 23 个问题上都选择了同样公司。每一对策略作出同样预测的比率大约为 92%(所有可能成对策略的平均比率)。

由于如此,我们又分别以达维斯规则和词典编纂规则作为可代偿和不可代偿策略的代表,有选择地构建了 30 个问题,以减少二者之间预测的重叠。我们采用的选择问题的方法是:逐个地随机生成每个四选一问题,而只保留达维斯规则和词典编纂规则作出不同选择的问题。* 在最后选出的 30 个四选一问题中,只有

* 而且,我们力图避免挑选出的一套问题与其他可能的问题在线索效度上有太大差异。为此,凡是引起线索效度失真的问题均被排除出去。此外,最后确定的一套问题中也包括了达维斯规则与词典编纂规则作出同样预测的 2 个问题。请注意:最后挑选出来的 30 个问题均为该推断环境中可能形成的问题,因此两种策略的成绩不依据所采用的选择程序加以考虑。

简捷启发式

两个问题是所有策略均作出同样预测的问题。与上述“有代表性”任务不同,每对策略平均作出同样预测的比率仅为 50%。与有代表性任务相比,构成有选择性任务的各个选项间的类似程度更高,即有代表性任务各个选项间线索值的平均标准差大于有选择性任务各个选项间线索值的平均标准差(分别为 1.3 和 1.2)。这意味着在有选择性任务上作出一种选择将会更加困难。在本实验中,50 名被试完成有选择性任务,25 名被试完成有代表性任务。 157

仍以选择利润最高公司次数所占比例为指标,不同策略在两种任务上取得的成绩如何呢?正像表 7-2 显示的那样,在有代表性任务上这些策略平均获得了 73% 的正确预测,而在有选择性任务上,不仅它们的平均成绩更低,而且策略之间成绩的差异也更大。在两种任务上,简捷启发式的平均成绩均与寻找所有线索值并需要大量计算的富兰克林规则的成绩持平。特别是,在有选择性任务上,加权正面线索规则和半词典编纂规则均比富兰克林规则作出了更加准确的选择。在有代表性任务上,词典编纂规则和好特征规则作出了像富兰克林规则同样多的正确选择。

上述研究结果是很有意思的。我们不妨将其与在偏好判断研究上取得的结果加以比较来说明这一点。在我们的研究中简捷启发式所以能够取得像富兰克林规则一样好、甚至更好的成绩,是因为我们采用了一种外部标准(公司利润)来衡量一种策略的预测准确性。而在偏好判断研究中,此种标准是不存在的。对于它来说,最基本的观点是:一种期望的理想模型(类似于富兰克林规则)是在不同选项间作出选择的最佳方法,并被用作评价其他所有策略的标准(例如,Payne et al., 1988)。结果是,在偏好判断研究中,其他启发式的成绩至多只能像富兰克林规则的成绩一样好,而不可能比它更好。

表 7-2 不同策略在有代表性和有选择性任务上的成绩

158

	富兰克林规则	戴维斯规则	好特征规则	加权正面线索规则	词典编纂附加规则	词典编纂规则	半词典编纂规则	逐项排除规则
有代表性任务	77	75	77	73	70	77	73	73
有选择性任务	57	37	43	60	53	53	63	47

注：成绩用正确预测利润最高公司所占的比例来表示（策略的含义参见文本框 7-1 的解释）。

另一方面，这些策略在完成两种任务时的节俭性如何呢？这里节俭性用每种策略作出决策时打开的信息方框数目所占比例（在所有 24 个方框中）来表示。表 7-5 显示了完成所有项目时不同策略打开方框的平均比例。就多数可代偿策略而言，为了作出一种决策，所有信息都是必要的。词典编纂策略最为节俭，仅使用了大约 1/4 信息。

表 7-3 不同策略在有代表性和有选择性任务上的节俭性

160

	富兰克林规则	戴维斯规则	好特征规则	加权正面线索规则	词典编纂附加规则	词典编纂规则	半词典编纂规则	逐项排除规则
有代表性任务	100	100	100	100	60	26	49	60
有选择性任务	100	100	100	100	63	24	47	63

注：节俭性用预测利润最高公司时所使用的线索数目的比例来表示（数字越小表示一种策略越节俭；策略的含义参见文本框 7-1 的解释）。

最后，不同策略预测被试作出选择的情况如何呢？我们首先看一看不同策略正确预测所有被试在所有项目上作出选择的情况。如图 7-2 所示，相对于有选择性任务来说，在有代表性任务上各种策略正确预测被试选择的比例更高，而且不同策略间正确预测率的差异更小。这也许是由于在有代表性任务上不同策略作出同样预测的重叠率较高造成的。

159

图 7-3 显示了采用不同策略的被试比例。就每一个被试来说，如果某种策略预测了其作出选择中最多的数目，那么该种策略就被分派给该被试。如果有两种以上策略对一个被试的选择作出

让我们更精明

简捷启发式

了同样多的预测,该被试就被归入“模糊不明”策略之列。在有代表性任务上,能够最经常地分派给被试的策略依次是半词典编纂

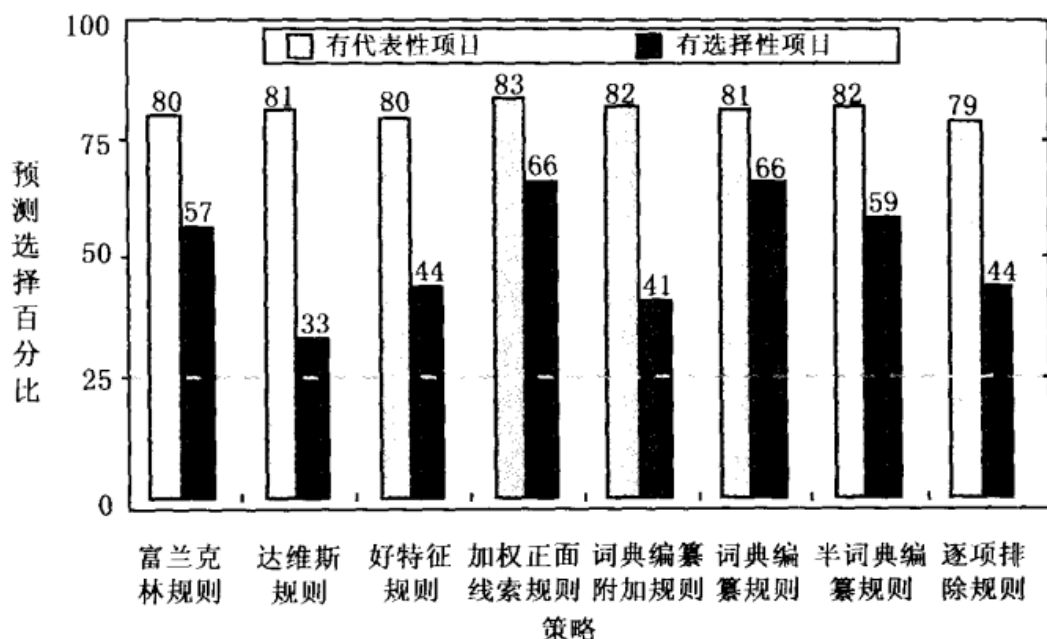


图 7-2 在有代表性和有选择性任务上不同策略预测被试选择的准确性。水平虚线代表凭借机遇(25%)可能达到的准确性。

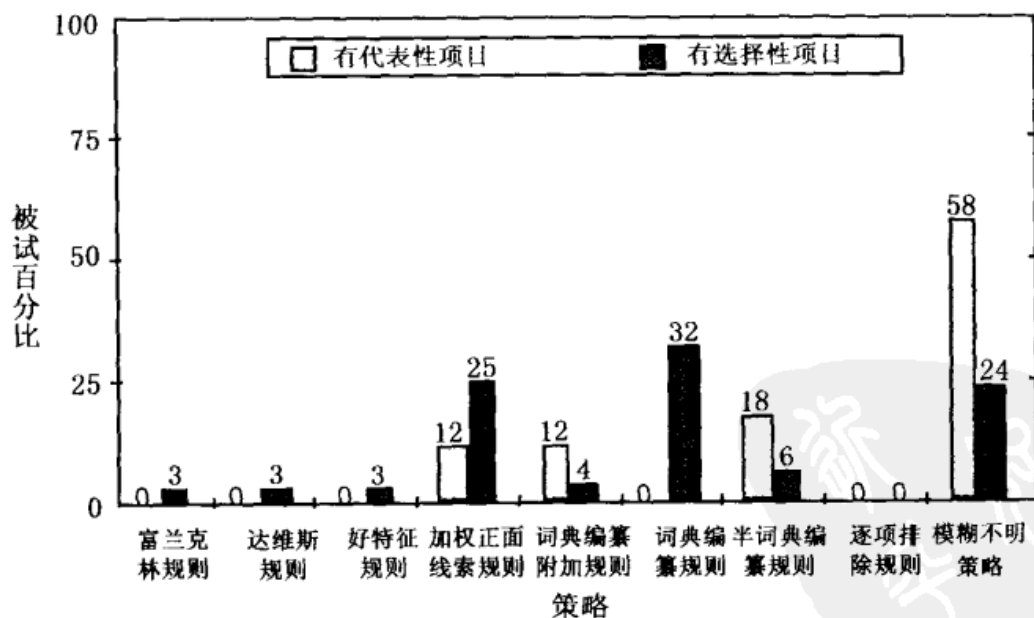


图 7-3 在有代表性和有选择性任务上分派给每种策略的被试比例(时间充裕和紧迫两种条件下的累加值)。

规则、加权正面线索规则和词典编纂附加规则；而在有选择性任务上，能够最经常分派给被试的策略依次是词典编纂规则、加权正面线索规则和半词典编纂规则。不能够被明确地分派给任何一种策略的被试所占比例是相当高的，在有代表性任务上更是如此。

如前所述，该项研究的主要目的是检验构建一套能将不同策略鉴别开来的任务的重要性。因此，在考察时间充裕和紧迫两种条件的影响时，我们仅分析了完成有选择性任务的 50 名被试的成绩。此外，我们还采取了一种方法将采用模糊不明策略的被试作了划分，即：每当两种或两种以上策略在预测某个被试的选择上打成平手时，我们就将该被试平分给有关策略（如，假如两种策略在预测一个被试的选择时打成了平手，我们就给两种策略各增加 0.5 个被试）。图 7-4 显示了时间充裕和紧迫两种条件下“被试—策略”的匹配情况。能够最经常地分派给被试的策略是词典编纂规则和加权正面线索规则。时间紧迫条件下词典编纂规则描述了更多被试采用的策略，而时间充裕条件下加权正面线索规则说明了更多被试采用的策略。这反映了一种趋向，即：从时间充裕条

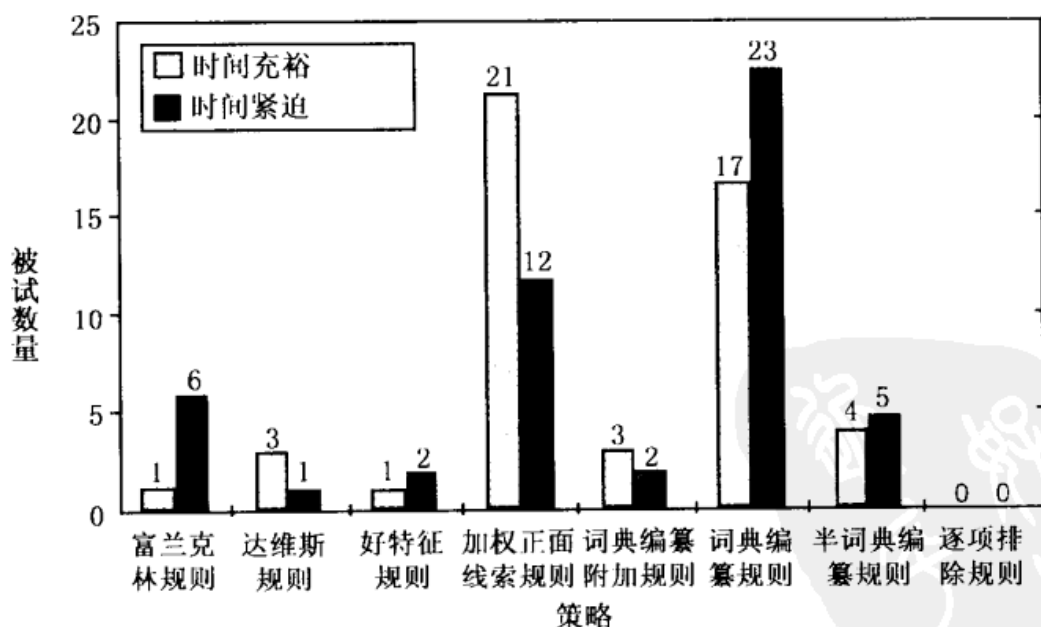


图 7-4 时间充裕(50 秒)和时间紧迫(20 秒)条件下被分派给不同策略的被试人数。资料仅限于完成有选择性任务的 50 名被试。

件下的可代偿启发式(加权正面线索规则)向时间紧迫条件下更为简单的不可代偿启发式(词典编纂规则)转移的趋向。它与我们的如下假设是一致的:相对于时间充裕条件而言,时间紧迫条件下可以用不可代偿策略对其决策过程加以最恰当描述的被试所占比例更大。 162

在所有策略中,词典编纂规则和加权正面线索规则对于描述被试选择来说同样好。那么,如果我们仅对这两种策略加以比较,会有什么结果呢?为此,我们用有选择性任务对词典编纂规则和加权正面线索规则作了直接比较,仍按照每种策略预测每个被试选择数目的多少来将被试分派给不同策略。在这种直接对比中,时间充裕条件下有 26 名被试分派给了加权正面线索规则,15 名被试分派给了词典编纂规则;而时间紧迫条件下情况恰恰相反,有 27 名被试分派给了词典编纂规则,15 名被试分派给了加权正面线索规则。(对时间充裕条件下另外 9 名被试和时间紧迫条件下另外 8 名被试来说,两种策略预测了他们同样多的选择)。总之,这种直接比较与上述用所有策略获得的结果是一致的,可以用类似的假设来解释,即:时间紧迫条件下,在预测被试选择方面不可代偿的词典编纂规则要比可代偿的加权正面线索规则更加有用。

现在我们可以尝试将前述来自于过程取向方法的变量与对决策结果的分析整合在一起,以便对实验结果作出更加清晰的解释。表 7-4 显示了分派给加权正面线索规则和词典编纂规则的被试在一些过程变量上的测定值。如果对被试所采用策略的分类是有效的,那么不同被试在有关过程变量上获得的分值就应该有所不同。表 7-4 中所示的过程变量值与我们的理论预期是一致的,尤其在时间紧迫条件下是如此。分派给词典编纂策略的被试平均搜索的信息更少一些。他们在信息搜索过程中也表现出了较高的选择性(较大的标准差)。例如,他们花费在每条线索上的时间与线索的效度呈现了更高的相关。然而此类差异都不太大,而且其中的一些不符合我们的理论预

期,尤其在时间充裕条件下是如此。

表 7-4 时间充裕和紧迫两种条件下对分派给词典编纂或加权正面
线索策略的被试过程变量的测定(有选择性任务)

163

过程变量	结果分析			
	时间充裕		时间紧迫	
	词典编纂 ($n = 14$) [*]	加权正面线索 ($n = 18$)	词典编纂 ($n = 18$)	加权正面线索 ($n = 7$)
打开的信息方框的数目	32.9 个	33.3 个	18.5 个	20.5 个
打开一个信息方框所用的 平均时间	0.45 秒	0.35 秒	0.32 秒	0.28 秒
花费在最重要线索上的时 间所占的比例	20%	21%	28%	23%
花费在不同选项上的时间 比例的标准差	$sd = 0.12$	$sd = 0.12$	$sd = 0.13$	$sd = 0.10$
花费在不同线索上的时间 比例的标准差	$sd = 0.08$	$sd = 0.07$	$sd = 0.11$	$sd = 0.09$
范型变量值	-0.38	-0.01	-0.53	-0.13
花费在每条线索上的时间 与线索效度的相关	$r = 0.70$	$r = 0.73$	$r = 0.91$	$r = 0.80$

* 时间充裕条件下 14 个被试被分派给了词典编纂策略,余类推。

总之,我们通过对结果的分析找到了几种策略中对预测个体选择来说最为适当的策略。我们特别强调一种重要的方法学观点:如果一种决策任务的选项是被随机地加以选择的,那么不同策略在预测被试选择时将会做得同样好。如果两种策略总是对被试的选择作出同样预测,那么我们就不能利用对结果的分析来确定个体究竟采用了哪种策略。正像沃斯腾(Wallsten, 1980)指出的那样,一种模型与一种决策之间的好的拟合并非一定意味着潜在于这种决策背后的认知过程是符合该种模型的。即便一种模型能够预测个体作出的选择,也不能由此肯定个体就是以这种模型为策略来作出他的决策的。潜在于决策活动背后的认知过程仍然是不清楚的(Luce, 1995)。解决这个问题的一种方法是构建一

让我们更精明

套任务,使不同策略作出预测的重叠率降至最低。在我们的实验中使用了一种方法,它使得我们能够确认两种启发式规则——词典编纂规则和加权正面线索规则——作为预测被试策略的最准确的模型。

人们何时使用简捷启发式?

164

经过一段时间的实践和探索,K先生对一般决策制定问题,尤其是简单决策启发式有些人迷了,进一步提出了他的问题。他现在想要弄清楚什么条件下人们更加经常地使用简捷启发式。如同我们已经表明的那样,条件之一就是时间紧迫性。关于时间紧迫性我们能够作出什么样的预测呢?我们能够从其他人的研究中知道些什么呢?

可以设想,当面临时间限制时,为了有选择地进行信息搜索,人们可能采用两种方式作出反应:要么不改变使用的策略,但更快地检索有关信息,以提高使用该种策略的效率;要么改变他们所采用的策略(如,缩短信息搜索的时间或使用完全不同的策略)。第一种反应加快了获得信息的速度(但不减少信息获得量)。许多研究已经证明了时间紧迫条件下这种信息获得的加速现象(Ben Zur & Breznitz, 1981; Edland, 1994; Kerstholt, 1995; Payne et al., 1988);证据通常来自于对同样时间内获得信息总量的测定。也有证据支持上述第二种反应,即改变决策策略。例如,一些研究者已经证明,时间紧迫条件下人们增加了对重要属性的注意力,对信息的搜索更具有选择性(Ben Zer & Breznitz, 1981; Bockenholt & Kroeger, 1993; Kerstholt, 1995; Payne et al., 1988; Wallsten & Barton, 1982)。另一些研究者表明,时间紧迫条件下个体使用了更多基于线索的信息搜索(Maule, 1994; Payne et al., 1988)。在艾德兰德(Edland, 1994)的一项研究中,人们通过更加频繁地使用简单决策策略来对时间限制作出反应。在培恩等人(Payne, Bett-

man & Luce, 1996)的研究中,以失去良机的代价为指标考察了时间紧迫性的作用。他们主要关心这样一个问题:当延迟作出决策将会导致利益损耗时,人们是否改变他们的策略?被试对有可能错失良机的反应类似于对时间紧迫性的反应,即采用了更为简单的策略。关于时间紧迫性对判断和决策制定可能产生影响的更为详细的分析,请参见斯文森和穆勒(Svenson & Maule, 1993)的有关评述。

促使人们更加频繁地使用简单决策策略的其他条件是什么呢?多数研究者都非常关心任务特征的影响(更详尽的讨论,请参见 Einhorn, 1970)。一种重要的任务特征是任务的类型,诸如选择还是判断等。判断常常需要对特定值作出估计。一些研究者假定,不可代偿启发式更加经常地被使用于选择任务上(Billings & Marcus, 1983; Hertwig & Chase, 1998; Montgomery et al., 1990)。韦斯腾伯格和库勒(Westenberg & Koele, 1992)发现,在选择和排序任务上人们使用的信息搜索方式不同,并得出结论:在选择性任务上人们更加经常地使用不可代偿的启发式规则。

充分理解不同类型决策任务的作用是非常重要的。结果取向的研究主要关心判断问题,而过程取向的研究通常涉及选择问题。在关于 45 项过程取向研究所作的评论中,福德(Ford, 1989)发现其中 35 项研究使用了选择任务,只有 7 项研究使用了判断任务,另外 3 项研究既使用了选择任务,也使用了判断任务。有关文献中结果取向和过程方法的这种分歧反映了它们所关心问题的不对称性。

选项和线索数量究竟起什么作用呢?当拥有更多选项和线索时,决策任务的复杂性也许就增加了。福德等人(1989)的结论是:更加复杂的任务导致了更有选择性的信息搜索。许多研究者已经证明,随着选项数目增加,简单不可代偿策略的使用也增加了(Billings & Marcus, 1983; Johnson & Meyer, 1984; Payne, 1976; Timmermans, 1993)。同样地,线索数目的增加减少了人们对信息

165

让我们更精明

简捷启发式

的搜索(Shields,1983)。然而,培恩(Payne,1976)并未发现随着线索数量增加从可代偿策略向不可代偿策略转移的现象,并指出线索数量究竟起什么作用尚不清楚。

线索值的取值范围、线索间的相关程度以及线索权重的不同效度又起什么作用呢?与较大取值范围相比,当线索取值范围较小时,可能会导致个体主观上赋予用来作出决策的线索更高的权重(Mellers & Cooke,1994)。线索间的负相关迫使人们投入更多精力到有关任务上,对信息的搜索也会相对地缺乏选择性,并更多地采用基于选项的信息搜索方式(Bettman et al.,1993)。在线索权重效度很高的条件下,人们倾向于更多地采用选择性的、基于线索的信息搜索方式(Payne et al.,1988)。

社会场合也可能影响决策策略的使用。一种重要的场合因素是策略的可解释性,即必须向别人证实自己观点的压力。塔特洛克(Tetlock,1983)设想,需要对自己的决策负责的人们将会付出更大努力以求作出更加准确和更好的决策,除非他们能够不费吹灰之力就使自己的决策符合别人的期望。这种设想植根于如下更为普遍的假设:采用需要付出更多努力的策略有望获得更加准确的决策。的确,麦克里斯特、米切尔和比奇(McAllister, Mitchell & Beach,1979)已经证明,必须对自己决策加以说明的被试更加经常地选择需要相当努力、但更有可能获得正确决策的策略。关于决策任务的重要性,也发现了同样的结果。在毕林斯和斯罗(Billings & Scherer,1988)的一项研究中,被试用逐渐增多的信息搜索来对重要性逐渐增加的任务作出反应,尽管这种效应仅表现在选择任务上,而没有出现于判断任务上。

迄今得到证实的影响所采用策略的个体特征主要有知识经验、情绪和动机、认知能力和性别等。单第欧(Shanteau,1988)在广泛意义上总结了专家的策略是如何不同于新手的策略的。专家更加经常地对起初的决策进行调整,力图从早期决策的成败中了解更多东西。他们能够接受微小的错误,但力戒较大的错误。单

第欧(Shanteau, 1992)指出,专业知识与所使用信息量的多少无关。将专家作出更好判断归功于他们使用了更多信息的想法显然是没有根据的。相反,他假定专家和新手的差异表现在他们区分有关和无关线索的能力上。专家似乎使用了同样多的线索,但他们很可能使用了对作出适当决策更加有用的线索。克雷梅(Klaymay, 1985)考察了认知能力对所采用决策策略的影响。他证明,当作出关于几个选项同样适用的重要决策时,具有较高认知能力的儿童获得了更多信息。

我们知道了什么?

本章有两个目标:第一个目标是考察人们是否使用简捷启发式;第二个目标是讨论用来识别人们所采用简捷启发式策略的两种可能方法——过程分析和结果分析。这两种方法对于探索人们的决策制定过程都是有用的。尽管其中任何一个都不足以单独对人们采用的决策策略作出精确预测,但这两种方法的使用能够增进我们对人类决策过程的理解。

过程取向方法监控决策前的信息搜索活动,并记录人们用于检测特定信息的时间长短。本章报告的实验表明,这种方法的确触及到了决策过程的一些方面。多数被试使用基于线索的信息搜索方式获得信息。在时间充裕和紧迫两种条件下他们的注意力都集中于重要线索上。花费在每条线索上的时间与线索的效度高度相关。这种结果支持这样的假设:被试采用了不可代偿的启发式策略,尤其在时间紧迫条件下。然而,过程取向的研究资料不可能鉴别哪一种特定启发式是最为适当的决策模型。我们所考察的多种不可代偿启发式在所预期的信息搜索和赋予特定信息的注意力上只有些微差异。

结果取向方法的主要想法是建立一种能够对被试选择作出预测的决策模型。通常采用的是线性模型,尽管这种方法不仅仅适

合该种模型。在我们的实验中,利用多种附加策略来预测被试的选择。对于有代表性任务来说,将一种特定策略分派给一个个体是困难的,因为不同策略所作出的预测有很高的重叠率。采用将重叠率降至最低的有选择性任务,我们发现:时间充裕条件下能够对被试选择作出最好预测的是加权正面线索规则,而时间紧迫条件下能够对被试选择作出最好预测的是词典编纂规则——采纳最佳启发式的延伸。这导致我们得出如下重要结论:时间有限时人们改变了他们的决策策略。这与文献中对决策过程的有关分析是一致的。

加权正面线索规则和词典编纂规则均逐个使用每条线索(基于线索的搜索)来比较有关选项。而且,二者都只检测一个选项是否比另一个选项有更高线索值,而忽略了线索值之间的绝对差异量。因此,这两种启发式对线索值之间的差异是非常敏感的。然而,尽管词典编纂规则是不可代偿的,但加权正面线索规则却是可代偿的,因为它对信息进行了整合。另一方面,相对于富兰克林规则而言,加权正面线索规则进行线索代偿的机会却要少得多。

168

值得注意的是,启发式只是为实际推理过程建立的模型。我们提供的证据表明以基于线索的信息搜索为基础的启发式规则能够较好地描述个体完成选择任务时的决策策略。个体似乎使用了拥有终止规则的、快速节俭的不可代偿策略(词典编纂规则)。尽管特定条件下也有证据表明人们采用了整合信息的策略(加权正面线索规则),但这种信息整合仅以适用信息中的一部分为基础。这与谢泼德(Shepard, 1967a)的看法是一致的,他指出任何时候人们都很难顾及到一种或两种以上的因素,尽管他们主观上觉得能够做到这一点。与其他研究发现(Simon, 1990)相一致,我们的被试力图回避需要利用所有信息和进行大量运算的策略。即便人们使用了大量信息,并试图将它们整合在一起,他们似乎也采用了简单的认知操作方法。总之,本章报告的研究结果强烈地表明:人们的确使用了精明简单的决策策略!

第8章

快速节俭启发式的贝叶斯基准

劳拉·马蒂冈(Laura Martignon)

凯瑟琳·布莱克蒙德·拉斯凯(Kathryn Blackmond Laskey)

当灵感来临时,不要再刻意去思考。

——鲁德亚德·基普林(Rudyard Kipling)

本章用来说明贝叶斯模型在推理活动上所具有的神灵般的优越性。首先,我们叙述了过去3个世纪贝叶斯模型和人类推理理论之间不断变化的关系;其次,我们描述了贝叶斯模型的基本范式;第三,我们提出了三种贝叶斯模型,作为较少神性化色彩的线性回归模型及过于朴素简单的采纳最佳启发式的竞争对手;最后,讨论了两个敏感问题:如何测定本书所讨论的推断策略的复杂性?贝叶斯模型作为快速节俭启发式的基准所起的作用究竟是什么?贯穿这一章,我们强调了一个总的问题:概率论在有关快速节俭启发式的心理学理论中究竟居于何种地位?

概率规则和推理规则

本书第一章概略地叙述了西方思想中概率论的发展和把概率作为推理基础的观念的变化。首先把概率推理当作理想的理性推理模型确立下来是启蒙运动时期的事情,那是一个用人类自身拥有的推理潜力将人类从教条和迷信的束缚和桎梏下解放出来的充满热情和活力的时代。启蒙运动的思想家们认为是他们发现了

让我们更精明

简捷启发式

人类推理活动的基本规律,并使之确立下来。他们把概率——逻辑演算的一种延伸——看作是一种推理演算过程,并将其视为理性选择必不可少的基石。概率规则不仅被认为是规范的,而且被认为很好地描述了现实生活中开明人士的思维活动。按照拉普拉斯的说法,如果大街上的普通行人尚不足以借助于概率演算进行推理的话,那么至少“开明人士”(L'homme éclairé,即 the enlightened man; Daston, 1998)将会利用概率演算进行推理。概率论以及由贝叶斯(Bayes)、雅克比(Jakob)、尼古拉斯·贝努利(Nikolas Bernoulli)、拉普拉斯和其他一些人在此基础上建立起来的决策理论成为人类推理活动的基本范式。但是,这种把人类推理时遵循的规则等同于概率规则的做法并没有持续很长时间。 170

许多人认为法国大革命是过分强调理性的结果。在此恐怖时期之后,概率论不再被看作是普通意义上的推理规则,而被看成是一种特殊演算规则。焦点慢慢地转移了,概率论变成了从事自然科学研究的数学工具之一。从19世纪中叶开始,概率论逐渐在物理学中确立了不可动摇的重要地位。它首先出现在热学理论中。1871年,路德维希·波尔兹曼(Ludwig Boltzmann, 1844/1906)把热过程的不可逆转性解释为分子均匀分配其能量的倾向性,他的计算借用了概率论。1900年,马克斯·普朗克将概率引用到了他的辐射理论中(Planck, 1958)。随着量子理论的进一步发展,概率被引入到了原子论中。量子力学的产生对于概率理论的革命性变化是非常关键的(Krüger et al., 1987),从那时起,由牛顿(Isaac Newton)、笛卡尔(Rene Descartes)、拉普拉斯等人倡导的关于自然界的决定论观点让位于概率论。

作为对人类实际推理过程的描述,从19世纪初往后概率论开始遭到来自各方面的非难。但是,总的来看,除了一些冥顽不化的理性主义者外,这种非难只持续到了20世纪上半叶。概率论的频率学派(Mises, 1957)试图把概率作为随机可重复事件的一种客观的、可测量的特性置于坚实的基础之上。直到20世纪,把概率论

作为理性思考活动潜在演算机制的观点才重新浮出水面(de Finetti, 1937; Sawage, 1954)。然而,这种复兴传统概率论者理想化理性的尝试刚开始就遭到了怀疑,甚至是公然的敌对。萨维奇(Savage, 1954)的理论放弃了具有相同知识的开明人士定会在“一贯正确的”概率论上达成一致意见的假设,其主观性令科学家们感到不安,因为科学家们要寻找的是关于自然界客观的、可证实的事实。当代实验心理学开辟了反对将理性的人类视为概率演算者的另一个战场。心理学家们通过大量研究证实,人类实际的推理过程在许多方面偏离了概率模型所假定的机制。这种偏离常常被一些人视为认知上的失误,还被作为没有外来帮助的人类推理活动充满谬误的证据(例如, Tversky & Kahneman, 1974)。

但是,不久人们便证明当信息以频率格式呈现时,普通人的贝叶斯推理能力将会得到改进,同时认知上的失误也会减少(Gigerenzer & Hoffrage, 1995)。然而,如果存在许多线索,而正在作出决策的人手头又没有计算器的话,运用概率模型(如,贝叶斯模型)或许是行不通的。随着线索数量继续增加,即使有计算机帮助,基于贝叶斯理论的概率模型也将变得难以实施。正是概率推断策略的难以实施和不可行性,严重阻碍了人们接受其作为人类认知和机器智能通用模型的步伐。

正如本书中我们指出的那样,人类采用的简单启发式规则是一种权且如此的适应性策略,它符合人们用有限知识和时间作出推断和决策的需要,并在迎合这种需要的过程中得以发展。认知资源是有限的,所以好的推理依赖于具有生态合理性的快速节俭启发式构成的适应工具箱。科学家们作为通用推断工具加以发展的复杂方法,与日常生活中人们作出选择和决策时发生的似乎合理的实际推断过程几乎没有共同之处。

这是否意味着正规概率推断理论已经变得无关紧要,对于它们的研究应该被放弃呢?曾几何时,不仅在认知心理学领域,而且在人工智能和机器学习领域,的确出现过一种向该方向努

简捷启发式

力的倾向。20 世纪 80 年代,里维斯特(Rivest,1987)介绍了他的决策表方法,对人工智能研究产生了重大影响。这种方法与采纳最佳启发式之类的决策规则密切相关,已经成为理论计算机科学的概念。几年以后,郝特(Holte,1993)证明:非常简单的分类规则——甚至比采纳最佳启发式还要简单——也取得了优异成绩。

然而,近年来概率推断理论始终处于缓慢的复苏过程之中。它失去了当初独一无二的地位,以一种更为有限、更为谦卑,但却更为适当的方式在悄悄地发挥作用。尽管有人预言概率论将让位退出,还有人假设计算上的难处理性将注定会使概率推断成为一种实际应用中难以付诸实施的理论怪物,但过去十年中该种推理模型却出人意料地重新焕发了生机。在人类推理、参数学习,甚至结构学习领域,贝叶斯模型和贝叶斯技术的迅速崛起致使一些乐观者甚至宣称:21 世纪将是贝叶斯的世纪!沃德·爱德华兹(Ward Edwards)——贝叶斯决策模型的主要倡导者之一——正在准备将贝叶斯模型的拥戴者们召集在一起,于 2001 年在伦敦贝叶斯墓前召开纪念大会!

在概率方法重新焕发出勃勃生机的情形之下,本章试图探讨这样一个问题:提倡快速节俭启发式的认知科学家们怎样才能恰当地使用概率理论?在完成比较、估计或分类任务时,如果受到了计算能力的限制,即便最为忠实虔诚的贝叶斯模型拥戴者也会采用快速节俭策略作为一种捷径或近似方法,其前提是由此造成的成绩(准确性)上的损失不至于太过明显。决策论的基本原则之一(Clemen,1996)就是设法平衡计算上付出的代价与所获得准确性之间的关系。另一方面,即便最为拥护有限推理方法的人也能看到贝叶斯范式的价值,因为它为估价快速节俭启发式的成效提供了最佳的可利用的基准。因此,在算法规则演化和发展的“元水平”(the meta level)上,贝叶斯模型和快速节俭启发式观点是能够和平相处的。尽管各自的目标和期望不同,但却是可以相容的。

172

有限理性的支持者们需要神灵的帮助以建立一种比较标准,借助于它来评价快速节俭启发式的性能。

贝叶斯神灵

第一章在两种理性观之间作出了区分:神灵模型和有限理性模型。我们所谓的神灵并不像拉普拉斯设想的超级神灵那样,生而具有确定无疑和无所不包的广博知识。相反,它们更像是一个计算能力不受限制的智者,可以置节俭性于不顾,情愿且能够通过长时间深思熟虑来解决手头的问题,并能够运用最先进的计算工具来解决它。在本书所涉及的诸多神灵中,贝叶斯模型也许属于神性化程度最高的一类。它们是我们所考虑过的模型中最不节俭、运作起来最为缓慢的模型。然而,从其成效及理论上的一致性和透明性来看,它们又是最富有魅力的神灵。

前面的章节已经证明采纳最佳启发式之类的简单快速节俭策略可以与多元回归之类的普通神灵模型相媲美,甚至做得更好一些。无论在拟合训练项时,还是在推论到测验项时,采纳最佳启发式均表现出了一定优势。本章的目的是将采纳最佳启发式的成绩与更为复杂的贝叶斯神灵的成绩加以比较。我们首先考察一种不受限制的贝叶斯神灵,其能量非凡人所能及。然后考察三种可操作的贝叶斯模型:“配置记忆法”(profile memorization)、“朴素贝叶斯”(naive Bayes)和最近由弗里德曼(Friedman)提出的一种特殊“贝叶斯网络”(Bayesian network)。

贝叶斯模型开始于关于线索和效标之间关系的先验分布,用一个训练样本来校正或更新先验分布,从而获得一个后验分布,然后运用这种校正过的后验分布来对测验项进行评价。最为烦琐复杂的贝叶斯神灵是不受任何限制的贝叶斯策略。对于它来说,凡是不能确定的变量都要事先给出一种先验分布。首先,它不能确定哪种模型——回归、逻辑回归(logistic regression)、分类和决策

简捷启发式

树(CART)、贝叶斯网络或其他模型——适合所面临的问题,因此需要针对模型类别建立一种先验分布;其次,它不能确定同一类别模型内哪种结构假设适合目前的问题——哪些回归变量应该进入回归方程、分类树中何处进行分割、哪些弧线隶属于贝叶斯网络。对于这些不确定性,它也必须给出一种先验分布。最后,即便给出了模型结构,它还必须为该模型结构的参数——回归权重、分类树中各叶结点的效标概率、贝叶斯网络中的局部条件概率表——提供一种先验分布。因此,该种贝叶斯神灵实际上涵盖了大量可能的模型。每一种模型都被指定一种先验概率,而每一种概率都要得到校正以获得一种后验概率。最终的判断取决于每种模型得出结论的权重的平均值,而每种模型的权重又根据其后的验概率来度量。很显然,该种神灵所面临的任務即便对于最为快速的计算机而言也过于繁杂了。

173

配置记忆法

能够被实际运用于第5章所描述的现实环境中去的第一类贝叶斯模型是配置记忆法。该方法仅适合于训练项与测验项完全一致的情形,其目标是拟合训练样本,而不是推论到尚未见过的情形。它的具体运作方法是:设法记住每种线索配置和与之相应的效标值。对每一对线索配置来说,该方法都将选择记忆中效标值较大的那一个对象。如果有几对对象具有相同的成对线索配置,那么配置记忆法需要检测所有成对对象,并确定拥有第一种线索配置对象在当前效标上得分高于拥有第二种线索配置对象的频率。该频率就是第一个对象得分更高的概率。如果必须找到一种确定答案的话,该方法将选择在给定效标上获得高分概率较大的那个对象。该种贝叶斯模型采取了频率主义者的态度!从概率论的观点来看,这种策略属于最优化方法。

配置记忆法的价值在于它代表了一种最优化方案。也就是说,对于给定的一套资料,它取得了能够取得的最佳成绩。但是当

训练项与测验项不同时,即当我们从已知资料到未知资料进行推论时,目前还没有哪种策略可以被证明为最优的。

与配置记忆法不同,下面介绍的两种贝叶斯模型既能被用于拟合,也能被用于推论。

朴素贝叶斯

朴素贝叶斯[又以“傻瓜贝叶斯”(idiot Bayes)而著称]假定:对于给定效标而言,线索是相互独立的。朴素贝叶斯是更为通用的贝叶斯网络的一个特例(Pearl, 1988)。让我们说明该模型是如何被用来解决其他几章所描述的推断问题的,即:当提供了一套与效标有关的线索时,如何确定两个对象中哪个有较高效标值。我们假定有一定数量的对象,每个对象都用一个效标值和数量为 M 的一套线索来描述。每条线索的值均可用两种线索值(1 或 0)中的一种来表示。效标值或者按基数方式排列,或者按序数方式排列。

让我们看一看如何才能为这种任务构建一个贝叶斯网络。设有一套对象,每个对象都用效标值 X 和 M 条线索($C_1, \dots, C_M$)来界定。假设 a 和 b 是从该套对象中随机选取的两个对象,其效标值分别是 X_a 和 X_b , 线索值分别是 $C_j(a)$ 和 $C_j(b)$, j 介于 1 和 M 之间。我们的任务是根据两种线索配置 $[C_1(a), \dots, C_M(a)]$ 和 $[C_1(b), \dots, C_M(b)]$ 推断 X_a 大于还是小于 X_b 。具体而言,我们的目标是:根据给定的 a 、 b 两个对象的线索配置,推断出在效标 X 上 a 的得分高于 b 的得分的概率,可用公式表示如下:

$$\text{Prob}\{X_a > X_b \mid a = [C_1(a), \dots, C_M(a)], b = [C_1(b), \dots, C_M(b)]\} \quad (1)$$

作为一个具体例子,我们不妨看一看第 4 章所采用的任务,即比较成对德国城市,以确定哪个城市有较多人口。在这里共有 9 条线索:“这个城市是国家首都吗?”(NC);“这个城市是州府吗?”(SC);“这个城市有参加全国甲级联赛的足球队吗?”(SO);“这个

简捷启发式

城市曾经是展览会所在地吗？”(EX)；“这个城市在城间铁路线上吗？”(IT)；“这个城市在牌照上的缩写只有一个字母长吗？”(LP)；“这个城市是大学所在地吗？”(UN)；“这个城市位于工业区吗？”(IB)；“这个城市在前西德吗？”(WG)。

图 8-1 表示的是为比较两个德国城市而设计的贝叶斯网络。一个贝叶斯网络是一组变量间关系的概率模型图解。网络中的每个结点都代表一个变量，每个变量都在一套相互排斥的可能性中取值。例如，一个城市可能有城间火车，也可能没有，二者必居其一。图 8-1 所示的贝叶斯网络共有 10 种变量：一种变量是关于效标的，其他变量分别是关于 9 条线索的。效标变量只能在两种可能性中取值：第一个城市 a 较大，或第二个城市 b 较大。每条线索变量只能在以下四种可能性中取值：

175

- (0,0)两个城市在一条线索上均取值为 0
- (1,0)城市 a 在一条线索上取值为 1,城市 b 取值为 0
- (0,1)城市 a 在一条线索上取值为 0,城市 b 取值为 1
- (1,1)两个城市在一条线索上均取值为 1

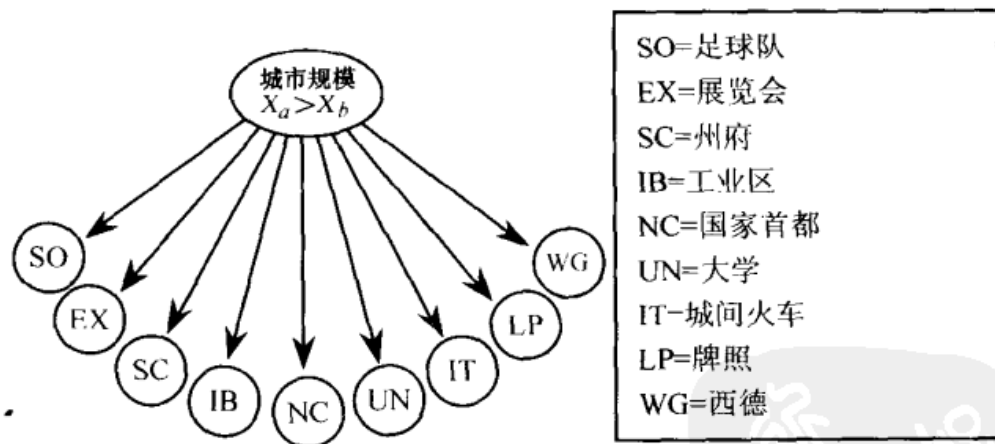


图 8-1 依据 9 条线索比较德国城市人口规模的朴素贝叶斯网络

贝叶斯网络中的变量是由结点表示的，结点之间用带箭头的弧线(或线条)相连，表示变量之间可能的依从关系。如果一条弧线开始于结点 g ，而结束于结点 h ，那么 g 被称为 h 的“亲体”(parent)，而 h 被称为 g 的一个“子体”(child)。如果有一条直接的路

径连接 g 和 j , 那么结点 j 被称为结点 g 的“子代”(descendant)。如果变量的亲体是已知的, 我们就可以说每一个变量都与其非子代变量相互独立。也就是说, 在一个人已知亲体变量值的条件下, 每个变量的值均独立于所有其他非子代变量的值。因此, 在图 8-1 中, 所有线索均依赖于效标变量, 但在此前提下, 各条线索之间却是相互独立的。例如, 一个城市是州府(SC)或有足球队(SO)的概率均依赖于人口规模大小这一效标变量(即较大城市是州府或有足球队的概率也较高)。然而, 如果我们知道了两个城市中哪个较大, 那么在图 8-1 所示的网络中, 这个城市是否是州府和它是否有足球队之间不存在任何关系。

图 8-1 的网络是朴素贝叶斯模型的一个范例。这个特殊网络是将 9 条线索和所有 9 种效标联系起来的最为简单的贝叶斯网络。有实验证据表明, 人们实际上也是按照朴素贝叶斯模型设想的那样去做的, 因为在他们找到线索相互依从的证据之前, 他们倾向于假定不同线索是相互独立的(Waldmann & Martignon, 1998)。然而, 当有 2 条或更多线索时, 人们在对条件概率进行数量估计时做得相当差。

为了对图 8-1 的网络构建一个完整概率模型, 一个人必须具体给出 10 个局部概率表(local probability tables, 缩写为 LPTs), 一个概率表对应于一种变量。每一个局部概率表都是一种概率分布, 即: 与亲体变量每种可能取值相对应的变量分布。在图 8-1 中, 我们既需要为效标变量给出一种具体概率分布, 又需要为每条线索给出两种具体概率分布: 一种是第一个城市较大时的概率分布, 另一种是第二个城市较大时的概率分布。效标变量的概率分布是简明易懂的: 如果随机从成对城市中加以选择的话, 那么有 50% 的几率第一个城市是较大的。为了计算给定效标条件下每条线索的局部概率表, 我们用训练样本来矫正或更新始终如一的先验分布(Looper & Herskovits, 1992)。当样本容量很大时, 这种方法从本质上说等同于使用样本本身的频率, 但当样本容量较小

简捷启发式

时,这种方法将会显得更加耐用。

在完成了训练之后,现在我们设想观察到了拥有一套线索配置的一对城市。我们的问题是在给定这些线索值条件下计算效标变量的条件分布。也就是说,我们希望算出公式(1)所示的概率。一般地说,可以使用任何一种贝叶斯网络软件包来完成这种计算,计算时令观察到的线索值为证据,求效标结点的后验分布。就图8-1所示的朴素贝叶斯网络而言,这里有一个计算效标变量后验概率分布的简单表达式:

$$\begin{aligned} & \text{Prob}\{X_a > X_b \mid [a = C_1(a), \dots, C_M(a), b = C_1(b), \dots, C_M(b)]\} \\ &= \alpha \times \text{Prob}[C_1(a), C_1(b) \mid X_a > X_b] \cdots \\ & \times \text{Prob}[C_M(a), C_M(b) \mid X_a > X_b] \times \text{Prob}(X_a > X_b) \quad (2) \end{aligned}$$

上式中 α 是为确保概率相加为1的一个标准常数。这里我们仅使用了贝叶斯的“反演规则”(inversion),假定在给出了如我们的朴素贝叶斯网络所示的效标情况下,所有线索都是相互独立的。由于当 $C_i(a) = 1, C_i(b) = 0$ 时, $\text{Prob}[X_a > X_b \mid C_i(a), (C_i(b))]$ 就相当于线索 C_i 的生态效度 v_i (参见第6章),而当 $C_i(a) = 0, C_i(b) = 1$ 时,它又相当于 $1 - v_i$,所以我们能够证明在朴素贝叶斯模型中,概率成了对线索生态效度的一种简单表达方式。

一种聪明的贝叶斯网络

朴素贝叶斯和采纳最佳启发式及达维斯规则有一些共同之处(参见第6章):即在环境的结构上“下赌注”(bets)。它盲目地假定线索在特定条件下是相互独立的。如果在给定效标上线索之间有很强的相互依从性,那么朴素贝叶斯将会失去其准确性。现在我们的目标是从一套数据中抽取出有关线索相互依从性的信息。我们采用了一种方法,称为“贝叶斯网络学习法”(learning Bayesian networks),它最初由库普和赫特维茨(Cooper & Herkovits, 1992)的资料发展而来,而其更为有效的版本是由弗里德曼和哥德斯密特(Friedman & Goldszmit,

1996)提出来的。尽管这些算法规则比上面描述的不受限制贝叶斯神灵简单得多,但却要比我们考虑过的其他神灵模型更加复杂、更富有表达力,而且从其潜在能量来看也更为准确。

更为通用的贝叶斯网络(如图 8-2 所示的网络)的结点之间包含了更为复杂的相互依从性。网络的联结程度越高,它所表示的相互依从结构越复杂。一般地说,越是复杂的网络,在给定了其他结点之后,计算某一特定结点的条件概率分布所需要的计算也将越复杂。更加复杂的网络也包含了更多的局部概率表,因此需要更多资料以获得对概率表的可靠估计。在网络充分联结(即每对结点之间均有弧线相连)的极端情况下,运用公式(1)进行计算相当于运用配置记忆法,这将导致对概率的不准确估计,除非样本容量足够大,以至能够对每种线索配置作出多次观察。一般地说,要想对概率表作出准确估计,需要能够提供许多观察机会的大容量样本。因此,对于这类推断问题来说,存在一个权衡问题,即:不得不在大样本更高的可推论性(结点之间更多联结代表可能的依从关系)与小样本计算的简便性和推断的耐用性之间加以权衡。

177

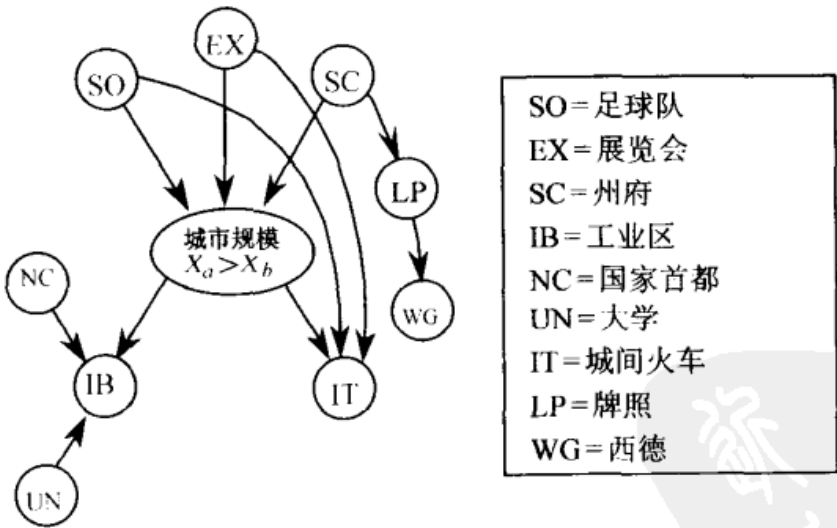


图 8-2 德国城市人口资料的完整贝叶斯网络。该网络(根据弗里德曼搜索法获得)中“城市规模”结点的“标识范围”(markov blanket)容纳了除 LP 和 WG 两条线索外的其他所有结点,这意味着在给出其他线索条件下, LP 和 WG 两条线索与计算城市规模大小无关。

让我们更精明

当然,有些结点之间的依从关系更加密切,因此相对于其他结点之间的依从关系而言更加重要。忽略较弱依从关系有助于避免匹配过当。正像第6章详细讨论过的那样,“匹配过当”一词描述的是为推论目的而拟合一种模型时产生的一种普遍现象:一种模型也许拟合了比它需要“捕捉”资料的内在结构更多的参数。这些额外参数拟合的是资料的“噪音”,而不是资料的内在结构,从而导致在先前未曾见过的资料上的成绩陡然下降。贝叶斯模型的使用者需要找到一些方法来决定哪些依从关系应该被包括在网络中,哪些依从关系应该被忽略。为了在大量可能的贝叶斯网络中搜索一种堪称良好模型的网络,即能够“捕捉”到训练项各个对象间强烈依从关系的网络,人工智能和统计学领域的研究者们已经发展了许多这样的方法。

用来进行搜索的一种方法是把寻找良好贝叶斯网络问题看成是一个贝叶斯推断问题(Cooper & Herskovits, 1992)。这类方法将能够把变量紧密联结起来的网络视为好网络。由于选择贝叶斯模型过程中有一种“自然的奥卡姆简化论”(Natural Occam's Razor; 参见第6章提到过的“奥卡姆简化论”,形容此种情况下人们有一种尽可能简化模型所包含参数的自然倾向),相对较弱的依从关系将会被忽略(Kass & Raftery, 1995)。先验分布产生了一种偏向于节俭性的倾向,这表现在:只有当成绩改进不可能仅仅归因于机遇时,较为复杂的模型才会比更加简单的模型获得更高的后验概率。这些搜索贝叶斯模型的方法首先找到一种“好”的贝叶斯网络,然后再确定与网络中的链接相联系的条件概率。换言之,它们既是学习结构的方法,又是学习参数的方法。

我们曾经运用弗里德曼和哥德斯密特(Friedman & Goldszmit, 1996)提出的贝叶斯网络搜索法来解决第5章讨论过的20种环境下的比较问题。该种算法规则扩展了库普和赫茨维特(Cooper & Herskovits, 1992)的方法,不仅能够利用贝叶斯网络的整体结构,而且能够利用局部概率表内的局部结构[称为“非对

称独立性”(asymmetric independence)],因此提高了学习效率(即能够使人们用较小的样本容量掌握更复杂的整体结构)。图 8-2 所示的是用弗里德曼搜索法获得的网络,该网络是用来解决根据人口规模大小比较德国城市问题的。

给定线索值条件下,根据图 8-2 所示的网络推断哪个城市较大需要运用公式(1)。可以利用标准的软件包来完成这项工作。但需要注意的是,在所有线索均为已知情况下,公式(1)可按照下述定理加以运用:

定理(Pearl,1988): 在给定网络中其他结点有关知识条件下,处于特定状态的结点 j 的条件概率是给定亲体条件下该结点的条件概率与给定亲体条件下其每个子体的条件概率的乘积,再乘以一个标准常数。

用公式表示为:

$$\begin{aligned} & \text{Prob}(\text{结点 } j \mid \text{其他所有结点}) \\ &= \alpha \times \text{Prob}(\text{结点 } j \mid j \text{ 的亲体}) \\ & \quad \times \prod_k \text{prob}(j \text{ 的子体 } k \mid k \text{ 的亲体}) \end{aligned} \quad (3)$$

上式中 α 是经过标准化的常数。

由一个结点和它的亲体、子体以及它的子体的其他亲体组成的集合被称为该结点的“标识范围”。上述定理所表达的意思是: 一个结点的概率仅仅通过其标识范围而依赖于网络中所有其他结点。在给定其标识范围的情况下,它便独立于网络中所有其他结点。

图 8-2 表明了城市规模结点的标识范围,该结点代表了德国城市人口问题的效标变量。它有两种可能状态:“城市 a 的居民多于城市 b 的居民”和“城市 b 的居民多于城市 a 的居民”。该结点的标识范围由除了“牌照”(LP)和“西德”(WG)2 条线索外的所有

简捷启发式

线索所组成。LP 和 WG 线索既不是城市规模结点的亲体,也不是它的子体,亦非其子体的亲体。这意味着当已知另外 7 条线索的值时,这 2 条线索与效标无关。为了算出效标结点的概率分布,我们使用公式(3)所示的定理:

$$\begin{aligned} & \text{Prob}(\text{城市规模} | \text{UN, NC, IB, SO, EX, SC, IT, WG, LP}) \\ &= \alpha \times \text{Prob}(\text{城市规模} | \text{SO, EX, SC}) \times \text{Prob}(\text{IB} | \text{城市规模, UN, NC}) \\ & \times \text{Prob}(\text{IT} | \text{城市规模, SO, EX}) \end{aligned} \quad (4)$$

上式中 α 是为确保城市规模变量的两种状态的概率相加为 1 而选用的常数。

公式(4)右边所涉及的三种概率分别来自于图 8-2 所示网络的局部概率表。计算概率分布时,利用局部概率表将会大大减少运算量。就整体概率分布而言,需要估计 2^{19} 种概率,每一种概率对应于线索/效标值的一种联结。当考虑到条件的非依从性时,计算的复杂性将大大降低。在给定线索条件下,城市规模变量的概率分布是由三种因素决定的,每一种因素对应于一种与城市规模变量有关的局部概率表。每一个局部概率表都有 2^7 个条目,因此在给定了城市规模变量标识范围之后,为了推断其概率,仅需要估计 3×2^7 种概率。条件的非依从性越强,需要计算的概率数量将会减少得越多。确切地说,正是这种计算繁琐性的降低导致了过去年中贝叶斯网络在统计学和人工智能领域的普及和推广。

需要估计的概率数量仍然是每个结点亲体数量的指数,因为每个结点对于其亲体变量值的每一种联结都存有一种概率分布。我们可以再一次通过作出结构假设来限制概率数量,从而降低问题的复杂性。我们采用的算法规则利用一个决策树来估计局部概率表,它可以大大减少计算量。这里面临一个问题:如何找到一个好的决策树?这个问题类似于如何找到一个好的贝叶斯网络来描述线索间的依从关系,因此可以借助于前述搜索贝叶斯网络的方法来解决它。

图 8-3 所示的是利用上述方法为“ $\text{Prob}(\text{城市规模} | \text{SO}, \text{EX}, \text{SC})$ ”生成的决策树。沿着该决策树的弧线可以获得城市规模变量的概率分布。从图 8-3 我们可以看到, 第一步是检测展览会 (EX) 线索。如果两个城市都不曾是展览会所在地, 那么概率分布将决定于两个城市是否有足球队 (SO), 州府 (SC) 线索是无关的。相反, 如果两个城市中有一个曾经是展览会所在地, 概率分布将取决于州府 (SC) 线索, 足球队 (SC) 线索成为不相关的。因此, 对于城市规模结点而言, 用图 8-3 的决策树加以表征只需要 $2^4 = 16$ 种概率分布, 而不再需要 $2^7 = 128$ 种概率分布。

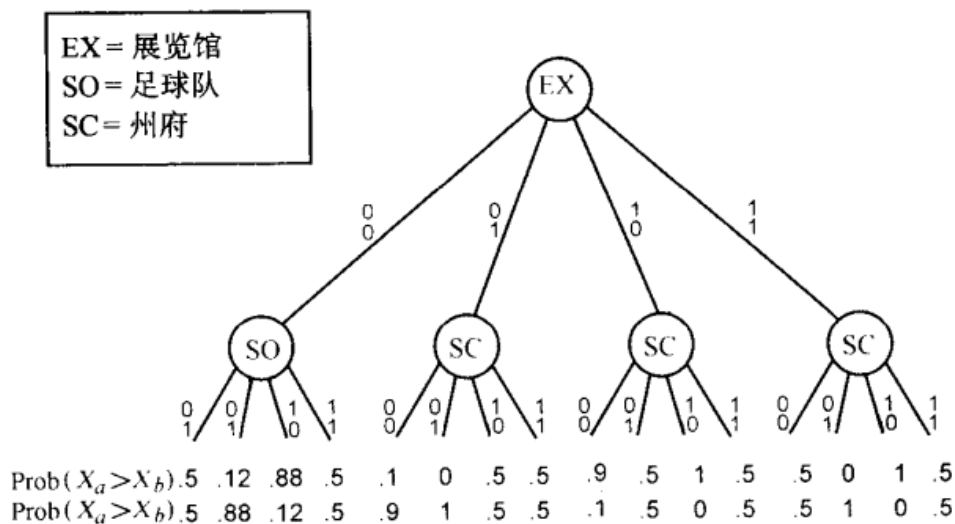


图 8-3 为计算图 8-2 中城市规模结点的局部概率表而设计的决策树。举例来说, 如果城市 a 和 b 均非展览会所在地 (用最靠左边分枝上的两个 0 来表示), 那么唯一有关的线索是足球队线索, 即一个城市是否有参加全国甲级联赛的足球队 (州府线索是无关的)。如果城市 a 有足球队而城市 b 没有 (a 为“1”, b 为“0”), 那么 $\text{Prob}(X_a > X_b | \text{SO}, \text{EX}, \text{SC}) = 0.88$ 。

总之, 为了找到一种能够很好地描述线索与效标之间关系的概率模型, 我们采用的方法涉及到了以下几个步骤:

1. 对能够表示线索和效标变量关系的贝叶斯网络空间进行充分搜索。
2. 对能够表征某结点内局部概率表结构的决策树模型进行充分搜索。

3. 估计局部概率表的概率分布。
4. 给定线索值条件下利用公式(4)计算效标变量的后验分布。

这种方法在计算上比多元回归还要复杂精细得多,更不用说采纳最佳启发式了。事实上,搜索贝叶斯网络和结点内决策树整个空间的复杂性——根据基本的操作步骤数来估算,每个操作步骤都是 M 条线索和 N 个对象的一种函数——可以用以 M 为幂的指数来表示。这个空间太大了,即便对于预测德国城市人口规模这种相对较为简单的问题而言,也难以对其进行详细搜索。因此,该神灵实际上也不得不在一定程度上利用启发式搜索。我们还大胆运用了由复杂算法规则鉴别出来的高概率网络(如图 8-2 所示的关于城市人口资料的网络),而不是运用现实的神灵模型,试图算出效标变量的完整后验分布。完整后验分布是用每种网络的后验概率对其预测结果进行加权后的平均值——很显然,这是一种难以实施的运算。

181

竞赛

快速节俭的采纳最佳启发式的性能甚至能够接近于贝叶斯网络(一种高级神灵)的成绩吗?为了回答这个问题,我们考察了四个参赛者在两种任务上的成绩。四个参赛者是:采纳最佳启发式、朴素贝叶斯、贝叶斯网络和多元回归。两种任务是:拟合给定资料(训练项等同于测验项)和推论到新资料(训练项不等同于测验项,即交叉验证)。在推论任务上,从所有对象中随机抽取 50% 对象作为训练项,进行 10 000 次测试,然后用剩下的 50% 对象来检验根据训练项得到的模型。这两种任务类似于第 5 章中的两轮“竞赛”,因此我们用第 5 章所使用的 20 种环境来进行比赛。此外,当测验项就是训练项(拟合任务)时,用配置记忆法来确定其最佳成绩。竞赛的条件与第 5 章的两轮竞赛相同。

首先向每个参赛者提供训练样本,以便它们掌握线索和效标间的关系。例如,采纳最佳启发式利用训练样本了解线索效度的排列顺序;线性回归利用训练样本估计回归系数等。之后,用测验项来对每个参赛者估计成对对象效标值大小的能力加以测定。这些参赛者必须作出的推断如下:两个德国城市中哪个有较多人口?两个美国城市中哪个无家可归者占的比例较高?两条北极嘉鱼中哪个产卵较多?中西部大学两个教授中谁的收入较高?等等(关于任务的完整描述参见第5章)。

首先看一看在推断城市人口数量任务上得到的结果(表8-1)。当用训练项测定模型的成绩(拟合)时,采纳最佳启发式的准确性仅比贝叶斯网络低2个百分点。最好的成绩是用配置记忆法计算出来的,其推断准确性达80%,比贝叶斯网络的准确性高出4个百分点。当测验项不同于训练项(推论)时,理论上不存在最好的方法,但是参赛者之间可以相互比较。在这种情况下,多元回归取得的成绩比采纳最佳启发式和贝叶斯网络的成绩略低。

表8-1 不同算法规则在20种资料上的成绩

环 境	对 象 数	线 索 数	拟 合					推 论			
			PM	TTB	Reg	NB	BN	TTB	Reg	NB	BN
旧金山的臭氧水平	11	3	85	84	85	84	84	79	77	80	78
奶牛耗氧量	14	6	83	79	79	79	80	76	72	78	79
洛杉矶的氧化水平	17	4	93	84	84	84	84	80	76	81	82
死亡率	20	15	100	77	83	78	79	62	54	66	67
房价	22	10	96	86	86	87	87	84	68	86	86
雨量	24	6	71	67	71	68	68	53	56	57	59
物种差异	26	6	88	84	80	83	83	80	72	80	82
女性魅力	30	3	80	71	71	71	71	66	67	68	59
男性魅力	32	3	75	73	73	73	73	71	69	72	70
哺乳动物的睡眠	35	9	95	78	79	77	83	75	65	76	80
交通事故	37	13	93	71	79	75	75	64	64	71	71
18岁时的胖瘦情况	46	10	70	74	74	77	79	71	63	71	69
燃料消耗	48	6	87	78	79	78	80	73	74	76	76

环 境	对 象 数	线 索 数	拟 合					推 论			
			PM	TTB	Reg	NB	BN	TTB	Reg	NB	BN
无家可归者	50	6	82	69	70	68	77	63	62	64	65
教授薪水	51	5	87	80	83	80	84	80	80	80	81
高中失学率	57	18	90	65	72	65	65	60	54	61	60
地租	58	4	82	80	81	80	81	77	80	77	78
城市人口	83	9	80	74	74	74	76	72	71	72	74
身体发胖	218	14	87	59	61	80	82	56	55	79	80
鱼类生殖力	359	3	75	73	75	73	75	73	75	74	75
20 种环境的平均值			85	75	77	77	79	71	68	73	74

注：PM=配置记忆法，TTB=采纳最佳启发式，Reg=回归，NB=朴素贝叶斯，BN=贝叶斯网络

在推断无家可归者的比例时，拟合任务上贝叶斯网络取得的成绩比采纳最佳启发式高出 8 个百分点(表 8-1,“拟合”)。当测验项不同于训练项时，这个差距缩小到了 2 个百分点(表 8-1,“推论”)。

作为最后一个例子，我们看一看参赛者在预测中西部大学两个教授中谁的收入高时的成绩。拟合已知资料时，采纳最佳启发式的成绩比贝叶斯网络的成绩低 4 个百分点。当测验项不同于训练项时，其差距缩小到 1 个百分点，采纳最佳启发式几乎取得了匹等于贝叶斯网络的成绩。

183

对整套资料进行观测，可以得出以下结论：

1. 一般地说，当测验项与训练项相同(拟合)时，采纳最佳启发式在准确性上仅比贝叶斯网络低 4 个百分点。采纳最佳启发式只为它的快速性和节俭性付出了很小代价。
2. 一般地说，当测验项与训练项不同时(推论)，采纳最佳启发式在准确性上仅比贝叶斯网络低 3 个百分点。贝叶斯网络在解决小样本问题时可推论性较差，而采纳最佳启发式在整套资料上均相当耐用。

3. 朴素贝叶斯不愧为简单、快速和耐用的模型。像采纳最佳启发式一样(参见第6章),朴素贝叶斯也敢于在环境的结构上打赌:它赌定在给定效标值条件下线索是相互独立的。总体上看,在跨越20种环境的竞赛中,更加节俭的采纳最佳启发式的成绩几乎匹等于朴素贝叶斯模型的成绩。

上述一般结论是令人惊奇的。简捷的单一理由决策模型——如采纳最佳启发式——几乎像计算上相当繁琐的贝叶斯网络一样准确。更为重要的是,与神灵模型相比,采纳最佳启发式几乎不需要在快速节俭性和准确性之间进行任何权衡。

复杂性:神灵模型和启发式

在选择解决问题的方法时,充分考虑准确性与计算负荷之间的平衡是明智的。因此我们在对不同算法规则进行比较时,不仅要比较它们的准确性,还要比较使用这些规则时所需要的计算负荷。这种计算负荷既包括先前学习一种模型需要付出的努力,又包括把算法规则运用到具体情境时所付出的努力。

我们采用逐步分解的算法形式(而不是平行加工方式)形成了快速节俭启发式的计算模型。因此,当谈到一种算法规则的复杂性时,就意味着对实施这种算法规则涉及到的操作步骤数量加以计算。我们不妨从计算机科学的角度来看一看本书所涉及的各种算法规则的复杂性程度。在计算机科学中,复杂性被界定为完成所有必要操作步骤所需要的时间。

理论计算机科学提供了许多方法来测定算法规则的复杂性。一种可能性是数出一种算法规则所包含操作步骤的数目,再将这个数目作为算法规则所包含各种参数的函数来表示。这样,我们就能够确定该函数的复杂性程度了。经常使用的一种复杂性指标被称为“Landau 符号”,记为“ $O(\quad)$ ”,定义如下:当给出两个关于自然数的函数 $F(N)$ 和 $G(N)$ 时,如果存在一个常数 K ,且对于所

简捷启发式

有不为 0 的自然数 N 来说,都有:

$$\frac{F(N)}{G(N)} < K$$

那么,我们说 F 与 G 同阶,记为“ $O(G)$ ”。举例来说,函数 $F(N) = 3N + 1$ 与函数 $G(N) = N$ 同阶,但 N^3 与 N^2 不同阶。“ $O(1)$ ”意味着不因 N 的变化而变化(即不是 N 的函数)的恒定时间。由 N 构成的每个多项式,即 N 的多种幂的线性组合,都与其中 N 的最高幂同阶。由于数出来的是一个“连续”加工过程所涉及的步骤数量,所以通常也可以把得到的效标“ $O(\quad)$ ”理解为算法规则的“时间复杂性”。当涉及到一种以上变量时(如表 8-2 所示的情形),Landau 指标“ $O(\quad)$ ”允许作出一种直截了当的推论。在计算机科学中,通常要考虑“最糟情形复杂性”(worst-case complexity),正像这里我们马上也要做的那样。它有助于估计出实施算法规则可能花费的最长时间限度。我们当然知道“通常情形复杂性”(average-case complexity)是更有意义的,但它更加难以估计,需要作出关于问题偏向的假设。

185

对于以 N 个对象和 M 条线索为参照进行操作的算法规则来说,对所需操作步骤数目进行排序取决于上述两个参数。

确定算法规则的“ $O(\quad)$ ”复杂性的一种直观方法是算出实施该算法规则过程中每种输入成分必须被操纵多少次。例如,如果 N 个成分中的每一个都仅被观察一次,就像求一系列数据的平均值时所做的那样,那么时间复杂性为 $O(N)$ 。就采纳最佳启发式的运作程式来看, M 条线索必须先按照其效度加以归类。标准的计算机分类程序需要 $M \log M$ 个步骤。花费在每次运算上的实际时间是一个恒定因素,不影响对时间复杂性的测定。这种恒定因素对于决定实际成绩以及“问题大小”(the problem size)是非常重要的。所谓问题大小是指具有不同复杂性的两种算法规则之间发生交叉的次数,在其交叉处可以确定哪一种算法规则在实际中将花费更多时间。然而,不论恒定因素是什么,对于足够大的问题来

说,时间复杂性程度较高的算法规则运作起来将会比时间复杂性程度较低的算法规则更为缓慢。

当一种算法规则的复杂性是两种或两种以上度量的和时,那么其时间复杂性用其中较大的一个度量来界定。举例来说,假如有一种算法规则,它首先寻找一套分数相互间的距离,时间复杂性为 $O(N^2)$,然后按大小顺序列出一个分数表,其时间复杂性为 $O(N)$,那么这种算法规则最终的时间复杂性为其中较大的那一个,即 $O(N^2)$ 。

能够被称为“简单的”算法规则的时间复杂性应该落在 $O(N)$ 与 $O(N^3)$ 之间。例如,在一个列表中查找一个项目的时间复杂性为 $O(N)$,最优的分类算法规则的时间复杂性为 $O(N\log N)$,而利用列表中所有配对完成计算的时间复杂性为 $O(N^2)$ 。遗憾的是,为了得到最优化解决方案,常常要面对许多不同种类的问题,需要大量算法规则,以至于时间复杂性达到了 $O(2^N)$ 甚至 $O(N^N)$ 。许多旨在优化的设计和配置方案的复杂性都呈现这种“指数”形式。如果一种算法规则的复杂性在阶别上比 N 的任何幂都高,那么它显然在阶别上也要高于任何一个关于 N 的多项式(因为,根据有关定义,一个多项式与 N 的最高幂同阶)。这些算法规则被称为“NP-complete”,它的意思是“非多项式可完成的”算法规则。在实际中,如果 N 非常大,那么“非多项式可完成的”算法规则是难以实施的。在这种情况下,就需要一些可代偿的算法规则,它们允许在问题解决方案的质量和時間复杂性之间进行权衡。贝叶斯推断通常面临的问题都是“非多项式可完成的”问题(Cooper, 1990)。也就是说,当没有关于贝叶斯网络结构的任何限制时,在给定的线索值与多项式不同阶条件下推断哪个效标是较大的。

表 8-2 对几种较为复杂的算法规则和本书提出的几种启发式规则的时间复杂性作了总结。该表对于比较不同算法规则在最糟情形下的成绩是很有用的。我们将总体时间复杂性分成两个阶段:一个是“运作程式”,另一个是“决策”。然而,我们注意到两种

简捷启发式

具有相同最糟情形复杂性的算法规则在实际中却有很不相同的成绩(反之亦然)。我们还注意到,即便在要求尽可能保证算法规则认知合理性(正像对本书研究的某些算法规则提出的要求那样)的前提下,执行计算机程序的时间复杂性与个人完成有关任务所花费的时间长短也几乎不存在什么关系。从我们对最糟情形时间复杂性的分析可以得出的明确结论是:本书提出的某些快速节俭启发式即便在最糟情形复杂性这种相当严格的结构中也称得上是快速的。

表 8-2 算法规则的“最糟情形复杂性”

算 法 规 则	运 作 程 式	决 策
贝叶斯网络	结构搜索的 $O(2^M)$	$O(2^M)$
配置记忆法	存储的 $O(MN)$	$O(1)$
估计树	搜索的 $O(2^M)$	$O(N \log^N)$
朴素贝叶斯	效度的 $O(MN \log^N)$	$O(M)$
分类的样例模型	存储的 $O(MN)$	$O(MN)$
神经网络	$O(M^3 N)$	$O(N^2)$
多元回归	$O(M^3 N)$	$O(M)$
富兰克林规则	$O(MN \log^N)$	$O(M)$
达维斯规则	效度的 $O(MN \log^N)$	$O(M)$
采纳最佳启发式	效度和线索分类的 $O(\max(MN \log^N, M \log^M))$	$O(M)$
最快者策略	效度和线索分类的 $O(\max(MN \log^N, M \log^M))$	$O(M)$
排除分类法	线索分类的 $O(M \log^M)$	$O(M)$
最少化启发式	效度的 $O(MN \log^N)$	$O(M)$

注:这里的 N 代表对象的数量, M 代表给定环境中线索的数量。若采用第 6 章中的公式(1b),那么计算每条线索效度的复杂性为 $N \log^N$ 。由于通常情况下对象数量多于线索数量($N > M$),所以 $\log^M$ 通常小于 $N \log^N$,以至于采纳最佳启发式和“最快者策略”的运作程式复杂性刚好为 $MN \log^N$ 。配置记忆法对于说明机器和人类之间的差异的是一个绝好的例子:有些算法规则对于机器来说只有较低的复杂性,但对于人类头脑来说却是难以企及的(如,记住所有 MN 条线索的值),反之亦然。(“最快者策略”将在第 10 章以及第 11 章的“排除归类”部分介绍)。

结论

我们将人类生活看成是由一系列决策活动构成的,有些是重大的、有意义的,有些是较不重要的、无意义的,但每一种决策都被限定在一个相对较小的具体领域内。萨维奇(Savage, 1954, p. 83)曾提醒我们:现实生活中我们需要作出的决策只有一个,即我们应该怎样生活。换言之,我们必须选择一种进行推断和作出决策的生活策略。其他所有决策都遵循和来源于这种策略。

那么,我们如何进行选择呢?是采用规范的贝叶斯方法进行推断和作出决策,还是依靠本书提出的快速节俭启发式之类的简单策略呢?

这并不是一个相互排斥的二择一选择问题:快速节俭启发式在时间紧迫和知识匮乏的日常事务上扮演着重要角色,而贝叶斯方法对于那些时间充裕、并有机会接近计算机的人来说不失为一种明智选择(Winterfeldt & Edwards, 1986)。在作出性命攸关或者对环境和社会有长远影响的重大决策时,哪怕是仅仅为了多得到几个百分点的准确性而在计算上付出更多代价也是值得的。受限制条件下寻求最优化的贝叶斯模型的拥戴者必须在综合权衡多种效标(如,计算成本、准确性,也许还有模型的透明度)基础上选择一种策略。因此,在一些问题上,一个贝叶斯模型的拥戴者也许碰巧选择了采纳最佳启发式或其他一种快速节俭启发式,而不是更为复杂的贝叶斯网络。贝叶斯推理本身只是告诉我们何时进行选择。

本章报告的结果是根据现实资料获得的。然而,需要指出的是,这些研究结果的应用价值仍然是有限的。首先,我们研究的是事先已给出问题表征的情景,即:使用一套事先给定的线索值对成对对象加以比较。我们并未研究其表征由任务本身来构造(即根据任务确定适当线索)的情景。其次,我们的研究结果在应用上

的局限性还表现在：我们仅仅在知识完整条件下研究了推断问题，而没有像吉戈伦尔和古德斯腾(1996a, 参见第4章)那样，在知识有限条件下研究启发式规则的成效。有限知识(如，仅了解所有线索值中的一小部分)对于许多必须作出预测的情景而言都是一种非常现实的条件。值得强调的是，当并非所有线索值均为已知时，也可以运用贝叶斯网络。方程(3)给出的简单公式必须被一个更为复杂的推演算法规则所代替(Jensen, 1996)，但效标分布仍然可以被获得，它取决于已知线索是什么。最后，我们作为理性范本提出的贝叶斯网络模型本身只是近似于完整贝叶斯模型的一种启发式。它使用的是启发式搜索，而不是穷举所有可能性的贝叶斯网络，因此它选择了唯一的最佳模型，而不是用不可忽视的后验概率对所有模型加以平均。而且，该种算法规则假定各次观察是相互独立的，但我们的资料并未满足这个条件。这是因为我们所用的资料是由成对对象构成的，每一个对象都参与了多次配对，从而导致了数据间的相互依从性。我们使用了一种简单的、易于得到的算法规则，希望能为简单启发式提供一个好的基准。相信更为复杂的贝叶斯算法规则不久将会在有关文献中多起来。它们之间又将如何比较呢？我们尚不得而知。

总之，本章强调了几个相关要点。首先，我们强调了贝叶斯模型作为评价快速节俭启发式之基准的重要作用。其次，在给定了启发式算法规则和基准性算法规则之后，我们就能够考查运算成本和准确性之间的权衡问题。这使得我们能够发展一些策略，以确定对于特定种类问题而言哪种解决方案是最为适当的。有人甚至主张在“元水平”上，即在选择大的方略——它可以被用来指导为特定种类问题选择具体方法——水平上推广贝叶斯学派的理性梦想。因此，在必须考虑计算时间和成本的决策情形中，用快速节俭启发式代替最优化策略，可能会引发“为决策或是为理论”之争。在对20种现实环境作出预测时，我们发现采纳最佳启发式几乎像贝叶斯网络一样准确，而且节俭性和准确性之间的权衡几乎可以

忽略不计。采纳最佳启发式是一种快速节俭策略，它在下述情况下是十分有用的：仅牺牲少量准确性便可节省大量计算成本。

附录

沃德·爱德华兹(Ward Edwards) 他对将采纳最佳启发式与贝叶斯模型加以比较的做法提出了疑义——以一种别出心裁的方式对表 8-1 中的结果作出了反应。下面是他的一首很有创意的打油诗：

要不要贝叶斯？

188

一首打油诗

哥德和他的朋友们这样认为：

采纳最佳能做出理想决断。

“让我们对采纳最佳进行严格检验，
以免遭学院派的讪笑。”

于是他们让采纳最佳与贝叶斯较量。

他们的发现让我困惑。

如果你是一个能干的人

你知道贝叶斯做得更好，

因为我们都知悉贝叶斯付出了。

但采纳最佳做得也很好。

尽管只用了一条线索。

如果事无巨细，样样考虑，

将会令你疲惫不堪，心力交瘁。

贝叶斯和采纳最佳仅打成了平手。

简捷启发式

采纳最佳令他们折服。
它甚至挤压贝叶斯。
其中的道理十分清楚：
你可无所畏惧地去行动，
做你喜欢做的一切。

沃德·爱德华兹

1997 年 9 月



第四部分

超越选择：记忆、估计和分类

LAZARUS

SMART

第9章

事后通偏向 ——值得为快速节俭记忆付出的代价

奥瑞齐·霍夫瑞格(Ulrich Hoffrage)

拉尔夫·赫特威格(Ralph Hertwig)

记忆不是无数固定不变、毫无活力的不连贯痕迹的重新激活。
它是富有想像力的建构或重新建构过程……

——弗雷德里克·巴特莱特爵士(Sir Frederic Bartlett)

多数人都曾为容易出错的记忆而烦恼：“但是凡人毕竟是凡人；最美好的事情有时会被遗忘。”(Shakespeare, Othello)对过往事件的回忆并不只是从仓库中把有关经验提取出来，仿佛从图书馆中查找书籍一样。记忆中的经验可能会丧失或失真，而且对从未发生过的子虚乌有事件的“记忆”也常常会在人们的头脑中闪现出来（例如，Loftus, 1997；Schacter, 1995）。与拉普拉斯们所谓的神灵拥有的记忆不同——我们不能够完整无缺地回忆出我们曾经思考过、说过或经历过的每一件事情。本书前面的有关章节已经讨论过有限时间和知识对推断活动造成的限制，本章着重探讨人类记忆的有限容量所造成的限制。既然记忆容量是有限的，它怎样才能生效呢？我们的答案是借助于“重建”(reconstruction)：当提取失败时，利用简捷启发式进行推断。这个答案一点也不新奇。它早已为弗雷德里克·巴特莱特爵士——现代记忆研究的先驱——所提出。在其经典性著作《记忆》一书(1932/1995, p. 213)中，

简捷启发式

巴特莱特强调指出：记忆是一个重建过程（参见——或者回想一下——本章开头引用的警句）。

然而，重建是有代价的。我们着重讨论其中的一种代价，即著名的“事后通偏向”（hindsight bias），并依据快速节俭启发式提出该种效应的计算模型。通常认为，事后通偏向只是人类信息加工活动的一种特殊错误而已。相反，我们指出，它是两种一般适应过程的一种“副产品”（by-product）：首先，人们接受新信息后将会对已有知识进行更新；其次，人们将会根据已被更新的知识作出快速节俭的推断。在详细描述该种模型之前，让我们首先以现代民主社会中每个公民均十分关心的一个话题——公众民意测验和选举结果——为例，来说明事后通偏向。

192

公众民意测验、选举和事后通偏向

美国政治性民意测验的历史与盖洛普（George Gallup）的名字紧密联系在一起。盖洛普相信他关于直接民主选举的理想需要来自公众的信息和不受金融财团影响的政策评估（Hamilton, 1995）。20世纪30年代早期，在其母亲作为候选人之一的依阿华州的一次地区性选举中，他通过民意测验这种手段实现了直接诉诸选民的理想。此后不久，他便开始运用这种技术去对选举结果进行预测。自从盖洛普开了民意测验的先河以来，民意测验已经逐渐“发展为美国竞选活动的中心”（Hamilton, 1995）。例如，盖洛普和哈里斯（Harris）——民意测验业的两个大亨——曾经引起了里查德·尼克松的兴趣，并成为他的行政班子操纵或攻击的主要候选对象（Jacobs & Shaapiro, 1996）。将民意测验作为一种政治工具接受下来并没有使其摆脱人们的批评。人们一再指出，对公众意见的测验并未产生积极的政治效果，而是被一些政治家们用来捏造他们所需要的所谓公众态度（参见 Jacobs & Shapire, 1996）。

民意测验公司的“阿基里斯之足”(Achilles heel)* 在于:公众能够反过来对其预测准确性加以检查或核对。如果它们的预测是准确的,那么将不存在什么问题。事实上,就盖洛普公司来说,对1997年英国议会选举的预测就是一个相当成功的例子。在由《今日电讯》(*The Daily Telegraph*)发起的一项民意测验中,盖洛普对选举结果的预测达到了近乎完美的程度。根据选举前一天和随机抽取的1810名有资格参与投票的选民进行的谈话,盖洛普预测工党将以超过保守党13%的优势取得选举胜利。这次具有历史意义的选举——它结束了保守党长达18年的统治地位——的实际结果是:工党以领先保守党14%的优势取得了胜利。

然而,这种事后真实性检验也常常置民意测验公司于非常困窘的境地。盖洛普和其他人的一次最为著名的预测“失误”是1948年美国总统选举中杜鲁门的意外获胜。在1948年11月份的最初几天里,几乎每个人都以为托马斯·杜威将在即将来临的总统选举中击败哈里·杜鲁门。民意测验专家和职业政治家也作出了同样预测。美国发行的各种日报以8对1的比例倾向于赞同杜威。为了抢到一条独家新闻,11月4日出版的《芝加哥每日论坛》(*Chicago Daily Tribune*)甚至以盖洛普和其他民意测验公司似乎合理的预测为依据,率先宣布杜威将是下一任总统(Hamilton, 1995)。(你也许还记得当年的那张照片:当选总统杜鲁门不无嘲讽地高举着这张报纸,上面有一个醒目的标题“杜威击败了杜鲁门!”)。

面对这种窘境,为了挽回面子,民意测验专家们通常喜欢说的一句话是:“我们早知如此——这才是我们真正预测的结果。”但是报纸、录像带以及其他媒体所记载下来的事实是不容民意测验专家们要滑头的——他们必须信守过去的预测。然而,日常生活中

* 阿基里斯(或译阿喀琉斯),传说中的希腊之神,出生后被其母倒提着在冥河水中浸过,除未浸到水的脚踝外,浑身刀枪不入。“阿基里斯之足”常用来比喻“唯一致命的弱点”。

简捷启发式

人们的谈话(或者只是“想想”而已)通常仅引起他们个人内部记忆的误差——不存在使他们感到困窘和与其记忆相抵触的外部记录。这可能导致这样一种情况:个体对其过去所作预测或声明形成了不准确记忆。例如,乔大叔(Uncle Joe)也许声称他早已经知道杜鲁门将会赢得选举,尽管他早先一直相信杜威将会入主白宫的。这种事后自欺欺人地相信自己早已正确地预测了事件结果的倾向通常被称为“事后通偏向”(关于重建过去事件时其他记忆失真现象的更为系统的阐述,请参见 Johnson & Sherman, 1990)。

近期的心理学实验研究表明:事后通偏向在普通人和专家(如,选民、医生、商人等)中都是普遍存在的,而且显然与多种判断活动(如,信心判断、选择、分类或数量估计等;相关的评论参见 Hawkins & Hastie, 1990)有关。毫不奇怪,这种现象在预测政治选举结果时也发生了。例如,1982年夏威夷州选举州长之前,西诺第诺斯(Synodinos, 1986)在其进行的一项研究中要求被试指出每个候选人赢得选举的可能性。选举结果公布后,要求另一组被试仿佛他们是在选举前接受测验那样作出同样预测。正像所预期的那样,被试表现出了明显的“早知如此”的倾向:选举后对获胜者赢得选举可能性的估计远远高于选举前的估计,而选举后对两名落选者赢得选举可能性的估计则远远低于选举前的估计。

西诺第诺斯(1986)通过比较两个不同被试组的预测证明了了解有关结果所起的作用。在一个单独的被试群体中也能够发现事后通偏向。例如,1972年尼克松总统访问北京和莫斯科前夕,费什霍夫和贝斯(Fischhoff & Beyth, 1975)曾经让一组大学生被试对此次访问可能产生的各种后果作出判断。各种不同后果以简短声明方式提供给被试,如:“美国将在北京设立永久外交使团,但不会给予其外交上的承认”和“尼克松总统将至少会见毛泽东主席一次”,要求被试在一个0%到100%的量表上表明他们对这些声明的信心程度。在尼克松总统访问之后,将这些声明重新提供给被试,要求他们对原先所作判断进行回忆。在这项研究中,被试也表

现出了事后通偏向：对他们知道已经发生了的事件的回忆成绩高于原先的判断成绩，而对于他们知道未曾发生事件的回忆成绩则低于原先的判断成绩。

关于事后通偏向的一些观点

人们已经从许多不同角度来理解事后通偏向。在此我们区分出了两类理解，并加上了我们的第三种理解。费什霍夫(1975)——其早期实验研究开了记忆领域事后通偏向研究的先河——强调指出：事后通偏向不仅是一种冥顽不化、难以排除的现象，而且具有潜在危害性：194

当我们试图理解过往事件时，我们会不经意间对我们通常用来理解和预测周围世界的假设或规则进行检验。如果事后回顾过往事件时我们系统地低估了它们曾经给我们造成的诧异之感，我们就使这些假设或规则接受了非常‘无力’的检验，结果是很可能找不到改变它们的理由。因此，正是对结果的了解本身——它使得我们自以为对过往事件了如指掌——使得我们不能够从中了解任何东西。(Fischhoff, 1982b, p. 343)

与上述强调事后通偏向有害性的观点截然不同，另外一些研究者(例如，Campbell & Tesser, 1983)指出了它的潜在适应价值。事后将我们自己标榜为智者使得我们显得更加聪慧、博学和有远见。事实上，只要不存在我们先前所作判断的客观记录(情况通常都是如此，不像对于民意测验专家那样)，以博学者自居所获得的直接利益都将远远超过被揭露为滥竽充数的骗子(这种情况不大可能发生)时可能付出的代价。除了在社会交往中有其益处之外，事后通偏向在建立并维持一个人自我概念同一性方面也具有重要作用。例如，当人们突然发现他们自己置身于一个价值系统完全

简捷启发式

改变了的社会中时,情况便是如此。20世纪90年代世界上发生了一系列急剧的社会变革,从南非种族隔离政权的垮台,到苏联和东德社会主义制度的结束。在过去政权中身居要职者的家人和朋友常常要求这些人对其过去的所作所为加以说明,或者某个正式团体(如,南非的“真相和调停委员会”)会对这些人进行审问,责令其说明过去的行为。在这些情况下,事后通偏向——即相信自己过去的信念和行为与新政权视为正确的信念和行为相一致——可能是维持个体人格(或许只是外表)同一性的有效途径。

我们提出了第三种观点(它并不排斥上述两种观点)。按照这种观点,事后通偏向只是适应过程的副产品,而不是适应过程本身(关于该种观点的更一般论述,参见 Campbell, 1959)。为了说明这种观点,我们首先提出这样一个问题:可用来代替人类记忆容量无限假说的理论究竟是什么?

不妨考虑如下情形:劳曼先生是一位推销商,经常开车去拜访他的顾客。每天他都要决定在哪儿停靠车辆,然后将这一信息存贮在记忆中,结束商业会谈后,再从记忆中提取出停车地点。日复一日、年复一年,他都要如此这般地重复许多次。如何设置记忆系统才能保证劳曼先生准确、迅速地提取出关于最近停车地点的知识呢?假如有这么一个记忆系统,它保持了关于所有过去停车地点的意识,那么这种记忆系统是足够有效的吗?当前一些有关人类记忆的观点似乎假定,我们实际上的确对曾经经历过的每一种具体事件都做了记录,当我们提取信息或对某个对象进行归类时,我们将其与所有现存记忆痕迹进行比较。例如,记忆的“范例模型”(exemplar models;例如,Estes, 1986; Hintzman, 1988; Medin & Schaff, 1978; Nodofsky, 1986; Ratcliff, 1978)就是建立在此种假设基础之上的。尽管这些模型为广泛领域的多种记忆现象提供了强有力的解释和说明,但是其在心理学上的合理性仍然受到怀疑,因为它们一方面需要付出巨大计算成本在众多相似对象之间作出审慎区分,另一方面又需要巨大记忆资源来储存有关信

195

息(Nosofsky et al., 1994; 亦可参见第 11 章的有关论述)。

带着这些疑问,我们完全同意安德森和斯库勒(Anderson & Schooler, 1991)的如下观点:“保持对无数项目的意识是不切实际的过高奢望”。(p. 396)此外,库存记忆信息(如,劳曼先生对以前所有停车地点的记忆)可能会干扰仅与当前行为有关的信息(如,目前劳曼先生的车停在何处)的提取。从这种意义上说,遗忘对于维持记忆功能是必要的,原因在于它避免了我们使用过时的旧信息,而更可能采用得到更新的新信息(Bjork, 1978; Ginzburg et al., 1996)。能够充分反映遗忘适应本质的一种著名现象是“柴格尼克效应”。柴格尼克(Zeigarnic, 1927)证明:相对于未被完成的任务而言,对已被完成任务的记忆迅速地衰退了(如,一个服务员对账单上钱数的记忆取决于顾客是否已付账)。因此,当某些信息不再有用时,遗忘最有可能发生。

可用来代替无边无际、连续延伸长时记忆存贮系统的是这样一个系统:它主要保持对最可能需要和最可能正确信息的意识。对于这样一个记忆系统而言,经常地、自动地更新知识是非常关键的。这一过程避免了项目过多的问题,同时也避免了以系列方式将新刺激与存贮的记忆痕迹进行比较时需时过多的问题。它使理性有限的记忆系统成为可能,该系统仅使最有可能需要的项目处于可调用状态。这种信息更新过程与巴特莱特(1932/1995)的经典发现——图式是不断变化和更新的——是完全一致的。

除了人们的多数经历在以后不需要回忆这一事实之外(Anderson & Schooler, 1991),我们所以没有必要为过去曾经想过、说过、经历过的每件事情保留一种记忆痕迹,还有另外一个原因:当需要回忆某些事件时,可以采取其他可供选用的记忆提取方法。举例来说,假设你在某公司拥有 25 股股票,报纸上登出每股股票的价格为 378.50 美元。为了算出这些股票的总面额,你用 378.5 乘以 25。几天以后,你想再次了解它们的价值。你还能记得它是 9 462.50 美元吗?也许不能。然而,这不是一个问题,因为你可以通过重复相同

简捷启发式

计算来弥补不能直接从记忆中提取它的不足：回忆可以用重新计算来代替。我们假定：当过去的判断（如，先前对竞选结果的预测）需要回忆时，可以做、且事实上也的确做了与上述过程相似的重新计算。如果不能够回忆起过去的某个判断，那么返回到导致原始判断的相同过程能够作出一个好的近似判断，甚至提供一个完美的替代判断。然而，在上述乘法运算和真实判断之间有一个重要区别。完成数学运算是一项专门技巧，我们都受过训练，能够可靠地完成它。因此，进行第二次相同运算将产生完全相同的结果。相反，作出一种判断常常意味着以知识为基础进行推论。如果像上面所指出的那样，知识经常地被更新，那么以更新后的知识为基础作出的推论也许不同于以过去知识为基础作出的推论。

知识更新是下面提出的事后通偏向模型所依据的关键假设。该模型适用于如下情形：原始判断是以知识为基础作出的推论。如果直接回忆原始判断的尝试失败了，那么便可通过重复导致该判断的相同过程来重建这个判断。然而，有关事件结果的知识，或有关某种推断是否正确的反馈信息，导致了相关知识的更新。结果是，以更新后的知识为基础的重建可能会系统地偏离以原始知识为基础的判断。这种偏离导致了我们所谓的事后通偏向。因此，在我们看来，所谓事后通偏向是一种适应过程——知识更新过程——的副产品。

以前我们曾提出过一种模型，用来说明事后通偏向研究中一种令人迷惑不解的现象，即：如果对一种断言的反馈为“真”，那么人们表现出来的事后通偏向非常明显；相反，如果对一种断言的反馈为“假”，那么人们的事后通偏向不太明显（Hertwig et al., 1997）。那种模型把上述现象解释为事后通偏向和“重复效应”（the reiteration effect）共同发生作用的结果。所谓重复效应是指这样一种现象：一种断言重复多次将会增强人们对其正确性的信心。然而，那个模型并没有对为什么会出现事后通偏向作出解释。目前的模型做到了这一点。尽管它并非唯一一种“借助于对已有

结果的‘重新估价’重建先前判断”(Hawkins & Hastie, 1990, p. 321)来解释事后通偏向的理论,但是,公正地说,我们的模型是将基于知识的推断过程详尽化的唯一模型。它允许我们在个别被试的个别反应水平上说明事后通偏向为什么发生、不发生,甚至产生逆转。

快速节俭地推论过去的判断

当人们试图重建他们的原始判断时,他们经历了什么样的过程呢?我们认为,提出这一问题等同于提出如下问题:潜在于原始判断后面的过程是什么?“概率心理模型”(probabilistic mental model,缩写为PMM;Gigerenzer et al., 1991)理论为这个问题提供了一个答案。它适用于这样一些任务:必须依据一种量化效标在两个选项间作出一个选择,并作出关于被选择选项为正确的信心判断。(此类任务的一个典型例子是问被试:“哪一个城市有更多居民:海德堡还是波恩?”“你对所选择城市为正确的信心有多大?”)现在我们将概率心理模型引入到已给出正确答案反馈信息,需要在头脑中重建原始反应(包括选择和信心判断)的情景中来。我们将这一模型称为“利用采纳最佳启发式的反馈后重建”(reconstruction after feedback with take the best,缩写为RAFT)。

原始反应

一个具体的例子有助于说明这种任务的性质和潜在于其背后的机制:我们有一个朋友,名字叫潘多莎(Patricia),她来自于美国的南加利福尼亚州,一直想要降低她的胆固醇摄入量。然而,她偏偏又喜欢甜食,在餐馆里想点一份甜点:巧克力牛奶蛋糕,或者南瓜鸡蛋馅饼。她想弄清楚这两种食物中哪一种的胆固醇含量更高(目的是选择含量较低的)。由于潘多莎不知道正确答案,所以她试图根据自己有关这两种食品的知识进行推论。我们假定:为了

简捷启发式

进行推论,她将建立一个概率心理模型。该模型包括参照评定等级、或然线索、利用这些线索获得的有关参照等级中各个对象的知识以及用来加工这些知识的启发式规则。

关于线索的知识 按照概率心理模型理论,知识被界定为一套线索(如含有的脂肪量)和这些线索在有关对象(食物)上所具有的值。当比较两种食物的线索值时,就量化线索来说,有四种可能关系:“较大”(如,蛋糕比馅饼含有更多脂肪)、“较小”、“相同”或“未知”。下文中我们将这些关系称作“对象关系”(object relations)。[请注意:关于对象关系仅有一种直觉(无论正确或错误)已经足够了;至于这些关系是直接从记忆中提取出来的,还是从绝对的线索值推断出来的尚待探索。]

198

概率心理模型还假定:人们具有关于一条线索预测力的直觉。线索预测力可以用它的生态效度来测定。生态效度被界定为在给定参照等级中一条线索正确预测哪个对象在效标上得分较高的相对频率(第4章)。生态效度可以仅仅通过比较彼此不同的线索(即对象关系为“较大”或“较小”)来决定。让我们假定潘多莎的参照等级由她从当地超市得到的食品所构成,并将饱和脂肪作为胆固醇的量化线索。当我们从一个超市随机抽取36种食品作为样本,并双配对进行检测时,我们发现在大约80%的配对检测中饱和脂肪(线索)较高的食品也含有较多胆固醇(效标),那么饱和脂肪线索的生态效度就是大约80%(在我们选取的超市样本中)。

启发式 潘多莎如何运用这种知识来推论哪种食物含有较高胆固醇呢?我们用在概率心理模型中被称为“采纳最佳”(Gigerenzer & Goldstern, 1996a;亦可参见第4章)的启发式规则来说明她的推论过程。如果两种食物均为已知的,那么采纳最佳启发式首先按照线索生态效度估计出它们的顺序,并依据区分两种食物时排列等次最高的(最佳的)线索作出推论。假定潘多莎的概率心理模型包含了三条线索:饱和脂肪、热量和蛋白质含量,而且已经按照其生态效度(分别为80%、70%、60%)排出了顺序。她原有的关于在这些线

索上蛋糕和馅饼之间关系的心理模型示于图 9-1(原始反应栏)。排列等次最高线索(饱和脂肪)未能将两个对象区别开来,于是采纳最佳启发式尝试第二条线索(热量)。由于蛋糕比馅饼含有更多热量,所以该种启发式终止搜索,选择蛋糕作为含有更多胆固醇的对象。对该决策为正确的信心判断就是导致该决策的那条线索的效度(这里为 70%,即热量线索的生态效度)。

这种快速节俭启发式的一个明显特点在于其简单的终止规则:当发现一个选项优于另一个选项的第一条好理由时,立即终止搜索。此后,无需搜索任何其他线索,也无需进行任何代价—收益分析,以尝试确定“最佳的”记忆搜索终止点。这样一种简单终止规则对于基于记忆的推论是十分重要的,因为此种推论需要花费大量时间和努力在记忆谜团中搜索信息。在下面所报告的研究中,我们仅利用三条线索(示于图 9-1 中的三条线索)来对被试进

	信心水平上的 事后通偏向			
	原始反应		回忆反应	
	蛋糕	馅饼	蛋糕	馅饼
饱和脂肪 (80%)	?		>	
热量 (70%)	>		>	
蛋白质 (60%)	>		>	
选择	蛋糕		蛋糕	
自信心	70%		80%	

图 9-1 在信心水平上的事后通偏向。概率心理模型包含了按照线索效度排列的三条线索(括号内给出了其具体值)。“>”和“?”表示在这些线索上两个对象之间的关系。例如,在左边的一栏中(它表示原始反应所依据的知识),饱和脂肪线索的对象关系为“未知”。正如箭头所表示的那样,得到反馈后对象关系改变了:蛋糕比馅饼含有更多胆固醇。在更新后的心理模型中(右栏),对象关系向反馈方向转化,即从“?”变成了“>”。结果是事后通偏向发生了。请注意:采纳最佳启发式在到达阴影所示的对象关系之前就终止了对线索的搜索。

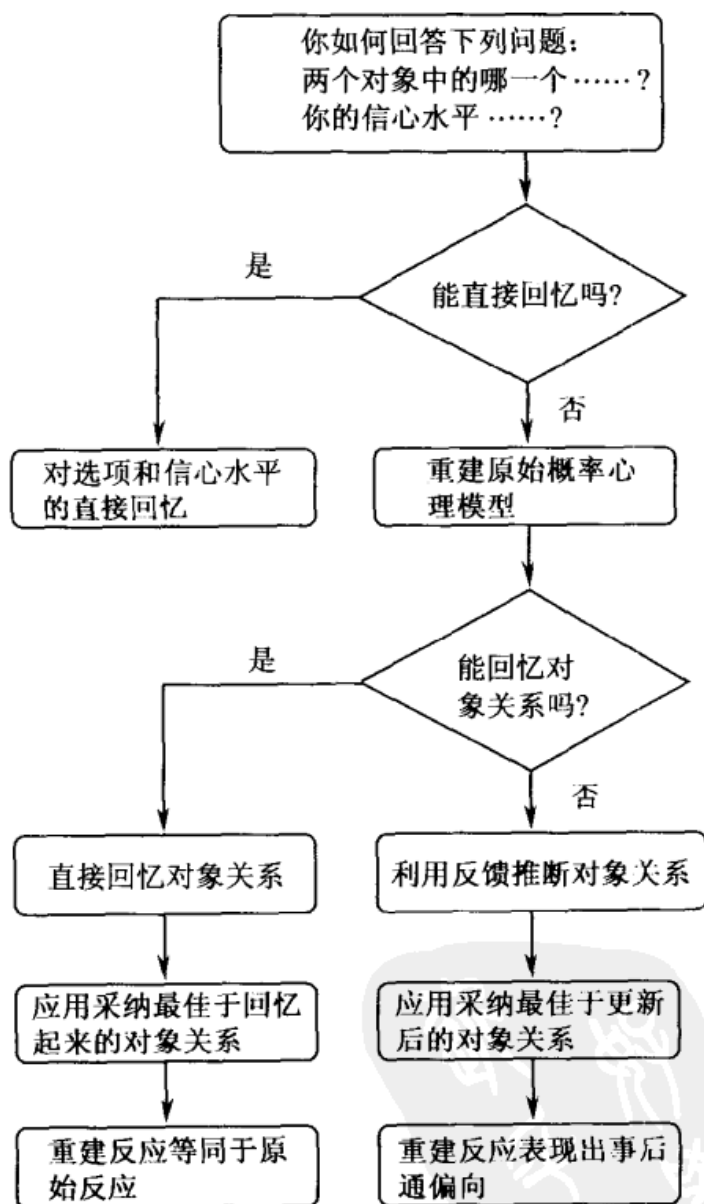
行训练,这样便人为地限制了搜索活动。在现实世界中,关于食品胆固醇含量的推论,通常会有更多适用线索,因此搜索活动将远远超出少量线索而延续下去,除非用终止规则来结束它。

反馈和重建

199

在餐馆用过晚餐后,又过了几周,潘多莎来到市场,发现巧克力奶油蛋糕比南瓜鸡蛋馅饼含有更多胆固醇。她力图回忆起她过去所做的选择。那么,回忆过去反应的潜在机制是什么?图9-2说明了由RAFT模型假定的认知过程。首先,试图直接从记忆中提取原始反应。成功地做到这一点的几率取决于原始判断和回忆之间的时间间隔(Fischhoff & Beyth, 1975; Hertwig, 1996)以及原始反应的编码深度(Hell et al., 1988)等因素。如果原始反应能如实地从记忆中直接回忆出来,将不会产生事后通偏向(图9-2中的左上框)。

如果原始反应



200

图9-2 事后通偏向的RAFT模型流程图

不能从记忆中提取出来,就会试图重建导致这一反应的原始概率心理模型。在下列各项条件均得到满足情况下,将会获得一种完全等同于原始反应的重建:(a) 原始反应与它的重建过程采用了相同的策略类型(如,词典编纂策略、线性模型、神经网络,或贝叶斯网络);(b) 这种策略利用了相同参数(如,相同线索排列次序、权重或概率)进行运算;(c) 这种策略使用了相同线索;(d) 提取出来的这些线索的值相同。违背了这些条件中的任何一个都会导致原始反应与重建反应之间的差异。事实上,RAFT 主要强调条件(d)的变化,即潜在于原始反应和重建反应背后的线索值的系统差异。它并没有排除违背条件(a)、(b)、(c)的可能性,实际上也的确有人用此种方式来解释事后通偏向(例如,Hawkins & Hastie, 1990)。尽管如此,我们指出并提供证据表明,对条件(d)的违背已足以解释事后通偏向。

知识更新 为什么反馈后在有关线索上的对象关系会发生变化呢?通常,有一条以上线索可用来推断效标,因此,如果有关一条线索的信息无法得到,那么该条线索就会被其他线索所替代。艾冈·布伦斯维克(Egon Brunswik, 1952)称这一现象为“代偿机能”(vicarious functioning)。而且,不仅线索是可以互换的,而且线索和效标也可以相互转换。许多情况下,从一条线索到一种效标作出推论的概率也可以被逆转。例如,你不仅可以根据饱和脂肪含量推论胆固醇含量,而且可以根据胆固醇含量来推论饱和脂肪含量。设想你既不知道巧克力奶油蛋糕中的饱和脂肪含量,又不知道它的胆固醇含量。如果现在你了解到蛋糕中含有大量饱和脂肪,那么你就可以把它作为线索来推论蛋糕中也含有大量胆固醇。类似地,如果你获悉蛋糕内含有大量胆固醇,你就可以利用这一线索推论它也含有很高的饱和脂肪。因此,有关效标的新信息能够被用来更新语义记忆中的相关知识——同样也可以被用来更新情景记忆中的过时信息(如同在上述劳曼先生停车地点的事例中那样;亦可参见 Bjork, 1978)。知识更新是具有适应价值的:它

简捷启发式

增强了我们知识的连贯性和推论的准确性(因为更为近期的信息通常更为有效和相关)。

因此,我们的假设是:记忆中储存的知识处于一种流动状态,常常发生变化,一方面因为新信息被获得了,另一方面因为与这种新信息相关的知识被更新了。下面,我们来说明知识随时间发生的这些变化是如何导致我们预测和解释事后通偏向的(如同图 9-2 右半部分的底部所表明的那样)。

预测

同一种变量既可以作为一种效标,又可以作为一条线索,这一事实为我们提供了一个非常有趣的分析问题的角度。在餐馆里,潘多莎担心胆固醇含量过高,她的原始心理模型将饱和脂肪作为一条线索(为了推论胆固醇含量)。在她试图重建原始心理模型时,饱和脂肪——或者更确切地说,当她在餐馆时,她所具有的关于饱和脂肪的知识——变成了效标。在一些情况下,如实地提取出过去线索和当前效标是可能的,而在另一些情况下它也许是不可能的。作为对此种记忆缺失的替代,潘多莎可以运用她现在拥有的知识。然而,在她发现两种食物哪一种含有更多胆固醇的同时,该种新信息也许会导致相关知识(如有关饱和脂肪的知识)的更新。作为效标——线索切换的必然结果,更新后心理模型中的知识将会表现出朝向反馈的系统转移(预测 1)。由于 RAFT 假定知识更新的发生具有一些偶然性,所以这种预测并不一定适合于每一个单独项目。

按照 RAFT 模型,关于线索知识的系统变化能够解释回忆先前选择和信心判断时所出现的系统改变。也就是说,事后通偏向的产生以重建后的知识为转移。排除了原始反应与回忆出来的反应完全相同(因此可以被归因于直接回忆)的情形之后,RAFT 应该能够对个别回忆作出解释:不论是否观察到了事后通偏向或逆转的事后通偏向,这种观察都应该符合 RAFT 的预测——它导源于为该项目回忆出来的对象关系(预测 2)。

对 RAFT 模型的说明

现在我们来说明 RAFT 是如何解释由事后通偏向形成的回忆的。当潘多莎在餐馆里试图推断两种食品哪一种含有更多胆固醇时,她并不知道饱和脂肪线索(即最有效线索,参见图 9-1)的值。而在她发现蛋糕比馅饼含有更多胆固醇后,这个值已经被更新了。结果是,当潘多莎力图回忆原始判断时,在事后通偏向影响下,饱和脂肪线索——在餐馆时不可得到的线索——发生了变化,于是潘多莎推断:她早就认为蛋糕含有更多的胆固醇。她还推断她作出这种选择的信心程度为 80%(饱和脂肪线索的效度)。这样,她重建的选择就等同于原始选择。但是,相对于原始信心程度来说,她重建的信心程度却增强了。这是事后通偏向的表现形式之一。更一般地说,如果一种选择被正确地回忆出来,但在获悉原始选择正确时,回忆起来的信心程度增加了(或在获悉原始选择为错误时,回忆起来的信心程度降低了),那么在信心程度水平上事后通偏向就产生了。

不仅回忆起来的信心程度不同于原始信心程度,而且回忆起来的选择也会不同于原始选择。如果有关反馈信息表明原始选择是错误的,而回忆起来的选择是正确的(如,原始选择是馅饼,反馈信息表明蛋糕才是正确选择,且回忆起来的选项恰好也是蛋糕),那么在选择水平上事后通偏向便产生了。RAFT 同样能够解释这种事后通偏向。图 9-3(板块 A)提供了一个例子:在餐馆里,只有蛋白质线索能够区分两种食品,且指向馅饼。然而,反馈后饱和脂肪线索能够区分二者,且指向蛋糕。如果原始选择根据这种更新后的知识得到重建,那么 RAFT 预测:将在选择水平上产生事后通偏向。

我们还有望获得逆转的事后通偏向。如果反馈信息证实原始选择是正确的,但回忆起来的选择是错误的,逆转的事后通偏向便在选择水平上产生了;如果回忆起来的选择与原始选择相同,但反

简捷启发式

反馈信息证实了原始选择时信心程度降低了,或反馈信息否定了原始选择时信心程度增强了,那么逆转的事后通偏向便在信心水平上产生了。RAFT 如何解释这种逆转的事后通偏向呢?它假定对象关系可能会发生一些随机变化,以此来解释上述现象。也就是说,除了反馈信息引起的系统变化外,RAFT 假定个人知识的不完善性也会引起非系统的随机变化。此种随机变化不受有关反馈信息的影响。由于这个原因,这些变化可能与反馈信息的方向相一致,也可能与其相抵触。在图 9-3(板块 B)中,一种随机变化改变了在饱和脂肪线索上的对象关系,使其背离了反馈信息指明的方向。此种随机变化导致了在信心水平上逆转的事后通偏向。

总之,我们以有关人类记忆容量有限的发现为基点提出了自己的观点。可用来代替不受限制记忆系统的是这样一个系统:它只维持对最可能需要和最可能正确信息的意识。对于这样一个系统而言,经常地、自动地更新信息是至关重要的。我们指出,事后通偏向是这种适应过程(信息更新)的一种副产品。RAFT 假定人们在根据更新后的知识重建过去判断时采用了一种快速节俭启发式,从而对事后通偏向作出了解释。

204

	板块A: 选择水平上的 事后通偏向				板块B: 信心水平上的逆 转事后通偏向			
	原始反应		回忆反应		原始反应		回忆反应	
	蛋糕	馅饼	蛋糕	馅饼	蛋糕	馅饼	蛋糕	馅饼
饱和脂肪 (80%)	?	→	>		>	→	?	
热量 (70%)	?		?		>		>	
蛋白质含量 (60%)	<	→	?		>		>	
选择	馅饼		蛋糕		蛋糕		蛋糕	
信心水平	60%		80%		80%		70%	

图 9-3 选择水平上的事后通偏向(板块 A)和信心水平上的逆转事后通偏向(板块 B)。有关符号的解释参见图 9-1。

实验证据

我们做了一项研究,用来对预测 1 和预测 2 进行检验(参见 Hoffrage et al., 1999)。在这项实验的起始阶段,被试了解多种食物中饱和脂肪、热量和蛋白质三条线索的值。他们还被告知了这三条线索在推断哪种食物含有更多胆固醇方面的效度(分别为 80%、70%和 60%)。学习阶段结束后,立即给被试一个食品配对表,并就每一种配对提出两个问题:“你认为哪种食物中含有更多胆固醇?”和“你对所做选择为正确的信心有多大?”(信心程度在 50%到 100%间。)被试作出他们的反应后,要求他们回忆他们记得的每一种食物中饱和脂肪、热量、蛋白质的含量,或者让他们指出每对食物在每条线索上的相互关系(这是反馈前他们具有的知识)。

在实验的第二阶段,实验组被试首先得到他们在第一阶段回答过的每个问题的正确答案(反馈)。在控制条件下,不向被试提供这种反馈信息。然后要求所有被试回忆:(a) 他们原来选择哪种食物作为含有更多胆固醇的食物;(b) 当时他们在多大程度上相信他们的选择是正确的;并增加了一个新问题(c): 原来所知道的线索值或在有关线索上食物之间的关系(这是反馈后他们具有的知识)。这里记录被试的知识是相当重要的,因为按照 RAFT,事后通偏向的产生依赖于这种知识。

我们首先看一看被试是否表现出了事后通偏向。对于正确选择而言,如果回忆时信心程度增加了(相对于原始信心程度),就说明产生了事后通偏向;对于错误选择而言,如果回忆时信心程度降低了,就说明产生了事后通偏向。为了能够对正确和错误选择的信心判断进行单一分析,我们在整个量尺上将对错误选择的原始信心判断和回忆起来的信心判断进行匹配。例如,在 70%信心水平上将一个错误选项判断为正确的被编码为 30%(即将正确选项

简捷启发式

判断为正确的信心水平)。依据这样一种量尺,事后通偏向将总是在信心程度增加时显示出来(从而避免了信心程度降低时考察事后通偏向的麻烦)。实验结果是:有反馈信息条件下信心程度平均增加 3.4 个百分点,无反馈信息条件下信心程度平均下降 0.6 个百分点。该差异的主效应处于中等水平($d = 0.54$, Cohen, 1988, p. 20),大于克里斯腾森-斯赞兰斯基(Christensen-Szalanski)和符宾·威尔哈姆(Fobian Willham)在有关元分析(1991)中所报道的平均效应水平。 205

反馈后在有关线索上对象关系发生了系统变化吗(预测 1)? 变化可能朝反馈信息指明的方向转移,也可能背离反馈方向。当线索原来指向错误选项,现在却指向正确选项或不能区分有关选项时,便发生了朝反馈方向的转移;当线索原来不能区分对象,现在却指向正确选项时,也发生了朝反馈方向的转移。同样的逻辑也可以用来界定背离反馈的变化。如果一个被试不能明确指出在一条线索上的对象关系(或两个对象的值),那么该条线索就没有区分力。在有反馈条件下,反馈后仍有 66% 的对象关系未发生变化(以所有被试、线索和项目为基础进行计算)。那么,剩下的其他对象关系朝反馈方向发生了系统变化吗? 图 9-4 显示的是朝着

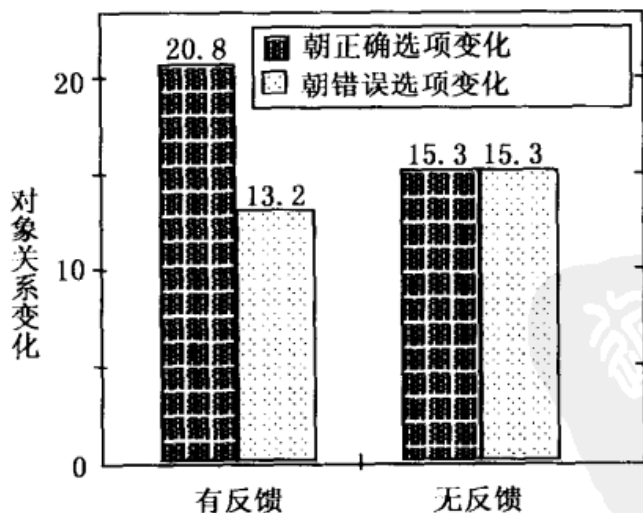


图 9-4 对象关系朝向和背离反馈变化的百分数。当给予反馈时,朝向正确或错误选择的变化等同于朝向和远离反馈的变化。

和背离反馈方向发生变化对象关系所占的比率。与预测 1 相一致,在有反馈条件下,朝着反馈方向变化的对象关系超过了背离反馈方向发生变化的对象关系;在无反馈条件下,两种变化的比率完全相同。

我们能够更加具体地阐明何时发生了朝反馈方向的转移吗?我们认为,当一条线索在原始反应中未能区分对象时,反馈后的知识更新过程更有可能发生。为了说明其中的道理,让我们首先考虑一下那些有区分力的线索。一条线索能够区分有关对象意味着在原始心理模型中知识是适用的。这些知识的存在使得以后某个时刻再次接近它们并直接提取有关对象关系成为可能——即使在反馈后,也是如此。相反,如果原始对象关系是未知的,那么反馈信息就成为一目了然的,无需克服现有知识的影响就能了解其含义。当反馈前、后对象关系“相同”时,也存在着类似情况。此种条件下,一条线索值的变化就足以改变对象关系。而对于原来就有区别的对象关系来说,线索值的变化可能减小两个对象线索值的差异,但不一定能引起对象关系的变化。

当一条线索在原始反应中未能区分有关对象时,反馈后最有可能发生知识更新吗?为了回答这个问题,我们算出了朝向和背离反馈方向(以所有被试、项目和线索为基础进行计算)转移所占比例之间的差异,结果示于图 9-5。当原来线索没有区分力时,反馈后 38.7%的情况下有了区分力,其中 27.7%朝向正确选项,11.0%朝向错误选项。二者之间 16.7 个百分点的差异用图 9-5 最靠左边的框条来表示。相反,当原来线索能够区分对象时,变化几乎是对称的——朝向和背离反馈方向转移之间的差异降至 2.8 个百分点。在无反馈条件下,无论有区分力的线索,还是无区分力的线索,两种转移间的差异都是微不足道的。这些结果有力地证实了上面的预测:当一条线索原来没有区分力时,反馈信息的作用最为显著。

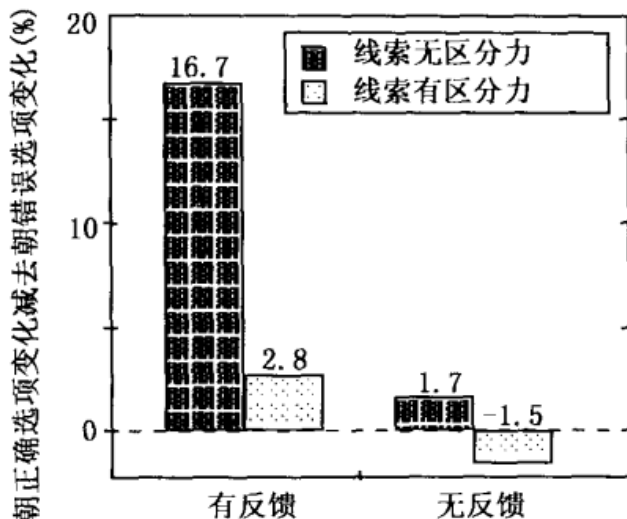


表 9-5 对象关系朝着正确选项转移的比例减去朝着错误选项转移的比例,依赖于作出原始反应时线索是否有区分力。

依据回忆起来的对象关系能否预测事后通偏向和逆转事后通偏向呢(预测 2)? 为了检验这种预测的合理性,我们首先排除了回忆反应等同于原始反应的情况,因为此种情况下总是可以用准确记忆(而不是重建)来解释有关反应,无法判定是否产生了事后通偏向。其次,我们用每个被试的反应检验 RAFT 的预测。对每种食物配对来说,我们都用采纳最佳启发式来更新被试的知识,并将其所导致的选择和信心判断与原始选择和信心判断加以比较。这种比较使我们能够确定 RAFT 究竟预测了事后通偏向、逆转事后通偏向,还是无事后通偏向(预测结果)。然后,再将这一预测结果与观测结果(事后通偏向或逆转事后通偏向)相比较,最终我们就能够确定每个被试在所有项目上正确预测的比例。

就所有被试的平均结果来看,正确预测比例为 76.3%。RAFT 正确地解释了有反馈和无反馈两种条件下几乎同样多的观测结果,分别为 76.6%和 75.9%。因此,当 RAFT 也能够说明基于未被更新线索值(如,由于未提供反馈信息)而重建的判断时,也就不足为奇了。为了了解 RAFT 究竟有多好,我们将其与一种“机遇模型”(chance model;详细内容参见 Hoffrage et al., 1999)

作了比较。在有反馈和无反馈两种条件下机遇模型正确预测所有被试成绩的百分比为 67.9% (比 RAFT 的成绩低 8.4 个百分点; $t = 5.0, p = 0.001$)。

回顾

我们提出了一个关于事后通偏向潜在认知过程的模型。该模型假定:有关正确答案的反馈信息导致了线索值的更新。如果原始反应无法直接得到,将会依据已被更新的线索值重建它。结果是,重建反应表现出了事后通偏向。与预测 1 相一致,我们发现,有关效标的反馈信息系统地影响了被试关于线索知识的回忆;与预测 2 相一致,运用采纳最佳启发式来回忆(或更新)线索值准确地预测了多数事后通偏向或逆转事后通偏向。在我们的有关论文中(Hoffrage et al., 1998),我们报告了进一步的研究结果,它验证了当前的结果,为前述第三种预测(对线索值的回忆将会降低事后通偏向)提供了证据,并详细讨论了 RAFT 是如何解释事后通偏向研究中所得到的其他发现的。

快速节俭的推断

我们提出的事后通偏向模型综合运用了弗雷德里克·巴特莱特、艾根·布伦斯维克和赫伯特·西蒙的思想。像巴特莱特一样,我们将回忆看成是一个重建过程。巴特莱特本人并未进一步详细说明这种重建过程究竟是如何被模式化的。在我们看来,重建以不确定线索为基础,这与布伦斯维克(1952, 1957)的理论框架是一致的。然而,与新布伦斯维克主义的拥戴者们(Cooksey, 1996; Doherty, 1996; Hammond, 1955)的观点——线索是通过多元回归方法得到加权和被整合在一起的——不同,我们的假设是:推断机制本质上是快速节俭的。采纳最佳启发式就是这样一种快速节俭机制。由于它有一个终止规则,所以毋需搜索所有适用信息,

简捷启发式

而且与多元回归相比,它在计算上相当简单。因此,RAFT 的推断机制是有限理性的表现形式之一(Simon, 1982; 第 1 章)。

正如本书所报告的许多结果所表明的那样,采纳最佳启发式可以与更加复杂的策略相媲美。由于其在心理学上的合理性,我们选择这种简捷启发式来为回忆过程建构心理模型。假如我们运用计算上更加有力,但从心理学上看较不合理的策略,我们能够在预测反应和实际反应之间获得更佳拟合吗? 为了回答这个问题,我们对资料进行了再分析,并采用其他几种方法检验了预测²。这些方法包括单元加权线性模型(达维斯规则,第 4 章)、以线索效度为权重的线性模型(富兰克林规则,第 4 章)和朴素贝叶斯模型(第 8 章)。其中没有任何一种模型在模拟实际反应方面比采纳最佳启发式做得更好。这些策略获得了几乎同样好的成绩。造成这种情况的原因在于它们仅仅依据三条线索形成反应,所以对多数线索值的联结来说,它们做出了相同的推论(参见第 6 章和第 7 章)。由于从心理学上看更加合理,而且有证据表明人们的选择可借助于仅加工少量信息的启发式策略来模拟(第 7 章),所以我们将采纳最佳启发式视为人们进行记忆推论时更有可能采用的候选策略。

结论

知识更新的适应过程缓解了我们对于储存过去所想、所说、所经历事件的迫切需要。知识更新过程还避免了我们使用那些由于环境变化而过时的信息,从而使我们变得更加精明。正如巴特莱特指出的那样:“在不断变化的环境里,逐字逐句地回忆是完全没有必要的。”(1932/1955, p. 204)适应性更新过程有一种不期而至的副产品——事后通偏向。但是,这种副产品也许是为快速节俭记忆付出的相当低的代价。

第10章

快速估计——让环境来做这项工作

拉尔夫·赫特威格(Ralph Hertwig)

奥瑞奇·霍夫瑞格(Ulrich Hoffrage)

劳拉·马蒂冈(Laura Martignon)

我们仿佛隔着一块有色玻璃窥视到了那扇窗户(心灵之窗)里面的东西,但我们在那里看到的東西很像是这个世界的真实写照。

——罗杰·N·谢泼德(Roger N. Shepard)

“1659

年9月30日,我,可怜而不幸的鲁滨逊·克鲁索,在一场可怕的风暴中,所乘的船在离岸不远的海上失事,船上的同伴均遭灭顶之灾,只有我九死一生,来到这个荒凉小岛上。——这个岛,我称之为‘绝望岛’。”(Defoe, 1719/1980, p. 74)鲁滨逊·克鲁索的传奇经历就这样开始了。丹尼尔·笛福的经典小说《鲁滨逊漂流记》(Robinson Crusoe)一直被视为人类征服大自然的传奇故事,甚至被视为资本主义制度下人们生存状态的寓言故事。然而,从更加现实的意义讲,克鲁索的冒险经历揭示了这样一个道理:在与大自然斗争中,能否准确估计周而复始的自然事件出现的频率是相当重要的。关于他第一次播种谷物时的情形,他在航海日志中写道:“这回我播下去的种子到头来颗粒无收;因为干燥的月份到来了,种子播下去后,大地上没有任何雨水的滋润。”(p. 106)从那以后,克鲁索对每月中的雨天和晴天做了详细记录,只在雨量最充足时播下种子。利用这种方法,他得到了丰厚回报,后来写道:“我成了种田的行家里手,知道什么时候播种最合适,而且也知

让我们更精明

简捷启发式

道每年可有两次播种季节，两次收获季节。”(p. 167)*

笛福给他所虚构的人物克鲁索配备了一本航海日志，可以帮助他预测雨量。那么，在没有任何书面记录帮助情况下，现实生活中人们仍具有估计环境各种数量关系的能力吗？我们有望为该种能力——假如它存在的话——找到适当证据的领域之一是人类觅食活动。在人类长期的进化历史中，狩猎—采集式的觅食经济占有极其重要的地位，在此类活动中人们必须决定什么东西是值得捕获的。生活在加拿大的“因纽特人”(Inujjuamiut)——爱斯基摩人的一支——为我们提供了观察当代狩猎和采集者如何选择觅食策略的机会(Smith, 1991)。因纽特人的食物来源之一是白鲸。在捕猎白鲸时，因纽特人首先围住一群白鲸，将他们赶进浅水区。由于白鲸对噪音十分敏感，猎手们便利用这一点，敲击独木舟的船舷上缘，将白鲸赶拢到一起。白鲸被火力猛烈的猎枪击中，并漂浮起来，成为因纽特人的猎物。接着，下一次捕猎活动又开始了。

因纽特人的觅食策略——即他们选择并捕获猎物的方法——可以归结为“意外猎物模型”(contingency prey model)。根据人类学家艾瑞克·奥登·史密斯(Eric Alden Smith, 1991, p. 237)的说法，这种模型是用来解释狩猎—采集式猎物选择的最佳工具。它很好地说明了因纽特人为什么选择了耗时长、且危险性大的猎鲸，而不是去捕猎鸭子、鹅和海豹之类的更易捕获的动物。像其他猎物模型一样，该模型的基本观点是：当捕猎者遇到一种食物(猎物或其他食物)时，如果猎取该种食物单位时间内获得的回报比继续搜寻另一种食物得到的回报更大，那么他就会选择该种食物作为猎取对象。因此，猎物选择依赖于由回报率决定的食物排列顺序

* 克鲁索的故事也许并非完全虚构。在出版《鲁宾逊漂流记》之前，笛福也许读过关于水手亚历山大·塞尔可克(Alexander Selkirk)的故事，他在智利附近的约翰·费尔南德斯群岛中的一个荒凉小岛上独自生活了五年。塞尔可克是在同船长发生争执后自己要求留在小岛上的。这件事被公布以后，曾经轰动一时。这名水手是如何生存下来的？人们对此很感兴趣。有些学者相信，当时笛福甚至去拜访了塞尔可克本人(参见 Swados, 1719/1980)。

(参见第 15 章)。抛开该模型的细节不谈(详见 Smith, 1991), 其关键假设是: 为了依据净回报量排列食物(来源于猎物)顺序, 必须按照在严格统计学意义上计算出来的各种不同回报率(如, 单元操作时间, 即: 花费在追逐、捕获和加工等活动上的时间)以及遭遇率(即遭遇某种猎物的频率)对它们进行分类。因此, 如同笛福将航海日志赋予克鲁索, 以帮助他预测雨量一样, 意外猎物模型赋予了人类估计环境数量关系(如, 遭遇某种猎物的频率)所必需的认知能力。

但是, 这仅仅是纸上谈兵式的策略和理论上的设想, 问题依然存在: 人类果真拥有这种能力吗? 如何才能建构它的理论模型呢? 按照布朗和西格勒(Brown & Siegler, 1993)的说法, 关于现实世界数量估计的心理学研究“在任何一种理论中都尚未臻于完善, 甚至连思考该过程的连贯一致的理论框架都还不存在。这种差距在该领域研究的异乎寻常的分歧上充分体现出来: 关于启发式的研究并未表明什么时候数量估计(如果曾经有的话)受到了特定领域知识的影响; 而关于特定领域知识的研究并未表明什么时候数量估计(如果曾经有的话)受到了启发式的影响”(p. 511)。在本章中, 我们设计了一种适合于在具有特殊统计结构环境中作出快速节俭估计的启发式, 试图用这种启发式在上述分歧之间架起一座桥梁。在描述该种启发式之前, 我们首先回顾一下先前有关数量估计的研究, 主要关心人们是如何估计事件(包括类型和表征)的频数的。我们所谓的事件可能是物体、人和情景等。^{*} 我们回顾了两类估计机制: “借助于直接提取的估计”(estimation by direct retrieval)和“借助于推断的估计”(estimation by inference)。

^{*} 本章未回顾有关下列变量估计的研究: 心理刺激(例如, Haubensack, 1992; Mellers & Birnbaum, 1982; Parducci, 1965)、概率(例如, Kahneman et al., 1982; Peterson & Beach, 1967)以及集中趋势、变异性及相关等统计参数(例如, Busemeyer, 1990)。

借助于直接提取的估计

苏格兰启蒙哲学家大卫·休谟(David Hume)确信:头脑无意识地、自动地记取了事件发生的频率,因而也无意识地、自动地记取了关于不同事件的信念水平。休谟断言(1739/1975),将观察到的频率转化为信念的潜在心理机制是相当和谐同步的:“当一方的机会和实验达到一万次,另一方达到一万零一次时,我们的判断就会因为后者略占优势而偏向于它。”(p. 141)

近期有关人类对事件发生频率监控能力的研究(Hasher & Zacks, 1979, 1984)支持休谟的观点。这些研究表明,记忆对不断流动的信息的发生频率是非常敏感的,虽然不像休谟所说的那样完全同步和谐(Hasher & Zacks, 1984, p. 1379)。人们对自然事件频率的敏感性已经被研究者们用多种不同刺激所证明。例如,一些研究者证实,人们对字词出现频率的主观判断表明他们对这些字词的实际出现频率是相当敏感的(例如,Attneave, 1953; Hock et al., 1986; Johnson et al., 1989)。^{*}

哈舍和扎克斯(Hasher & Zacks, 1979, 1984)假定:人们对偶然发生的事件进行了自动编码,并精确地储存了其频数。当需要对频数加以估计时,直接从记忆中提取该频数即可。他们指出,人们之所以能准确地估计频数,是因为记取偶发事件是一个相当自

212

^{*} 如,阿特尼乌(Attneave, 1953)要求被试判断词表上所有字母的相对频率,发现在被试所判断的频率的中数与实际相对频率之间的相关达到 0.79。

斯(Jonides & Jones, 1992)所得出的如下结论,人们的意见还是相当一致的:“如果问被试多种事件的相对频数是多少,你会发现得到的答案相当真实地反映了事件的实际相对频数。”(p. 368)在有关概率学习的研究中也得出了类似的结论,对此艾斯特斯(Estes, 1976)评论道:“被试显然能够相当有效地获取有关事件相对频率的信息。”(p. 51)

借助于推断的估计

虽然哈舍和扎克斯深信人们能够直接通达和提取事件频数,但与其观点相反的研究者们坚持认为人们是依据相关线索推断出事件发生频数的。支持这种观点的研究者们又可以根据对线索本质的假设不同分为两派:生态(ecological)派和主观(subjective)派。

基于生态线索的推断

按照布伦斯维克(1952, 1955)的观点,知觉系统利用可能与某种“远程”(distal)变量(如,距离)相关的“近程”(proximal)线索(如,知觉一条线上物体的大小)对其作出估计。布伦斯维克指出,为了成功地作出反应,一个系统应当根据线索的生态效度来利用它们(参见 Hammond 的有关讨论, 1966, p. 33),而且生态效度最好根据变量间的相关来测定。因此,生态效度被界定为一条近程线索与一种远程效标之间的相关(Brunswik, 1952)。

与哈舍和扎克斯的理论不同,布伦斯维克主义的拥护者们(例如, Gigerenzer et al., 1991; Hammond et al., 1975)假定:效标——如,五月份罗马晴天的频数——通常不能够从记忆中直接提取出来。相反,它将根据手头的线索——如,罗马位于欧洲南部这一事实——被推断出来。尽管如此,布伦斯维克的理论与哈舍和扎克斯(1984)的理论之间仍有一种有趣的联系:后者假定并提

简捷启发式

供证据表明人们贮存了关于事件频数的精确记录,前者则假定并提供证据表明人们保持了有关线索生态效度的准确记录(例如, Arkes & Hammond, 1986; Brehmer & Joyce, 1988)。然而,对线索生态效度的了解和掌握需要准确登记每种事件频数及其同时发生频率的能力,除非进化已经赋予了人们有关线索效度的知识(如,在深度知觉研究中即是如此)。

基于主观线索的推断：可用性

在特威斯基和卡那门的一项经典研究中(1973),要求被试判断5个辅音字母(K、L、N、R、V)中的每一个在英文单词第一还是第三位置出现的次数多。尽管这五个辅音字母均更加频繁地出现在第三位置上,但近乎三分之二被试判断多数字母更有可能出现在第一位置。^{*}

特威斯基和卡那门(1973)提出了“可用性启发式”(availability heuristic),作为现实生活中人们进行数量估计的潜在机制,并用它来解释人们进行数量估计时产生的系统偏差。根据可用性解释,对频数(或概率)的评估以“能够进入头脑”(could be brought to mind; p. 207)的事件样例数量为基础。也就是说,其基本假设是:人们提取出所考察事件的一个样本(如,分别提取出字母“R”位于第一与第三位置的单词),或者对提取这样一个样本的难易程度加以评估,然后利用样本统计量来估计有关效标。然而,样本参

* 在讨论特威斯基和卡那门的研究时,罗普斯和奥登(Lopes & Oden, 1991)发现,20个英语辅音字母中的12个出现在第一位置上的次数多于出现在第三位置上的次数。这也许可以解释他们的研究结果。相反,如果假定人们接触了字母及其在特定位置出现次数的有代表性的样本(如,在阅读中),那么他们的心理模型就应该对应于主试呈现给他们的那个有代表性的样本。赛得梅尔等人(Sedlmeier et al., 1998)呈现给被试一个有代表性的辅音字母样本(即:一些字母在第二位置上出现频率较高,另一些则相反)和元音。在三项研究中,他们均发现,除了一个过高估计的低值和一个过低估计的高值外,估计出的第一与第二位置的相对频率与实际排列顺序非常接近。赛得梅尔等人检验过的两种可用性启发式均无法解释这些结果。

数也许会系统地偏离总体参数(如,假如提取出某个字母位于第一位置的单词比提取它位于第三位置的单词更容易,那么这个样本不具有对于总体的代表性)。从这种意义上看,可用性线索的使用可能会导致系统偏差。由于样本对效标的预测能力只能够用特定个体提取出来的样本来评价,所以可用性线索是主观的而非生态的。

自从特威斯基和卡那门(1973)将可用性和其他启发式作为潜藏于(相对)频率和概率判断背后的机制以来,他们的发现和提出的启发式规则已经激发了大量研究,并引起了人们对人类准确估计事件频率和概率之能力的广泛关注。在这一点上,采用可用性规则是“决策制定和社会判断研究中最广为人们接受的假设之一” 214 (Schwartz et al., 1991, p. 195)。例如,不少人指出,可用性启发式可以很好地解释人们夸大一些特殊死亡原因(如,龙卷风)出现频率的倾向(Lichtenstein et al., 1978),并能够说明人们在估计国家人口数量之类的人口统计参数时的成绩(Brown & Siegler, 1992, 1993)。

矛盾的假设和对立的发现

这样一来,就存在着一个令人困惑不解的问题。哈舍和扎克斯(1984)指出,人们对某一事件的偶然发生进行自动编码,储存其频数,当需要估计其频数时,直接提取出该数目。相反,特威斯基和卡那门(1973)似乎假定:人们并不保存事件频数的记录,而是建立一个所考察事件的样本,然后根据建立这个样本的难易程度来推断事件频数。哈舍和扎克斯(1984)认为他们的实验“令人信服地、明白无误地印证了关于迄今为止检验过的所有事件发生频率的知识”(p. 1373),而特威斯基和卡那门(1973)则将他们的研究结果看成是运用可用性启发式有可能导致“系统偏差”的证据(p. 209)。

简捷启发式

这两种对立的假设和发现在科学杂志和教科书中一直同时得到报道和介绍,很少有关于每条研究路线如何适合于解释其他发现的讨论(作为例外,可参见 Ayton & Wright, 1994; Holyoak & Spellman, 1993; Williams & Durso, 1986)。* 一个人或许想要调和二者之间的矛盾,于是假定这两种解释——基于记忆经验的准确判断和基于主观线索的(不)准确判断——分别适用于不同情境:每当人们作出判断前曾经逐个地接触过和编码了每一个事件时,前一种解释是真实合理的;而每当人们未曾直接接触过效标,因此不得不依据与其相关的(主观)线索作出推断时,后一种解释是真实合理的。

然而,这种试图调和二者矛盾的做法却不能奏效。特威斯基和卡那门的实验也包括了被试实际上相继经历了每种事件的情境。例如,在他们的一项研究中,相继向被试呈现著名男性和女性的名字(如,伊丽莎白·泰勒),然后要求一组被试判断这个名单中包含了更多男性还是女性名字(Tversky & Kahneman, 1973)。另一个例子是上文提到过的关于字母位置频率的经典研究,在这项研究中他们要求被试依次判断先前经历过的事件。这两项研究均表明在对经历过的事件进行估计时被试同样试图采用可用性规则。

在我们看来,关于人类频率判断准确性的相互矛盾的发现以及潜在于这些发现背后的相互对立的假设是难以仅仅通过增加现有实验数量来加以调和的。在这些实验中,人们的估计要么是正确,要么是错误的。而实际上我们毫无疑问可以找到既可能产生有偏差估计,又可能产生无偏差估计的实验情境。更为有趣的课题是:我们如何才能为数量估计的潜在认知过程建立更为有效的理论模型? 为了实现这一目标,我们提出了与上述两种方法均

* 这种矛盾没有引起更多注意,其中一个原因可能是哈舍和扎克斯对此只作了轻描淡写(1984)。他们在有关脚注中这样写道:“我们的看法与特威斯基和卡那门的观点之间的矛盾应该说是表面的而非实质性的。”(p. 1383; 参见他们在脚注 9 中的讨论)

有关联的两个相关问题：首先，为了达到适应性目标，人类需要计算的是什么？其次，需要量化的环境结构是什么，以及什么样的启发式规则能够充分利用这种结构？

需要计算什么？

世界可以被细分为无数个独立事件或物体。其中哪些值得监控呢？哈舍和扎克斯(1984, p. 1373)并未明确强调这个问题，而是假定：对于将要得到编码和储存的某个刺激物的频数来说，人们自然而然会给予其最低限度的“注意”。在此他们并未对“注意”这个概念给以明确界定和说明。^{*} 这样一种通用编码机制——即仅仅在需要的注意力(或“有意”注意，像扎克斯后来提出的那样，1986)上受到限制的编码机制——在多大程度上是合理的呢？

举例来说，试想一下我们边在街上行走、边进行有趣谈话时可能发生的信息加工过程。通常，我们都能成功地避开路边的物体或其他行人，这说明我们注意到了他们的位置。但是，以后我们能够判断他们在不同方位出现的相对频率吗？(如，多少对象出现在我们左边？多少对象出现在我们右边？)或者我们能够判断我们遇到过的戴帽子男人和女人出现的相对频率吗？如果当时我们不认为这些判断是有用的，我们有什么理由应该能够对此作出判断呢？更一般地说，难道我们会仅仅因为曾经经历了一件事情就对其进行编码，并记取其发生频率吗？ 216

这是英国经验论者与哈舍和扎克斯(1984)均未曾认真考虑过的一个问题。例如，戴维·哈特莱(David Hartley, 1749)曾经为频率计算设想一种通用生理机制，用来与牛顿的“振动理论”(theory of vibration)相类比(Daston, 1988, p. 203)。根据这种机

* 然而，我们可以排除一种可能的定义：有意监控。哈舍和扎克斯(1984)指出，即便一个刺激没有受到有意监控，它也能够得到自动的编码。这意味着有意监控并不包含在他们的“注意”定义之中。

简捷启发式

制,某物体的重复出现引起大脑振动,直到“心理习惯的槽道”(grooves of mental habit)铭刻于大脑之中。哈特莱假定的机制是一种通用机制,原因在于它并未对需要计算的对象类型作任何限制。在现代认知心理学中,我们还可以找到与哈特莱设想的不受限制生理机制相类似的“认知”模型,如“密涅瓦第二”(MINERVA 2)模型(例如,Hintzman, 1984, 1988)。* 它一直被用来为频率判断过程建立心理模型。该模型保存了(根据记忆痕迹)我们过去生活中经历过的所有事件的副本或拷贝(尽管一个人也许通过参数学习考虑到了特定领域具体内容的限制,如同 Hintzman 在 1988 年所做的那样)。

我们能够判断戴帽子男人和女人出现的相对频率吗? 玛西亚·约翰逊及其同事(Johnson et al., 1989)指出这是不太可能的。以一系列设计精巧的研究为基础,他们证明:要想使频数判断准确反映事件出现的频率,必须满足两个条件:第一,呈现时间必须 ≥ 2 秒钟;第二,加工过程必须涉及到对对象特性及其空间位置的注意。尽管他们的发现对前述的记数机制进行了初步限制,但在究竟需要计算什么问题这个上,这种机制仍然是不受限制的。

布瑞斯、考斯米德斯和托比(Brase, Cosmides & Tooby, 1998)提出了一种更加严格的限制。他们指出,限制记数机制的另外一条途径是充分考虑所要计算东西的本质;现实世界的某些方面是我们难以指望一种人类推断机制自发地加以计算的。按照他们的说法,个别整体对象而非人为地分解开的对象(即随机组块、无用碎片等)才是分析的自然单元:蹒跚学步的儿童可能自发地去数玩具熊,而不是去数玩具熊的耳朵(只要它们没有被从玩具熊身上剥离下来)。

布瑞斯等人所提出方法的一种变化形式是充分考虑所要计算东西的适应价值:当了解事件频数对于有机体似乎具有适应价值

* MINERVA,传说中的智慧、技术和工艺之神。

时,最有可能对事件发生频率加以记录。在选择配偶和觅食活动(如,因纽特人的猎食活动)等领域,很容易看到监控特定事件产生频率的适应价值所在。但是,充分考虑适应价值能够帮助我们在其他领域作出与我们直觉相反的预测吗?我们认为这是可以的。尽管特威斯基和卡那门(1973)似乎提供了与此相反的研究结果(在他们关于字母位置的研究中),但我们有一个很好的理由可作出这样的预测:人们完全能够对语言的统计结构加以量化,因为这种统计结构是具有适应价值的。

了解语言统计结构的适应价值是什么?

在任何一种特定语言中,一定发音顺序在某些词汇中发生的可能性要大于在其他词汇中发生的可能性。例如,不妨考虑一下“pretty baby”这个发音序列:从“pre”转移到“ty”的概率要大于从“ty”转移到“ba”的概率。因此,我们将会预期后面两个音节之间更有可能出现词汇间的中断。对于正在学习语言的婴儿来说,记取这种语音转移概率有重要作用,因为这种概率的大小能帮助他们辨别词汇之间的界限(这是编制“理解”口头语言的计算机程序时所遇到的主要障碍)。萨夫瑞恩、阿斯林和纽鲍特(Saffran, Aslin & Newport, 1996)最近的研究结果表明:婴儿对这种语音转移概率是十分敏感的。 217

为了检测婴儿是否意识到了这种统计信息,萨夫瑞恩等人对婴儿在整词和“部分词”(part-words)之间作出区分的能力进行了测试(两种情况下均用无意义刺激)。刺激词包括“bidaku”和“padoti”之类的发音序列。该种语流的一个例子是“bidakupadoti-golabubidaku...”。婴儿听两分钟连续的语流录音,该语流由三音节词的随机重复而组成。一种经过合成处理的女音发出这段语流。词汇之间没有转调或明显停顿,也免去了平常言谈中单词界限这一线索。唯一的线索是成对音节同时出现的相对频率。在这

简捷启发式

里相对频率较低是单词界限的信号。

听完语流后,婴儿逐个地听 4 个三音节测试词。其中 2 个包含在原来的语流中,另外 2 个是部分词。所谓“部分词”是由一个单词的后面一个音节和另外一个单词的前面两个音节组成的。这样,“部分词”就包含了婴儿曾经听到过的发音,但是这种发音与单词发音不完全一致。只有当婴儿对原先语流中的单词非常熟悉,以至于对打乱了单词界限的发音序列(即“部分词”)相对陌生时,他们才会将“部分词”识别为新的。实际上,婴儿听“部分词”的时间的确比听单词的时间长,这表明他们探测到了“部分词”比单词更加新奇。

这个例子提出了这样一个十分重要的问题:什么样的信息适合于编码、贮存和量化?带着这个问题,一个人也许会推导出一些与我们直觉相反的非常有趣的预言,比如,他也许会预测语言学习者会很快地掌握语言的统计结构。现在我们转向第二个重要问题:需要估计数量关系的环境究竟具有什么样的结构?

“生态结构”的重要性

“尽管判断失误仅仅来源于研究某些认知过程的方法,但方法已成为有关发现不可忽略的组成部分。”(Kahneman & Tversky, 1982, p. 124)这句引文说明卡那门和特威斯基意识到他们倡导的“启发式 偏差”方案忽略了人类认知成功的一面,而将注意力过多地放在了人类认知的失误上。事实上,生态观在他们最初的可用性启发式思想中占有重要位置,只是后来被放弃了。关于可用性启发式,卡那门和特威斯基(1973)写道:

对于频率判断来说,可用性是一种从生态学上看有效的线索,因为一般而言频繁发生事件比非频繁发生事件易于回忆和联想起来。然而,可用性也受到了与实际频率无关的多种因素的影响。若运用可用性启发式的话,这些因素将会影

响对事件发生频率和主观概率的知觉。结果是,可用性启发式的使用导致了系统偏差。(p. 209)

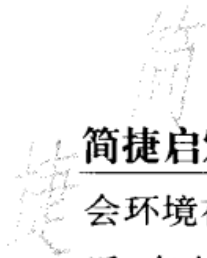
卡那门和特威斯基(1973)不仅把可用性视为频率判断的“从生态学上看有效的线索”,而且强调它充分利用了环境结构,因为客观上频繁发生的事件由于不断重复而得到了更好的表征,从而比非频繁发生事件更易于回忆。就其原始含义来看,可用性启发式本来可以发展成为一种既能够反映环境结构又能够反映头脑结构的认知策略的,但事实上却没有。

几十年以前,艾冈·布伦斯维克(Egon Brunswik, 1957)就已经强调了研究认知过程与环境之间拟合程度的重要性:“如果说有什么东西使心理学,特别是有关认知过程的心理学仍然处于相当困窘境地的话,那就是对有利于有机体(的结构和过程)发挥作用的环境或生态结构的漠不关心。无论从历史的角度看,还是从现实的角度看,心理学都已经忘记了它是一门关于有机体和环境之间关系的科学,而变成了一门关于有机体的科学”。(p. 6)

在下面的部分中,我们提出了一种数量估计的启发式规则。它在以下几个方面不同于“启发式—偏差”方案提出的启发式(如,可用性):首先,我们对它如何充分利用特定环境结构作了详细说明;其次,它有一种精确的中断记忆搜索的终止规则;最后,它被模式化了,以便我们可以模拟其行为。由于这些原因,它体现的是有限理性。在分析多种数量必须得到估计的各种环境结构之前,让我们首先看一看对此类数量之一——人口数量——作出估计的适应价值何在。我们推测:对人口统计学方面的数量(具体来讲就是对不同社会群体规模大小的估计)作出估计或许是从进化角度看非常重要的任务之一。

估计:在J型分布的现实世界里使用生态线索

由于人类总是生活在群体(如,家庭、宗族、部落等)中,所以社



简捷启发式

社会环境在塑造人类头脑过程中很可能扮演着重要角色。直到最近,在人类推理和决策制定研究中,这种可能性一直未受到足够重视。然而,王(X. T. Wang, 1996a, 1996b)已经证明社会线索是如何以令人惊奇的方式影响决策制定的。利用特威斯基和卡那门(1981)著名的亚洲人疾病问题,他发现当有关文句暗示某个决策将适用于一个大群体时,被试的偏好发生了转变[这通常被认为是不合理的,因为他们违背了“预期效益理论”(expected utility theory)的“不变性公理”(invariance axiom)]。然而,当有关文句暗示某个决策将影响一个小群体时,无论是得是失,多数被试都倾向于赞成具有风险性的决策。

219

王的发现(1996a)表明人类在制定决策时对群体规模大小是十分敏感的。有人也许推测这种敏感性依赖于已经很好地得到进化的估计群体规模大小的能力。实际上,这种准确估计社会群体规模大小的能力对于我们尚未得到很好进化的祖先适应种种环境也是很有价值的。例如,当他们必须迅速决定是否同其他家庭、种族或部落为争夺生活资源而战斗时,就必须拥有这种能力。自从进入狩猎—采集式社会以来,人类社会的结构发生了巨大变化。从游牧时代的散兵游勇,到较小的农业和畜牧业部落,再到成千上万人构成的以城市为经济和社会中心的现代社会,人类群体的规模越来越大(例如,Reynolds, 1973)。有趣的是,在美国人和中国人构成的被试样本中,王(1996a)发现决策制定对社会群体结构的某些特殊文化特征十分敏感。抛开进化方面的考虑,我们假定这里所提出的估计启发式适用于现代群体的人口规模。下面我们分析一下该种启发式能够发挥作用的环境的统计结构。

让我们从下面的问题入手来分析人口规模的统计结构。设若一个人想要绘制人们在完成下述任务时取得成绩的散点图,那么结果将会呈现何种分布呢?这项任务是:列举莎士比亚《错误的喜剧》(*Comedy of Errors*)中所有人物的名字。如果我们对人们在该项任务上的成绩(如,能够列举出0、1、2、3个人名的被试人

数)加以分析,我们很有可能会发现多数人得分均很低,只有少数几个人得分很高。教育研究者们通常假定人们在知识程度、智力水平和学习能力等方面的成绩一般都符合“钟形”(bell-shaped)分布(即正态分布),即趋中值出现频数最高。与这种假设相反,人类的成绩最好用“愈高愈少的经验法则”(empirical law of the higher, the fewer)来描述(Walberg et al., 1984, p. 90),或者换一种说法,用正偏态的J形分布来描述(该种分布的形状很像是按顺时针方向将“J”旋转 90 度后的形状)。^{*}

体育比赛成绩也符合此类J型分布。不妨以 1996 年夏天在亚特兰大举行的奥运会的最终奖牌分布为例。在这届奥运会中,共有 197 个国家或地区角逐 842 块奖牌。图 10-1 显示的是除去未获得任何奖牌的国家或地区外,每个国家或地区获得奖牌(金牌、银牌、铜牌)的总数。平均每个国家或地区获得 4.3 块奖牌。其中美国、德国和俄罗斯分别获得 101 块、65 块和 63 块奖牌,也就是说,参赛国家和地区中的 1.5%(世界人口的 8.5%)几乎获得了全部奖牌的三分之一。另一方面有 118 个参赛国家或地区未获任何奖牌。高度的正偏态分布还可以被用来描述生物学(如,以振幅来区分的神经冲动)、地理学(如,以震强来区分的地震)、心理学(如,以人们在多大程度上有可能需要它们来区分的记忆痕迹分布;Anderson & Schooler, 1991)以及其他领域的过程和现象。

以实际人口数量来界定的城市大小同样遵循J型分布。在世界上的任何一个地区,人口拥挤的大居民区都是很少的,而人口稀疏

* 这些分布与 Zipf 法则(Zipf, 1949)有关,该法则是指这样一种观察:当一些事件的“秩”(rank)是由它们发生的频率决定的时候,作为秩(i)的函数的这些事件的发生频率(p)是以近乎一致的 a 为指数的幂函数 $P^i = 1/i^a$ 。Zipf 法则最为著名的例子是英语单词的出现频率。假定“the”、“to”、“of”是最常出现的单词(即它们的秩分别为 1、2、3),那么当将它们出现的频数作为秩的函数绘制分布图时,其形态就是具有接近于 1 的指数的幂函数。Zipf 法则有几种变化形式,如 Pareto 法则,它在本质上形成了 J 形分布的基础。更一般地说,有人(Grüneis, 1989)已经证明:J 形分布属于能够依据“联合普尔松过程”(adjoint poisson process)建立模型的一类分布。

的小居民区却很多。赫伯特·西蒙(1955b)曾经指出,就城市人口规模这样一种特殊情况而言,如果仅仅根据人口出生数量超出人口死亡数量的净增量来计算人口增长率,而且该净增量与现有人口规模是成比例的,那么就有望得到一种J型分布。有人曾经建立了一种城市人口增长模型,它使用了原本用来为液体和气体中分子凝聚和运动建立模型的技术,该模型也预测了这样一种城市人口分布(Makse,1995)。图10-1还显示了超过100 000居民的83个德国城市的人口规模分布状况。该种分布从三个方面反映了“愈高愈少的经验法则”:最高值(这里是“柏林”的人口数量)是绝无仅有的例外情况;平均值(309 000)受这个极端数影响,因而远远高出了中位数(180 000);相对于平均值而言,标准差(428 000)过大。

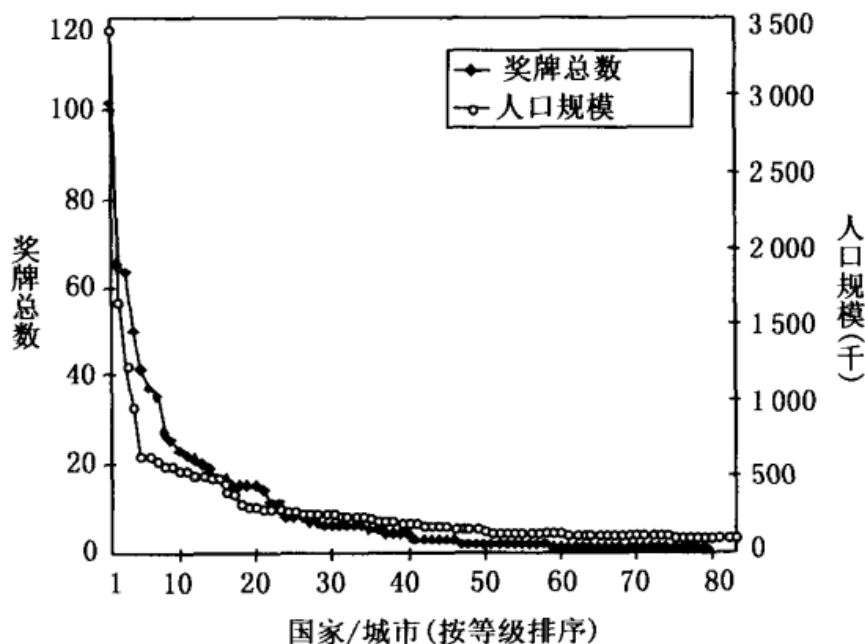


图 10-1 1996 年亚特兰大夏季奥运会上每个国家获得奖牌数及德国 83 个最大城市人口规模分布(Fischer Welt Almanach,1993)。

那么,普通人是否了解诸如德国城市规模之类的变量呈J形分布呢?如果我们假定他们实际上了解这一点,这种假设在多大程度上是合理的呢?为了回答上述问题,我们要求 74 名德国被试按照 25 种规模类别(100 000—199 999; 200 000—299 999; 如此

等等)来估计各个类别中所包含德国城市的数目。图 10-2 是他们作出的频数判断的平均值与实际频数分布的比较。(请注意:与图 10-1 相比,纵横坐标是相反的。)尽管被试低估了最小规模类别(100 000—199 999)中包含的城市数目,但这些结果表明他们意识到了城市人口规模是呈偏态分布的。既然我们已经证实人们对德国城市规模分布的“愈高愈少”特征具有直观知识,下面我们就转向下一个问题:一种启发式规则如何利用此种 J 型生态结构以减少做出一种估计所需要的繁琐计算步骤?

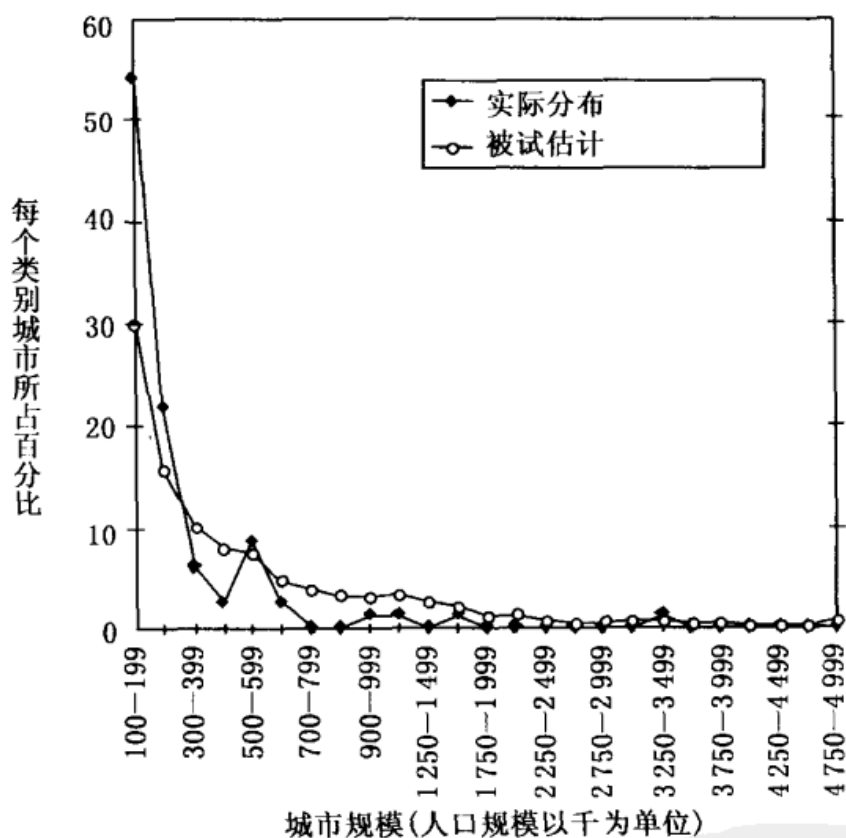


图 10-2 25 个规模类别所包含德国城市数的百分比,与被试作出的估计相比较(百分比的数值来源于频率估计)。

快速节俭估计: 最快者启发式

让我们首先考虑一个技术问题,即按照尺寸大小拣选煤块。拣选煤块的一种方法是使用传输带运送大小不等的煤快,使其经

简捷启发式

过网眼不断增大的过滤网。这种传输带经过特殊设计,以便最小煤块首先通过“小”过滤网落入下面的粉碎机里,接着中等大小煤块从“中等”过滤网掉下去,依此类推。那些通过所有过滤网而未掉下去的煤块被倾卸在一个杂物存放处。让我们假定煤块的尺寸大小符合J形分布,也就是说,多数煤块是小的,只有极少数煤块是(非常)大的。利用这个事实来设计传输带可以将拣选过程所需时间降至最少:首先拣出大量小煤块,然后拣出较少的中等煤块,最后拣出极个别的大煤块。图10-3表明了这种传输带的设计特点。现在我们提出一种估计启发式,即快速估计启发式(最快者)。与拣选煤块的传输带相类似,它充分利用了J型分布。

222

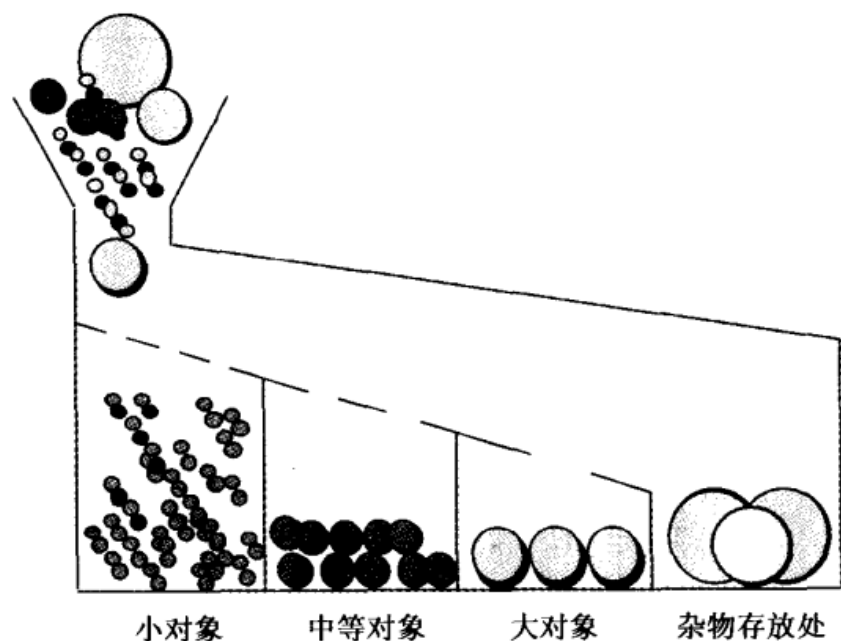


图10-3 根据尺寸大小拣选煤块的传输带图解(尽管这是一个虚拟例子,但它的设计与许多制造商在网站上推销的传输带相类似)。

224

最快者的设计特点

最快者的基本设想是:充分利用环境结构尽可能快地为绝大多数普通对象(如,小城市)作出估计。那么,该种启发式的哪些设计特点能够使它实现这种设想呢?

线索如何排列? 当要求一个人估计某个城市人口数量时,这个城市是州府的事实可能会作为一种潜在生态线索进入他的头脑。作为州府的城市(如,巴伐利亚州州府慕尼黑)很可能比不是州府的城市有更多人口(不符合这一规则的例外情况多数在美国)。对于任何一条双极线索 i , 一个人均可以计算出拥有这一特征城市的平均人口规模(S_j^+ , 如作为州府的所有德国城市的平均人口规模)和不具有这一特征城市的平均人口规模(S_j^-)。值得注意的是,为了便于计算机模拟,我们根据不具有(或具有)某种特征德国城市的实际人口规模计算 S_j^- (或 S_j^+)。然而,这种计算的输入不需要实际值,而可以代之以不甚精确的主观估计值。

因为正线索值被界定为较大城市,所以 S_j^- 小于 S_j^+ 。由于这个缘故,在最快者启发式中按照 S^- 排列线索顺序,即最小的 S^- 优先。这种设计遵循拣选煤块的类比,即线索(过滤网)根据其网眼大小来排列,网眼最小的线索优先。对于这种线索顺序来说,启发式无需确切地知道 S^- ; 它只需要根据 S^- 估计出一种相对线索顺序。

何时终止搜索? 每条线索均询问一个城市的一种特征,例如,“这个城市有大学吗?”最快者启发式有一个简单终止规则:当找到一个城市不具备的第一种特征(即对上述问题的回答是“否”)时,立即终止搜索。如果一个城市具有某种特征,那么搜索继续下去,下一个具有最低 S^- 的线索值从记忆中提取出来。这种终止规则表现出一种负值偏向,即负线索值而非正线索值终止搜索。这导致了一种重要后果:由于只有少数几个城市拥有正线索值,而绝大多数城市拥有负线索值,所以具有负值偏向的终止规则通常能够使该种启发式及早终止搜索,并迅速作出估计。*

根据其终止规则,最快者启发式的推论取决于一个城市不具

* 例如,在依据 8 种生态线索(足球队、州府、前东德、工业带、牌照、城间铁路、展览会和大学,见第 4 章)判断人口超过 100 000 的所有德国城市的大小时,这些德国城市平均拥有大约 6(5.7)个负值。

简捷启发式

备的第一种特征。与多元回归之类的计算起来相当繁琐的策略相反,最快者启发式不对线索值加以整合。最快者启发式的终止规则产生了一种重要后果:该种启发式是不可代偿的。进一步的线索值(即便它们全都是正值)不能改变依据所遇到的第一种负线索值作出的估计。由于其节俭性,不可代偿决策避免了对相互矛盾线索的加工和处理,无需对它们加以权衡。

此类估计是粗略的吗?最快者所估计的是一个城市不具备的第一种特征的 S^- ,按照四舍五入原则将其转换为最为接近的“自发数”(spontaneous number)。按照阿尔伯斯(Albers,1997)的观点,自发数是10的乘方的倍数 $\{a10^i: a \in \{1,1.5,2,3,5,7\}\}$,这里 i 是一个自然数(natural number)。例如,300、500、700、1 000 是自发数,但900不是。在构成自发数时,该种启发式考虑到了人们作出估计的两个特征。首先,自发数与阿尔伯斯所说的数的“突出性”(prominence)有关。所谓数的突出性是指这样一种现象:在使用十进位制的文化中,10的乘方“是最为突出的候选对象,最经常地被用来作出一种反应,或者被用作修饰给定反应的术语”(Albers,1997, p. 6)。其次,自发数与下述现象有关:当要求人们作出数量估计(如,一辆保时捷轿车的价格)时,人们倾向于给出相对粗略的估计(如,70 000 美元,即 7×10^4 ,而不是 75 324 美 225 元)。这种估计的粗略性,或“相对精确性”(Albers,1997, Part I, p. 12)的概略程度,反映了人们对其判断的不确定感(亦可参见 Yaniv & Foster,1995)。^{*}

* 由于其中有更多的小数字(1、1.5、2、3),而不是大数字(5、7),所以自发数似乎也可以根据 Benford 法则来预测。Benford 法则(1938; Raimi, 1976)指出:如果按照第一重要的数字对数字资料(如,原子重量)进行分类,通常有9种大小不同的数字类别。在随机生成的一套资料中,每一个数字都是产生频率为 $1/9$ 的第一重要的数字,在许多现实资料中,该频率约略等于 $\log_{10}(p+1)/p$ 。因此,数字“1”为第一重要数字的出现频率最高,大约为30%，“2”次之,依次类推,直到出现频率最低的数字“9”,其频率低于5%。与该法则相一致,57%人口在100 000以上的德国城市以“1”打头,只有1.2%的城市以“9”打头。

将某种特征的 S^- 四舍五入为最接近的自发数有两种意义：首先，对于数值估计来说，最快者启发式无需估计 S^- 。它仅需要估计临近 S^- 的两个自发数中哪一个最接近 S^- ，然后将该自发数作为给出的估计。^{*} 其次，该种启发式作出的估计只会尽可能逼近精确性，而在粗略性上不会超过自发数限定的范围。

该种启发式如何处理个别大城市？由于当前采用的终止规则一旦发现了正在考查的城市所不具备的一种特征，立即终止搜索，所以它加快了估计速度。然而，仍然有极少数具备大多数特征的“漏网”(outlier)大城市。为了避免花费不必要的时间去搜索这些大城市可能不具备哪种特征，最快者启发式备有一个“杂物存放器”。也就是说，一旦大多数城市(超出了该种启发式所“知晓”的范围，即训练项)得到了过滤，该种启发式随即停止增加更多线索到它的线索序列上。就我们的模拟研究来说，我们设定：一旦所有城市中的五分之四已经被启发式过滤掉了，立即终止对线索的搜索。剩余的五分之一城市被存入“杂物存放器”中，并被自动地分配给一种估计 S_j^+ (这里线索 j 是将剩下大城市一网打尽的线索)，然后四舍五入为最接近的自发数。

从生态学上看最快者启发式合理吗？最快者启发式从两个方面利用了城市人口环境特征：首先，在负线索值占支配地位的环境中，它的终止规则——找到第一个负线索值即终止搜索——有效地限制了搜索过程；其次，它按照 S^- 排列线索顺序，最小的 S^- 首先得到检索，这使它产生了一种偏向，将任何给定城市估计为相

^{*} 假定 S^- 介于自发数 300 000 和 500 000 之间。为了确定 S^- 应该向上或是向下四舍五入，该种启发式仅需要知道它位于中介点(400 000)左边还是右边。这仅仅需要作出一种选择(即： S^- 大于还是小于 400 000)。

简捷启发式

估计。

例证

我们的一位美国同事——瓦勒瑞(Valerie)——根据先前对德国的访问大体上知道五个德国城市的人口规模(慕尼黑, 1 000 000; 法兰克福, 700 000; 纽伦堡, 500 000; 波恩, 300 000; 海德堡, 150 000)。她还知道这些城市在三条线索(展览会、州府、大学)上的值。在给定上述关于参照等级(德国城市)的有限知识情况下, 她怎样才能准确推断其他城市(如, 勒沃库森)的人口规模呢? 为了回答这个问题, 我们首先看一看: 作为指导瓦勒瑞作出推断的模型, 最快者启发式是如何学会掌握有关参数的?

训练 最快者启发式根据具有负值(S^-)的城市平均人口规模来排列线索顺序。就瓦勒瑞拥有的知识来看, 具有最小 S^- 的线索是“展览会”, 它提供的估计为 200 000。^{*} 下一条线索是“州府”, 它得出的估计为 500 000。根据这两条线索, 最快者启发式能够过滤掉瓦勒瑞所了解城市中的大多数: 5 个城市中的 4 个(80%)至少在这两条线索的一条上具有负值。因此, 唯一一个在展览会和州府两条线索上有正值的城市(慕尼黑)被投入杂物存放器。对它的估计源自能够“捕捉”到慕尼黑有关信息的最近线索。这里即州府线索。根据该条线索估计的人口规模数是 1 000 000 人(它完全等同于慕尼黑的人口数量)。

总之, 依据瓦勒瑞有关德国城市的知识, 实施最快者启发式涉及到了她所了解的三条线索中的两条(展览会与州府)以及一个“杂物存放器”。对特定参照等级内任何一个给定城市来说, 这种设计都允许最快者启发式仅从以下三种估计中选择其一: 200 000、500 000 和 1 000 000。当被推论到勒沃库森、汉堡等新城

^{*} 这个数值是这样计算出来的: 瓦勒瑞所了解的 5 个城市中有 2 个(海德堡和波恩)没有展览会。也就是说, S_{expo}^- 等于将海德堡和波恩的平均人口数量(225 000)四舍五入为最接近的自发数(200 000)。

市时,按照这种方式实施最快者启发式获得的成绩如何呢?

估计 为了估计勒沃库森的人口规模,最快者启发式首先提取该城市在展览会线索上的值。由于勒沃库森没有展览会,搜索被终止了,估计其人口数量是 200 000,接近于其实际人口数量(160 000)。为了对汉堡的人口数量作出估计,最快者首先查询其在展览会线索上的值;由于该值为正,它接着提取在州府线索上的值,它仍然是正的。结果,汉堡被放进“杂物存放器”,估计其人口数量为1 000 000。该估计与其实际上拥有的人口数量(1 650 000)相距较远。

这一成绩是好还是坏呢?相对于其他启发式而言,最快者的节俭性如何呢?

成绩的检验:环境和竞争者

为了更加全面地检验最快者启发式的成绩,我们用它对实际上超过 100 000 居民的德国城市人口数量进行估计。自从 1990 年两德统一之后,德国有 83 个这样的城市。所有这些城市(柏林除外)及其在人口规模上的 8 条生态线索(除国家首都线索外,与第 4 章所用过的那些线索相同)被包括在本测验中。(柏林被排除在外,是因为它是一个漏网者,估计其人口规模时产生的误差过大,与估计其他城市时产生的误差不成比例。)为了评价最快者的成绩,我们将其与两个需要相当多计算和知识的竞争者——多元回归和“估计树”(estimation tree)——进行了比较(关于估计树复杂性的量化问题,参见第 8 章)。

多元回归能够作为一种严格基准,原因在于它需要计算能够将权重降至最低的最小平方和,以此来反映线索与效标之间的相关以及各条线索之间的协方差。长期以来,多元回归一直既被作为一种描述性认知模型,又被作为一种规范性认知模型来看待,尽管它所假定的计算过程过于复杂,以至于对认知活动的描述引起

简捷启发式

了争议(有关这个问题的论述,参见第4章)。

第二种基准是一种估计树(关于以树形图为基础的方法的更多论述,参见 Breiman et al., 1993)。借助于计算上相当复杂的贝叶斯搜索过程(如,第8章;Chipman et al., 1998),该种估计树被视为一种可望获得好成绩的方法。^{*} 它将拥有相同线索配置——即在8条生态线索的每一条上均有相同线索值——的城市“折叠”为一类。为每个城市估计的人口数量均用该类别中所有城市人口数量的平均值来表示。(如果某个城市具有独特的线索配置,那么为它所作的估计就是它的实际人口规模。)只要训练项和测验项是一样的,这种算法规则就是最优化的,而且等价于皮尔逊(Persson, 1996)提出的基于样例的算法模型。^{**} 当训练项与测验项不同时,该估计树将会遇到具有不同线索配置的新城市。如果一个新城市与一种旧的线索配置相匹配,那么对这个城市人口规模的估计值就是拥有这种线索配置城市(在训练项中)人口数量的平均值。如果一个新城市拥有一种新的线索配置,那么就将这种配置与最接近于它的配置匹配起来。这个过程是如何完成的呢? 228

首先,在每种线索配置内按照各条线索的效度(关于线索效度的更多论述,参见第6章)排列它们的顺序,效度高的线索优先。其次,按照词典编纂规则排列各种线索配置的顺序,以便在最有效线索上呈正值的那些线索配置排列在前。如果两种或多种线索配置在最有效的那条线索上的值相等,那么就按照它们在第二条有效线索上的值排列它们,以此类推。采用同样逻辑分别将新线索配置归入按照词典编纂规则排列好的老线索配置序列中去。当估计一个具有新线索配置的城市人口规模时,该估计树采取在词典

^{*} 贝叶斯搜索仅限于树的子集。树形图按照一定的间隔对新的线索配置加以归类。间隔之间的界限由已知城市的线索配置来确定。

^{**} 该种最优化方案是:记住所有的线索配置,将具有相同线索配置的城市“折叠”为同类。在统计学上,这种最优化方案以“真回归”(true regression)而著称,它类似于在选择任务上获得最优成绩的配置记忆法(参见第6和第8章)。

编纂序列中线索配置高于它的那些城市人口数量的平均值。该估计树是一种基于样例的模型,它记录了学习阶段呈现过的所有样例以及它们的线索值和人口数量。因此,当训练项很大时,它需要庞大的记忆资源(关于基于样例模型正、反两方面的论述,参见 Nosofsky et al., 1994)。

我们设定了关于估计环境的多种知识程度,在此基础上模拟了对人口数量的估计过程。共有 10 种训练项规模,即分别有 10%、20%、30%、40%、50%、60%、70%、80%、90%、100% 的城市(及其各自的人口规模)为已知的。在训练阶段,三种策略——最快者启发式、多元回归和估计树——尝试掌握数据(如,城市及其线索;权重, S_j^- , S_j^+ , 等等)的范型(或参数)。为了获取可靠参数,每种类型的训练项都抽取 1 000 个随机样本。例如,我们从 82 个城市中随机抽取出了 41 个城市(50%训练项),然后用它们形成 1 000 个样本。

在测验阶段,我们分别将上述三种策略应用于所有城市(即测验项,它包括了训练项)。这些策略的任务是估计所有城市的人口规模(假定这些城市在各条线索上的值是已知的)。为了使我们的模拟从心理学上看更为合理,我们假定一个城市属于训练项的概率与其人口规模是成比例的。该假设符合如下事实:人们更有可能了解较大城市,而不是较小城市。

最快者的节俭性如何?

设计最快者启发式是为了快速地作出估计。那么,在搜索被终止之前该种启发式需要考虑多少条线索呢?图 10-4 表明了对不同规模训练项来说每种策略必须提取出来的线索数量。平均来说,最快者启发式作出每种估计需要 2.3 条线索——这个数目在不同规模训练项间保持相对稳定。相反,不管训练项规模是大是小,多元回归总是使用所有 8 条适用线索。随着训练项规模增加,

估计树使用的线索越来越多——就所有 10 种训练项规模来看,它平均使用了 7.2 条线索。因此,最快者启发式作出估计所需要的信息大约分别为多元回归和估计树需要信息的 29% 和 32%。

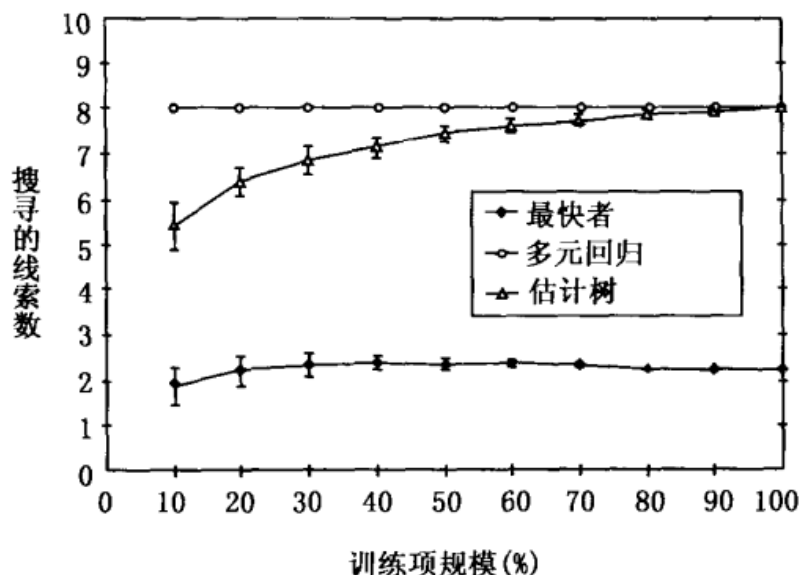


图 10-4 作为训练项规模的函数,最快者启发式、多元回归和估计树所搜寻的线索数目。垂直线代表标准差。

最快者的准确性如何?

最快者启发式在计算上使用了简单的均分和约简方法,而多元回归则使用了复杂精细的计算方法。那么,相对于多元回归来说,最快者启发式的准确性如何呢? 我们使用两种不同的准确性测定方法对三种策略的成绩作了比较。首先,遵循布朗和西格勒 (Brown & Siegler, 1993) 的做法,我们采用了一种最常用的测定估计准确性的方法,即(平均)绝对误(实际人口数量与估计人口数量之间的绝对离差)。其次,对全套对偶比较中的 $(82 \times 81)/2$ 种城市配对,我们模拟出了其选择(“两个城市中哪一个较大?”),然后计算正确推断的比例。

绝对误

以绝对误来衡量,最快者启发式由于仅考虑少数几条线索付出的代价是什么?图10-5显示的是作为训练量(即训练项的规模)之函数的绝对误。10%训练项代表了典型的知识匮乏情形(在多数领域这都可能是常见现象,而非绝无仅有的例外情况)。在这种情形下,最快者启发式的估计值平均产生了大约132 000的误差(约为本模拟情境下所有德国城市平均人口数量的一半),而多元回归和估计树平均产生的误差分别为303 000和153 000。也就是说,当在心理学意义上处于知识匮乏状态时,最快者启发式凭借较小的估计误差,取得了远远高于多元回归和估计树的成绩。

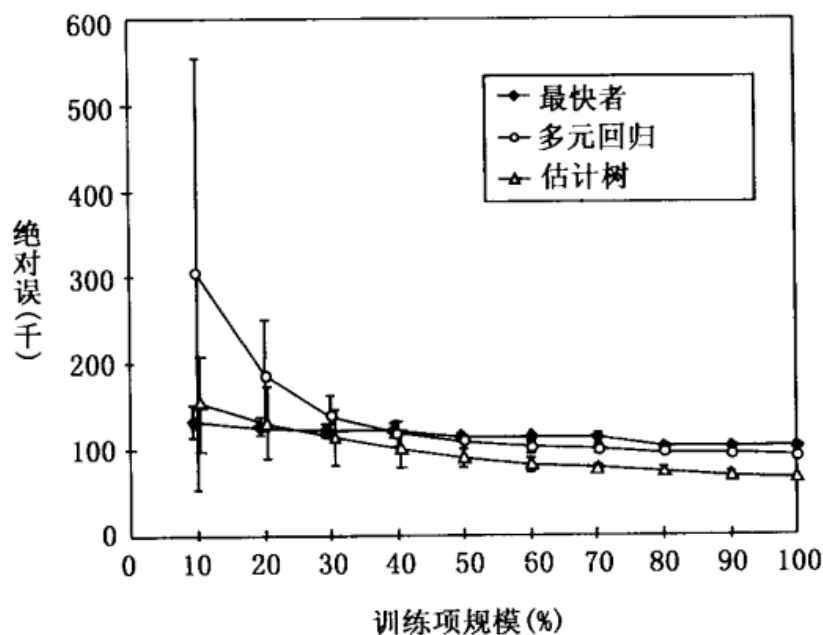


图 10-5 作为训练项规模大小之函数的平均绝对误(即预测人口规模和实际人口规模之间的绝对离差)。垂直线代表标准差。请注意:在水平方向上,为了使误差条框之间易于分辨,一些点略微有些偏移,但它们对应于横坐标上相同的训练项规模。

那么,估计成绩是如何因训练量变化(即已知城市的增多)而变化的呢?举例来说,当50%城市为已知时,最快者启发式和多元回归取得了大体相当的成绩,而估计树则以其较小的估计误差

简捷启发式

超过了这两者的成绩。当这些策略拥有完整知识(所有城市均为已知)时,多元回归以相对较小的估计误差超过了最快者启发式的性能——它们各自的绝对误差分别为 93 000 和 103 000。而估计树取得的成绩更是超过了它的两个竞争者(绝对误差约为 65 000,等同于最优化成绩)。这就是说,在拥有全部知识这种从心理学上看并非不可能的情形下,最快者启发式的性能仅略微落后于多元回归的成绩,而与估计树相比,差距就比较明显了。(甚至当多元回归仅使用那些权重明显不同于零的线索——平均 7.3 条线索而非 8 条线索——时,其绝对误差的改进也很小,如果将其绘制在图 10-5 中,几乎看不出有什么差异,当然 10%的训练项除外。)

231

这种结果与查特等人(Chater et al., 1997)报告的结果相类似。他们以四种计算起来相当复杂的策略——包括神经网络模型和范例模型——为竞争者,对采纳最佳启发式(第 4 章)这种快速节俭策略(最快者启发式的亲属)进行了检验,要求各种策略完成的任务是判断两个德国城市中哪个有更多人口。查特等人发现:当训练项少于测验项的 40%时,采纳最佳启发式取得的成绩超过了所有其他竞争者。只有当训练项超过了测验项的 40%时,竞争者的成绩才高出了采纳最佳启发式的性能。

最快者启发式在哪里犯下了实质性的错误呢?图 10-6 显示了每种策略估计的人口数量与实际人口数量之间的离差是如何因实际人口数量不同而变化的。每种策略都有一种明显不同的错误范型。最快者启发式非常准确地估计出了许多小城市的人口数量,但在估计极少数大城市人口规模时犯了实质性错误,这是由于它将这些大城市全都放入了“杂物存放器”。相反,多元回归在估计所有城市人口数量时都出现了实质性错误。估计树在对小城市和大城市人口规模进行估计时,均犯了相对较少的错误。

232

图 10-6 的另外一个方面也是值得注意的。与估计树和多元回归作出的估计不同,最快者启发式的估计呈现了向平均数回归的趋向:一般而言,它低估了大城市的人口规模,而高估了小城市

的人口规模。在人类数量估计中,此种回归效应是非常典型的(例如,Attneave, 1953; Lichtenstein et al., 1978; Sedlmeier et al., 1998; Varey et al., 1990)。在图 10-6 中,相对于对大城市的过低估计来说,对小城市的过高估计显然是微不足道的。然而,如果用预测人口规模和实际人口规模之间的离差除以实际人口规模,那么小城市的回归效应远远大于大城市的回归效应。在 100% 训练项条件下,最快者启发式、多元回归和估计树所作估计对所有城市人口数量中位数的回归分别为 56%、45% 和 23% (我们使用了在塞多梅尔等人的研究中所采用的分析方法)。因此,最快者启发式的回归量最接近于人们在其他任务上作出估计时产生的大约 70% 的回归量(Sedlmeier et al., 1998)。

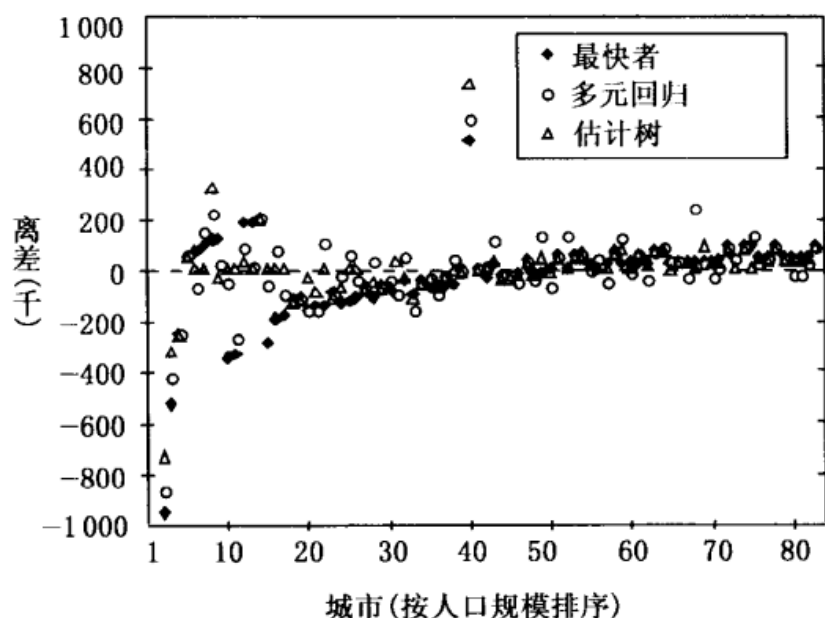


图 10-6 三种估计方法在估计所有城市人口规模时的估计值和实际人口数量之间的离差(100%训练项条件下),按照人口规模排列城市顺序。

最快者启发式仅将自发数作为它的估计值。如果多元回归也不得不在受到同样心理限制条件下进行估计,它将会付出多大代价呢?不妨回顾一下:在拥有全部知识(即所有城市均为已知)情况下,多元回归取得的成绩超过了最快者启发式(绝对误分别为

简捷启发式

93 000 和 103 000)。然而,如果多元回归也将其估计值四舍五入为最接近的自发数,那么它的成绩将不如最快者启发式的成绩(绝对误分别为 114 000 和 103 000)。

总之,尽管最快者启发式仅利用了其竞争者所使用信息的三分之一,且采用了比多元回归更简单的计算方法,但在知识匮乏条件下其成绩却超过了多元回归和估计树的成绩。此外,最快者启发式的成绩在不同学习量条件下是相对稳定的:10%训练项条件下的绝对误仅仅比拥有全部知识条件下的绝对误高 1.3 倍。相反,多元回归和估计树在 10%训练项条件下的绝对误分别比它们在拥有全部知识条件下的绝对误高出 3.3 倍和 2.3 倍。只有在拥有丰富知识(即 50%以上城市为已知的)——这种情况从心理学上看不太可能——的条件下,最快者启发式的成绩才(略微)被其竞争对手所超过。

正确推断比例

当比较成对城市时,该种启发式作出了多少正确推断呢?图 10-7a 显示了排除凭借猜测进行选择的情形(即由于为两个城市预测的人口数量相同,启发式只能随机选择其中之一的情形)后作出的正确推断比例,而图 10-7b 显示了包括猜测成分在内的结果。当不必进行猜测时,最快者启发式的成绩是非常好的:在所有训练项规模条件下,其作出正确推断的比例是 81%,而多元回归与估计树作出正确推断的比例则分别为 73%和 77%。 234

当所预测的两个城市人口规模相同时,每种策略都要随机地在两个城市中进行猜测,所以此种情况下有望获得的正确推断比例为 50%。由于该值低于无需猜测情况下每种策略取得的成绩,可以预期:当包括了猜测成分时,总体成绩降低了(参见图 10-7b)。而且,最快者启发式将会表现出更为明显的成绩降低,因为它可以用来作出不同估计的数目范围更小,不得不更加经常地进行猜测:在所有不同训练规模条件下,当包含了猜测成分时,最快

者启发式作出正确推断的比例为 66%，而多元回归和估计树作出正确推断的比例分别为 71% 和 75%。

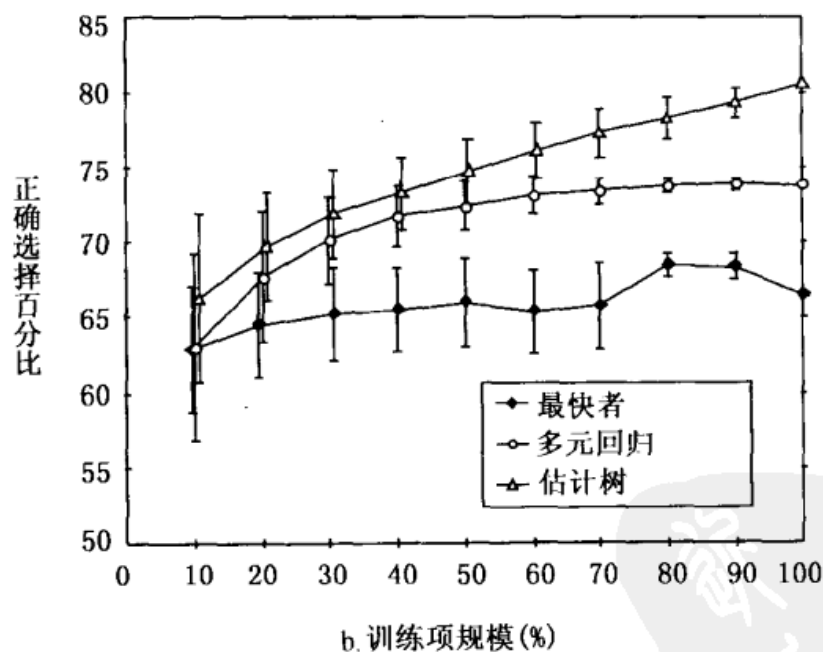
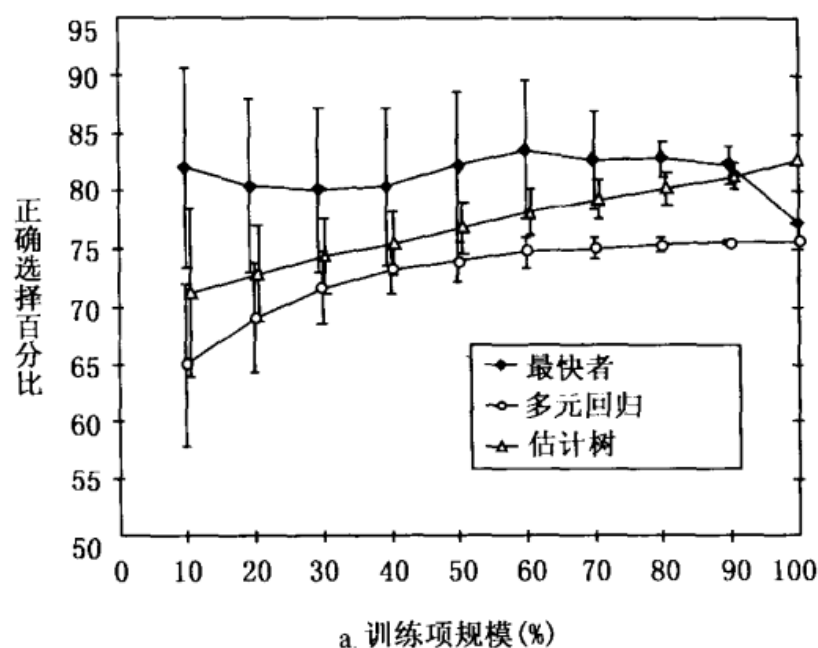


图 10-7 作为训练项规模之函数的正确推断比例,含排除猜测(a)和包括猜测(b)两种情形。垂直线代表标准差。请注意:在水平方向上,为了使误差条框易于分辨,一些分值点略微有些偏移,但它们对应于横坐标上同样的训练项规模。

结论

让我们像本章开始时所做的那样,用鲁滨逊·克鲁索的日记摘录作为本章的结束。一旦克鲁索意识到经常有野人光顾他的荒岛,他便开始为某一天可能发生的冲突做准备。一天清晨,他不由得大吃一惊:

看见至少有五只独木舟紧靠在一起泊在我这一侧的岸边。船的主人已经全部上岸,不知去向。他们的人数一下子使我的计划全部落空;我看见过他们许多次,知道他们来时每只船上一般都是四个或六个人,有时候还要多些。一时间我脑子里一片空白,不知道如何实施我的计划,赤手空拳去攻击二三十个身强力壮的男人;因此我只得老老实实地待在我的寨子里,感到一筹莫展,颇为沮丧。(Defoe, 1719/1980, p. 198)

对于许多进化上相当重要的任务——从选择何处去觅食,到决定是否去争斗——来说,富有适应性的行为部分地依赖于有机体估计数量的能力。此类决策常常不得不在短时间内迅速作出,而且只能依靠不完整的信息。快速节俭估计启发式利用现实世界中何种信息结构作出准确估计呢?我们提出了一种启发式,即最快者启发式,它利用了一种特殊环境结构,即J型分布。我们通过模拟研究证明:在知识匮乏情况下——自然决策情境中通常都是如此(例如, Klein, 1998)——快速节俭的最快者启发式取得的成绩好于、至少匹等于多元回归和估计树的成绩。最快者启发式是一种从生态学上看合理的策略,它的成功从一个侧面突出了研究环境结构的重要性。

第11章

排除归类法——用少量线索做选择

帕特丽夏·M·贝雷蒂(Patricia M. Berretty)

彼得·M·托德(Peter M. Todd)

劳拉·马蒂冈(Laura Martignon)

归类,或者死亡

——F·格利高里·阿斯比(F. Gregory Ashby)

当你徒步漫行于巴伐利亚州的阿尔卑斯山上时,偶尔遇到一只栖息在篱笆上的大鸟。你拿出你的《欧洲鸟类手册》来辨认它。从它的头形,你推断这是一只食肉鸟,于是你翻到手册上介绍猛禽的部分。然后你根据身体大小把搜索范围缩小到几种类别的鹰上,接下来根据颜色进一步限制了搜索的可能性。当你考虑到鸟尾长度时,你终于获得足够多的信息来确定它的类别。仅使用四种特征(线索),你最终正确地把这只鸟确认为鹞。

现在,设想你发现一只鸟正在头顶飞翔,而不是栖息在篱笆上。它飞得很快,而且阳光映照在它的背上,因此你只能觉察到它的少数几种特征。你首先注意到的是它的大小,其次是它的侧面轮廓,最后是它多姿多态的滑行动作。再一次,你仅利用有限的信息就能够把这只鸟归类为鹞,但这一次使用了一套不同的线索,而且比鸟静止时用的线索要少。

如果是一只野兔而不是人在观察这只鸟,那么归类过程又如何进行呢?野兔对在头顶飞翔的鸟属于什么鸟类不感兴趣,相反它只想尽快知道这只鸟是不是食肉动物。在野兔的“鸟类手册”

让我们更精明

(如果有的话)上只有食肉动物和非食肉动物这两个类别。虽然野兔也可以利用多条线索作出类别判断,但是一旦它找到了足够信息——如,鸟是否在滑翔——确认鸟为食肉动物,它将不会再操心费神去搜集更多信息,而是马上奔向隐身之处。很显然,对于野兔来说,在对鸟作出食肉动物或非食肉动物的区分时,速度是至关重要的。

像动物一样,人类也时常面临着需要迅速作出归类的情境,而且只能利用立即可以得到的少量信息。例如,迅速地把逼近自己的动物或人的意图归为善意的或是敌意的可以使我们有足够时间作出适当反应,以确保满意的后果。事实上,归类本身可以被视为迅速作出关于环境中某些事物之决断的一种手段:观察了某种事物的少数几种特征之后,准确地将其归入某种类别,可以使一个人超越这些特征作出推论,并对该事物的其他特征作出预测(Ander-son,1991; Medin et al., 1987; Rosch, 1978)。如果说归类的目标在于预测,那么下结论前必须对所有相关特征进行详尽搜索的归类过程将使该目标无法实现。也就是说,假如每一种特征都需要花费时间去寻找或观察,那么需要利用所有特征的归类过程将会比无需利用所有特征的归类过程更为缓慢。因此,快速归类对于以归类为基础的预测来说最为有利。但是如何才能实现快速归类呢?我们提出了一种快速节俭的算法规则:一旦找到足以据其作出决策的信息,便立即终止对特征的搜索。该规则将实现我们的目标!

在本章中,我们考虑此种快速节俭归类模型得以运行的一种环境,它类似于第一段所描述的基于线索排列顺序的归类过程。我们将这种模型称为“排除归类法”(categorization by elimination,或简称CBE; Berretty, 1997)。它仅使用必需的适用线索或特征作出一种具体归类。结果是,在对某个给定对象加以归类时,它所使用的线索常常要比需要整合线索的标准模型少得多——因此它是快速而节俭的。尽管如此,但正像我们将会证明的那样,这

种方法的准确性通常一点也不比涉及更多计算过程的算法规则差。因此,我们把排除归类法作为一种节俭的心理学模型来对待,并将其作为一种完成某些应用性归类任务可供选用的有用模型。

在下面的部分中,我们首先描述了心理学和机器学习领域可作为排除归类法竞争对手的归类模型。由于排除归类法与特威斯基关于选择过程的“逐项排除模型”(elimination by aspects model; Tversky, 1972)密切相关,所以我们对排除模型的主要假设进行了较为深入的讨论。接下来我们对 CBE 模型作了详细的介绍和说明。近期多数关于人类归类过程的其他模型均是在仅有两三条线索的十分有限(小范围内)环境中得到检验的,此种环境中 CBE 的优势难以表现出来。因此,我们的研究使用了“多线索归类任务”(multiple-cue categorizing task),在这种任务上我们能够依据解释人类成绩方面的成效将我们的模型和其他模型加以比较。我们的研究涉及到了利用 7 种线索维度将机体的动作轨迹归入不同意图类别。在叙述研究结果(它被更为详细地包含在第 12 章中)的第 5 部分中,我们简单地谈到了该项研究。在那里 CBE 像线性归类方法完成得一样好,而且不像神经网络模型有时所做的那样——对有关资料匹配过当。在那一部分中我们还考察了相对于心理学和机器学习领域(依据于来自该领域的标准数据)出现的一些线索整合归类方法而言,我们的算法规则做得有多好。这种比较显示:即便在使用更少线索情况下,CBE 在准确性上也与许多领域较为传统的方法相匹等。最后,我们考虑了发展快速节俭归类模型仍然面临的一些挑战。在本章的剩余部分中,我们使用了“线索”(cues)、“项目”(aspects)、“维度”(dimensions)、“特征”(features)等术语作为适当概念,它们大致上代表同样的意思。

现有归类模型

在心理学和机器学习领域归类模型本身的类别还需要进行分

简捷启发式

类。心理学家主要关心制定一些能够最好地描述人类归类成效的模型,而机器学习领域研究的目标是发展达到最高归类准确性的模型。这两种目标并不一定是相互排斥的;实际上,迄今人类归类研究领域的一个主要发现是:人们常常能够达到接近完美的准确程度(即:以最少的错误对一套刺激作出归类;参见 Ashby & Maddox, 1992)。结果是,一些模型(如,Miller & Laird 的符号概念获得模型,1996)的目标就在于实现这两种功能。

归类模型可以按照第1章介绍过的启发式构筑块的三个维度加以分析:一种特定模型是怎样搜索信息以作出一种归类判断的?该模型何时终止信息搜索?一旦它拥有了所需要的信息,它将采用什么样的决策过程?然而,此种分析并不总是能够使我们走得很远,因为在心理学文献中研究者们很少将注意力集中在人们是如何获得所需要信息的这个问题上(前两个维度)。有关归类问题的多数心理学研究使用了仅在几个(两到四个)维度上变化的简单人工刺激(如,两维空间内的半圆形;Nosofsky, 1986)。* 由于所有(有限的)信息均已经提供给被试,所以没有必要去搜索信息,当然亦无需终止搜索信息——只有决策规则需要明确化和具体化。当对有多种维度的对象加以归类时,仍然需要证明人类做得有多么准确。这同机器学习领域所用的更为自然的多维归类应用问题——如,品酒(Aeberhard et al., 1994)或笔迹辨认(Martin & Pittman, 1991)——形成了鲜明对照。在这一领域,适当的信息搜索可能是非常重要的。这里我们本着三种启发式构筑块的想法,对当前在人类研究和机器学习领域相当流行的一些归类模型加以简要回顾。但在许多情况下我们必须作出推测或默认有关做法,因为在充分考虑这些构筑块所体现的时间、记忆容量和计算能力的有限性时,我们将难以设计出任何一种算法规则。

* 波斯纳和基勒(Posner & Keele, 1968)曾经使用多维的随机圆点刺激来测试人类的归类能力,但其确切的维度数量是难以确定的。

心理学文献中的归类理论主要以“范例模型”(exemplar models; 例如, Kruschke, 1992; Nosofsky, 1986)、“决策一分界模型”(decision-bound models; 例如, Ashby & Gott, 1988)和“基于规则模型”(rule-based models; 例如, Miller & Laird, 1996; Nosofsky et al., 1994)为代表。这些模型假定任何一个对象都可以被表征为线索值(可能有些不确定性)多维空间内的一个点。而且, 这些模型还假定在归类决策过程中人类整合了所有特征——也就是说, 我们将多种来源的线索结合在一起以作出最终判断。由于依赖于线索整合过程, 所以正像后面我们将证明的那样, 当只有很少线索可以利用时, 这些模型的成绩会蒙受巨大损失。最后, 这些模型假定归类者使用了现有的所有线索——也就是说, 不存在明显的终止规则来限制信息搜索, 我们不能忽略任何可以利用的信息。当有许多线索可以利用时, 这将导致时间的紧迫性。为了进一步说明心理学归类模型的某些重要特点, 下面我们将罗索夫斯基(Nosofsky)的“概化场合模型”(generalized context model)和阿斯比(Ashby)的“决策一分界理论”(decision-bound theory)作一比较。

范例模型(Brooks, 1978; Estes, 1986; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1986)假定: 人类利用所有适用线索对一个新对象与其可能隶属的每一个类别包含的所有范例之间的相似性加以计算, 从而对其作出归类。因此, 不存在任何终止信息搜索的规则。有两类主要决策规则可以采用: 按照类似性尺度将一个对象确认为它与之最为相似的那个类别, 或者用标准化相似值作为可能性基准将其安排到某个类别(正像在多数范例模型中所做的那样)。罗索夫斯基(1986)的“概化场合模型”(GCM)设想: 归类过程中给予不同特征的注意量是各不相同的(亦可参见 Medin & Schaffer, 1978)。由于一些线索被给予了权重为0的注意量, 被忽略了, 所以不同线索可能在不同任务上得到利用。然而, 得到利用的特定线索及其权重对于每一种特殊归类

简捷启发式

任务的整套刺激来说都是一样的。(相反,正如我们将要看到的那样,CBE 可以利用一套不同线索来对每一个新对象进行归类。)如果时间紧迫,以至于利用该种模型的人无法对所有适用信息加以评估,那么这些权重可以被用来排列线索的搜索顺序——尽管坚持范例模型的理论家们并未明确强调,但他们的一种潜在假设是:具有较高注意量权重的线索可能优先于具有较低注意量权重的线索而得到评估,并被利用。 239

决策一分界理论(或 DBT;参见 Ashby & Gott, 1988)假定:在多维空间内每个类别都有一个与其密切相关的特殊区域,因此不同类别被区域之间的界限所分开。决策界限由每个类别内各刺激点的平均值、方差和协方差来确定。DBT 按照一个对象所隶属的知觉空间区域,并利用一种决定性反应规则,对其作出归类。当归类将要被完成时,如果一些线索不适用,那么原来已经习得的知觉空间(由所有适用线索来确定)的低维部分被使用。这里再一次不存在明确的搜索顺序或终止规则,而且决策是通过联结所有适用线索值作出来的。

虽然这些心理学模型均借助于对所有适用线索(除了范例模型中一条线索的注意量权重为 0 的情况外)的整合作出归类,但他们对记忆容量的要求却是不同的。GCM 假定所有曾经接触过的范例均得到储存,并在对一个新对象进行归类时得到利用,而 DBT 只需要储存用来确定每个类别界线的参数。在本章最后一部分我们将回到线索的记忆和权衡问题。

基于规则模型采用了一种不同的归类方法。米勒和莱德(Miller & Laird, 1996)的“符号概念获得模型”(symbolic concept acquisition, 缩写为 SCA)逐步建立一套规则,以便按照具体的离散特征对刺激加以归类。最初建立起来的规则是很笼统的,只检测一种单一特征。如果当前的一套规则不能区分有关类别,那么就要增加更为详细的规则,它必须依据更多特征来匹配有关刺激。在采用这种方法学会一套规则之后,当遇到新刺激时,就要首先用

涉及所有适用特征的具体规则来检验它们。如果在所有觉察到的特征和具体规则之间不能一一匹配起来,那么这些特征就要被相继放弃,直到发现能够匹配剩余特征的一种较为通用的规则为止。因此,搜索线索的顺序、终止搜索的条件以及最后的决策均借助于模型的规则而明确化。这种方法与排除归类法之间有一些相似之处,特别是其加工特征的顺序与我们的线索效度测定有关。二者之间的主要差别在于:在归类过程中,CBE 是从某种单一特征开始的(而不是所有适用线索),如果必要的话才增加新的线索,因而减少了计算量。另外一种用更有效的“从一般到具体”(general-to-specific)顺序对规则加以检验的模型是菲根鲍姆和西蒙(Feigenbaum & Simon, 1984)提出的“EPAM 符号辨别网络模型”(EPAM symbolic discrimination-net model);然而,它与我们所考察的两种方法之间的明显差异无关。

在发展他们的“规则加例外”(rule-plus-exception, 缩写为 RULEX)模型时,罗索夫斯基、鲍默里和马肯莱(Nosofsky, Palmeri & McKinley, 1994)充分考虑到了其心理学的合理性和计算上的简便性。在“RULEX”中,有关双值特征维度的简单逻辑规则被学会了,它把特殊的特征值与特殊的类别匹配起来。然而,这些规则是不够完善的,所以还必须记住那些例外情况。罗索夫斯基等人发现,在他们所研究的资料中准确归类可以利用很少几种得到储存的规则加例外而实现。这种简单方法还能够解释归类成绩的个体差异。然而,“RULEX”是针对被双值线索限定了的两种类别情景而设计的(虽然现在正在得到拓展);我们的目标是寻找更为通用的模型。

在机器学习领域,占主导地位的归类理论包括“神经网络模型”(neural network)、“贝叶斯模型”和“决策树模型”。这些机器学习模型的目标通常是使给定的一套有用资料的归类准确性达到最高。发展机器学习模型时,限制模型的复杂性通常并非最为重要的事情,因此常常导致从心理学上看不甚合理的、计算上相当复

简捷启发式

杂的模型。

以神经网络为基础的归类和分类模型(如,可参见 Hertz et al., 1991)能够洞悉类别之间的超平面(hyperplane)界限,它在各个类别可训练的权重上猎取这种知识。某种特定刺激所有适用线索的值通常借助于一种非线性决策(激活)规则得到加工,以便确定该刺激在学习(训练)和推论(测验)阶段均落入其中的多维空间的区域或由该区域界定的类别。神经网络通常被设想为平行加工系统,因此它们有望同时拥有所有适用的输入信息,毋需搜索或终止规则。一旦用一套特定线索对某种网络进行了训练,那么测验阶段任何线索的丢失都将导致归类成绩的下降;然而,此类分区模型的特点之一是“优雅退步”(graceful degradation),即成绩随着输入信息的减少而逐渐降低,而非陡然降低。

解决归类问题的贝叶斯方法提出了一个概率判断问题:在给定对象线索值及有关线索效度知识前提下,设法找到一个类别,以便使一个新的(未受训练的)对象隶属于它的概率达到最高。朴素贝叶斯归类者假定:所有线索在所关心的类别上均是相互独立的,这就大大地简化了安排每一条线索时的计算过程。所创立的归类模型的结构确定使用哪些线索,但是这些线索的顺序未被明确化和具体化。已经证明:在一些归类领域,朴素贝叶斯方法的成绩接近于神经网络和决策树的成绩(Blum & Langley, 1997; Domingos & Pazzani, 1996)。安德森(1991)曾经提出过一种朴素贝叶斯方法,并将其作为人类行为的标准模型。他原先并未打算把他的贝叶斯方法作为说明心理机制的一种模型,因为它所涉及的计算过程实在太复杂了:人们甚至难以处理每次涉及到一条线索和一个类别以上的条件概率。但是安德森分析了多种多样的实验结果,发现:人类的归类行为接近于该朴素贝叶斯标准的行为。尽管朴素贝叶斯模型从心理学上看或许是不合理的,但完整的贝叶斯方法——即充分考虑到所有相互依从性的贝叶斯模型——甚至是更不合理的。(参见第8章关于此种神灵方法复杂

性的讨论。)

在归类的决策树模型中,主要目标是通过减少作出归类时所需要的维度数量来简化实际归类过程。但是为了实现此种简化决策过程的目标,通常在决策树的多个支点都需要利用非常复杂的方法来选择拟使用的维度。例如,在昆兰(Kuinlan, 1986, 1993)的“ID3 决策树”的建立阶段,需要设法找到能够为判定正确类别提供最多信息的特征,并将该种特征作为决策树的根(即第一个决策支点)。决策树的剩余分支通过反复计算随后每一个决策点最有信息量的特征加以选择。一旦一种决策树被建立起来了,它的结构就决定了线索的确切数目和顺序,随后必须依据这些线索及其特定顺序作出最终的类别判断。所有决策都是在决策树限定了的特殊路径连续支点上作出来的。

布里曼等人(Breiman, 1993)的“归类 and 回归树”(CART)模型需要在成对特征之间进行更为繁琐的统计检验,以便确定哪种特征被用在决策树的哪个位置。对 ID3 和 CART 来说,先前检测过的特征值可能会影响下一次寻找什么特征,以便不同对象用一套不同特征加以归类(参见第 1 章开头提供的关于心脏病患者归类的例子)。在一种归类决策能够被作出来之前,必须进行大量计算来建立这些模型(即使随后的每一种决策本身是简单的),这使得它们从心理学角度看是相当不现实的。郝特(Holte, 1993)只考虑一种线索维度,将其分成不同区段,每个区段对应于一种特定类别,从而大大简化了决策树方法。如果有足够多区段的话,这种简单方法可能是十分准确的,但它对记忆容量的要求甚高。

一些研究者试图提出具有心理学合理性的机器归类模型,例子之一是根纳雷(Gennari, 1991)的“CLASSIT”。它将选择性注意应用于无人管理的概念形成系统。该系统使用必需的少数几种最突出线索对一个对象作出初步归类。然而,在作出最后决策之前,所有线索仍然必须得到考虑,以确保进一步的信息不会改变最初的归类。此种检测线索的“最糟情形”(worst case)终止规则是

第1章所界定的“受限制条件下最优化”的表现形式之一。

尽管许多机器学习模型(如,CART 和 CLASSIT)在给出一种归类时仅使用了少量线索,但事先建立算法规则决策机制以及决定使用哪些线索的过程却是相当复杂(如,CART 使用了 t 检验——参见 Breiman et al., 1993)或耗费时间的(如,神经网络的训练迭代)。相反,我们能够设计出一种易于建立的、仅利用有限线索和简单决策规则的、同时又有高度准确性的启发式归类方法吗?

242

排除模型

出于上一部分所提出的心理学合理性方面的考虑,加之人类和其他动物常常必须在很短时间内依据少量线索对物体作出归类,我们试图发展一种快速节俭的归类方法。这种方法集心理学和机器学习领域现有归类模型的优点于一身:简单的决策过程和有限的线索利用。此种特征的结合在被称为“排除模型”(elimination model)的心理学理论中得到了充分体现。

排除模型最初是为选择和偏好判断任务而设计的(Restle, 1961; Tversky, 1972)。在“相继排除选择模型”(sequential elimination choice model)中,不断根据进一步的考虑从一套对象中排除其子集,逐步减少剩余子集存在的可能性,直到某种单一选择被保留下来,作为从一套对象中选择出来的一个对象。子集中的对象是由许多特征来界定的。为了排除第一个对象子集,采用一些方法挑选出一种特征,将具有这种特征的所有对象从可能的子集中排除出去。采用同样方法相继挑出具有其他特征子集,将其排除掉。经过此种排除过程仍然得以保留的对象就是最后的选择。这样,排除模型必须指定一种搜索顺序,一旦仅剩下了单一选择,立即终止信息搜索,因此它们的终止和决策规则是密不可分的(像在决策树模型中一样)。

心理学中广为人们所知的排除模型是特威斯基(1972)关于概率选择的“逐项排除模型”(EBA)。将 EBA 发展为关于选择活动描述性模型的促动因素之一是:经常有许多相关线索可以被用来在复杂选项间作出选择(Tversky, 1972)。因此,特威斯基想要用他的模型把从心理学上看合理的程序整合在一起,去挑选和排列从多种选项中作出选择时使用的线索。在 EBA 中,根据线索或项目对作出特定选择的有效性,排出一种可能性顺序,并据此对它们加以选择(如,若想从附近的许多餐馆中选择一个,最重要的项目是它的服务特色及价位),把与当前正在考虑项目不一致的可能性(如,不供应海鲜的餐馆)从选项中排除掉。而且,只有剩余选项中包含的项目才得到进一步考虑(如,如果附近所有海鲜餐馆均不太贵,那么价位就不能再被用作区分它们的项目)。另外附加的项目被用来排除剩余的可能性,直到仅留下一个选项为止。这种仅使用必要项目的做法与上一部分描述的使用所有适用线索的归类模型截然不同。(参见第 7 章关于 EBA 和其他基于排除过程选择启发式的更多论述。)

243

从选择过程到归类过程,排除模型随着目标的变化而得到了发展。汤森德和阿斯比(Townsend & Ashby, 1982)的“序列特征检测法”(sequential feature testing algorithm)是在字母识别过程中使用视觉特征进行决策的一种排除模型。像逐项排除模型一样,每一种特征都要被加权(这里使用与人类成绩最接近的值)。然而,不再从当前选项中选择下一个项目,而是将具有最高权重的项目作为下一个选项。如果该项目在所呈现的字母中被觉察到了,那么不包含该项目的所有剩余字母从选项中排除掉[例如,如果在字母“b”里发现了曲线部分,那么没有曲线部分的所有字母(包括“x”、“k”等)将从进一步的考虑中被排除掉]。同样地,如果该项目未被觉察到,那么所有含有该项目的字母被排除掉。项目逐一得到检测,直到仅有一个字母被保留下来。假如所有字母都被排除了,那么就随机地作出一种选择。

简捷启发式

需要重申的是,潜藏于排除模型(当然也包括 CBE)背后的假设主要有两个:首先,所使用线索必须采取一些方法加以排列(即便只是一种可能性),即必须有一种搜索顺序构筑块。其次,不必总是使用所有线索来作出一种决策——当唯一的决策被确定了时(或当没有更多线索可利用时),简单终止规则中断了搜索。尽管这些假设尚未在归类任务上得到明确检验(据我们所知),但它们已经在有关其他决策形式的多种研究中得到了检验。

先前的研究已经表明人们是按照线索对完成特定任务的有效性顺序来加工线索的。有关假设检验中信息搜索过程的研究表明:人们更喜欢使用具有“诊断力”(diagnostic)的线索,而非不具有诊断力的线索(Skov & Sherman, 1986; Trope & Bassok, 1982)。一条双极线索在区分两个假设时的诊断力被界定为:该条线索存在时,第一个假设为真的概率与第二个假设为真的概率之间的比率。因此,一条线索的诊断力表明了它在多大程度上能够将两个假设鉴别开来。排除归类法能够使用多种方法将线索从最无效到最有效排列起来(参见第 6 章关于不同类型线索顺序的更多论述)。这里我们根据线索得到使用时作出正确归类的频次数排列它们的顺序。

关于第二个假设,有关研究已经表明人们在多种决策任务上均未使用所有适用于他们的线索(参见第 7 章)。萨德和鲁索(Saad & Russo, 1996)曾经调查了人们使用多少条线索在两套公寓之间作出选择。当允许他们按照自己喜欢的顺序挑选线索时,他们仅使用了候选线索的 1/4(中数为 24%);即便在实验者控制了线索呈现顺序条件下,作出决策前人们仍然只需要大约一半线索(中数为 52%)。在另外一项决策任务上,人们宁可根据有限的线索作出一种他们并不完全确定的决策,也不愿意进一步检测适用于他们的更多信息(Jacoby et al., 1994)。就我们目前的目标来说,最重要的发现或许是:如果线索有较高诊断力,那么人们使用较少线索;如果线索有较低诊断力,那么人们使用较多线索

(Van Wallendael & Guignard, 1992)。这些发现进一步表明：当人们能够按照自己期望的特殊顺序评价线索时，他们通常使用较少线索。

排除归类法

为了依据排除过程建立一种归类模型，我们首先提出下面的一般假设：待归类对象拥有可觉察的线索，这些线索具有连续的或离散的值。对于给定的一项任务来说，有一套可能的类别，可以依照各个对象的线索值分别将它们归入其中。与特威斯基有关偏好问题的“逐项排除模型”(EBA)和有关归类问题的“概率匹配模型”(probability-matching models)不同，这里我们关心的是：待确定的正确答案在哪里？何处我们能够依据一种外部准确性标准测定所提出算法规则的成绩？结果是，我们的模型是一种确定性模型，而非概率模型(除了需要作出猜测时)。而且，由于毋需将一个对象的所有线索值与记忆中储存的某种线索配置确切匹配起来，以准确识别该对象，所以我们的模型具有概括化的特点。这意味着在某些场合我们的模型也许不能正确预测一个物体所隶属的类别。

更具体地说，CBE 的运作要领如下：为了归类某个对象，该对象在多种线索维度上的值按照事先确定的特殊顺序得到加工。所有对象均使用同样的线索加工顺序，但由于模型的排除性质，并非所有对象都要加工所有线索。每一条得到加工的线索均对保留下来以供进一步选择的一套候选类别产生限制，通过进一步考虑，从中排除一些类别后，作出最终选择——这就是所谓的排除过程。(请注意：一条给定线索有可能使候选类别数目保持不变，但不能增加候选类别数目。)当只有一种类别保留下来时，或没有更多线索可以利用时，终止线索加工过程。在后一种情况下，从剩余可能性中随机作出一种选择。在给出了上述假设和运作要领之后，

简捷启发式

下面我们就能够具体考察排除归类法究竟是如何运作的以及它的成绩如何。

排除归类法的一个例子

为了说明 CBE 模型是如何运作的,我们可以将其应用于确定酒类品种的任务上去——利用品酒过程中搜集到的一套线索,确定一种酒是 zinfandel 酒、Pinot Grigio 酒,还是事先指定的一套可能性中的另外一种酒。我们首先提供一套可供考虑或选择的类别(品牌)和一套可用来将酒归入这些类别的线索。其类别(品牌)和线索值均已知的一些酒被用作训练项。我们的启发式规则首先对这些训练项进行加工,此后对一些新酒作出归类。鉴于我们可以使用多种可能线索对酒进行评价,这里我们只考虑四条明显的线索:甜味、色泽、芳香、有无酒精。首先,我们的启发式需要一条规则来给出搜索方向,即用来确定首先检测哪条线索,其次检测哪条线索,如此等等。正如前面提到的那样,单独使用时作出更多正确归类的线索应当排在作出较少正确归类的线索之前。就品酒任务来说,我们设想色泽是最佳线索,随后依次为芳香、甜味和有无酒精。 245

其次,对已知其品牌的酒的各条线索值加以分析,以便将每条线索维度划分为与酒的品牌数相对应的“区段”(bins)。例如,训练项中所包含的酒的线索值分布或许可以把“色泽”维度分成三个“区段”:一个区段拥有从浅黄色到金黄色的线索值,一个区段拥有粉红色线索值,最后一个区段拥有深红色线索值。(在下一部分中,我们将讨论如何根据训练项的有关数据划分这些区段。)每一个线索值区段都有一套与其相对应的类别。举例来说,整个红色线索值区段可能对应于 Beaujolais、Burgundy、Cabernet Sauvignon、Pinot Noir、zinfandel、port、merlot 等酒类品牌。(更细微的区段划分可以增加归类准确性,但就目前的任务来看,我们认为三个区段已经足够了。)*“芳香”维度可能包括直接与线索值相对应的几个区段:树香型的、泥土香型的、草香型的、水果香型的和无

法指认型的。为了划分这些区段,我们对已知其品牌的酒作了分析,根据这种分析可以将无法指认的芳香与 Sparkling wine、champagne 和 Beaujolais 等品牌对应起来。(请注意:这些类别可能也出现于同一线索维度的其他区段——如,假如在我们所知道的品牌中有两种类型的 Beaujolais,一种是水果香型的,一种是无法指认型的,那么 Beaujolais 就可能落入芳香维度的无法指认型和水果香型两个区段内。)

一旦排列好了线索顺序,并将线索值的范围分成了代表不同类别的区段,就可以利用这种知识对新酒进行归类了。设想我们面前放着一个盛满酒的神秘玻璃杯,我们想要使用 CBE 来确认这是什么种类的酒。我们首先觉察到了颜色线索,看到酒是深红色的。因此,我们知道这种酒一定包含在颜色线索维度的深红色区段可能对应的酒的类别中,这些类别包括:Beaujolais、Burgundy、Cabernet Sauvignon、Pinot Noir、zinfandel、port 和 merlot。这样,颜色线索已经把许多酒的品种排除在进一步考虑之外,如 golden Pinot Grigios 和 pink roses 等。接下来我们用鼻子嗅酒的气味,发现它不容易鉴别,因此芳香线索表明这种酒应该属于无法指认香型对应的酒,即 sparkling wine、champagne 或 Beaujolais。我们几乎排除了前面的深红色线索确定下来的所有酒类,看看还有哪些选项可以保留下来。它实际上就是两条线索各自对应酒类的交集。当我们这样做时,我们发现只有一种品牌的酒同时匹配于两条线索,那就是 Beaujolais。因此,我们推断待归类的这种酒属于此种类别。如果既属于深红色、又属于无法指认型的品种有好几种,那么我们就继续检测下一条线索——甜味,看看它能否进一步缩小可能性。我们可以用排列等次更低的线索继续该过程,直到仅留下唯一一种类别或用尽了所有线索为止,后一种情况下我们不得不在仍然剩下的类别中进行猜测。

算法规则

排除归类法是归类过程的一种快速节俭的、不可代偿的、基于

简捷启发式

线索的模型。这意味着它按照一种特定顺序使用线索,而且由前面线索确定了类别不能够被后面线索所改变(即代偿)。在CBE中,线索是按其成功几率得到排列和使用的。我们把成功几率界定为一条单独线索准确地对一些刺激作出归类的百分比(参见第6章关于排列线索多种方法的更多论述)。仅使用待检测的单一线索运行CBE,然后看一看该种算法规则作出的正确归类比例,我们就能够很容易地算出这个百分比。如果用单一线索使得CBE不能区分一种特殊刺激的类别,就像在多数情况下一样,那么该种算法规则就随机地从这些类别中选择一个类别。举例来说,在对鸟进行归类时,如果个头大小线索比形状线索更为准确(或者说在缩小可能类别范围方面更为成功),那么大小线索将比形状线索具有更高的成功几率,因此也将被CBE优先使用。

正如上一部分所描述的那样,为特定任务建立CBE涉及到了识别有关线索维度和把每个维度分成多个区段。每个区段都涵盖了对应于一定类别的连续线索值(称名的或数字的)范围。这些区段形成了CBE用来把线索值和可能类别一一匹配起来的知识基础。要么可以在测定CBE归类成绩前将区段充分地建立起来(就像我们这里所做的),要么可以逐步地进行区段学习和归类测验。在这两种情况下,区段的建立都是通过确定每一种线索维度上各类别的高、低线索值界限来实现的。然后利用这些界限把每一个线索维度分成若干个代表不同区段的线索值范围(参见图11-1)。这意味着CBE具有非常明确和节俭的记忆需求:它只需要为每个线索维度的每种类别贮存两个值,不管遇到的对象数目有多少。 247

我们曾经探索过两种方法,以便确定与每种类别相联系的高、低两种线索值界限。首先,在假定特定类别线索值呈正态分布的情况下,我们可以使用比平均值低1或2个标准差的数值作为该类别的线索值下限,并使用比平均值高1或2个标准差的数值作为该类别的线索值上限。此种“高斯式”(Gaussian)分段法(在论述单一标准差界限的下一部分中,我们介绍的第一个例子就使

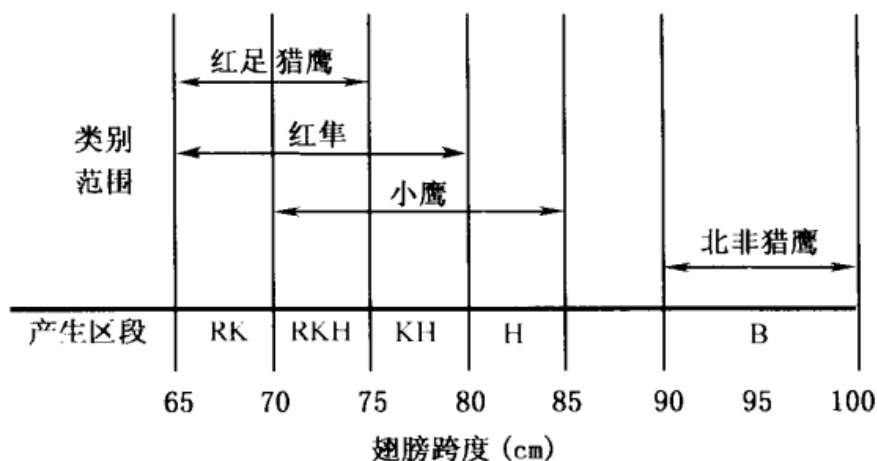


图 11-1 一种线索维度上线索值区段的建立。这里四类鸟的翅膀跨度维度被分成六个区段(见维度线的下方,对应于不同的鸟类)。区段间的界限由每类鸟翅膀跨度分布的上限和下限来确定,正如垂直线所表示的那样。

用了此种方法)使一个特定类别的多数成员落入一个特定线索维度内。但我们也可能发现该类别中的对象具有超出此范围的线索值。这意味着此种分段结构对在该维度上处于非中心位置的对象作出了错误归类;然而,其他维度依然能够对这些对象作出正确归类。

其次,在一些特殊情况下,如果类别的线索值被认为是不均匀地分布的,那么该类别的线索值下限就只能是其最低线索值,而上限只能是其最高线索值。在这种情况下,线索值区段将把一个特定线索值和与其相关的全部可能类别(在训练项内)匹配起来,而高斯式分段只匹配特定线索值最典型的类别。这种“完全分段法”(perfect binning; 也将在下一部分使用)将不会对训练阶段(区段建立阶段)见到过的对象作出错误归类,除了最终需要进行猜测的情况外。因此,相对于高斯式分段法而言,完全分段法通常在训练项上将产生更高的归类准确性。

CBE 怎样使用建立起来的线索值区段对一个特定对象作出归类呢? 它所采用的加工步骤流程图见图 11-2。当需要对一种特定刺激进行归类时,我们从一套初步的候选类别 S 和一个排列

简捷启发式

好顺序的待用线索维度表入手。首先挑出具有最高成功几率的线索维度 C , 归类过程就开始了。接下来, 使用已经建立好的区段将该刺激在维度 C 上的值与其所对应的候选类别中的一套 S^* 匹配起来。如果只有一种类别对应于该线索值(即: 如果 S^* 中只包含一种类别), 那么归类过程就结束于这种单一归类。

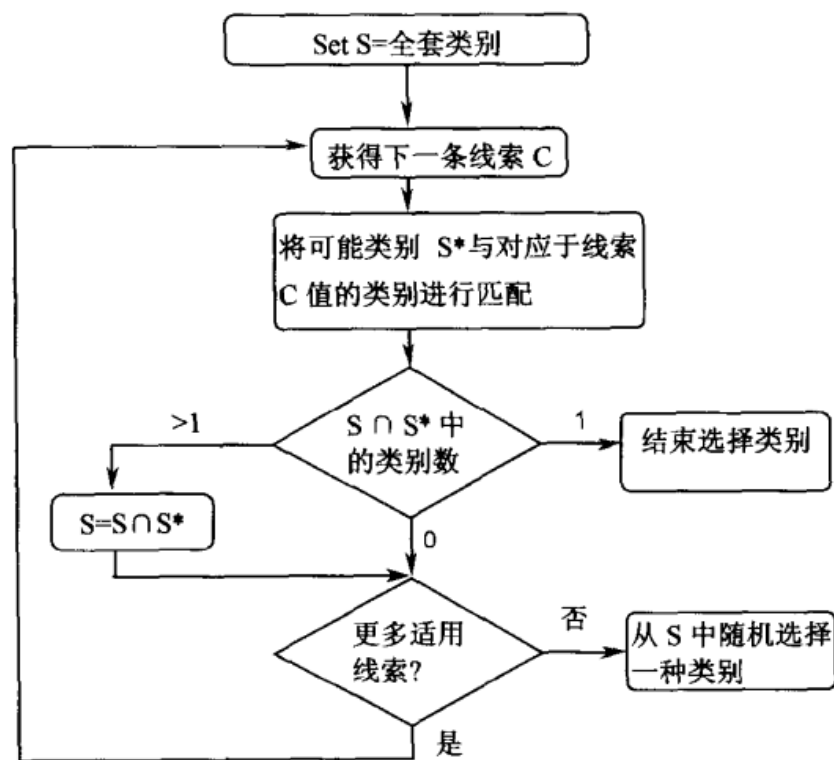


图 11-2 排除归类法流程图

然而, S^* 中通常包含不止一种可能类别。在这种情况下, 候选类别中的那套 S^* 就同先前的那套(全部的)可能性 S 相交, 在此交集中保留下来的可能性作为候选类别的一套新的 S , 以便下一次接受新的检测。CBE 的排除步骤如下: 既不在新的一套可能性中、又不在旧的一套可能性中的所有类别被排除掉。按照这种方法, 只有能够匹配就该对象来说迄今可见的所有线索的那些类别才能作为可能性保留下来。为了继续此种排除过程, 挑出下一个具有最高成功几率的线索维度作为新的 C 。与该对象在维度 C 上的值相对应的那套类别被指定为 S^* , 它与先前保留下来的那套可

能性相交。再一次,如果它们的交集包含了唯一类别,我们就完成了排除过程;如果不是这样(而且仍然有剩余的线索需要检测),那么该交集就成了下一轮排除的候选类别,排除过程将继续下去。

因此,排除归类法有一种简单的、直截了当的终止规则:如果在新形成的一套可能性中只有一个类别保留下来,那么它就是该模型的最终决策;或者,如果所有线索均得到了加工,仍然有不止一种可能类别保留下来,那么从剩下来的这套选项中随机选择一个类别。只要这两个条件没有被满足,该算法规则将采取两种方法中的一种继续下去:如果有多于一种可能类别被保留下来,那么正像上面表明的那样,检测下一条最成功的线索。相反,如果 S^* 与 S 的交集是空的,未留下任何可供选择的类别,那么忽略当前线索 C ,先前的一套类别 S 得以保留,检测下一条线索。

这种算法规则有几个有趣的特点。它在信息上是节俭的,只使用了对于作出决策来说必不可少的线索;它是不可代偿的,用前面线索排除了某些类别后,后面线索无法给予补救;用来把可能类别与特定线索值联系起来的分段方法也是相当简单的,使用了很少的记忆资源,且易于掌握。得到使用线索的确切顺序显然不太重要:在初步的测验中,不同的线索随机顺序仅使得该算法规则的归类准确性改变了很少几个百分点。但非常明显的是,得到使用的具有不同顺序的线索数目对归类成绩造成了较大影响。然而,当这些线索以其成功几率大小顺序得到使用时,这个数目是很小的。

排除归类法在几个方面与特威斯基的逐项排除模型密切相关,但两者也有一些重要区别。首先,EBA 是关于选择问题的一种或然模型,而 CBE 则是关于归类问题的确定模型(但也包括作出最后猜测时的或然性)。其次,在 CBE 中,线索是在开始归类前通过成功几率测量来排序的,所以同一种线索顺序被用来评价每一个对象。而在 EBA 中,项目是按照它们的权重随机选择的,所以对每一个对象来说,考虑线索的顺序不一定相同。第三,正像前

简捷启发式

面提到过的那样, EBA 只选择目前保留下来的那套选项中的项目, 因此加工过程不会由于一套项目的清空而终止。然而, 这需要检测每一个给定项目是否存在于剩余选项中。CBE 压根不需要对将要使用的适当线索进行预选, 而是采用从心理学上看更简单的方法弃用所见到的上一条线索, 即每当它导致了类别之间的空白交集时, 即弃用它。 250

CBE 还与吉戈伦尔和古德斯腾(1996a)提出的关于双项选择问题的采纳最佳启发式密切相关(参见第 4 章)。采纳最佳启发式通过对基于生态效度(参见第 6 章)而排列好的线索进行搜索来作出选择。能够据其作出一种决策的第一条线索将终止对更多线索的搜索。CBE 采用了类似的终止规则: 一旦能够作出一种类别选择, 立即终止对更多线索的搜索。采纳最佳启发式可以被视为 CBE 的一种特例, 倘若用它完成归类任务的话, 待归类项目就应该以如下形式来表述: “A 怎样与 B 比较?” 可能的类别是“A 较大”和“B 较大”, 线索成对得到排列(A 的线索 x 、B 的线索 x)。CBE 将采纳最佳启发式的离散线索和双类别问题扩展或延伸为连续线索和多类别问题。

排除归类法的成效如何?

当然, 如果我们提出的这种算法规则不能够获得好成绩的话, 那么上述所有有趣特征都是枉然的。如何将快速节俭的 CBE 与标准可代偿模型进行比较呢? 我们首先将 CBE 作为人类归类过程的一种模型, 检验了它是如何把特定动作轨迹归入特定意图类别的(Blythe et al., 1996; 参见第 12 章的完整描述)。在该项实验中, 被试观看两只人工模拟的昆虫在一个空白背景上四处游动的“影片”, 必须尽快确定带有记号的那只昆虫正在追逐、逃跑、争斗、追求、被追求, 还是正在与另一只昆虫玩耍。我们用四种不同的算法规则来考察人类在这种“从动作到意图”(intention-from-

motion)归类任务上的成绩:CBE、达维斯规则(使用单元加权的线性模型——参见第4章)、富兰克林规则(使用不同权重的线性模型——参见第4章)和“三级神经网络”(three layer neural network)模型。其中每一种都利用7条基本动作线索作出一种决策。无论在训练项(拟合)或是测验项(推论)上,CBE对动作轨迹的归类都几乎像达维斯规则和富兰克林规则一样准确,尽管它只用了一半线索(参见表12-3)。神经网络模型比其他三种算法规则做得更好,但当排除了一些线索时,它在测验项上的成绩蒙受了巨大损失,说明它与CBE相比缺乏推论上的耐用性。所有这些算法规则的成绩均超过了人类被试的成绩(部分原因在于这些算法规则在接受测验前均受过训练,而人类被试未曾接受任何训练)。然而,即便在平均仅考虑不到一半适用信息情况下,CBE仍能取得像线性规则同样好的成绩,这一事实已足以使其成为一种快速节俭的、具有心理学合理性的分类规则。

为了在一种更为恶劣的竞争环境中考验CBE,下面我们转向在机器学习领域中作为算法规则“试金石”(testbeds)的那些资料*——“艾林机器学习库”(Irvine machine learning repository)中的多维对象数据库(Merz & Murphy, 1996)。我们以该数据库中的三套资料(下面将描述)为基础,把基于完整分段法的CBE与基于各变量维度权重(由递减法来确定)的一种范例模型(即Nosofsky在1986年提出的GCM)加以比较,并将其与一种三级神经网络模型——即用“反向增殖法”(backpropagation)训练过的“三级前寻逻辑激活神经网络”(three-layer feed-forward logistic activation neural network)——作了比较。对于每一套资料,我们都进行10次尝试。每次尝试均将全套资料分成两半:一半被用来训练有关规则(记录拟合成绩),另一半被用来测试推论成绩。

* 将CBE与现有的人类多线索归类模型相比较是很困难的,因为使用3或4条以上的线索进行的实验还很少见。我们希望将来可以看到更多的研究者从事此种现实的多线索实验。

简捷启发式

表 11-1 显示了在三套资料上这些比较的结果。首先是费舍 (Fisher, 1936) 著名的“鸢尾花”(iris flower) 数据库, 用 4 个连续取值特征(花瓣的长度和宽度)对 150 朵花进行测定, 以便分别将其归入三个不同类别(不同鸢尾属)。第二套资料由 178 种酒所组成, 它们分属于三类不同的意大利红葡萄酒, 每一种酒都用 13 种连续的化学成分线索来测定。第三套数据由分属于两类蘑菇(有毒的和可食用的)的 8 124 种蘑菇组成, 用 22 种假想线索值维度来测定。对于每一套资料来说, 我们均给出了三种相互竞争的算法规则在拟合训练项和推论到测验项两种情况下的准确性和使用的线索数量(节俭性)。此外, 我们还包括了(在两种适用情境中)郝特的一项测验(1993)所报告的机器学习系统的平均成绩(可推论性), 以及我们在机器学习文献中所发现的每套资料的最好成绩。

表 11-1 不同模型在三套资料上的归类准确性(正确百分比)和信息 252
节俭性(所用线索数目)

	鸢 尾 花				酒 类				蘑 菇			
	拟 合		推 论		拟 合		推 论		拟 合		推 论	
	节俭性	准确性	节俭性	准确性	节俭性	准确性	节俭性	准确性	节俭性	准确性	节俭性	准确性
CBE	1.2	96%	1.2	93%	2.3	94%	2.2	87%	5.8	91%	5.7	91%
范例模型	4.0	99%	4.0	92%	13.0	94%	13.0	84%	22.0	100%	22.0	100%
神经网络	4.0	99%	4.0	96%	13.0	100%	13.0	94%	22.0	100%	22.0	100%
ML 平均值	—	—	—	94%	—	—	—	—	—	—	—	97%
ML 最好	—	—	—	98%	—	—	—	100%	—	—	—	100%
节俭范例	—	—	1.2	88%	—	—	2.2	80%	—	—	5.7	52%
节俭网络	—	—	1.2	40%	—	—	2.2	54%	—	—	5.7	91%

注: 在训练项上的拟合成绩和在测验项上的预测(推论)成绩均示于表中。第四行 (ML 平均值)显示了几种机器学习规则(Holte, 1993)的平均推论成绩, 而第五行 (ML 最好)显示了迄今所报告的机器学习规则在三套资料上的最好成绩 (James, 1985, 鸢尾花资料; Aeberhard et al., 1994, 酒类资料; Holte, 1993, 蘑菇资料)。最后两行显示了当受到限制, 仅能使用 CBE 所用的线索时, 范例模型和神经网络模型的推论能力。

总起来看, CBE 在这三套多特征自然对象上做得相当好。它

在三套资料的两套上均比范例模型产生了更好的推论成绩,而仅比机器学习模型的平均成绩落后几个百分点。我们并不奢望 CBE 的成绩超过多种专门机器学习模型的成绩;只要处于同一水平上就足以有力地表明:这种简单方法的准确性或许可以推论到其他多种领域。令人感兴趣的关键之点在于:尽管只使用了很少一部分适用线索,CBE 却取得了与更为复杂竞争对手相似的成绩。事实上,平均而言,CBE 能够仅使用其他方法所使用信息的大约 1/4 而作出准确推论。

但是,其他方法或许并不真正需要所有信息——情况或许是:即便仅有 CBE 所使用的那些信息,它们也能做得同样好。为了检验这一点,我们再一次在三套资料上运行了范例模型和神经网络模型,但这一次我们只让它们使用 CBE 检测每个对象时所使用的那些线索。这样,假如 CBE 只看到前 2 条线索后,就对某种酒作出了归类,那么我们也让范例模型和神经网络模型仅使用这两条线索对该种酒进行归类。加上这种信息输入的限制之后,这两个竞争对手蒙受了巨大损失,正如从表 11-1 最后两行(“节俭范例”和“节俭网络”)可以看到的那样。在每套资料上,范例模型通常都落后于 CBE 几个百分点,而神经网络模型受到的影响更大,几乎丧失了其原有准确性的一半。因此,CBE 的优点——尽可能少地使用线索作出归类——恰恰是另外一些算法规则的致命弱点所在!

未来的工作

我们这里提供的结果表明:快速节俭归类法是(传统的)线索—整合代偿模型可行的替代规则。由于仅使用作出一种类别决策所必需的那些线索,CBE 能够在时间紧迫条件下对特定刺激作出归类。而且,即便错失了某些线索(即:一些特征值未知或不能觉察到),CBE 依然能够利用适用线索达到一种类别判断,正是在此

简捷启发式

种意义上,迄今已经证明它要比简单范例模型和神经网络模型更加耐用(尽管超出表 11-1 所示资料之外的进一步比较仍有必要,这些比较需要在训练和测验阶段使用不同数量的线索)。然而,尽管 CBE 限制了知识的使用,但它依然取得了相当准确的成绩,足以与更为复杂和精致的算法规则相媲美(未涉及人类被试)。

关于快速节俭归类法与传统方法之间的关系,本研究提出了几个有待进一步探索的问题。一个重要问题是:一种归类规则的准确性是如何受到它用于作出决策的线索数量和它用来贮存参数的记忆容量[这里我们主要关心用来贮存学习或训练阶段所获得知识的长时记忆容量,而不是史密斯等人(Smith et al., 1998)讨论过的算法规则的工作记忆需求]之间的交互作用的影响的。这两种因素可被用来将归类规则分为四种,并可用本章讨论过的算法规则来说明它们。例如,范例模型通常使用大量线索(所有适用线索)作出归类决策,同时也需要巨大记忆容量来贮存所有范例。决定分界理论也需要大量线索,但却有较低的记忆需求。郝特(1993)的模型只需要单一线索来归类一个对象,但它却需要巨大记忆容量来贮存该线索维度上的许多线索值区段以及对应于每个区段的类别。惟有排除归类法仅需要少量线索,同时又只有很低的记忆需求。关于准确性、线索使用和记忆需求之间是如何相互作用的更为系统的分析将有助于我们进一步弄清楚这样一个问题:如何设计归类模型以强调一些因素而不是其他因素?而且,我们预期:不同环境的信息结构将使不同的归类方法更加准确或更不准确、更节俭或更不节俭,因此快速节俭归类法的生态合理性也必须得到探索。

其次,需要搜集和分析更多有关人类在多维对象归类任务上的资料,以便为比较 CBE 和其他归类模型建立数据库。我们对比较 CBE 的学习曲线、错误范型、决策时间和人类的实际成绩尤为关心。CBE 的归类准确性似乎不因线索顺序变化而改变,此种有趣发现也应该在实验上得到进一步研究,尽管我们预期人类行为

将不会明显地表现出此种效应,假如他们所考虑的线索数量有一个绝对上限的话。

第三,正确和错误归类的基础比率及其代价和收益也应该被纳入到模型之中。例如,如果所有线索均得到检测之后,仍然无法确定一种蘑菇是有毒的还是可食用的,那么宁愿犯过于谨慎小心的错误而猜测它是有毒的似乎更为合理。这种额外信息不仅可以被用来指导出于安全考虑的猜测,而且可以在建立区段结构时事先考虑进来,这样就可以使代价和收益分析问题贯穿于整个决策过程之中。

归类的实质在于从有限的信息量到达较大的知识实体——一旦我们把一个对象指定为一种类别,我们就触及到了该类别所包含(其他)范例的知识,然后我们就能够使用这种知识作出关于该对象的更为精明的决断。快速节俭归类法与此种适应性推断机能直接相关,它使得人们能够从少数几种特征跳跃到记忆中贮存的更为有用的知识,再根据这种知识迅速地作出决策。在辨别白鸽和老鹰或梅克多葡萄酒(Medoc)和格拉夫葡萄酒(Graves)时,毋需积累和联结所有适用信息——快速节俭归类法将使你获得成功!

第五部分 社会智能

THAT MAKE US
SMART

数字智能
PDG

第12章

动作如何揭示意图——社会互动归类

菲利普·W·布利兹(Philip W. Blythe)

彼得·M·托德(Peter M. Todd)

乔夫雷·F·米勒(Geoffrey F. Miller)

不要走别人走过的路,去尝试开辟一条新的道路吧!

——拉尔夫·沃德·埃默森(Ralph Waldo Emerson)

假

如你正在加利福尼亚州的内华达山(Sierra Nevada)上艰难跋涉,不幸遇上了一头美洲狮。按照美洲狮的习性,这时你切忌做两件事情:背对它或撒腿就跑。这两种行为中的任何一种都会激起狮子的捕猎性追逐行为,从而使你从受惊的行人变成其意中的猎物。只有消除了通常引起狮子食欲的那些知觉线索,你才有可能摆脱这一厄运。因此,了解其他猛兽是如何根据动作线索(motion cues)对各种行为作出判断的,生死之事也就不再系于毫发之间了。

人类也非常善于根据简单动作模式对行为作出判断:当两个孩子在一块场地上来回奔跑时,他们的父母亲一眼就能分辨出他们是在玩耍还是在打架;当一只鸽子在另一只鸽子面前迅速旋转或高视阔步时,后者无视这种表演,转身走开,我们立即就能判定前者对后者的追求失败了。在诸如此类的情况下,仅仅根据两个有机体相互间的动作,我们这些局外人便常常能够判定正在发生什么事情——谁正在对谁做什么。而且,此种情况下作为当事者的人类或动物也能够分辨他(它)们之间正在发生的互动行为类

让我们更精明

简捷启发式

别,所依据的也是动作线索。那么,有机体如何才能仅根据观察到的动作模式将行为归入特定类别呢?换言之,人类和其他动物怎样才能将纯粹的身体动作转译为追逐、玩耍、追求之类的富有意向性或目的性的行动呢?最重要的行为类别是什么?什么样的动作线索对于区分行为类别最为有用?我们如何利用这些线索快速准确地识别其他有机体的意图呢?

258

对这些问题的回答需要分三个阶段进行。首先,我们必须详细说明一个有机体的整体身体运动能够实现的主要行为机能,如追逐、争斗或追求等;其次,我们必须确定可观察的动作线索,依据这些线索我们能够将一个有机体在特定时间正在完成的动作归入特殊机能类别;最后,我们必须找到一些“认知算法规则”(cognitive algorithms),用这些规则来指导基于动作线索的行为机能归类。

这三个阶段需要截然不同的研究方法。第一阶段需要对动作本身进行进化意义上的任务分析,即在已知其生存与繁衍需求的情况下,找出动物产生特定运动的基本理由;第二阶段需要分析环境的信息结构,以便确定哪些有用动作线索可以通过观察正在运动的动物而获得,以及哪些动作线索在区分动物运动理由或意图方面最为有用;第三阶段需要对不同决策规则的成绩加以比较,这些算法规则将观察到的动作线索与动物运动的基本机能类别一一匹配起来,以避免对动物和人类实际可能采用的所有算法规则进行过多搜寻。由于从动作推断意图时许多算法规则取得了类似的准确性(参见第7章关于算法规则比较问题的更详细论述),所以我们的目标不是在这个问题上作出最后结论。毋宁说,我们主要关心如何判定具有重要机能的的行为类别,以及可以用来区分它们的动作线索。

从动作推断意图并不是一件轻而易举的事情。当我们将辞典上与“动作”有关术语的标准定义和较高水平的机体行为加以比较时,这一困难尤为明显。比如,《牛津英语辞典》(Oxford English

Dictionary)将“跑”界定为“迅速移动双腿(在一只脚落下前抬起另一只脚),以比‘走’更快的步伐前进”;将“追逐”界定为“为了捕获而奔跑”。这样一来,“跑”是参照可观察的动作线索来界定的,而“追逐”则是借助于指向某种未来目标的意图来界定的,该目标涉及到了某种未明确表述的对象。在“追逐”的定义中没有任何迹象表明一个缺乏经验的观察者如何才能将“追逐”与其他有意图运动类别区别开来。由于我们无法直接知觉运动的意图和目的,而且行动目标还有可能被故意隐藏起来,所以我们必须运用可直接知觉的动作线索间接地推断它们——这是第1章所描述的不确定条件下推断活动的典型例子之一。

一般地说,关于人们如何才能从可观察行为推断目的与意图的一般描述,形成了关于人类“心智理论”(theory of mind; Baron-Cohen, 1995)和社会心理学诸多问题的相当完整的理解。这里我们无意提供此种一般性理解——我们仅将注意焦点集中在这样一种非常简单的情形上:运用动作线索对两个有机体之间最为典型的“意图驱动互动行为”(intention-driven interaction)加以归类。尽管如此,这种特殊情形对几乎所有动物种类来说都具有生物学上的重要性,并作为社会认知的最初级形式具有心理学上的重要性。不同动物类别在身体结构、动作方式、动作受到的环境限制、运动知觉能力、交配系统以及在当地食物链上所处位置等方面有所不同。但是,我们仍然希望通过分析“从动作到意图”进行推断的个别特殊情形,得出一般性的结论,以帮助我们鉴别那些具有普遍适应价值的动作线索和意图的归类策略。

我们还必须限制机体性动作类型、线索以及我们将要考虑的“线索加工规则”(cue-processing algorithms),来进一步限定本研究所要考察的特殊情形。

首先,我们假定人类和其他动物具有特定的动作知觉和意图推断能力,且与其祖先遗传下来的典型动作模式和意图范型相一致。这些与追逐、逃跑、争斗、追求、玩耍之类的机体行为密切相关

简捷启发式

的典型动作模式之所以得以保留,是因为恰好有那么几种生存和繁衍目标需要它们去实现。动物之间产生互动行为只有少数几条理由,这使得心理学家们开展研究活动成为更加容易的事情。他们将不受限制的推断任务(“那个动物为什么运动?”)转换成了一种简单归类决策任务(“那个动物正在追逐、逃跑、争斗、追求,还是正在做别的事情?”)。

其次,我们限制了可用来判断其他有机体意图的信息数量和范围。这些潜在线索不仅包括整个身体的动作(相对于其他有机体而言)——如同人们背对狮子时,狮子就会推断出人的可食性那样——而且包括身体局部的动作,如威胁与服从的姿势、面部表情(如,嗥叫的狗或大笑的孩子;参见 Darwin, 1892/1965)以及喉咙、舌头、嘴唇这些能够使人清晰发音部位的细微动作。这里我们仅将注意力集中在判断意图时首要的、最为简单的信息形式上,即两个有机体相对运动的完整动作轨迹。

第三,可以采用上述动作信息作出推断的规则也是无限的,但是我们感兴趣的仍然只是简单的算法规则:仅利用少量适用信息、并迅速完成任务的快速节俭启发式。当判断(可能怀有敌意的)其他有机体的意图时,仅利用手头信息迅速作出决策(而不是等到所有可能证据全部搜集上来、狮子已经开始袭击时才作出决策)就显得尤为重要了。

本章我们首先对人们怎样才能根据动作线索推断出有机体的意图进行了一种进化意义上的分析,然后提出了一种新颖的实验方法,用来探索人类是如何作出这种推断的。这种方法的首要步骤是让被试玩要求他们模仿各种动作(追逐、逃跑、争斗、追求、被追求、玩耍)的互动式电脑游戏,以此来搜集六种最为典型机体性动作模式的具有生态代表性的样例。被试用鼠标控制屏幕上的昆虫,使其产生真实物理运动。我们将产生的动作模式记录下来,呈现给另一组被试,让他们据此判断昆虫之间正在发生什么事情。根据被试的判断,我们确定哪些意图最经常与哪些其他意图相混

淆。据此我们构建了多种模型,并测试了它们,这些模型揭示了人们是如何根据能够直接从动作模式推算出来的简单动作线索来对其加以归类的。其目的在于识别有用的客观动作线索和能够加工这些线索并推断有机体意图的启发式规则。我们将这种“意图—推断任务”(intention-inference task)视为更为高级的社会认知和归因活动的基础。

先前关于从动作线索推断意图的研究

对于运动知觉与社会认知研究来说,人们如何从动作线索推断意图似乎是一个最基本的问题,但这一点很少引起主流心理学家们的注意。那些采取关于知觉生态学观点的人似乎不在此列,他们的工作引起了人们对可观察环境结构的更直接的研究——这些环境与有机体的内在目标密切相关。

弗利兹·海德(Fritz Heider)也许是该领域研究的先驱,他系统地探讨了引发不同归因的特定刺激,从而开了社会事件知觉研究的先河。20世纪40年代,他与西门尔(Heider & Simmel, 1944)所做的实验表明:如果圆点、三角形之类的毫无个性色彩的几何图形在卡通片中以动物示爱或争斗的动作模式四处移动,或者至少让人联想到了动物的此类动作模式,人们就会自发地赋予它们特定的意图和个性。海德和西门尔并未明确指出哪些动作线索易于诱发人们作出此种理解或解释。但是,他们已经具备了关于这些线索的足够多的隐性知识。正是这些线索使得旁观者仅凭借简单情境中的某些动作,就能够把一个游动的三角形视为诡计多端的恶棍,或者把一个轻飞的圆盘视为薄情的荡妇。

为了发现动物之间相互表达意图的动作线索(如,狼崽弯腰鞠躬表示嬉戏意图,参见 Fagen, 1981),动物行为学家们进行了一些研究,除此之外,人们对该领域的兴趣沉寂了大约四分之一个世纪。20世纪70年代,吉布森(J. J. Gibson, 1966, 1979)关于“直接”

(direct)知觉具有生态重要性的视觉运动线索的研究引起了考斯劳斯基和卡亨(Kozlowski & Cutting, 1977)的极大兴趣,使得他们对“步态”(gaits)和其他生物性动作模式开展了大量研究。这些研究旨在识别“点—光演示”(point-light displays)影片中动态和结构的多种不变模式。演示影片中人们在完全漆黑的空间里走动和活动,只有一点微弱光源附着在他们的四肢和身体上,作为仅有的运动线索。然而,即使这样一种微弱刺激,也包含了足以使视觉系统作出复杂推断(如,那是一个男子在举一个重物,或那是一个女子在行走)的信息。

261

在另外一个很不相同的研究传统中,法国心理学家阿尔伯特·米考特考察了人们是如何根据动作线索知觉原因和结果的(Albert Michotte, 1963)。他原先的研究计划主要强调因果推断的心理学问题(受到了亚里士多德和休谟哲学思想的启发),但他被自己的如下发现深深吸引:人们倾向于把矩形解释为仿佛人或动物那样,在没有任何身体接触情况下发生互动作用(Michotte, 1950)。这与海德和西门尔“强烈拟人论”(rampant anthropomorphism)式的发现相类似。他的研究影响了当代关于婴儿知觉意图的“先天预备性”(innate preparedness)的研究(Premaik, 1990; Spelke et al., 1995),更一般地说影响了关于机体性动作知觉的研究(Freyd & Miller, 1992; Gelman et al., 1995; McBeath et al., 1992)。米考特对因果关系的强调使近期的研究更多地集中于人们是如何区分“机体性”(animate)和“非机体性”(inanimate)动作的,而非人们是如何区分不同类型机体性动作的。

令人遗憾的是,这种对区分“机体性—非机体性”动作的执迷是不合时宜的,因为它强行终止了对更为有趣问题的求证:将运动着的对象识别为机体性的只是确认其意图的第一步,决定采取什么行动对其作出反应才是问题的关键所在。斑马群不得不与时常徘徊在它们饮水源周围的狮子同生共存,只有当狮子表现出追赶它们的意图时才开始逃命。对一只斑马来说,关键的区分不是

机体性动作(一只行走的狮子)和非机体性动作(一阵吹过草地的微风)之间的区分,而是有关机体性动作(一只全速向它扑来的饿狮)和无关机体性动作(正在追赶发育成熟雌狮的一只雄狮)之间的区分。知觉机体性动作的主要目的在于指导自己的机体性动作。区分机体性和非机体性动作对于决定自己做什么来说只是一条无力的线索;还有更为有力、更富有信息量的适用线索,应该使用它们来判定其他有机体的意图,并决定自己作出何种反应。基于这种观点,本章将对机体性动作的一些基本目标、普遍适用于不同类别动物和整个生态圈的一些动作线索,以及基于这些动作线索对意图加以归类的一些简捷启发式规则进行详细考察。

机体性动作的基本目标

运动需要付出能量代价,因此只有当它所得到的“适当收益”(the fitness benefits)超过其所付出的能量代价时,才能期望动物产生运动。像其他行为的适当收益一样,运动的主要收益是增进生存与繁衍能力。因此,我们有可能根据自然选择与性选择的基本原则推导出机体性动作的几个典型目标。

进化使得动物能够利用环境中的多种“适宜载体”(fitness affordance)——可能影响其基因复制的一些资源或东西(Miller & Freyd, 1993; Todd & Wilson, 1993)——进行适应性互动。积极适宜载体,比如食物和性伙伴,会改进其生存与繁衍能力;消极适宜载体,如天敌、病菌、寄生虫以及性伙伴的竞争者,会干扰其生存与繁衍活动。动物发展了特殊的感觉运动系统以趋利避害。如果两个动物彼此间存在互惠的潜在可能性,通常就会出现相互趋近的行为(如,生物学上的共生关系)。如果它们之间彼此威胁到了对方的利益,就会产生相互规避行为。这些情形中的动作模式只适合于多种“正面的”(positive-sum)合作博弈情境。更有趣的情形是当动物彼此间具有相反适宜载体的情形,即一只动物想接

简捷启发式

近另一只,而另一只动物却想躲开它。斑马对狮子来说是积极载体(食物),而狮子对于斑马来说却是非常消极的载体(死亡)。这种利益冲突导致了更为复杂的互动关系,使简单趋近变成了无情追猎,简单规避变成了拼命奔逃(Miller & Cliff, 1994)。

从上述分析,我们可以得知两个有机体间的机体互动行为有两种基本类别:彼此趋近和规避(无趣的)及追猎和奔逃(有趣的)。在生存方面,追猎与奔逃通常发生在食肉动物和其猎捕对象之间,或者发生在争斗着的同类动物之间。在繁衍方面,雄性通常是追逐者,雌性通常是逃避者,至少在一段时间内是如此(Ander-son, 1994; Darwin, 1831)。因此,几乎所有动物都需要掌握以下五种基本机体性动作中的一些:追逐、逃跑、争斗、追求与被追求。我们另外加上了第六种动作:玩耍。它被广泛地用来(尤其是哺乳动物)练习和掌握上述五种动作类型。虽然并非所有种类动物都需要具有上述全套行为技能,但一般而言他们需要具备识别每种行为的能力。下面我们依次讨论六种行为类型。

追逐 动物向它们意中的对象逼近。如果这种意中对象是无生命的,我们就得到了一种退化了的简单“目标—指向性”(goal-directed)行为;但如果该对象是有生命的,而且不想作为适宜载体(如,食物或配偶)被利用,那么它将会离开原地(逃跑)。最简单的追逐策略是朝着意中对象的方向全速奔袭,若意中对象左躲右闪,追逐者也必须随时改变自己的方向。然而,如果事先考虑到了环境对动作的限制(障碍物和地形地界条件),并对追逐对象的前进方向、运动轨迹和意图进行预测,追逐行为也许可以更为有效。

263

逃跑 动物避开对它们有威胁的对象。如果威胁它们的对象是无生命的,我们就得到了一种退化了的有害对象规避行为,或“一蹴而就的逃避行为”(one-step evasion);然而,如果威胁者是有生命的,而且不希望其猎物跑掉,就会展开追逐,那么疲于奔命式的奔逃就成为必要的。就逃跑行为来说,欺骗性的伪装和疾进突停,不可预料、“变化多端”的锯齿形行进路线(Driver & Hum-

phries, 1988; Miller & Cliff, 1994), 以及有效地避开环境中的陷阱和绝境等, 都是非常有用的策略。

争斗 同一类别动物之间常常为了争夺地盘、资源、性伴侣和社会地位等适宜载体而产生争斗。争斗行为过于精细, 以至于难以解释, 因为双方在保证自身不受伤害的前提下, 为了胁迫和征服对手, 必须交互追逐和逃避、攻击和防御。由于动物的身体结构各不相同, 一些部位专门用于攻击, 另一些部位则易于受到伤害, 所以争斗通常包含针对特定对手的大量精确而灵活多变的身体姿势。

追求 动物(通常是雄性)朝着它们希望与之交配的异性(通常为雌性)逼近(Anderson, 1994; Darwin, 1871)。但是, 由于选择配偶的主动权几乎总是掌握在异性(通常为雌性)手中, 所以简单逼近是远远不够的。因此, 求偶的动物常常需要发展一些很有特点的复杂示爱行为, 以展示它们的健康、力量、身材大小、地位、智慧, 或创造性(Anderson, 1994; Miller, 1998)。这种表演通常在距离意中配偶足够近的地方进行, 以便对方能够清晰地知觉到它们, 但又不能太近, 以免对方受到惊吓而不是引诱。在经过短至数秒(对某些昆虫来说)、长至数年(对某些人类来说)的表演之后, 如果意中配偶发出了她(或他)有兴趣交配的信号, 最后的接近和交配行为才有可能发生。

被追求 作为交配对象受到追求的动物(通常为雌性)有充分理由非常谨慎地在它们的求爱者之间作出选择, 因为一旦发生了交配行为, 它们所承担的风险将远远大于其追求者。而且, 它们选择与其交配的求爱者的基因质量将决定它们后代基因质量的一半。机体交配是愚蠢之举(参见第13章关于非随机交配和配偶选择的更详细论述)。因此, 当受到追求时, 被追求者应该表现出足够兴趣, 使各个追求者充分展现自己, 以便评价它们的交配质量, 但也不宜表现出过于浓厚的兴趣, 以免求爱者跳过示爱阶段而直奔交配主题。因此, 被追求者应在互动式鼓励与矜持冷漠之间把握好分寸。一般地说, 受到追求的动物都能够与其追求者保持足

简捷启发式

够近的距离,以便观察它们的素质,但又不至于过于接近以甘冒性虐待和被强暴的风险。

玩耍 玩耍包括各式各样的动作,通过这些动作幼小动物对上述所有动作类别加以练习,以游戏方式表示它们正在进行无伤害性或无性意图的追逐、逃跑、追求或争斗(Fagen, 1981)。在最基本的游戏中,动物不断在追逐者与逃跑者、攻击者与防御者之间转换角色。在较为复杂的游戏中,脑容量较大的灵长类动物能够以假想伴侣和相互模仿等更为抽象的方式产生互动行为。 264

我们无意将上述机体性动作的六种目标作为某些通用运动规则的构筑块。还有一些动作目标无法归入这六种类别,如父母保护无防御能力的后代以免遭同类或天敌袭击、寻觅食物或筑巢材料、迁徙到新栖息地、相互梳理皮毛、聚集在一起抵御食肉动物,等等。这里我们仅仅涉及到了两个有机体之间发生互动的最基本的情形。我们希望未来的研究能够对有助于区分其他行为类别的动作线索进行更详细的分析。

诱发六种典型行为的动作轨迹

遵循海德和西门尔将简单运动刺激归因于有意图行为的早期研究传统(1944),我们仅依据两个互动有机体在空间中相对运动的轨迹来探索人们是如何对它们的机能性意图进行归类的。这就要求我们搜集上述六种行为机能类别的一些有代表性的运动轨迹,这些轨迹能够得到客观测量,以便发现有用的动作线索,而且能够作为刺激以标准化方式提供给人类被试。但是,有关这种运动轨迹的资料是相当罕见的。在生物学和心理学文献中,一方面有大量的关于动物大范围运动(远行、迁徙、来往飞行)的研究资料,另一方面也有许多关于动物小范围肢体运动的研究资料。然而,公开发表的关于这两个极端之间行为轨迹的可用资料却很少见。因此,我们决定创建自己的行为轨迹样本,所用被试是我们很

容易找到的数量丰富的在校大学生。

问题是怎样才能获得六种典型行为类别的具有生态代表性的样本。尽管可供利用的被试很多,但我们不可能像用人造地球卫星记录地球上发生的事情那样,将无线电跟踪探测仪悬于被试的头脑之上,然后静静地等待追逐、争斗、追求、玩耍之类的行为样本出现。尽管高架摄像系统已经从我们慕尼黑实验室附近举行的“Oktoberfest”(慕尼黑节——慕尼黑十月份的一个盛大民间节日)上抓拍到了所有这些行为的样例,但它们的动作轨迹某种程度上显然已经被“Weissbier”(一种小麦制的烈性啤酒)弄得面目全非了!

相反,我们让成对被试通过计算机网络发生互动,指导他们玩各种运动性电脑游戏,借助于这种方式诱发其动作轨迹。每个被试都坐在计算机前,用鼠标控制屏幕上出现的像昆虫一样的小动物的动作。被试可以看到自己控制的昆虫,也可以看到另一只昆虫——它是由另一个房间里的另一个被试通过计算机控制的。两只昆虫出现在计算机屏幕上一个四周有围墙的普通矩形区域内,从上方看该区域无任何障碍物(参见图 12-1)。我们安排了 10 对被试,让他们在这种简单电脑互动游戏中演示前述六种基本行为类别。

265

屏幕上昆虫的活动不像直接反映鼠标移动的屏幕指针那样,仅通过手部的运动就能直接加以控制。相反,它们的运动受到了一些模拟物理学规则的限制,包括使昆虫慢加速和减速的动力、使两个昆虫相互弹开或从墙上弹回的碰撞力、昆虫移动的最高速度等。在设计上,被试的鼠标移动实际上被作为昆虫下一次运动应该朝向的目标(更详细的论述参见 Blythe et al., 1996)。这些“半现实”(semirealistic)的物理限制使得我们的运动游戏更富有挑战性,增强了对其实意性和动态性的知觉,我们希望由此使所得到的动作轨迹对于自然的动物运动来说具有更好的生态代表性。

简捷启发式

我们之所以选择昆虫形状作为刺激形式是因为它看起来比单纯使用“圆点”更加生动有趣。而且,如同前面讨论过的那样,先前有关动作知觉的实验研究表明:将一个物体的运动归因于某种有意图动作显然是人类的一种自然倾向,不管运动的物体是形状上类似于人的东西,还是一种几何图形(Heider & Simmel, 1944; Michott, 1963; Rime et al., 1985)。然而,我们认为采用昆虫状刺激物有助于激发被试的兴趣,而又不至于使他们像面对人物肖像时那样,偏向于采用类似于人的动作。使用头、尾分明的图形刺激(如,不是无头无尾的圆形刺激)也是非常重要的,这样动作游戏中双方的方位就不会发生混淆。此外,正像生物学家哈丹(J. B. S. Haldane)观察到的那样,“大体上说,所有动物都是甲虫”。 266

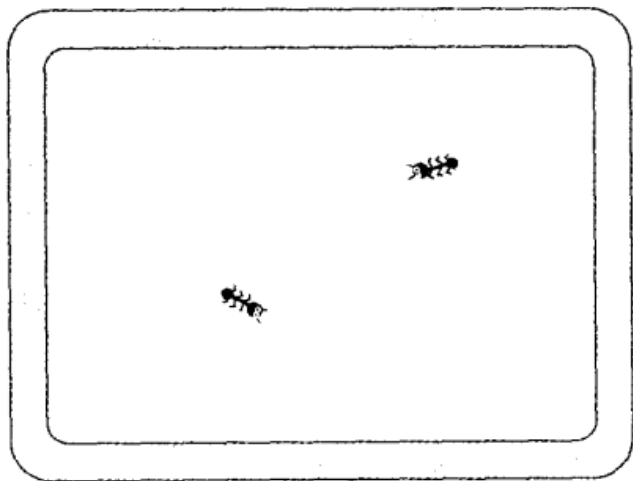


图 12-1 产生动作轨迹时,每个被试均可看见的两只昆虫的方位(在同一瞬间摄下的镜头)。昆虫的颜色互不相同。

每个实验进行过程中,两个互不相识的被试被分别带入两个配有电脑的房间,要求他们根据计算机屏幕上的指示语依次练习和正式玩运动游戏。首先给他们 2 分钟的练习时间,以便他们了解如何使用鼠标控制昆虫的动作。预备研究表明,对于多数熟悉计算机鼠标操作的被试来说,2 分钟练习时间已经足以使他们达到适当技能水平了。此后,他们参与六个单元的实验,每单元持续

时间为2分钟。*“追逐—逃跑”和“追求—被追求”这些双方地位不对等的游戏需要游戏双方互换角色来完成,而“争斗”和“玩耍”只需要每对被试做一次。

具体地说,要求两个被试(这里假定为A和B)完成的六种动作游戏如下:

1. A追逐B:被试A被告知尽可能快和尽可能频繁地移动他的昆虫去拦截另一被试操纵的昆虫,而被试B被告知尽可能避免被拦截。
2. B追逐A:仅在被试A和B之间调换追逐者与逃跑者的角色。
3. A追求B:被试A被告知移动他的昆虫去追求另一只昆虫,可以采取任何一种有可能使后者感到有趣、兴奋或被诱惑的方式与它接触;被试B被告知扮演被追求者的角色,移动他或她的昆虫以表示感兴趣或不感兴趣,可以采取自己所喜欢的任何方式诱使追求者做进一步的表演。
4. B追求A:仅调换了追求者与被追求者的角色。
5. 争斗:两名被试均被告知从背后攻击另一只昆虫,同时要避免受到攻击。特别是,他们被告知尽可能用自己昆虫的前部去打击另一只昆虫的尾部,同时防止后者“以其人之道还治其人之身”。这种争斗更像是第一次世界大战期间的空中战斗(在不让敌机前部的枪炮对准自己的同时,用本机前部的枪炮去攻击敌机),而不像拳击或摔跤。
6. 玩耍:两名被试均被告知以他们愿意采用的任何方式与另一只昆虫玩耍。

267

很显然,这些运动性游戏对被试的要求是不够明确和具体的:我们没有给出任何反馈信息、没有对被试的成绩提供任何物质奖

*“追求”单元的持续时间为3分钟,因为对于一些被试来说,它是一种需要更多时间的较为缓慢而渐进的行为。在现实生活中,虽然“争斗”和“追逐”可能在数秒钟之内结束,但“追求”行为将会花费相对更长的时间。

励,而且我们允许被试几乎不做任何练习。我们对被试是如何通过长时间考虑而采取一种均衡策略来完成这些动态性游戏不感兴趣,而只对他们是如何将这种游戏作为一种快速节俭方法来生成六种典型机体性动作轨迹感兴趣。我们预期:被试对每种类别的运动看起来应该是怎样的都有一种稳定的、易于实现出来的观念,这也正是我们想要调用的知识。就追逐、逃跑、争斗这些动作来说,将被试的此类知识诱发出来是有可能的,而对于追求、被追求、玩耍之类的动作来说,诱使被试调用此类知识就不那么容易了。

描记得到的动作轨迹

在每一种动作类别的实验单元中,电脑都以极高的时间和空间分辨率记录了每个被试所控制昆虫的运动轨迹。然后,我们采用多种方法对所得到的精确细微的轨迹进行分析,看一看能否找到将一种行为类别与另一种行为类别区分开来的动作线索。首先,我们采用简单的直接描记法看一看在2分钟的实验单元中两个昆虫的互动轨迹。图12-2到图12-5显示的是成对被试生成的典型的追逐—逃跑、追求—被追求、争斗、玩耍等动作轨迹的“时空区域图”(space-time plots)。图的水平面代表昆虫在两维环境中的水平和垂直位置,而纵轴代表在90秒钟时段内的时间变化情况。这些区域图揭示了关于动作轨迹的一些基本信息。较高速率(单位时间跨越了更多空间距离)导致了较为平滑的轨迹;不移动的昆虫导致了区域图中某一时段的垂直线;两昆虫间较近的距离导致了紧密混合的轨迹。

在不同动作游戏的区域图中,一眼就能看出一些非常明显的特征。在追逐与逃跑的区域图中(图12-2),我们可以看到跨越较大环境区域的非常平滑(很高速率)的运动,与追求动作较为缓慢的、区域有限的运动(图12-3)形成了鲜明对比;追逐—逃跑(图

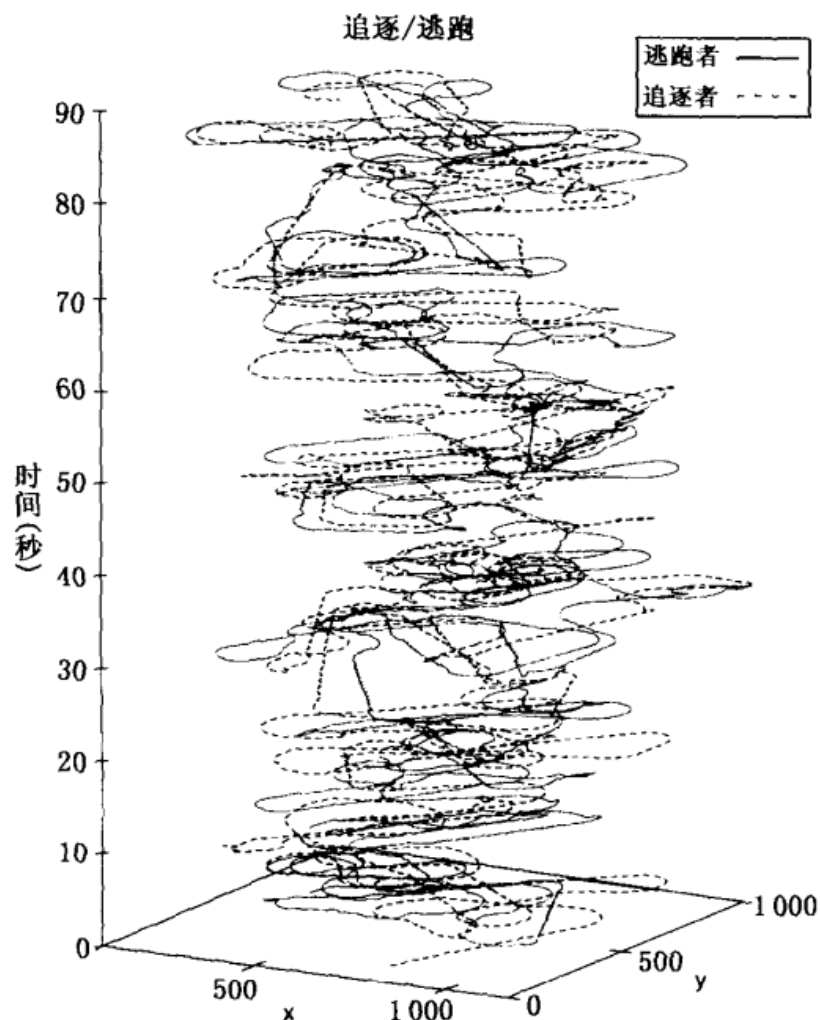


图 12-2 一只昆虫追逐另一只、而另一只逃跑时形成的动作轨迹样本。这里只显示了 90 秒钟的互动行为,纵轴表示时间的延续,X-Y 平面上的点表示每只昆虫在屏幕上的位置。通常应注意高速率(平滑的截面)和大面积覆盖的区域。

268

12-2)和争斗(图 12-4)均显示出了很高的速率,并伴随着大量的转折与环结,但争斗(图 12-4)以两昆虫间更短的平均距离、更紧密纠缠的环结而与追逐—逃跑区别开来,说明争斗时两昆虫有更频繁的身体接触(其时两昆虫的动作轨迹重合了);在追求—被追求的动作轨迹中(图 12-3),追求者的动作远远多于经常处于静止状态的被追求者的动作,有时围着被追求者转圈,偶尔逼近被追求者,引起短暂的追逐和逃跑。在追求—被追求关系中,只有很少身

让我们更精明

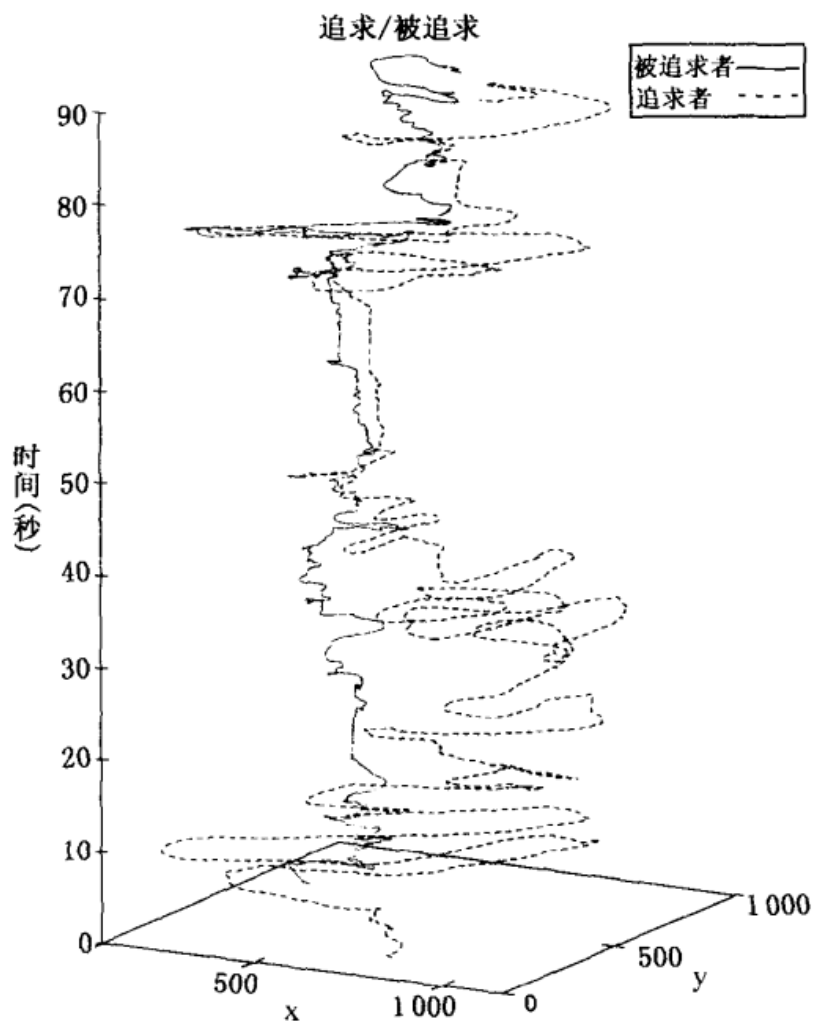


图 12-3 一只昆虫追求另一只、而另一只作出反应时形成的动作轨迹样本。注意追求者更为精细的动作和被追求者偶尔的迅速逃逸动作。

269

体接触。玩耍(图 12-5)看上去好像是追逐、逃跑、争斗、追求与被追求等动作的有机结合,其运动轨迹包含了环结、迅速冲撞、长时间停滞等动作。后面我们将利用这些观察表明哪些动作线索有助于区分上述 6 种动作类别。

272

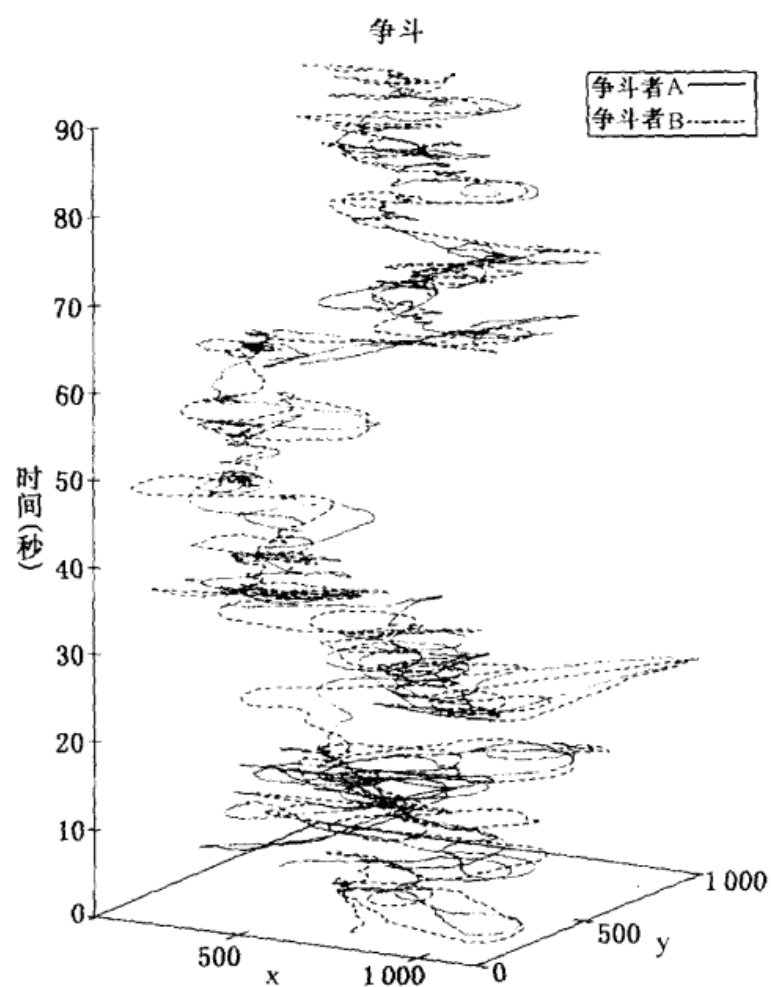


图 12-4 两只昆虫争斗(每一只都试图从背后袭击对方)时形成的动作轨迹样本。注意高速率和紧密纠缠的环结。

270

271

让我们更精明

玩耍

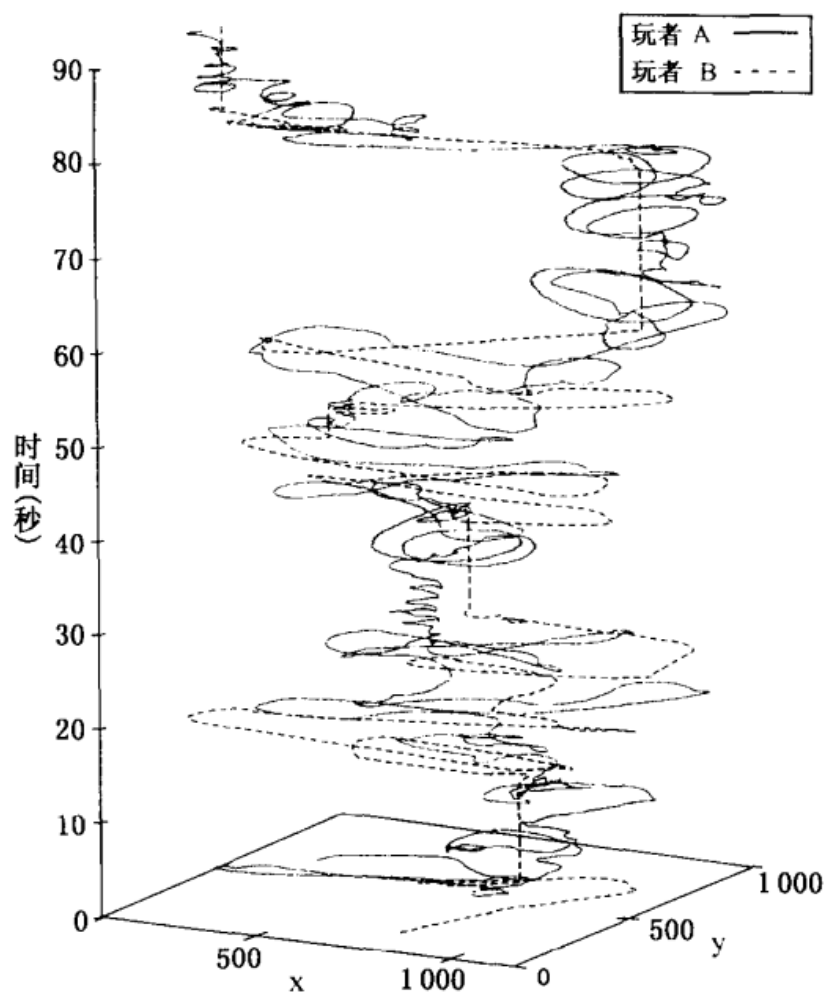


图 12-5 两只昆虫相互玩耍时形成的动作轨迹样本。这里，一只昆虫在大范围活动，而另一只在结环。

271

人们如何对动作轨迹进行归类?

我们的目标是找到能够区分 6 种行为类别、且人类在作此种归类时可能采用的动作线索。然而，在详细分析为不同种类意图行为搜集到的动作轨迹之前，我们必须确保这些动作轨迹确实包含了作出适当类别区分所必需的线索。为了做到这一点，我们对一组新被试进行了测试，看一看他们能否准确地对动作模式加以归类。根据给定的一部分动作轨迹范例，这些未受训练的被试能

否很好地将它们分派给它们原本所属于的行为类别呢？弄清楚了这一点就对这些动作轨迹的生态效度进行了粗略检验，也使得我们有可能弄清楚哪些动作类别是容易混淆的，哪些是容易分辨的。一种机体性动作归类模型可能会取得不错的成绩，但除非它作出了与真正的人类被试大致相同的判断（不管是正确的，还是错误的），否则它就难以称得上是一个好的心理学模型。而且，当我们将同时提供两只昆虫动作轨迹的情形与单独提供一只昆虫动作轨迹的情形加以比较（参见下一部分内容），以考察单昆虫动作线索与互动相关动作线索对于轨迹归类的相对重要性时，我们就能够更加清楚地看到究竟发生了什么事情。也就是说，通过系统地排除一些动作线索并记录归类成绩的下降情况，我们就可以知道哪些线索是最重要的了。

双昆虫实验情形：相关的和个别的线索

在这种归类实验中，以随机顺序向 10 名被试出示第一个实验中记录下来的 30 幅动作轨迹中的一部分，一只昆虫的动作轨迹用蓝色表示，另一只用红色表示。要求被试尽快判断相对于蓝色昆虫来说红色昆虫正在从事 6 种行为类别（追逐、逃跑、争斗、追求、被追求、玩耍）中的哪一种。一俟被试觉得他已经能够确定红色昆虫正在从事的行为类别，立即停止动作轨迹的重放，并作出一种选择。为了说明 6 种行为选项的含义，向轨迹归类组被试出示了与原来的轨迹生成组被试相同的指示语。然而，预先并未向这些被试呈现有关动作类别的任何样例，也没有提供关于其选择的任何反馈信息。尽管如此，在所有 30 次测试中被试的归类成绩波动甚微，前、后两半部分测试取得了同样的判断准确性，表明不存在明显练习效应（我们还记录了每次判断的反应时，以便以后在不同归类模型之间进行比较，但在本章中我们将不讨论这些数据）。

总之，在所有测试中被试有近乎一半机会（49%）选择原先的有意动作类别。这远远高于 6 个选项的机遇值（大约 17%）。（相

273

让我们更精明

简捷启发式

比较而言,3位熟悉该种任务并看到过许多动作轨迹的专家被试——即本章的3位作者——获得了72%的匹配率。)为了看一看被试的错误究竟出在哪里,我们可以建立一些混淆矩阵,以表明被试最经常在哪些动作类别之间相互混淆。表12-1是为被试作出的300种归类判断建立的一种混淆矩阵。行表示原先要求轨迹生成者演示的实际意图类别,列表示轨迹归类者作出的选择。这样一来,表中的主对角线就代表了正确反应。每个被试作30次归类,共有10名被试,他们作出的300次反应分布在六个行为类别中,每行包括大约50种反应。*

表 12-1 被试对两昆虫动作轨迹归类的混淆矩阵

实际类别	选 择					
	追逐	逃跑	追求	被追求	争斗	玩耍
追 逐	29	1	1	0	9	17
逃 跑	5	20	0	5	6	7
追 求	2	0	38	2	4	8
被追求	0	2	5	29	4	6
争 斗	5	6	3	3	10	23
玩 耍	2	0	9	10	7	21
合计值	43	29	56	49	40	82

注:每一条记录均表示被试将依据实际行为类别(行)生成的动作轨迹作为特定行为类别(列)的次数。主对角线(加粗)表示正确的归类。

表12-1中各列的合计值(如果被试准确地匹配了所有动作轨迹类别,这一数值就会等于行的合计值)表明:被试对玩耍出现次数的估计比其实际出现次数高出大约70%,而对逃跑出现次数的估计比其实际出现次数低了大约30%。(被试还略微高估了追求与被追求出现的次数,低估了追逐和争斗出现的次数。)这些倾向反映了被试关于自然界中这些动作发生频率的潜在信念,特别

* 表12-1和表12-2中多数的行并非刚好包括50条记录,因为一些数据丢失了,而且在追逐—逃跑或追求—被追求配对之间是随机分派测验序列的。

是玩耍要比更为严重的冲突(尤其是逃跑)更为经常地出现。对玩耍出现次数的高估还表明:当被试无法明确区分有关动作轨迹时,他们就使用玩耍作为一种默认的“包罗万象”(catch-all)的选择。这与我们前面提出的观点相一致:玩耍是对其他行为类别的练习或训练。

通读表中各行数值,我们还能发现一些更为具体的混淆来源。追逐最经常地被误认为是玩耍,争斗被错误地当成玩耍的次数甚至高于被准确地归为争斗的次数。争斗也常常被错误地判断为追逐和逃跑。除了将其他行为错误地归为玩耍这种一般趋向外,实际的玩耍也常常被误认为是追求、被追求、争斗之类的其他行为。例如,当被试首先看到构成整体玩耍行为的部分追逐动作,并据此迅速(错误地)作出“追逐—逃跑”的决策时,此类错误就出现了。在自然界中,其他非轨迹线索(如,人类的笑)可能被用来对玩耍进行更加准确的归类。

另一方面,被试也的确经常作出正确归类:追逐与逃跑极少相互混淆,追求也很少与被追求相互混淆。而且,如同人们很容易预期的那样,追逐与被追求从未被混淆过,追求与逃跑也从未被混淆过。这与追逐和追求之间以及逃跑和被追求之间存在的某种程度的相似性是一致的。然而,重要的区别显然也是存在的,因为追逐几乎从未被知觉为追求,追求几乎从未被知觉为追逐,被追求也很少被误认为是逃跑。这些差异与相似性在我们最终确定的用来归类动作轨迹的基本线索中应该有所体现。

尽管我们的数据显示存在较高比例的系统误差(高估玩耍、低估逃跑)和许多个别动作混淆现象,但必须强调的是,即便在任何反馈信息、没有任何练习,仅给出最低限度指示语条件下,对实验目标一无所知的被试亦能够仅依据记录下来的两个机体之间的互动轨迹,在远高于机遇水平之上把行为归入从生态学上看相当重要的类别。我们排除了所有场合信息、所有机体步态和姿势信息、所有面部表情信息和所有交流信息。尽管如此,单纯的整体

身体动作线索已足以对行为作出相当准确的归类。据此,我们认为:我们搜集到的动作轨迹事实上的确包含了足够多的信息以表明意图类别。这样,我们就可以继续分析这些信息是如何在动作轨迹中得到体现的:我们可以用来从动作推断意图的线索究竟是什么?

单昆虫实验情形:无相关线索动作轨迹

要确定人们(尤其是我们的被试)究竟使用哪些种类的线索从动作推断意图,一种方法是从动作轨迹中排除某些适用信息,看一看这将如何影响根据修改过的动作轨迹作出类别判断。在上一项归类实验中,被试获得了两个昆虫完整互动轨迹中包含的所有信息,从而有可能使用与两只昆虫相对位置、相对方向及速度有关的所有动作线索。为了估量这类相关线索的重要性,我们希望了解在排除了它们的情况下,被试在归类任务上的成绩如何。仅回放记录下来的两只昆虫中某一只的动作轨迹,以便被试无从知道另一只昆虫在做什么,就可以成功地排除这种相关动作信息。

275

为了实现上述目标,我们进行了第二项归类实验,采用的方法与上一项实验相同,但使用了 10 名新的被试,而且在每次呈现的动作轨迹中,只让两只互动昆虫中的一只是可见的,要求被试对这只昆虫正在完成的动作加以归类。表 12-2 是根据本实验结果绘制的混淆矩阵,呈现方式与表 12-1 相同。现在总体的正确比率从双昆虫实验条件下的 49% 降低到了 30% (仍然远远高于 17% 的机遇值)。这种大幅度成绩下降表明:在我们有关这一领域重要线索的列表中应该包括相关动作信息。对此种情况下所出现的错误范型加以分析,我们还能够发现一些更为有趣的现象。

对表 12-2 各列的合计值加以检查表明,本实验中被试较为均匀地选择了 6 种行为类别;高估玩耍行为出现次数的倾向消失

了,而代之以略微高估追求行为出现次数的倾向。多数错误归类是由于某些行为类别本身极容易混淆。虽然在另一只昆虫可见条件下(表 12-1),能够明确地将追逐与逃跑、追求区别开来,但在仅呈现单只昆虫动作轨迹情况下却常常将其与后两者混淆起来;逃跑被错误地归为追逐的次数像对其作出正确归类的次数一样多,这可能是因为这两种行为都必须在特定空间中以不可预料的路线迅速兜圈。追求常被错误地理解为被追求,但奇怪的是,被追求却很少被误解为追求;将争斗误解为追求和玩耍的次数几乎像对其作出正确归类的次数一样多。

表 12-1 和 12-2 之间的差异表明:相关线索对于作出某些区分是重要的,但对于作出另一些区分却不太重要。特别是,对于区分侵犯性意图(如,追逐、逃跑、争斗)和较为被动消极的意图(如,被追求)来说,单昆虫(无相关)信息通常已经足够了。但是,当要判断一只昆虫是追求者(追逐或追求)还是规避者(逃跑或被追求)时,相关线索似乎是必要的。这两项发现告诉我们,作出不同归类需要不同类型(和数量)的线索——某些情况下不使用全部适用信息也可以作出决策。本章后面寻找适当归类规则时将充分利用这一事实。

表 12-2 被试对单昆虫动作轨迹归类的混淆矩阵

实际类别	选 择					
	追逐	逃跑	追求	被追求	争斗	玩耍
追 逐	15	8	12	3	7	7
逃 跑	13	14	9	1	5	6
追 求	2	5	16	17	5	9
被追求	5	7	4	21	2	5
争 斗	7	4	12	0	13	14
玩 耍	5	4	10	5	14	10
合计值	47	42	63	47	46	51

注:每条记录均表示被试只看到两只昆虫中一只的动作轨迹后选择实际行为类别(行)作为特定行为类别(列)的次数。主对角线(加粗)表示正确归类。

识别可从动作轨迹中抽取出来的有用线索

刚刚描述的归类实验表明,要想根据两个互动机体的动作轨迹对其意图加以归类,我们不仅要考虑由单个机体性动作抽取出来的个别线索,还要考虑通过比较两个机体的动作得到的相关线索。为了缩小可利用线索的范围,我们必须对这些动作轨迹中蕴涵的信息加以分析。

有些显而易见的线索可能与行为目标有关。比如,在我们的实验中,追逐的目标是拦截另一个机体,争斗的目标是用自己的前部去打击另一机体的尾部。因此,一个人或许试图数出成功拦截或打击的次数,来将追逐、争斗与玩耍区分开来。然而,另一机体不情愿受到拦截或打击。成功的逃跑意味着不成功的追逐。因此,作为一条线索,成功次数对于意图类别只有较弱的鉴别力,特别在面对正在发生的互动行为的一个片段时,更是如此。而且,事实上在我们的实验情境中该条线索并未被证明具有诊断功能。

根据给定的有意动作轨迹,对可能从中抽取出来的简单客观(非意图性)线索加以考虑是更为有用的。从我们的实验记录下来的动作轨迹发生在两维空间中,所以我们可以将注意力集中在一些简单牛顿式力学线索上。一种动作轨迹可以依据昆虫在两维空间每个维度上的位置和速度来测量。由于昆虫有头部和尾部,所以它们还有一个自由旋转度(反映了它们的方位或朝向)和一种旋转速率(等同于它们变换方位的速率)。不计加速度之类的高级参数,我们为每个机体找到了6种基本动作参数:水平位置、水平速度、垂直位置、垂直速度、方位和旋转速率。这样,由两个机体构成的系统就可以用12种独立参数来加以描述。

对于一个有兴趣为“双体系统”(two-body system)位置变化建立模型的物理学家来说,上述分析或许是很好的,但是它是否有助于对有机体的行为和意图作出归类呢?我们的直觉以及前面提

供的实验结果都告诉我们：答案是否定的。关于某个机体和其他机体在世界中绝对位置的信息对个体来说是不重要的，重要的是自己与他人之间的比较。社会行为，如同进化本身一样，是由有关的差异驱动的。在知觉两个机体如何互动（不包括环境中水源或遮蔽物之类的特殊位置的作用）时，绝对位置信息在很大程度上是无关紧要的。重要的是彼此之间的相对位置。同样地，在分析他们彼此间的意图时，他们个体的方位提供的信息量也远远没有相对方位提供的信息量大。从任何一个机体的角度来看，都可能有3种重要的位置和方位参数：（1）它与另一机体之间的距离（即我们所谓的**相对距离线索**）；（2）它目前的头部朝向与另一机体位置形成的角度（**相对角度**）；（3）它目前的头部朝向与另一机体头部朝向之间的角度（**相对朝向**）。

我们还可以把描述一般双体系统的6个速度参数减少为4个，用来说明我们想要对其行为加以归类的机体：**绝对速度**（机体向前推进的速度，而不是分别在两个成直角的空间方向上的速度——正像我们将要看到的，我们可以忽略对应于边缘动作的另一种速度）、**相对速度**（一个机体相对于另一个机体的运动速度）、**绝对转速**（机体改变朝向的速率）、**相对转速**（相对于另一机体而言一个机体改变朝向的速率）。我们忽略了一个机体在靠近自己的边缘上的运动速度（在自然界和我们的昆虫实验中通常为0）和机体的相对边缘运动速度（通常也为0），从而剔除了两种速度参数。这样一来，我们就只保留了3种有关的位置参数和4种有关的速度参数。取其在一些“时窗”（temporal window）内的平均值，就可以将它们转换为动作线索。与我们在上述第二项归类实验上获得的结果——它证明了相关线索的重要性——相一致，我们最终拥有的7条线索中有5条是相关线索（仅给出一个机体的动作轨迹难以抽取出来的线索）。

总之，我们提出了7条简单的、从生态学上看有关的线索，在一个机体与另一机体发生互动时的意图作出归类时，这些线索

可能是有用的。

相对距离 两机体之间的距离。在昆虫的动作轨迹中,争斗和追逐—逃跑往往比追求和玩耍产生的相对距离更小。

相对角度 “焦点”(focal)机体当前的朝向(面对的方向)与另一机体之间的角度。该条线索表明另一机体在焦点机体前方或是后方,据此很容易将追逐与逃跑区分开来。 278

相对朝向 一个机体与另一机体之间朝向的差异,表明它们是朝向同一方向(追逐和逃跑时更常见些),还是面对面的(争斗、追求、玩耍时更普遍些)。

绝对速度 焦点机体相对于背景的前进速度。追逐—逃跑和争斗比追求和玩耍产生的绝对速度更高些。

相对速度 两机体间速度的差异。在整个追赶过程中追逐者和逃跑者的相对速度接近于 0,但追求过程中追求者与被追求者之间的速度有更大的差异。

绝对转速 机体相对于环境的转动(改变朝向)速率。争斗和追求的转速高,而玩耍和被追求的转速低。

相对转速 两机体之间转速的差异。追逐者和逃跑者常常以同样的频率转向,因此它们的相对转速为 0,而追求者与被追求者通常有更大的转速差异。

在最终把从动作轨迹中抽取出来的上述 7 条线索作为归类依据之前,我们必须确保它们包含了作出准确意图归类所需要的足够信息。如果我们发现它们做不到这一点,就必须重新寻找一套更为适当的线索。检验这个问题的一种方法是看一看“与算法规则相联系的某种通用模型”(a general-purpose pattern-associating algorithm)能否将 7 条线索值与动作轨迹样例的正确类别一一匹配起来。依据被试在前面归类实验中见到过的 300 种动作轨迹样例,以 7 条线索值为输入项,以适当类别为目标输出项,我们对“简单逻辑—激活三层神经网络”(a simple logistic-activation three-layer neural network)进行了训练。经过训练后,该神经网络可以

对 300 种样例中的 247 种(约为 82%)作出正确归类。[当使用另外 300 种不同轨迹样例测试该网络的可推论性时,它仍能对其中的 200 种(约为 67%)作出正确归类,参见表 12-3。]这使我们联想到了前面的有关实验结果:被试仅能对 300 种样例中的 49%作出正确归类。据此,我们可以断言:虽然这 7 条线索也许没有涵盖对动作轨迹具有鉴别力的所有信息,但它们的确抽取出了足够多的信息,容许我们据其建立一个有关此类任务的合理模型。

某些独行动物任何时候都只能遭遇到一个单独的同类个体,所以只需要对与己有关的其他动物的行为进行归类,也就是说从自我中心的角度对其意图进行归类。对于处于更多社会类别之间的动物来说,经常可以目睹其他动物之间的互动行为,记取不同群体成员相互间的意图也是非常重要的。在具有主导性等级关系和家族网络的动物类别中,这一点显得尤为重要。而且,不同种系动物间的互动往往需要作出准确判断。比如,一只小羚羊应该能够辨别一头母狮正在追赶自己的同伴,还是这头母狮自己正在被追赶(或许被一头雄狮追赶)。无论从自我中心的角度来看,还是从旁观者的角度来看,上述 7 条线索中的每一条都应该是很容易推算出来的。虽然我们实验中提供的动作轨迹基于生态学上相当不合理的(但计算上是简单的)“自上而下”(top-down)的考虑,但我们使用的 7 条线索却如同直接从现实中得到它们一样很容易推算出来。

像 12 条牛顿式力学参数一样,刚刚讨论过的 7 条线索原则上是相互独立的,但实际上却有很高程度的相关。也就是说,自然界中有机体产生的动作轨迹不可避免地会使一些线索值同时出现。这些动作线索间的自然相关对于意图归类有两层含义:第一,仅用少量线索就足以准确地作出一种决策,因为线索包含的信息内容之间有很大程度的重叠;第二,如果无法获得某条线索,可以用另一条线索来代替它:在决策制定过程中,相关线索间可实施“代偿机能”(vicarious functioning)——从本质上说即所谓的“线索可

精

明

让我们更精明

替代性”(cue substitutability; Brunswik, 1943)。

代偿机能是非常重要的,因为在许多现实情境中,潜在线索中的一些或多数在任何给定时段内都难以觉察得到。而且,登记线索的速度对决策制定过程也造成了另外一种限制。对于时间紧迫的任务或问题来说(比如,确定一头美洲狮是否正在悄悄地向你逼近),等到所有线索都搜集齐全后才作出决策很可能为时已晚。此类时间紧迫问题需要快速节俭的不可代偿启发式,它最为有效地利用了当前适用的动作信息(参见第7章)。下面我们将要讨论的正是此类基于动作线索对行为加以归类的决策规则。

280

行为归类的决策启发式

人类和其他动物利用什么决策机制来加工动作线索以作出意图类别判断呢?正如我们已经指出的那样,从动作推断意图常常是在时间紧迫条件下需要完成的一种关键性任务。而且,我们的归类实验表明,不同线索适合于对不同类别行为作出归类。这些发现促使我们寻找一种简单的快速节俭决策机制,它仅使用对于选择某种行为类别来说是必要的信息作出决策。“排除归类”(the categorization by elimination, 缩写为 CBE)规则就是这样一种机制(参见第11章的完整介绍)。在对一种给定刺激加以归类时,CBE以所有可能类别为起点,然后利用该种刺激的各种特殊线索值排除越来越多的类别,直到仅有一种可能性保留下来为止。因此,在对不同刺激作出归类之前,需要加工的线索数量多寡不等。该规则总是以特定顺序来检测线索,而且仅使用能够侥幸得到的少量线索作出类别决策,从而使其成为一种快速节俭的策略。然而,它在一些标准化测试项目上的归类准确性却非常接近于需要将所有适用信息整合起来(如同第11章所表明的那样)的更为传统的算法规则。

为了使用 CBE 对动作意图作出归类,我们首先依据每条线索

将我们使用过的把 300 种动作轨迹样例分别归入 6 种意图类别时的成绩,确定了 7 条线索的顺序。我们发现:当单独使用时,绝对速度线索在归类这些动作轨迹时最为准确,所以我们把它排在线索列表的第一位,接下来依次为相对角度、相对速度、相对朝向、相对转速、绝对转速和相对距离。(我们发现严格的线索顺序对整体归类成绩的影响并不大,但它的确对算法规则需要寻找的线索数量有较大影响。)然后,为了对一种特定动作轨迹作出归类,CBE 从所有 6 种意图类别开始,考虑该种动作轨迹在绝对速度线索上的值,以确定从现有可能性中排除哪种意图类别。例如,每秒钟 400 像素是相对较低的速度值,它表明一种动作轨迹可能对应于追求、被追求、玩耍之类的较为轻松的行为类别,于是其他 3 种较为紧张粗野的行为类别就应该从现有 6 种可能性中被排除掉。由于检测了绝对速度线索后仍然有不止一种行为类别被保留下来,所以必须使用预定线索列表中的下一条线索(相对角度)来排除更多行为类别。检测线索的过程持续下去,直到仅有一种行为类别保留下来为止,该种行为类别就是排除归类规则最终作出的决策。

作为一种基于线索进行归类的算法规则,“排除归类法”有几个有趣特点。首先,它是非整合性的,仅使用那些对于作出一种决策来说必不可少的线索。如果每条线索都需要花费一定时间加以评估的话,那么排除归类法的这个特点就使其能够比利用所有线索的其他算法规则更快地作出决策;其次,它是不可代偿的。前面的线索一旦排除了某些类别的可能性,后面的线索将无法使其恢复,也就是说,后面的线索将无法改变或取代依据前面线索作出的决策。这种特点是否合理取决于 CBE 出错的类型、数量以及这些错误导致的后果。例如,如果 CBE 过于轻率地排除了“争斗”这一类别,并愚蠢地将其误解为玩耍,那么它将是难以适应环境要求的;最后,CBE 总是以特定顺序使用线索。这里我们是按照线索作出正确归类的能力(“线索成效”,参见第 11 章)来排列其顺序的。以一种好的顺序使用线索是非常重要的,因为正是这一点使

简捷启发式

得 CBE 能够迅速地作出决策。举例来说,如果一种算法规则从最无用的线索开始检测线索,那么作出一种决策将会“浪费”更多线索(因此也浪费更多时间),而且作出的决策也更有可能是不正确的。

我们利用 7 条动作线索测试了 CBE 对动作轨迹进行意图归类时的成绩,并将其与正确(原来要表示的)类别和被试实际选择类别(参见表 12-1)作了比较。结果示于表 12-3 中。为了看一看这种快速节俭启发式相对于较为传统的决策方法而言做得如何,我们还构建了两线性联结模型(达维斯规则,使用单元加权;富兰克林规则,用线索成效作为权重——参见第 4 章),它们采取整合的可代偿方式对线索进行加工。像上一部分提到过的那样,我们还对一种有关该种任务的三层神经网络模型进行了训练,以考察这种非线性可代偿系统是如何运作的。

表 12-3 的第一行表示上述 4 种算法规则在 300 次归类试验上的成绩,同时也列出了被试在同一套资料上的成绩。(300 次试验中每次试验的线索值均依据被试在前述有关实验中看到过的同一时段的轨迹数据来计算,以确保相对于被试来说这些算法规则没有特别优势。)结果发现,每种算法规则的成绩均显著地超过了被试的原始成绩。但这并不值得惊奇,因为这些算法规则全都通过反馈信息在有关资料上受过训练,以便作出适当归类,而在我们的实验中,被试没有得到任何反馈信息,必须在第一次看到每种动作模式时对其作出归类。尽管如此,令人惊奇的是,快速节俭的 CBE 做得像较为传统的线性信息整合方法一样好(参见第 11 章更多的此类发现)。而且,CBE 平均仅使用了其他算法规则所用线索的一半多一点——因此它是用较少的资料取得其良好归类成绩的。

CBE 和其他算法规则是如何匹配于被试作出的决策的呢? 282
对于被试作出正确归类的 147 种动作轨迹来说,我们算出了每种算法规则也选择(同样的)正确类别的次数,并将其作为

147 次试验的百分比示于表 12-3 的第二行中。(不同算法规则与被试所犯错误的匹配具有较大随机性和不规律性——此类任务中错误的来源要远远多于作出正确归类的来源——因此这里未包括此种资料。)结果是, CBE 再一次做得几乎像其他线性规则一样好——匹配了被试正确答案的近乎 3/4。考虑到 CBE 比其他两种线性模型快速节俭得多, 所以我们有理由将其作为一种值得进一步探索的心理学模型。(相比较而言, 作为基准的神经网络模型匹配了被试正确反应的 90%。)

表 12-3 对于单或双昆虫动作轨迹, 我们的被试和其竞争者的归类准确性, 同时列出了每种规则使用的线索数量

标 准	竞 争 者				
	被试	CBE	达维斯规则	富兰克林规则	神经网络
使用 7 条线索在训练项上正确百分率	49%	65%	62%	68%	82%
匹配于被试的正确归类百分率	—	73%	72%	77%	90%
使用线索的平均数量	?	3.6	7	7	7
使用 2 条线索在测验项上正确百分率	30%	34%	33%	34%	26%
使用 7 条线索在测验项上正确百分率	—	57%	60%	60%	67%

注: 除了第二行——它表示与被试正确归类(147)的匹配率——以外, 这里的准确性均用正确归类百分率(对 300 种轨迹)来表示。

那么在此种任务上 CBE 究竟有多么节俭呢? 在表 12-3 的第三行中, 我们列出了不同算法规则在对 300 种动作轨迹进行归类时平均使用的线索数。就线性模型和神经网络模型来说, 该平均数是一目了然的: 这些算法规则总是使用所有适用线索, 所以它们平均使用的线索数为 7 条。但是 CBE 在对不同动作轨迹进行归类时可以使用不同的线索数目, 通常总是使用尽可能少的线索。平均起来看, 每作出一种归类 CBE 仅使用 3.6 条线索, 略多

让我们更精明

于适用线索总数的一半。

CBE 尽可能少地使用线索的优势在表 12-3 第四行列出的数据上体现出来。在那里我们严格限定了被试或各种算法规则用来作出归类的线索数量。特别是我们仅向被试呈现了两只互动昆虫中一只的动作轨迹(也就是从前述第二项归类实验得到的资料),而且我们允许各种算法规则仅使用从此种动作轨迹中抽取出来的两种非相关线索(绝对速度和绝对转速)。CBE 和两种线性模型再一次做得至少像人类被试一样好,但强有力的神经网络模型(它在全部训练项上做得何其好!)由于缺乏所期望的信息而受到了严重削弱——成为唯一一种比被试作出更少正确归类的决策模型。在表 12-3 的最后一行,我们获得了这些模型对新情境适应性的证据:这里我们列出了将它们推论到另外 300 种动作轨迹时的成绩。所有模型的归类准确性均有相当程度的下降,其中神经网络模型下降最多,线性模型最少,CBE 居中。

这些结果有助于我们在脑海中形成这样一种印象:神经网络模型的优势仅在线索完整的归类任务上表现出来。诚然,神经网络模型有更多不确定参数,这有助于提高它在拟合测验上的成绩。此外,当所有线索均为已知时,非线性的线索整合在该领域或许具有一定优势,而且此种情况下使用可代偿线索或许更为适当。然而,使用可代偿线索需要同时通达所有适用线索,而在一些情境下 283 线索是在不同时间或以不同顺序自然而然地被觉察到的。此种情况下,相对于其他快速节俭的竞争对手来说,标准神经网络模型可能会处于相当不利的地位。相反,CBE 能够利用以自然顺序觉察到的线索,并尽可能根据给定的每条输入信息作出归类。比如,较之平均速度或转速而言,相对位置通常更易于得到估计,因此能够以快速节俭方式对适当意图类别作出初步估计。通过创建一种能够在具有生态现实性的时间限制条件下对有机体的意图加以归类的“在线动态决策模型”(on-line dynamic decision model),这些问题有望得到进一步的探索。

总之,作为从动作到意图作出归类的一种人类推断模型,*快速节俭的 CBE 启发式几乎获得了像传统的“线性线索—整合机制”(linear cue-integrating mechanism)一样准确的成绩。尽管它所用的信息仅是其他规则的半数,但已取得了令人惊叹的成绩。这种节俭性允许 CBE 作出更快的决策,尤其是在估计每条线索值均需要一定时间的现实情境中。与神经网络模型的参数过剩相比,使用较少线索还有助于增强 CBE 的耐用性——环境中的一些线索难以觉察一点也不会影响 CBE 的成绩!这些优点在实验室情境中也许无关紧要,但对现实世界中的生死决策却是至关重要的。由此,我们认为:CBE 在仅通过机体性动作确定其意图时是一种具有生态合理性的算法规则。

动作的图灵测试及其扩展

作为对上述归类研究的补充,我们设计了以动作为基础的“图灵测试”(Turing test)试用版。在此项测试中,用一只智能型机器昆虫代替了前述轨迹归类实验中的被试之一。测验后,要求仅有的一个人类被试回答:他们在与另一只人类控制的昆虫互动,还是在与一只计算机控制的昆虫互动?一项预备研究的结果表明:即便是最为简单的智能算法规则也能令人惊奇地让人信服,特别是当它们对由人控制的昆虫作出反应时(与完全独立的运动不同)。也就是说,如果一只昆虫(如,追逐者)是由人控制的,而另一只昆虫(如,逃跑者)是由一种简单的计算机程序控制的,那么所导致的动作模式看上去与两只人工控制昆虫的动作模式几乎无法分辨。这一简单测验进一步证明了有限动作线索对揭示机体意图的

284

* 像在其他情境中一样,将此种快速节俭的算法规则与基于选择活动的其他简单方法从实验上区分开来是很困难的。关于此种困难,以及除了依据有关选择结果的数据外,利用反应时之类的有关选择过程的数据区分算法规则的一些方法,请参见第7章更为详细的论述。

简捷启发式

效力。类似地,我们可以把从动作推断意图的另外一些线索运用到新智能昆虫的行为中去,并考察哪些线索组合对于人类观察者来说最有说服力,从而测定这些线索的效力。

诸如此类的机体性动作实验为分析和考察有关人类心智理论——将意图、信念和愿望等内部状态归因于基于可观察行为的其他现象——提供了一些非常有用的新方法。一些研究者指出,“孤独症患者”(the autistic people)持有一种关于心智成分的不适当观点,以至于他们对意图的归因能力被严重地削弱了(Baron-Cohen, 1995)。果真如此的话,那么考察一下孤独症患者如何根据动作对意图加以归类将会是很有趣的:不同个体能够从动作轨迹可靠地识别出来的意图类别也许可以作为他们是否缺乏“移情”(empathic)能力或“社会归因”(social-attribution)能力的有效证据。

“从动作推断意图”的启发式还可以被应用于其他实际领域。许多国家都把闭路摄像系统作为探测犯罪行为的一种重要手段。这种犯罪检测手段实际上就是要在给定动作线索条件下,把犯罪性意图及行为模式与良性行为模式区分开来。即使我们用来在追逐、逃跑、争斗、玩耍等行为类别间作出区分的简单方法也可以运用到这一系统中去。现有的自动化犯罪检测系统本身还不完善,它只是帮助安检员在需要他们监视的许多屏幕画面中识别最值得引起注意的画面。制药公司也常常要记录药品对动物行为模式造成的影响,以此来检测药物效果。对鼠类行为(如,攻击、探究、玩耍等)进行归类的规则系统将有助于将此类评价过程自动化。现实生活中由动作线索推断意图的问题是如此普遍,所以简单的快速节俭启发式毫无疑问可以应用到成百上千个其他领域中去。

结论

本章我们成功地发现了一些动作线索,它们可以被用来推断

某些具有重要适应价值的意图。我们提出的 7 条动作线索对于揭示 6 种意图类别显然已经足够了。我们引入了一种新的简单算法规则,即“排除归类法”,它能够利用最少线索对动作模式作出准确归类,这表明我们可以用快速节俭方式作出此类具有重要适应价值的推断。

我们的主要目标是寻找人类和其他动物可用来彼此有效地产生互动的简单动作线索和快速节俭的线索整合机制,从而为本书前面提出的社会理性概念提供证据。这些机制充分利用了这样一个事实:机体性动作倾向于落入从进化与生态原则推导出来的几种相当刻板化的类别。我们根据人类被试在玩多种电脑互动游戏过程中产生的动作轨迹对相互竞争的多种算法规则进行了测试。我们将这种方法作为探索社会认知基本构筑块的第一步。未来研究应该将进化原则、生态刺激、人类实验以及计算机模拟等多种方法结合起来,进一步分析不同决策规则获得的成绩,并将其与来自人类的资料加以比较。我们一贯的指导原则是:自然环境提供了少数几种关键性动作线索及少数几种典型机体性动作类别,但却提供了过多的令人生畏的动物(包括美洲狮们!)——它们总是伺机消灭那些不能觉察它们动作意义的其他动物!

285

第13章

从骄傲、偏见到规劝 ——寻找配偶过程中的“满意性”

彼得·M·托德(Peter M. Todd)

杰奥弗雷·F·米勒(Geoffrey F. Miller)

婚礼是坟墓,如同上绞架。

——约翰·海伍德,格言(John Heywood, Proverbs)

我嫁给了我曾经吻过的第一个男人。当我把这一点告诉我的孩子们时,他们差点跳了起来。

——巴巴拉·布什,第一夫人(Barbara Bush, First Lady)

1611年,天文学家约翰尼斯·开普勒(Johannes Kepler, 1571—1630)的第一任妻子在布拉格死于霍乱。从包办、不幸的婚姻中解脱出来之后,开普勒马上开始按部就班地寻找继室。尽管开普勒出生于一个贫穷的雇员家庭,而且身材矮小,身体也不太好,但他有从蒂宾根(Tübingen)大学获得的神学硕士学位,又接替泰戈·布雷希(Tycho Brahe)成为神圣罗马帝国的皇家数学家,并由于一系列成就而闻名于世,如:他详细地解释了眼镜是如何校正近视眼的(Ad Vitellionem Paralipomena, 1604),证明了一颗“超新星”(supernova)的存在(De Stella Nova, 1606),还证明了火星的运行轨道是椭圆形的(Astronomia Nova, 1609)等。因此,他是一个令人向往的追求对象。本着严格的选择标准,在妻子死后的两年时间内,开普勒共考查了11位候选人。在1613年娶了第5

号候选人后不久,他给布朗·斯特伦多夫(Baron Strahlendorf)写了一封信,详细地描述了此种按部就班地寻找配偶的方法。朋友们曾经怂恿开普勒选择第4号候选人,该女士有很高的社会地位和诱人的嫁资,但是她以对她戏弄时间过长为由拒绝了他。于是开普勒安心地与他最钟爱的第5号候选人成了家。开普勒作出了一项很好的选择:尽管他的新妻子没有很高的社会地位和很多的财产,但受过良好教育,给他生了7个孩子,并为开普勒发表他的4部主要著作营造了良好的家庭氛围——这些著作作为牛顿的地球引力定律奠定了坚实的实验基础。顺便提及的是,1620年他的妻子还从火刑柱上救下了开普勒的母亲——她是由于施行巫术而被绑上火刑柱的(参见 Ferguson, 1989; Kostler, 1960)。

开普勒的经历与过去几十年频繁出现于统计学、经济学和生物学文献中的有关搜索策略的诸多问题密切相关。迄今为止本书主要关心所有选项同时出现时的决策问题,此种情况下个人只需要搜索有关信息,以便在现有选项中作出抉择。然而,在许多现实选择问题上,选项是相继出现的,而且是以随机顺序出现的,一个人不得不在一个拥有多种不确定参数的总体中进行选择,这些参数事先只能部分地得到了解。在这种情况下,对于选项本身的搜索(而不是对与选项有关信息的搜索)成了中心问题。尤其就选择配偶来说,搜索任务的结构本身就要求一个人按照现实的成功标准选择其意中人,而不是按照善意的人们所建议的不切实际的理想标准进行选择,这是因为考查每一个候选对象的时间都是有限的,况且作为一个拥有自己头脑和身份的活生生的人,候选对象还未必对你中意呢!迄今为止有三个学科对搜索问题做了最为充分的探讨,它们分别强调了不同的具体问题。

统计学家们(Ferguson, 1989; Gilbert & Mosteller, 1966)一直将注意力集中在“秘书问题”(secretary problem)上。在此问题上,一个人必须从随机出现的一系列申请者中挑选出最好的秘书,这些申请者来自一个才智状况分布未知的总体。一旦遭到了拒

简捷启发式

绝,申请者将不再被召回。秘书问题追求完美的结果,要么如愿以偿地挑选出最好的申请者,要么一无所获——没有任何申请者被选中。它忽视了时间之类的搜索代价,也忽视了互动式双向选择问题(你喜欢的候选人有可能不喜欢你),同时假定搜索者早已知道将会到来的申请者的确切数量。但它直接强调这样一个问题:当面对下一个申请者有可能比迄今所见过的最好申请者还要优秀的不确定情况时,应该怎样做?可以看出,解决秘书问题需要选出一定比例的申请者作为样本,记住其中最好的,然后挑选下一个比他(她)更好的申请者。该样本的最佳数量是 $1/e$ (37%)。遵循此种“37%规则”就可能在大约 37% 的时间里找到最佳申请者(参见 Ferguson, 1989)。

经济学家们提出了一些模型,用来说明人们对薪水最高工作和价格最低商品的搜索过程。这些模型强调搜索成本的重要性和“非最佳选项”(less-than-perfect options)的可接受性(Lippman & Mc-Mall, 1976)。像统计学家们所做的那样,这些搜索模型通常忽视了双向选择问题,但它们并不事先假定候选对象总数为已知的,也不认为只有最佳选项才是唯一选择。另一方面,这些模型通常假定你可以回头重新挑选前面见过的选项。由于其依据的假设明显不同于秘书问题依据的假设,所以解决问题的方法也很不相同。解决此类搜索问题的方法通常是:设定一种“保留价”(reservation price),在这一点上进一步搜索需附加的最小成本恰好等同于找到比迄今所见过的最佳选项更好选项时预期增加的最小收益。也就是说,你应该持续搜索更高薪水或更低价格,直到进一步搜索付出的代价有可能超过你能够从中得到的收益时为止,然后返回到曾经见过的最佳选项上来。这主要依赖于薪水或价格分布的标准差,而它们可以从先前观察到的选项上估计出来(Martin & Moon, 1992)。由于此类方法需要通过繁琐计算来确定何时终止搜索,所以它们隶属于第 1 章讨论过的受限制条件下的最优化方法。

20 世纪 80 年代,生物学家们做出了相当大的努力来累积证据,以支持达尔文(1871)的如下主张:动物具有足够的鉴别力和耐心来从事寻找配偶的活动,以便把性选择压力施加于另一个对象(参见 Andersson, 1994)。一些研究者提出了关于此种搜索行为的详尽模型(例如, Johnstone, 1997; Wiegmann et al., 1996),但相比较而言,他们不像统计学家和经济学家那样热衷于定理的证明,而是将更多注意力投向了模型的可验证性问题上。这些模型通常都将搜索成本考虑了进来,有时还缺乏有关潜在候选对象分布的知识(Mazalov et al., 1996)。最近的研究集中在这样一个问题上:寻找配偶过程中不同动物使用“N 中最佳规则”(best of N rule)还是“阈值标准规则”(threshold criterion rule)?(例如, Fiske & Kalas, 1995; Forsgren, 1997; Valone et al., 1996)。“N 中最佳规则”是指抽取出一个候选对象数量为 N 的样本,然后从中选择最佳对象,而“阈值标准规则”是指事先设定一种抱负水准,选择第一个超过该抱负水准的候选对象,就像前述的“37%规则”和“保留价规则”所做的那样。西蒙(1990)将后一种规则(抱负水准设定机制)称为“满意性原则”,并将其界定为“利用经验建立起一种关于我们可以合理地获得的方案究竟有多好的预期,一旦找到了一种符合该预期标准的方案,立即终止搜索”(p. 9)。正如第 1 章指出的那样,西蒙将“满意性原则”视为不知道或不可能知道所有候选对象情况下适用的有限理性的主要形式之一。

所有以上方法均倾向于认为一个搜索者仅需要对被动地等待选择的“货物”(goods)加以评估,但是寻找配偶时面对的一个主要问题是互动双向选择。为非互动单向搜索提出一种有效满意性规则是相当容易的,如购买西红柿或电视机时,本着满意性原则进行搜索是很容易做到的,因为这些东西不会拒绝或反对被买走。但是,关于不确定条件下如何找到互动双向搜索的满意性原则,目前的研究还很少。与此有关的文献主要来源于“经济博弈理论”(economic game theory)关于“双向匹配”(two-sided matching)的

简捷启发式

研究传统(Roth & Sotomayor, 1990),它主要是关于互动选择的研究,但却使用了确定和完整的知识。

正像多数博弈理论分析的那样,这种研究传统主要关心找到一个均衡点或对彼此来说均为最佳的一套策略。可以证明:如果一群数量确定的男女相互间具有一致的、可传递的偏好,那么至少存在一种“稳定匹配”(stable matching),即没有任何一个有配偶的人会选择某个喜欢他或她的其他人。然而,有关双向匹配文献还表明,对于给定的一群具有特定偏好的男女来说,常常有多种均衡点或不同稳定匹配方案。尽管从不存在离婚和再婚合理动因的意义上讲,每种均衡点(匹配方案)都是稳定的,但它们却在不同程度上实现了人们的偏好:一些是“男性最佳的”(male-optimal),即:使男性在给定女性偏好条件下尽可能幸福;一些是“女性最佳的”(female-optimal),即:使女性在给定男性偏好条件下尽可能幸福;一些二者都不是。有一种简单搜索方法称为“延缓接受程序”(deferred acceptance procedure),若给出了总体中每个人的完整信息,此种方法在互动选择条件下能够有效地确保产生一种稳定匹配(Gale & Shapley, 1962)。但是在现实的总体中是否存在此种均衡点,以及在信息不完整的真实情境中是否有一种算法规则可以找到该均衡点,尚未得到证明。 290

所有这些统计学、经济学和生物学模型是如何说明开普勒的求偶行为的呢?更一般地说,它们如何解释面对一系列选项时人类的选择行为呢?寻找配偶可以被视为不确定条件下相当困难且极其重要的一种决策行为。上面提及的模型已经涉及到了其中的几种困难:将会遇到的配偶品质分布状况的不确定性、对将会遇到的候选对象出现顺序的一无所知、与先前拒绝过的候选对象重修旧好的困难、搜索成本、时间限制和损失等。这还都不是最重要的,最为困难的是双向选择问题,即婚配必须建立在双方互相接受基础之上。

不同领域以不同方式强调或忽视了这些困难。统计学和经济

学家们倾向于把寻找配偶作为发展工作和商品搜索过程最优化模型的一种有趣尝试来对待,而未将其作为人类生活的中心适应问题来对待。生物学家则非常强调此种活动的重要性,因为寻找和选择配偶引起了性选择活动。这是一种其能量和创造力也许等同于自然选择的进化过程。由于20世纪70年代末以来性选择理论(参见Anderson,1994;Bateson,1983a;Cronin,1991)重新引起了人们的浓厚兴趣,加之20世纪80年代进化心理学(参见Barkow et al.,1992;Buss,1994;Miller & Todd,1998)的迅速发展,大量研究将注意焦点集中在了这样一个问题上:通过选择配偶而实现的性选择对于塑造人类精神世界所起的作用究竟是什么(Milller,1998;Ridley,1993;Wright,1994)?因此,研究寻找配偶问题时,我们实际上是在探讨一个有趣且困难的不确定条件下的决策问题。这个问题在此类决策问题中具有独特性,它很可能在人类进化过程中一直都扮演着重要角色。

由于人类决策制定还存在着诸多问题,所以人类选择配偶(包括寻找过程)的合理性和有效性也曾受到过质疑。富雷和艾森伯格(Frey & Eichenberger,1996)曾指出,人们寻找配偶时并未花费足够长的时间去进行搜索,离婚和不幸婚姻的频繁发生就是此种不充分搜索的明证。拉波特和特威斯基(Rapoport & Tversky,1970)曾经对已知候选对象品质分布和搜索成本条件下人们是否根据“保留价规则”进行搜索提出了疑问。然而,这些对人们偏离已知最优化策略的担心并未主宰随后的研究,部分原因在于就许多现实搜索情境而言最优化策略是未知的,部分原因在于相对于其他决策制定任务来说,心理学家们对搜索任务给予的注意较少。研究搜索问题的心理学和经济学家们常常将注意焦点放在了人们实际上显然在使用的简捷启发式规则上。海(Hey,1982,1987)已经确认了许多这样的规则,“单一弹性规则”(one-bounce rule)即是其中之一。实施这种规则时,为了寻求高值,人们密切注意值的变化,只要它们在增长,就保持对它们的检测,但是一旦它们开始

下降,立即终止检测,并回过头来选定前面的值。马丁和穆恩(Martin & Moon, 1992)通过计算机模拟评价了消费者搜索商品情境中不同简单策略的相对成效,他发现一些规则能在 3% 的误差范围内接近于正常水准。

踏着这些研究者的足迹,在本章中我们试图寻找特定领域的简单满意性搜索规则,这些领域均与生物学上相当现实的寻找配偶问题有关。我们还采用多种测量指标,在多种搜索条件下对所模拟的不同启发式作了评价。按照生物学上性选择理论的传统做法(参见 Cronin, 1991),我们首先探索了带有雄性中心化色彩的单向搜索情境,继而探索了更加现实的双向互动搜索情境——它像重视雄性选择一样重视雌性选择。(为简化分析过程,本章我们未考虑可能存在的配偶搜寻策略的性别差异及其影响,尽管这种差异可能会产生非常有趣而重要的后果。)

通过分析,我们发现:即便对于简单的非互动搜索(就像在秘书问题上所做的那样)来说,在许多标准上“37%规则”获得的成绩也被选取较少候选对象的启发式规则所超越。这些启发式规则甚至不需要知道一个人将会遇到的候选人数量:一种被称为“尝试 12”(try a dozen)的简单满意性规则在候选人数量不等的多种条件下均很有效。我们还发现:当处理互动双向选择问题时,此类搜索策略做得相当差。此种情境中只有那些自视甚高的个体才有可能采用“37%规则”和“尝试 12”启发式而侥幸获得成功。(在这方面开普勒是幸运的:他的高婚配值有助于确保其“尝试 11”策略产生良好结果。)

相反,充分考虑个人自身的婚配值的搜索启发式在互动选择上会做得更好一些——更迅速、更经常地为个体带来高质量婚姻。即便一个人起初不了解自身的婚配值,他也可以根据取样阶段从别人那里得到的拒绝和许可次数来确定一种调整自己抱负水准的简单适应性规则,采用这种规则也能收到良好搜索效果。如果一个人同时注意到了对自己表现出或未表现出兴趣的那些人的婚配

值,那么了解自己的婚配值就会更加容易。以此为基础,一个人就可以制定一种有效搜索策略,以便在互动双向选择条件下找到希望找到的最佳配偶。简言之,在非互动单向选择条件下表现极佳的“37%规则”和“尝试12”启发式之类的搜索策略在互动双向选择条件下的表现极差,远不如那些从别人的反应中了解自身婚配值的互动选择策略做得好。遵照生态理性观点,我们发现在给定情境中——不管是单向搜索,还是互动搜索——做得最好的寻找配偶满意性启发式均充分利用了环境的信息结构:前一种情况下它们仅依赖于婚配值,后一种情况下它们仅依赖于别人对自己表现出来的兴趣。

单向配偶搜索的算法规则:嫁妆问题

前面所描述的寻找配偶的理想化方法迥然不同于实际上正在寻找伴侣的男男女女们面对的情境,至少在当代许多西方国家的文化背景下是这样。现实生活中的伴侣选择通常包括对相继出现的候选对象进行系列搜索,依次对每个候选对象进行评价,并作出一种决策。这个过程需要花费几分钟、几小时、几天甚或几年时间。(这里的决策被界定为决定是否与某个特定的人成家生子,尽管还可能有其他定义)。在此种搜索过程中,对每个人进行审查当然都需要付出一定代价。但是最大的代价或许是你很难或根本不可能再回到前面被你抛弃的候选对象的怀抱里去(因为他们依然留在“配对库”中,很可能同时已与别人结成连理,如同无数爱情悲剧所证明的那样)。更为复杂的是,一个人事先并不知道候选对象的范围有多大:我们如何才能知道假如我们花费足够长时间寻找配偶的话,那么当我们第一次坠入爱河时,我们的配偶是否还能够激起我们更深厚的感情呢?我们甚至不能够断定我们会遇到多少个候选对象。鉴于搜索过程受到了种种限制,加之我们对正在搜索的空间知之甚少,看起来寻找配偶确实是一件令人望而生畏的

我们可以借助于前面提到过的秘书问题的一种翻版——“嫁妆问题”(dowry problem),来对此种情境进行更详细的分析,这样做时所涉及的主要是配偶选择(而非搜索)问题。该问题是统计学和概率论中著名的难题之一(Corbin,1980;亦可参见 Gilbert & Mossteller,1966;Mosteller,1987),其曾经有过的五花八门的名称说明了这一点(它还以“选美问题”甚至“大数目问题”而著称。)在其嫁妆版本中,故事大体上是这样的:苏丹(Sultan)希望对其首席顾问的智慧加以测试,以便决定是否保留其内阁成员的职位。碰巧首席顾问正在寻找配偶,因此苏丹决定利用这个机会来判断他的智慧。他从他的王国中找来100个女子,让她们依次从该顾问面前经过。要想保住自己的职位,顾问必须从中选出拥有最高嫁资(从其家庭得到的结婚礼物)的那个女子。若选对了,他可以娶那个女子,并保留他的职位;若选不对,他将被撤职斩首——糟糕的是他还是个单身汉呢!顾问每次只能见一个女子,可以询问她有多少嫁资;然后他要么马上决定她是否就是100个女子中拥有最高嫁资的人,要么放她过去,继续接见下一个女子。他不能返回到前面见过的女子——一旦放她们过去了,她们将一去不复返。更重要的是,在接见这些女子之前,顾问对嫁资高、低的界限一无所知。采用什么样的策略他才有可能使自己挑出拥有最高嫁资女子的几率达到最高呢?

正像前面提到过的那样,可以证明,该顾问应该用来确保作出正确选择几率达到最高的算法规则是“37%规则”。本例中该规则的运作步骤如下:顾问查看前37个候选对象(更一般地说是候选对象总数的37%),放走每一个对象,但要记住该样本的最高嫁资数——该值被称为 D 。然后,从第38位女子开始,他应该选定嫁资超过 D 值的第一个女子(关于此种方法的推导过程,可参见 Ferguson,1989; Gilbert & Mossteller,1966; Mosteller,1987)。

“37%规则”是该顾问能够作出的最佳选择——它找到最高值的几率比其他任何算法规则都高(在 37%的时间内),在此种意义上它是解决此问题的最佳方法。利用这一规则,该顾问有略微高于 1/3 的几率选中适当女子,并保住自己的脑袋。另外 2/3 时间,他就不得不听由苏丹去寻找另一个顾问了!

在许多方面,嫁妆问题显然都不是人类寻找配偶过程的真实反映——它仅涉及到了单向(而不是互动)搜索,并把搜寻活动限制在单一线索维度上,而未对我们通常借以判断其他人的多个侧面加以考察(Miller & Todd, 1998)。它还排除了同时比较多个候选人和返回到先前见过候选人的任何可能性,如此等等。但是它至少为我们测试某些特殊领域的一些特殊配偶搜索机制提供了一个合理切入点。而且,我们可以采取有效实用的方法修正它的一些假设,以帮助我们更好地理解更为适当的搜索机制,正如下面我们将要表明的那样。

嫁妆问题和现实世界之间的主要区别之一在于:后一种情况下,我们的婚配决策当然很少会如此具有戏剧性——我们通常将会(或不得不)履行我们作出的任何决定,即便它并非是“最佳的”。对于苏丹的顾问来说,未挑出最高嫁资的情况下,“37%规则”的成效就变得无关紧要了——无论如何他都将被杀头。但是对于每个人均使用这种规则来挑选配偶的一个总体来说,在另外 63%时间里获得的成效也是至关重要的。举例来说,倘若用“37%规则”在嫁资恰好为整数从 1 到 100 的一百个候选对象中进行挑选的话,平均获得的嫁资大约为 81(即:由该规则选择出来的嫁妆的平均值)。由该规则选择出来的对象中只有 67%的个体位于总体的前 10%,而有 9%的个体落在了总体的后 25%。而且,在挑选出一个配偶之前,“37%规则”平均需要对潜在候选对象进行 74 次检测(即:在开始进行选择之前,必须对两倍于 37 的对象进行检测)。(这些数据均将在下一部分加以讨论。)如果有其他一些指导系列选择的算法规则能够改进这些数据中的任何一种,而且能够很好地证明这些算法规则对于配

简捷启发式

偶挑选者总体来说是更加适用的,即能够使他们更经常地、更快地挑选到更好的配偶,并使他们挑选到不适当对象的几率更小,那么我们就有理由期待它优于“37%规则”。

如果说嫁妆问题本身具有非现实性的话,那么“37%规则”这种解决问题的方案也有许多特点使其成为一种关于人们实际上如何寻找配偶的难以置信的模型。这里我们仅关心其所涉及到的两种困难:首先,为了算出前37%候选对象有多少人,并设定个人的抱负水准,它需要知道总共有多少个潜在候选对象(N);其次,作出一种决策之前,这种规则需要检测大量个体——如,在上述例子中需要检测100个人中的74个。即便假定对一个人婚配值的考察或评价是相当快的(不妨设定为一个月多一点吧),所花费的搜索时间也过长了!

因此,使用“37%规则”寻找人类配偶需要获得一些难以得到的信息(N 的某种确切值),并检测大量个体,从而导致了过长的搜索时间。另一方面,富雷和艾森伯格(Frey & Eichenberger, 1996)指出,在婚姻的诸多问题中,最令人费解的地方在于:人们在寻找终身伴侣时进行的搜索过少——在将自己的终生托付给某个陌生人之前常常只检测了少数几个候选对象。他们援引的证据反对在人类配偶搜索中使用“37%规则”——但这些证据同时也表明,当搜索时间不够长时,人们作出的配偶选择甚至更糟。如果人们不采用像“37%规则”那样的耗时过长的算法规则,他们又能采用哪些规则呢?是否存在这样的可能性:有一些更快的搜索规则,但其取得的成效足以消除富雷和艾森伯格对糟糕伴侣搜索行为的担心?如果是这样,那么这些规则是否更为复杂呢?在下一部分中,我们探索了这些问题,并发现在配偶搜索过程中我们实际上可以花更少精力做更多事情。

搜索较少对象的后果

为了探讨是否有任何一种简单搜索启发式在标准秘书或嫁妆

问题的各项指标上均能超过“37%规则”，我们首先考查从“37%规则”派生出来的一套满意性规则。研究表明，即使在此类为数不多的相似的启发式规则中也有一些在许多维度上均超过了“37%规则”，因此这里我们把讨论的重点限制在这些规则上（尽管其他类型的简单搜索方法或规则在一些效标上或许有更好的成绩）。我们将这里要考虑的搜索启发式称为“采纳下一个最佳者”规则（take the next best, 缩写为 TNB；由第 4 章介绍过的快速节俭的“采纳最佳”启发式发展而来）。

TNB 规则的运作步骤与 37% 规则极其相似：对于给定的 C 来说，全部候选对象 N 的第一个 $C\%$ 得到检测（不作选择），记住其中最高嫁资 D ——这是搜寻者的抱负水准。在检测了全部候选配偶的第一个 $C\%$ 之后，下一个拥有比 D 更高嫁资的候选对象被选中。（如果随后没有出现更高嫁资，那么我们假定搜索者将接受该序列中的最后一个个体，这就是在后面将要出现的图形中成绩曲线最终落入非 0 值区间的原因所在。）这种简捷启发式（“37%规则”是它的一种特例）具有最低限度的认知需求：每次只需要记住一个值（当前最高嫁资），只需要知道 N 和 C 并计算 $N \times C\%$ ，并且每次只需要比较两种嫁资值。我们主要关心：当不断改变潜在候选对象的百分比（ C ）时，这些简单算法规则的成绩是如何发生变化的？由于我们还希望能够改变该问题的潜在假设，如嫁资值的分布、检查每个候选对象的成本以及 N 是否为己知的等，所以涉及的数学运算很快就会变得复杂起来，因此我们决定采用一种灵活的模拟方法来回答这些问题。

我们分别采用如下 C 值来检测 TNB 规则的成效：从 0%（即总是选择第一个潜在候选配偶）到 15%，以 1% 的速率递增；从 20% 到 50%，以 5% 的速率递增（除了 37% 周围的价值以外，在那里我们再一次进行了更细致的分解）；从 60% 到 90%，以 10% 的速率递增。（因为我们相信多数好行为——即好成绩——发生在 C 值较低的范围之内。）我们用不同潜在候选对象数量 N 来运行每种规

精

让我们更精明

则,每次运行都使用了 10 000 种随机创立的嫁资(或婚配值)表。我们搜集了每种算法规则选择出来的婚配值分布的统计量(包括平均数、标准差、四分位数分布以及唯一最高嫁资被选中的次数)和作出配偶选择的位置(平均数和标准差)。有了这些值,我们就可以回答上一部分最后提出的问题了。直白地说,这个问题是:37%规则能否被打败?

拥有 100 个候选对象时的搜索成绩

即使就这些简单 TNB 规则而言,对上述问题的回答也是肯定的。当然,37%规则最经常地挑选到最高婚配值。在图 13-1 中,“选中最佳配偶”曲线表明了就总体对象(潜在候选配偶)的不同百分比(C)来说最高婚配值被一种 TNB 算法规则选中的几率(参见 Gilbert & Mosteller, 1996, 图 1, p. 42, 那里有关于此函数的推导公式)。正如所预期的那样,C 值为(大

296

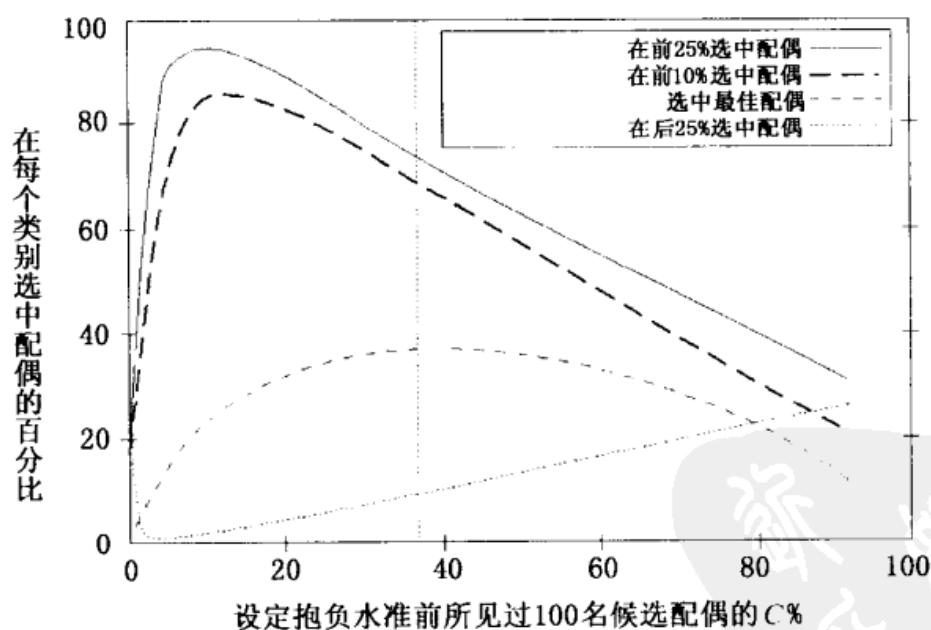


图 13-1 采纳下一个最佳者之前,给出不同待查对象比例(100 名候选对象中的一部分)以设定抱负水准,在此基础上特定种类的值找到一个配偶的几率。37%规则在多种效标上的成绩用垂直虚线来表示。

约)37%时,选中最高婚配值或嫁资的几率最高(由于我们所运行模拟程序的随机性质,该图中的最大值并未恰好出现在37%处)。但是该曲线也表现出了平缓的顶端,因此使用哪一种精确的C值似乎无关紧要——当C值在30%—50%之间变化时,结果几乎是同样的。而且这些策略中的任何一种找到拥有最高嫁资配偶的几率永远不会超过37%,就像上一部分中所提到的那样——这是一种并非很高的几率。

对于一个正在寻找配偶的动物来说,仅有1/3几率能够从异性中选出最佳成员或许并非一种值得投下的赌注——相反,它常常为了节省搜索时间和精力而选择其他“相当好”的对象(或许由于动物的知觉能力不足以在“最佳”和“相当好”之间作出区分)。为了拥有超出其他竞争对手的适应性优势,相对快速地找到一个婚配值处于所有候选对象前10%的配偶已经足够了。在图13-1 297中,我们看到,相对较低的C值(14%)在前10%候选对象中选中一个配偶的几率最高(83%)。如果选择标准再放宽一点,只要求找到一个配偶值处于第一个四分位数(前25%)内的配偶,那么只需要检测开头的7个($C=7\%$)对象就可以使实现该目标的几率达到最高,即在多于92%的时间内有望实现该目标。最后,一个动物也许不是一个风险追求者,非要在总体前列找到一个配偶不可,相反它或许是一个风险规避者,只要将在总体最后一个四分位数(后25%)内选中配偶的几率降至最低即可,那么情形就大不一样了。从表示“在后25%中选中配偶”的那条曲线,可以看到:实现该方法的方法是使用低于3%的C值,它使得在婚配值分布末端(最后一个四分位数)选中配偶的几率低于1%。“37%规则”将有高于9%的几率选中这种较差对象,相对于上述反冒险标准而言这是一种更糟糕的成绩。

相反,一个动物也许可以仅仅通过使其对所选择配偶的期望值最大化来获得超出竞争对手的适应性优势。图13-2表明如何来实现这个目标,它显示了所获得的平均婚配值是如何随

简捷启发式

着用来设定期望水平(D)的候选对象比例(C)不同而变化的。在候选对象总数为 100、婚配值分布为 1—100 的环境中,采用“9%规则”($C=9$)的搜索者将选出其平均值达到最高(接近于 92)的那些对象。相反,如果搜索者采用“37%规则”,该平均值降低到了 81。

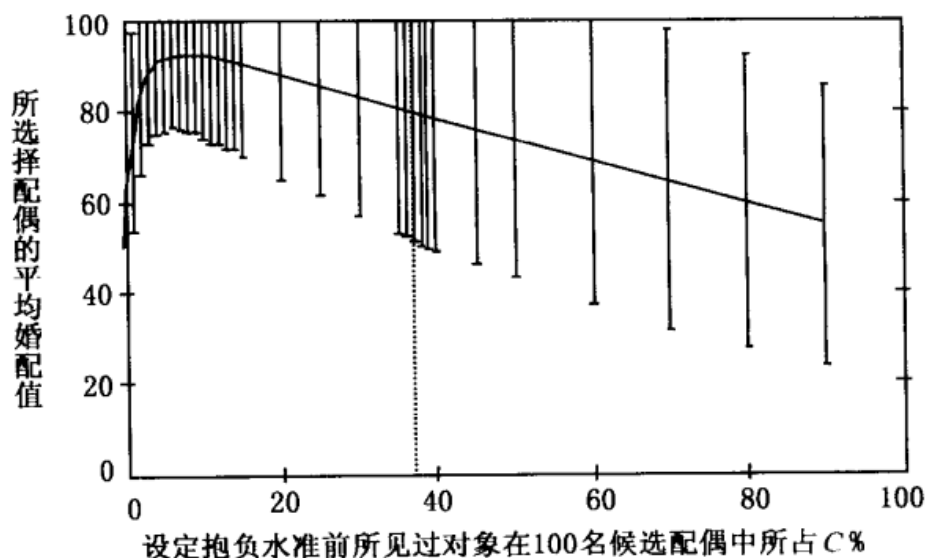


图 13-2 采纳下一个最佳者之前,给出不同待查对象比例(100 名候选对象中的一部分)以设定抱负水准,在此基础上所选择配偶的平均婚配值(竖条表示一个标准差)。“37%规则”在该效标上的成绩用垂直虚线来表示。

由这些搜索规则选择出来配偶的婚配值并非衡量搜索效果的 298
唯一标准——花费在搜索活动上的时间和精力也对所使用算法规则的适应性具有强烈影响(参见 Pomiankowski, 1987; Sullivan, 1994)。在图 13-3 中,我们看到,最终选定配偶之前平均需要查看的候选对象数量是如何随着开始进行选择前为设定抱负水准而得到检测的候选人数($C \times 100$)而变化的。在最终选定一个配偶之前,“37%规则”平均需要查看 74 个候选对象。随着 C 值的降低,需要查看的候选对象人数迅速地减少了,即:适应上的优势伴随着 C 值的降低而增加。若仅按照这种效标,最佳 C 值应该是 $C=0$,即总是选择遇到的第一个候选对象。当结合其他效标来衡量

时,就需要在找到一个好配偶和为此花费的时间和精力之间进行审慎权衡了。

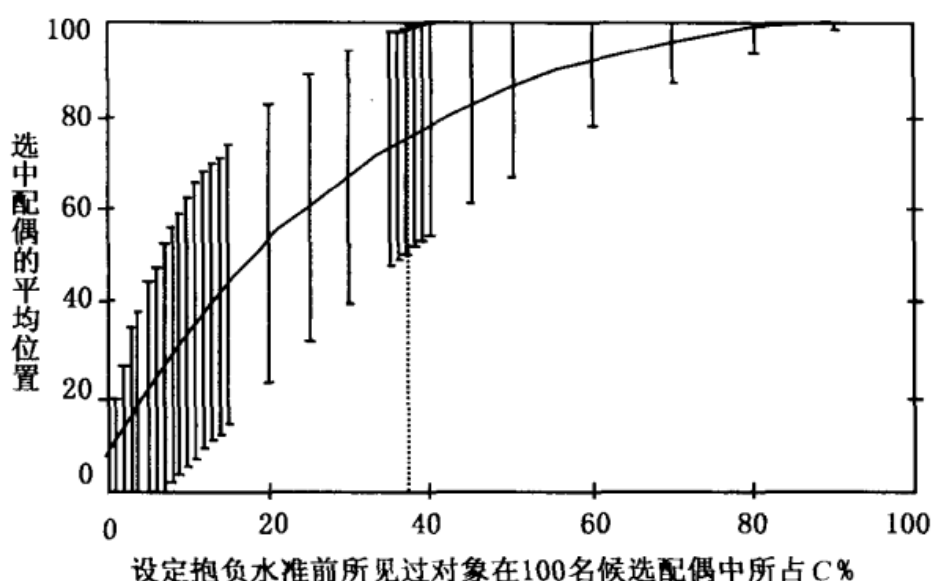


图 13-3 采纳下一个最佳者之前,给出不同待查对象比例(100 名候选对象中的一部分)以设定抱负水准,在此基础上选定一个配偶时所处的平均位置(竖条表示一个标准差)。“37%规则”在该效标上的成绩用垂直虚线来表示。

拥有更多候选对象时的搜索成绩

采用所有效标,而不仅仅是选中一个最佳配偶的几率,对于仅检测少于 37% 对象以设定抱负水准的“采纳下一个最佳者”规则更为有利。在选中其后拥有最高嫁资个体之前,仅检测候选对象的 10% 左右将导致最高平均婚配值,很有可能在第一个四分位数(25%)内和前 10% 候选对象中选出配偶,而且在作出最后选择之前仅需要搜索 34 个左右候选对象。在只需要检测 100 名候选对象中的最初 10 名的条件下,这看起来是一种很不错的成绩。但是我们的先人们或许有过比这大得多的有效婚配群体,当然在现代社会中一个人也会期望有 100 名以上的异性成为他的候选配偶。如果候选对象规模增加到了 1 000,使用我们的简单搜索规则将会导致什么样的后果呢? 此种情况下,检测

299

让我们更精明

简捷启发式

10%对象就意味着要考查 100 名个体,这看起来不再像是一件有趣的事儿,而更像是辛苦的工作了吧? 由于当参数 $C\%$ 一定时,依据 TNB 规则得到检测的个体数量随总体人数规模 N 的增加而直线上升,所以至少在总体规模较大时,采用这些规则不会十分快速节俭地结束搜索过程。

然而,如同图 13-4 表明的那样,当潜在候选配偶总数为 1 000、婚配值分布为 1—1 000 时,我们对搜索时间直线上升的顾虑是多余的。像在前面的例子中一样,当首先检测总体的前 37% 对象时,选中一个拥有最高嫁资配偶的几率最高。但是如果想要使在前 10% 候选对象中选中配偶的几率达到最高(97%),只需要检测 3% 候选对象来设定抱负水准 D ; 若只想使在前 25% 候选对象中选择配偶的几率达到最高(98%),只需要检测 1%—2% 的候选对象。同样地,为了使在后 25% 候选对象中选中一个配偶的几率(最低为 0.3%)最小化,只需要检测 1% 的候选对象。

300

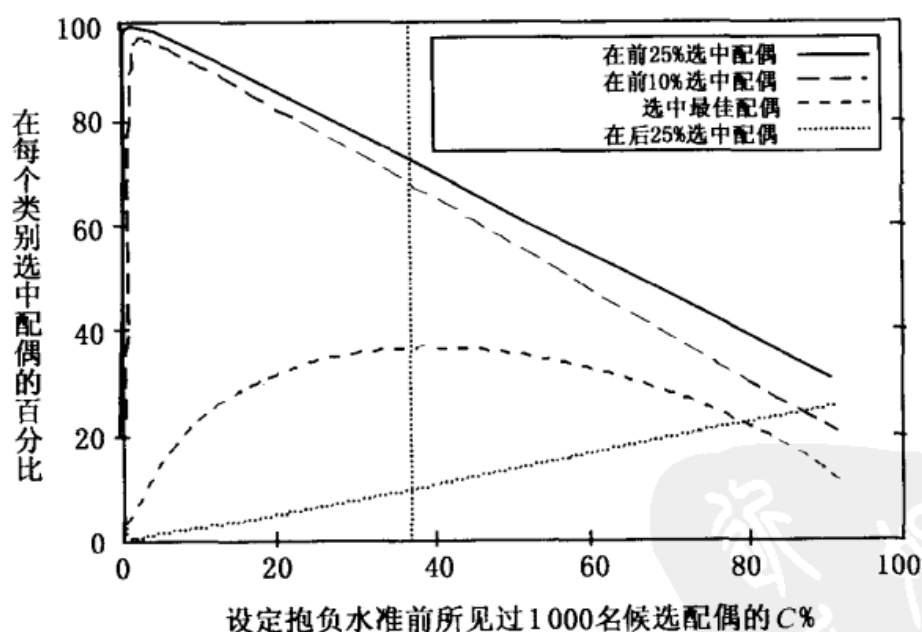


图 13-4 采纳下一个最佳者之前,给出不同待查对象比例(1000 名候选对象中的一部分)以设定抱负水准,在此基础上特定种类的值找到一个配偶的几率。请注意:为了使选中位列前茅配偶的几率达到最高,现在只需要检测较小比例的候选对象。“37%规则”在该效标上的成绩用垂直虚线来表示。

因此,为了使可能的婚配值达到最高,并把从1 000个候选对象中进行选择时所冒的风险降至最低,需要首先检测1%—3%的候选对象(即10—30人),以设定抱负水准值 D 。在前述候选对象总数为100的例子中,也需要检测其中的10个才能获得与此相同的成绩。因此,尽管候选对象总数增加了10倍,但需要检测的对象数量只有些微增加。这表明我们的TNB规则可以被简化。为了设定一种抱负水准,我们毋需按照特定比例来增加待检测对象数量,而只需要检测一定绝对数目的候选对象。该数目对于范围广阔的候选对象规模都是适用的——例如,“尝试12”(C=12)规则适用于从100到几千的候选对象规模。这种简化的搜索规则避开了早先针对“37%规则”提出的批评:它在多种效标上均获得了高于“37%规则”的成绩,毋需事先知道总体规模大小,亦不需要在作出选择前检测太多个体。这些结果均表明,富雷和艾森伯格(Frey & Eichenberger, 1996)关于短暂搜索不能找到适当配偶的悲观主义论调是不能成立的——即便很少搜索也能让你走得很远!

转向互动系列配偶搜索

这就是说,如果你是一个专制君主,能够迫使一群倒霉的候选配偶列队经过你的面前,直到你选中一个为止,那么即便只有很少搜索,你也能够做得相当好。尽管开始时几乎每个人都抱着青春期的幻想,试图找到最渴望得到的那个人,但是多数人很快就会发现,择偶活动的实际情形远非如此浪漫和简单。不妨设想你采用效果相当不错的自我中心式的“尝试12”规则来参与择偶游戏,目标是找到拥有最高婚配值的那个对象。你充分地考虑随机遇到的前12个对象,拒绝每一个,但记住其中最好的。从第13个开始,你审查随后的每个人,直到第20个人时,你发现你找到了要找的人:比你先前见过的候选对象都要好的人终

简捷启发式

于出现了。这满足了你所采用的规则,同时也很满你的意。你向你新找到的对象求婚,结果当场被拒绝了。问题究竟出在哪儿呢?

问题在于,在你评价这些候选配偶的同时,他(她)们也在评价你。如果你不能满足你所选中的那个人的择偶标准,那么无论你单方面的求婚行为多么殷勤,都是无济于事的(至少在此种受限制条件下是如此)。而且,假如你和该总体中的每个人都采用“尝试 301 12”规则来设定抱负水准的话,那么你们均将对自己与之婚配的对象有相当高的期望值。麻烦在于,如果你自身不具备很高婚配值的话,你将不会被其中任何一个人作为候选对象来选择,只能以独身而告终!

我们创设了一种新的模拟情境,以探索在互动双向搜索情境中不同配偶搜索规则是如何生效的。该情境包含了 100 名男性和 100 名女性构成的一个总体,男女双方的婚配值均为 0.0—100.0 的整数,而且每一方成员对方成员的婚配值都相当熟悉,但却不一定了解自己的婚配值。我们为这 200 个人中的每个人都提供了同样的搜索策略,首先让他们在“青春期”评价特定数目的异性。在此期间,每个人都可以调整自己的抱负水准,如果他们所采用的搜索规则需要的话。其后,将男性和女性随机配对,此时他们要么向对方提出求婚(主动提出婚配建议),要么拒绝这样做。如果每对中的两个人均向对方提出求婚,那么它们肯定会成为配偶,于是把他们从总体中排除出去。否则,这两个人将继续留在配对库中,以便再作尝试。这种“配对—求婚—成为配偶”的循环重复进行,直到每个人都有了配偶或每个人都有了评价每个异性成员和向其求婚的机会为止。我们感兴趣的是:采用不同搜索规则,谁在此种情境中找到了配偶?其他效标,如此种配偶搜索过程需要花费多长时间,也是我们关心的问题,但这里我们将不讨论它们。

图 13-5 表明当每个成员均使用同一种配偶搜索策略时,由

100 名男性和 100 名女性组成的总体能够婚配成功的对数。若每个人均使用 $C=1\%$ 的“采纳下一个最佳者”策略,即只检测一个个个体来设定抱负水准,那么大约一半人能婚配成功。但当我们增加了“青春期”长度(即增加了首先得到检测的候选对象数量 C)时,能够婚配成功的对数急剧地减少了。如果每个人均使用“尝试 12”策略,即在“青春期”仅检测 12 个候选对象来设定抱负水准,那么大约只有 8 对能够婚配成功。其原因可以从图 13-6 中看出来,它显示了所有婚配成功者的平均婚配值。对于使用“采纳下一个最佳者”策略的人来说,“青春期”越长(需要检测的候选对象数量 C 越多),婚配成功者的平均婚配值越高。也就是说,TNB 规则使总体中每个人都拥有过高的抱负水准,所以只有实际上具有最高婚配值的个体才能找到彼此满意的配偶。其他人只好抱着电视机度过周末之夜了!

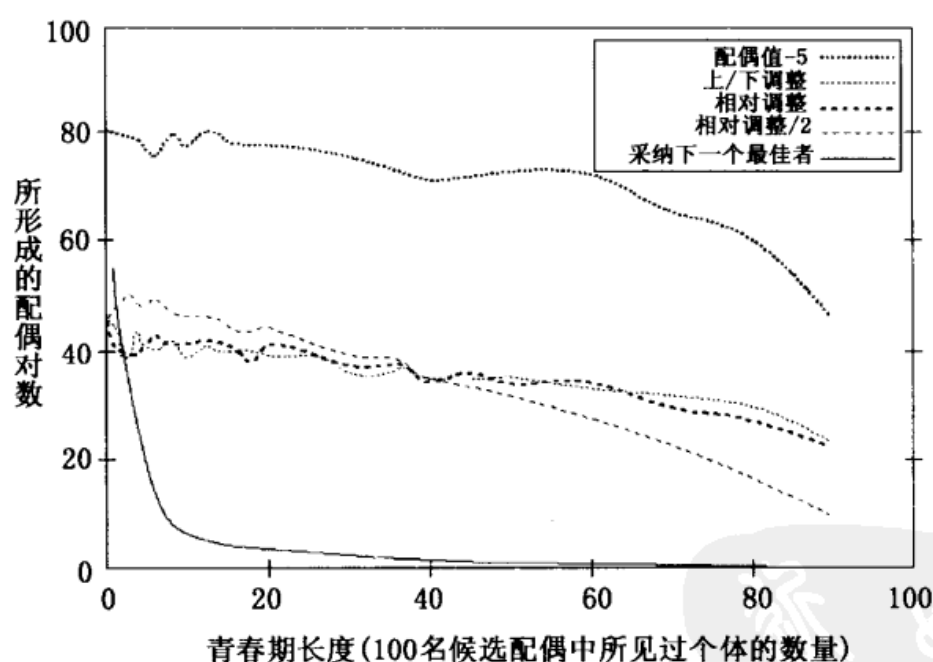


图 13-5 总体中所有个体均采用一种特定互动系列配偶搜索策略时,作为“青春期”(期间可以调整抱负水准)长度之函数的婚配成功对数。数值越大表明一种配偶搜寻策略越成功。

然而,为什么不采用 $C=1\%$ 的 TNB 规则呢? 这样抱负水准

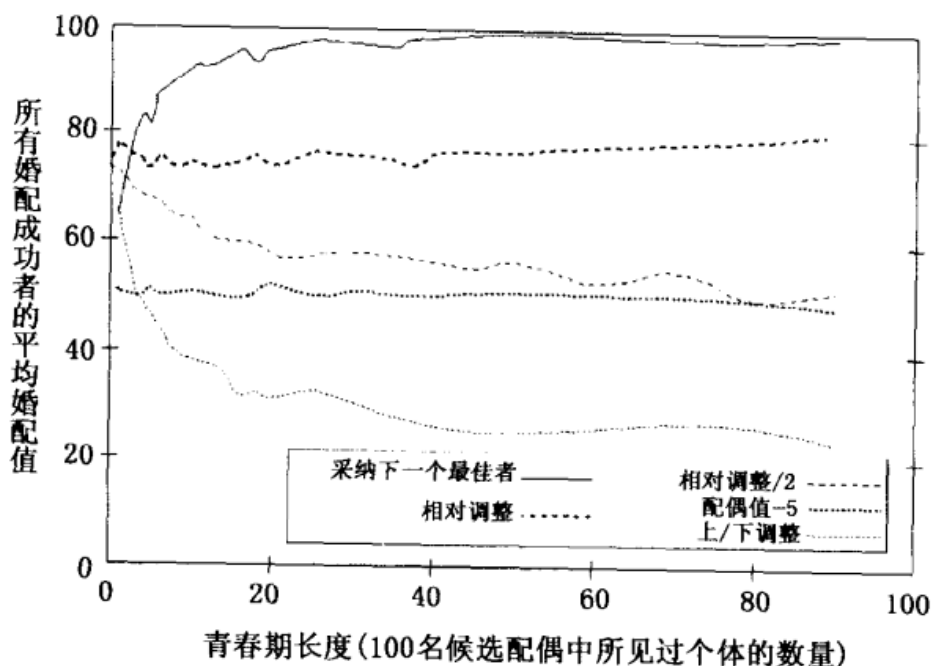


图 13-6 总体中所有个体均采用一种特定互动系列配偶搜索策略时, 作为“青春期”(期间可以调整抱负水准)长度之函数的婚配成功者的平均婚配值。中间的值(50 左右)表明更为成功的对等式配偶搜寻策略(如, 能够使总体中更多人而不仅仅是精英找到配偶的策略)。

不至于太高, 总体中将近一半人能够婚配成功, 它也许是更加合理的策略。问题在于用来衡量总体婚配质量的第三个效标上: 婚配成功配偶间婚配值的平均差异。这一点示于图 13-7 中。可以看到, 虽然仅检测总体中的少数个体(只有很短的青春期), TNB 规则确实可以产生很多配偶对数, 但这些配偶间是很不般配的——配偶双方的婚配值存在接近于 25 的平均差异。按照本章第一部分讨论过的博弈理论, 如此大的差异将会使双方的婚姻关系很不稳定: 很多个体倾向于更换配偶。那么, 我们如何才能找到这样一种互动系列配偶搜索规则: 它不仅能使总体中找到好配偶的人数比例增高(图 13-5 中的高值), 而且能使拥有各自婚配值的更多个体(图 13-6 中 50 左右的值)找到配偶, 同时在依据各自婚配值相互般配方面(图 13-7 中的低值)也是成功的。

现在设想, 你吸取了在婚姻问题上屡遭失败的教训之后, 这次你让自己变得更加现实了, 只希望找到一个与你有类似婚配值的

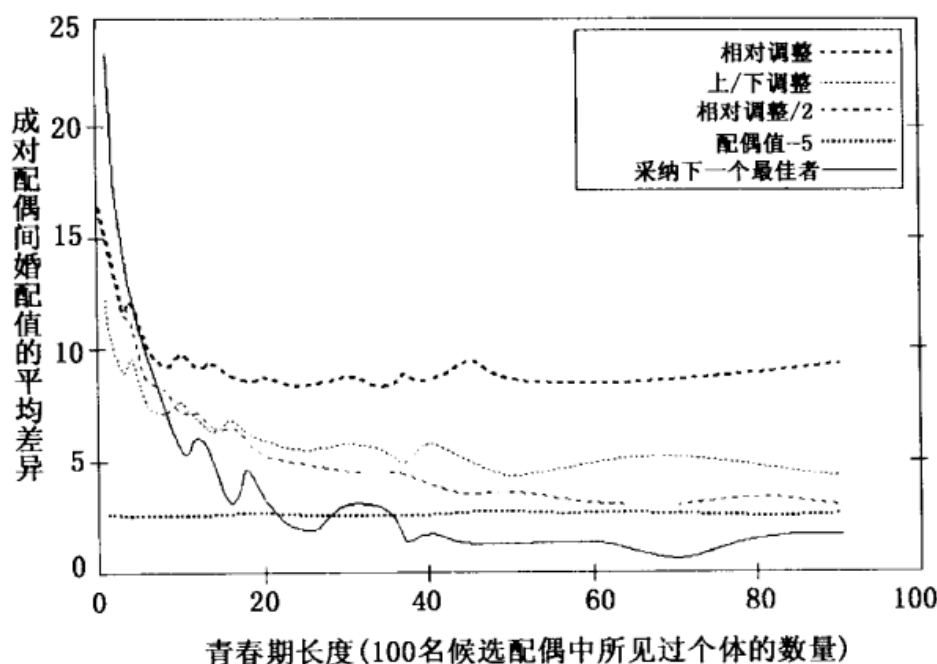


图 13-7 群体中所有个体均采用一种特定互动系列配偶搜寻策略时，作为“青春期”(期间可以调整抱负水准)长度之函数的婚配成功者之间婚配值的平均差异。较低的值表示在形成好的婚姻关系方面更为成功的配偶搜索策略。

配偶，而不再奢望找到某个高傲的好莱坞偶像。事实上，出于谦逊的考虑，你设定了一个比自己婚配值低 5 个百分点的阈值，向首先遇到的婚配值高于该阈值的任何一个个体求婚。现在你将会有什么样的遭遇呢？如果其他人也采取类似的谦逊阈值，他们又会怎样做呢？在图 13-5 中，我们看到，这种策略导致很高比例的个体找到了配偶。此种情况下，“青春期”不涉及了解或调整抱负水准的问题，因为你所设定的值是固定的，但它代表一个漫长的非生育期，期间你遇到了许多人却不能向他们求婚（以后你也不能再回到他们身边）。“青春期”长短对此种谦逊配偶搜索策略的成绩影响甚微。只有当“青春期”过长时，才会引起婚配成功对数的减少，这仅仅是因为不再有足够多的人数留下来等待你去搜索，以确保找到一位足够好的配偶。

在我们的其他测量指标上，这种基于婚配值的谦逊搜索策略做得也很好。由于大多数人均能够婚配成功，所以婚配成功者的

303

让我们更精明

平均婚配值处于 50 左右(见图 13-6)。它还成功地使具有相似婚配值的个体彼此成为配偶(参见图 13-7),从而有利于婚姻关系的稳定性。也就是说,该种策略能够成功地按照婚配值对个体进行归类,以满足其彼此婚配的要求。因此它似乎是一个可采用的有效互动系列配偶搜索策略。但是这里有一个问题:了解自己的婚配值并非一件容易的事情。我们不可能生来就了解这一点,因为它不仅容易受到环境条件的影响(它依赖于我们周围的其他人),而且伴随着我们的成长而随时发生变化。我们也不能仅凭对自己的观察来确定这个值,因为我们看待自己的方式与把我们作为候选对象来判断的其他人看待我们的方式不同。我们甚至不知道从异性角度评价我们自己时所依据的适当标准究竟是什么。在不具备这些基本知识情况下,我们必须设法估计出自己的婚配值,如果我们想要用它来设定我们的抱负水准的话。

304

因此,我们必须采取另外一个步骤,以使我们的配偶搜索策略具有越来越少的自我中心性。我们从仅考虑我们自己对他人的看法开始(采纳下一个最佳者规则),继而考虑到了我们是如何看待我们自己的(基于自我观察的抱负水准),现在我们将要考虑别人是如何看待我们的(基于反馈信息调整我们的自我知觉)。得到尝试的第一种基于反馈的方法为:每当我们得到来自于另外一个人的求婚时,就提高我们的抱负水准(类似于对我们自己婚配值的自我估价),而每当另外一个人没有向我们求婚时,就降低我们的抱负水准。我们将再次在特定时间的“青春期”里实现这一点(即使用从最初遇到的特定数量的个体那里得到的此类反馈信息)。每种反馈情境下我们对抱负水准的调整量取决于“青春期”的总长度,并与之成反比:如果我们有一个短的青春期的话,那么每次反馈后我们将作出更大调整(学得较快);如果有一个长的“青春期”,那么我们可能会学得较慢。因此,以抱负水准 50 为基准,我们使用如下公式来计算抱负水准调整量:调整量 = $50/(1+C)$ 。这里 C 是“青春期”内得到我们审查和审查我们的人数。这一规则使总体中

305

大约 40% 的人婚配成功(图 13-5 中的“上/下调整”曲线),但婚配成功者主要为总体中的后半个体(图 13-6 中婚配成功者的平均婚配值大约为 25)。这里究竟出了什么问题呢?

此种抱负水准调整策略的问题在于它使得寻找配偶者自视过高。每当有人求婚时,不管那个人的婚配值是多少,受到求婚的人都将会沾沾自喜,从而提高他或她的抱负水准。这样,婚配值高于 50 的个体将会得到许多次求婚,并将其抱负水准提升到过高的程度,而那些婚配值低于 50 的个体将会因更经常地遭到拒绝而降低其抱负水准,结果更加助长了另一半人的自我膨胀。但处于总体后半部分的个体,由于抱负水准较低,也很容易成功地找到配偶,而处于总体前半部分的自视过高的个体常常不能找到配偶。

关于你本人是否具有高的婚配值,请不要仅仅听信他人的言论,还应该考虑一下这话究竟出自于何人之口:正在评价你的那个人的婚配值是多少?假如那个人的婚配值比你对自身婚配值的估计要高,且他或她在向你求婚,那么你就应该提高自己的自我估价。其前提是:此人是一个良好效标,能够提供关于你的婚配值的准确反馈信息。(你将总是能够从婚配值低于你的个体那里得到求婚,因此你不应该因为他们的求婚而夸大自己的自我意象。)类似地,如果你从一个婚配值比你对自身婚配值的估计还要低的个体那里遭到了拒绝(未得到求婚),那么这将会使你对自己的自我意象产生怀疑,并导致你降低自己的抱负水准。(再一次,缺乏婚配值高于你的那些个体的求婚不应该对你的自我意象产生影响。)

如果我们参照其他人的婚配值来调整我们的抱负水准,并达到了像上一种策略同样多的调整量,那么总体中仍然有大约 40% 的个体能够婚配成功(图 13-5 中的“相对调整”曲线),但这次是总体中的前半个体相互配对(图 13-6 中婚配成功者的平均婚配值大约为 75)。然而,他们的配对并非最佳——这种策略使个体间产生了最糟糕的误配(大约有 10 个百分点的差异,参见图

13-7)。这次的问题在于我们对抱负水准所做的调整仍然与涉及到的婚配值无关,其调整量还是仅依赖于“青春期”长短的一种固定值。无论婚配值为 100 的个体向你求婚,还是婚配值为 60 的个体向你求婚,你都将抱负水准向上调整同样的量(假定你当初的自我意象值为 50 的话),这样做仍然无助于在个体间作出精细区分。相对于后一个求婚者来说,你应该对前一个求婚者感到更加兴奋,相应地你也应该把你的自我评估提升到更高水平。在下一策略中,我们正是这样做的。

当我们在所采用的策略中逐渐弱化了我们的自我中心倾向时,我们也就增加了更多的关于我们与之互动的其他候选配偶的信息。首先,我们看他(她)们是否向我们求婚;其次,我们仔细考虑他们的求婚,并考查相对于我们的婚配值而言他们的婚配值方向(即:只需知道其婚配值比我们自己的婚配值大或小,而不需要知道其婚配值的确切数量);现在我们来进一步考虑他们的求婚,并认真考查其婚配值与我们对自己婚配值估计之间的实际差异。如果某人向我们求婚,且其婚配值比我们对自己婚配值的估计要高,那么我们就将我们的自我估价(也就是我们的抱负水准)提高其差值的一半;如果“青春期”遇到的候选配偶没有向我们求婚,且其婚配值比我们对自己婚配值的估价还要低,那么我们就将我们的自我估价降低其差值的一半。采用这种方法,我们就把更大的权重放在了我们从婚配值与我们的自我估价相去甚远的个体那里获得的反馈信息上。

当我们这样做时,我们最终就找到了迄今所考虑过的抱负水准调整策略中的最佳策略。就较短或中等长度的“青春期”而言,该种策略使总体中大约一半个体婚配成功(图 13-5 中的“相对调整/2”)。学得越多(即“青春期”越长),它越有可能使配成对的个体趋于一种平滑均匀分布(图 13-6)。而且随着学习的增多,配偶之间的不般配现象逐渐接近于基于婚配值的谦卑策略的结果(图 13-7)。

这样,我们渐渐趋向于一种合理的互动系列配偶搜索策略。在假定我们了解异性成员婚配值的前提下,需要利用从他(她)们那里得到的求婚和拒绝求婚的反馈信息,来估计自己的婚配值。但值得注意的是,这种简单策略并未事先假定我们了解或能够计算出有关总体的任何信息。例如,我们毋需记录所遇到个体婚配值的平均值和标准差(如同 Martin 和 Moon 的一些策略所要求做的那样,1992);我们也毋需计算最优化的搜索时间(如同用来解决秘书或嫁妆问题的许多搜索方法所要求做的那样)。相反,在仅检测大约 20 个个体之后,多数效标就已经接近了正常值。而且我们不需要关于总体的先验知识,这一点将该种策略与解决双向匹配问题时考虑过的方法区别开来(Roth & Sotomayor,1990)。一个接一个地审视每一个个体,在此过程中了解我们自己,有了这些就已经足够了。

未来方向

这里我们提出了一系列单向和双向搜索的简单满意性启发式,它们能够在仅检测少数几种可能选项之后便确定适当的抱负水准。因此,这些启发式与我们关于有限理性的总体设想不谋而合:使用尽可能少的适用信息却产生了令人满意的成绩。从生态学角度看这些规则也是合理的,它们依靠环境的信息结构——这里即异性成员求婚和拒绝求婚的行为模式——来实现其适应性选择行为。

当然,关于配偶搜索还有许多问题有待进一步探索。总体永远不会是固定不变的,且择偶活动不可能在一个个互不连贯的孤立阶段进行,期间预先设定的一群人中的每个个体都必须婚配成功或者放弃婚配——相反,新的个体总是不断被引入,这对不同策略的总体婚配成功率将产生重要影响(Johnstone,1997)。我们这里使用的婚配值分布是均衡一致的,但在现实世界中它很可能接

307

让我们更精明

简捷启发式

近于正态分布。婚配值的不同分布将如何影响不同策略的成绩呢？这还是一个有待探索的问题。我们让总体中每一个人对所有异性成员形成完全相同的印象（所有女性采用同样方法排列男性的等次，反之亦然），这也是不现实的：关于谁是一个好的候选对象或谁是一个糟糕的候选对象，人们在一定程度上会有一些相同看法，但在配偶选择偏好上又存在着许多个别差异。在这些个人偏好差异中，有些仅与纯粹的审美标准有关，但也有一些具有重要的适应性后果（如，与远亲而非近亲结婚——参见 Bateson, 1983b）。

这引出了我们必须强调的另外一个问题：对于诸如此类的算法规则来说，我们应该用来对它们加以比较的最重要的维度是什么？这里我们已经指出，如果搜索时间、被选择对象的平均婚配值或其分布能够得到改进的话，那么从总体中找到唯一最佳配偶并不一定是最好的适应性目标。而且，在互动双向搜索情况下，最终找到一个合适配偶意味着选择婚配值与自己相仿的个体。但我们需要累积证据以支持这些主张。解决此问题的方法之一是创设一种进化模拟情境，在该情境中不同算法规则为了好的配偶和后代彼此展开竞争，看一看哪种算法规则随着时间推移而最后胜出。然而，这种方法仅仅在以下意义上有助于我们了解现实进化的人类（或动物）行为的一些事实：我们成功地将有关的生态细节（比如，婚配值是如何影响后代数量的）纳入到了我们的模型之中。

我们的最终目标是为人类和其他动物在实际进化过程中采用的特定搜索策略寻找证据，正如其他人在包括配偶选择在内的实验情境中所做的那样（如，Alatalo et al., 1988; Harrison & McCabe, 1996; Hey, 1982, 1987; Martin & Moon, 1992; Rapoport & Tversky, 1970; Sethuraman et al., 1994）。然而，一直有这样一种担心：实验情境也许不能触及到（tap into）真实行为所涉及的心理机制。因此，为实际观察到的动物或人类配偶搜索策略寻找证据也是非常重要的。我们的模拟研究试图揭示哪种搜索规则从心理

学上来看是可能的、简单的,同时又是有效的,以表明我们可以期待和应该寻找的策略究竟是什么,从而为此类关于现实行为的研究起到抛砖引玉之效。

所有这些并不是说爱在配偶选择中无关紧要,也不意味着配偶选择可以仅仅用百分数、抱负水准和适应性自我评价等概念来加以说明。爱可以成为坚持特定选择的理由,它可以减少或消除配偶间任何外在的不般配,并使进一步搜索变得完全没有必要。爱和其他情感是行为机制的重要组成部分,而不是与适应性行为相抵触甚或相敌对的独特而不明确的力量。但是爱可能——事实上也的确——妨碍了配偶选择过程中决策机制的正常运行,因此当某人深陷其中之时,整个配偶选择过程似乎是神秘莫测的。选择配偶不是科学研究活动,但我们希望科学研究能够被用来揭示潜藏于人们彼此搜索和寻找背后的某些行为模式。



借助于简单决策规则的父母投资

詹妮弗·奈瑞瑟·戴维斯(Jennifer Nerissa Davis)

彼得·M·托德(Peter M. Todd)

有一位老夫人生活拮据,太多的孩子使她不知所措。

——传说中的童谣

这位老夫人可能没有太多东西分给她的孩子们,但她仍不得不想办法把仅有的一杯残羹冷炙分配给他们。她怎样才能做到这一点呢?为人父母者如何决定怎样在孩子们身上分配他们的时间、金钱和精力呢?如果现实中此类东西可以加以量化的话,他们可能会尝试进行一些复杂分析,充分估计当前可能作出的所有选择从现在起到孩子们独立时止的所有代价和收益。由于需要对诸多不确定的未来事件进行大量计算和预测,所以找到解决此类问题的“最佳”方案是相当复杂甚或是不可能的。然而,这并不意味着父母们必须通过复杂计算才能进行明智的投资。相反,他们可以凭借一些简单规则来指导他们对孩子的投资。本章我们提供了一项研究的结果,该项研究就是为了检验此种简单规则的有效性。

父母投资

经济学家和行为生态学家都曾提出过父母如何在孩子们中间分配资源的问题。然而,乐观一点说,他们创建的模型通常需要难以数计的信息;悲观一点说,这些模型需要的信息实际上是不可知

的。例如,贝克(Becker,1991)曾对理性的父母如何在孩子们中间分配资源这一问题作了一种经济学分析。他假定父母们总是力图最大限度地提高“总体子女素质”(total children quality),它是以前所有子女成年后获得的财富总量来界定的。一个孩子的素质根据其童年期所消耗的资源、他或她自身的技能和能力,以及他或她成年后靠运气挣得的额外收入加以综合衡量。贝克假定,就对一个孩子投资的盈利而言,所得到的回报是递减的。也就是说,你每增加一份对一个孩子的等量投资,对这个孩子素质的影响将会越来越小。贝克的分析表明,只要对于所有孩子来说这种盈利曲线是等同的,父母们就应该均等分配其资源,以便每个孩子都能产生同样多的盈利。然而,如果对某些孩子每单位投资比其他孩子产生了更多盈利,那么父母当然应该更加偏爱他们。

类似的投资建议在有关父母投资问题的生物学文献中也能找到。在那里个体素质通常以其成年后可望生育的后代数量来界定,也就是说,父母总是期望其子女尽可能多地生育后代。只要所有子女有同样多的预期繁殖量,父母就应当平等地投资于它们(Lacey et al.,1983;Real,1980),但是如果后代存在着差异,并且影响到了父母对投资收益的预期,那么父母们就应该注意这些差异,并相应地使投资有所倾斜。

尽管这些结论听起来似乎是合理的,但在对特定情境中父母实际行为作出预测时却很少有人使用它们。这是因为这两种观点都假定:父母有办法计算每单位投资对他们将来有望从每个子女那里获得的盈利产生的影响。然而,当付诸实施时这种计算需要涉及难以获得的信息。例如,在贝克的做法中,子女的总价值是每个孩子成长过程中父母全部投资的价值,每个孩子成年后自身禀赋的价值,以及每个孩子靠运气在市场上获得的资产的价值之和。试图采用这种分析方法对行为作出具体预测时面临的最大难题是:确定这里涉及到的几种未来价值是不可操作的或不可行的。子女们并没有为了方便父母而佩带着“投资计量器”来到世

上。要想兑现诸如此类的投资机制,必须给出一些具体明确的预测未来回报量的方法,但这即便不是不可能的,也是难以实现的。

父母投资的机制

在行为生态学和进化心理学领域,许多业已完成的关于父母投资决策的研究都将父母不可能完成此类复杂等式运算作为前提。相反,这些研究将注意力集中在了父母可以用来取得成功的更为简单的决策规则上。在此成功仍被界定为尽可能多的后代数量。迅速作出粗略估计的一种通用方法是将抚养成年的后代数量作为测定父母投资成效的指标。正如前面提到过的那样,如果所有后代都对父母投资给予了同样多的预期回报量,那么父母就应当平等地对待他们。当然,即使在这种情况下,父母仍然需要作出关于在特定时间供应多少以及何时供应的决策。当每个后代都需要达到一种既定供应水平,且对所有后代的投资量均相同时,父母可以根据每个后代的需求进行投资。例如,在鸟类中,幼鸟饥饿时经常会讨食。如果讨要的次数诚实地、准确地反映了它们的需要状况(例如,Godfray,1991),那么我们会期望它的父母按照讨要的强烈程度给它们的后代喂食,以达到投资的公平性。这显然是一种非常简单的决策规则。

当然,在现实中所有后代对父母投资都给予同等回报的情形是很少见的。如果不同后代的预期回报量各不相同,父母应当敏锐地觉察到这一点,并采用一种有利于回报量较大后代的倾斜性投资决策。对此,行为生态学家再一次提出了一种充分体现这种偏向的简单决策规则:首先满足长子(女)的需要。为了理解这一规则为什么有效,有必要首先了解一下引起投资回报差异的因素。

在贝克看来,具有较高回报率的子女是那些成年后能取得巨额收入的具有突出天赋或才能的子女。而当行为生态学家谈到投资的高回报率时,他们指的是那些较之同胞兄弟姐妹有更高几率

存活到生育年龄、并能繁殖更多自己后代的子女。存活可能性的一个主要预测指标是后代目前的年龄。年龄越大,越接近于独立生活和繁殖期,也越有可能实现所有投资目标。因此,如果父母打算有偏向地进行投资的话,他们就应该优先照顾年龄最大的子女。于是,一种简单决策规则便形成了:优先投资于年长子女。

如果我们想要预测父母行为,就需要找到一些方法来确定父母面对的究竟是什么环境。他们应该根据子女的需要来投资呢?还是根据它们的年龄或其他一些尚不明确的变量来投资呢?例如,即使所有后代的年龄均相同,假如它们在身体大小之类的一些变量上不同,它们也不应该得到同样多的资源,因为身体大小之类的因素也很可能与将来的成功有关。

如果资源匮乏,父母无法成功地抚养所有后代,那么偏爱长子是具有适应意义的。如果你不能保全所有子女,就要保护那些最有可能长大成人的子女。然而,此种情形下作出的选择或许完全不同于有足够资源情况下作出的决策。在后一种情形下,没有哪种理论假设可以使父母作出决断:高额投资于某些后代而不是其他后代将会取得最大成功。因此,不是仅抚养一个投资回报特别好的后代或少数几个此类后代,相反,采取其他投资策略也许能获得更高回报。例如,父母均匀分配资源或偏袒那些需要较多资源才能达到同等满足程度的后代,获得的回报或许更高。因此,人们总是期望父母作出与环境要求相一致的投资决策:在恶劣的环境条件下,他们可以选择优先投资年龄较长后代的简单策略,而在充裕的环境条件下,他们最好依据其他规则进行投资。

父母用于投资的决策规则可能是非常简单的。在本章中,我们将要讨论这样一个问题:设若父母明确决定每个后代的投资效果是一个难以操作的问题,那么是否依然存在一些简单决策规则:父母们毋需计算无法计算的变量便能运用这些简单规则指导他们的投资活动。我们相信这些决策规则是存在的。

确定决策规则

借助于对投资问题的计算机模拟,我们解决了为不同资源水平寻找成功的父母投资策略问题。由于人类父母投资的情形很难加以刻画,而且我们正在研究的是一个适用于各种有机体的一般理论问题,所以我们选择了鸟类父母投资这样一种简单情形来建立模型。像人类一样,许多鸟类也面临着同时抚养多个不同年龄和不同发展水平幼鸟的任务。许多种类的幼鸟彼此之间孵化成熟的时间都要相隔一天。对于鸟类来说,这种年龄差异将会造成它们个头大小和成熟程度方面的明显不同。鸟类父母投资是易于量化的,而且已经有人对多种鸟类的投资策略进行了测量。相反,人类投资可能采取许多不同形式,诸如提供食物的热量、花费资金的数量以及人际互动的质量等,因此很难用某种单一的值来表示。而且,个体投资策略的变化也尚未得到系统测量。

更具体地说,我们研究的动物是西方蓝鸫,或者毋宁说是这种鸟的电子模拟物。我们模拟了孵化、喂养、成长、幼鸟离巢(羽翼丰满)等一系列过程。所以选择蓝鸫作为研究对象,是因为关于这种鸟代谢作用和发育速率的信息容易获得,这些信息对于建立一种精确模型来说是必需的。代谢率、消化率及胃部大小等资料来源于已经发表的关于蓝鸫整个雏鸟期发育和代谢率的文献(Mock et al., 1991)。蓝鸫喂给幼鸟的昆虫含有的热量以及这些食物包含的可代谢能量的比例等信息来自于迪克斯托(Dykstra)和卡拉索夫(Karasov)的研究资料(1993)。

我们的目标是看一看多种环境条件下何种父母投资策略对于我们模拟的蓝鸫抚养一窝幼鸟更为有效。在我们的模拟实验中,父母投资仅限于父母一方提供给一只幼鸟的食物总量,而投资成效最后离巢飞走的“幼鸟克数”总量来衡量。我们采用幼鸟克数而不是离巢幼鸟数量作为衡量标准,是因为我们预期幼鸟离巢时

体重的增加与其在第一年中的存活能力密切相关；也就是说，我们预期离巢幼鸟体重的增加与幼鸟顺利度过第一年和第一个繁殖期的能力有关。在我们的模拟中，父母总是尽可能使离巢幼鸟的数目更多，而不是有数量较少的胖幼鸟，因为数目不等的各巢幼鸟的总克数不可能相同，这是由每只离巢幼鸟可能的体重极限决定的（例如，三只幼鸟的总克数约在 78 克到 84 克之间，而两只幼鸟的总克数可能在 52 至 56 克之间）。

在我们的模拟中，食物是以碎块形式出现的，所以有时正在觅食的父母能遇到成排的几只昆虫，而随后的一段时间又找不到任何虫子了。为了探讨不同环境条件下各种投资策略是如何起作用的，我们操纵两个变量以创设不同环境条件：可得食物的平均数量和食物块大小。所以选择这两个参数，是因为它们似乎最有可能影响各种父母投资策略的效果。例如，正像上面提到过的那样，在只能得到较少食物的情况下，首先喂养年龄最大幼鸟可能是最好的选择。

我们检验了六种不同的父母决策规则。有研究发现，一些鸟类父母偏向于优先喂养年龄最小或个头最小幼鸟，因此我们首先检验了这两种规则（按照从幼到长的年龄顺序喂养和按照从小到大的个头顺序喂养，例如，Stamps et al., 1985）。虽然该种文献倾向于把这两种规则混为一谈，但在我们看来它们是很不相同的。如果总是先喂个头最小幼鸟，它就会迅速发育，而其他幼鸟则长得较慢，因此个头最小幼鸟是随时间发生变化的。然而，年龄最小幼鸟将始终是年龄最小的。此外，父母辨认个头最小幼鸟所需要的知觉能力不同于辨认年龄最小幼鸟所需要的知觉能力。前者仅需要作出大小判断，而后者却依赖于对后代的个体再认——这是父母可能具备、也可能不具备的一种能力。

我们还检验了另外两种投资策略：依据需要程度进行投资（即首先喂最饥饿的幼鸟）和更多地投资于年长后代（即按照从大到小的年龄顺序喂养）。我们曾考虑过对首先喂养个头最小者规

则的对立面——首先喂养个头最大者——加以检验,但是根据幼鸟的发育规律,只要后代在年龄上不同,首先喂养个头最大者就相当于首先喂养年龄最长者。我们还检验了按固定顺序循环喂养,以便每只幼鸟得到的喂养次数均相同的简单规则。该规则再一次要求父母能够识别它们的每只幼鸟——一种更加困难的知觉任务。此外,随机化投资也被作为比较基准纳入待检验策略之列。我们猜想父母们将不会浪费时间去选择一种其投资效果尚不如随机化投资效果的决策规则,而且在某些环境中随机化投资可能是最佳选择。

更确切地说,得到我们检验的六种启发式规则是:

1. 按照从幼到长的年龄顺序进行喂养

父母首先把昆虫喂给年龄最小幼鸟;如果这只幼鸟喂饱了,再将昆虫喂给年龄第二小幼鸟,以此类推。

2. 按照从小到大的个头顺序进行喂养

父母首先把昆虫喂给个头最小幼鸟,然后是个头第二小幼鸟,以此类推。(由于进食后幼鸟的个头会增长,所以该规则产生的效果不同于上一种规则。)

3. 按照饥饿程度进行喂养

父母首先把昆虫喂给最饿的幼鸟(即胃内相对空间最大的幼鸟)。

4. 按照从大到小的个头顺序进行喂养

父母首先把昆虫喂给个头最大幼鸟,然后是个头第二大幼鸟,以此类推。

5. 按照固定顺序循环喂养(轮流喂养)

父母按照固定顺序把昆虫喂给每只幼鸟,也就是说,首先喂给幼鸟1,然后喂给幼鸟2,以此类推。该固定顺序最初是随意选择的。

6. 随机喂养

父母把一只昆虫喂给随机选择的一只幼鸟。如果这只幼鸟吃