



数字缉凶

The Numbers Behind Numbers
Solving Crime With Mathematics

美剧中的数学破案

基思·德夫林 加里·洛登 著

陆继宗 译



NLIC 2970701414



解密热播美剧《数字追凶》

看数学家与罪犯斗智斗勇

数字缉凶：

美剧中的数学破案

◎ 基思·德夫林 ◎ 加里·洛登 著

◎ 陆继宗 译

◎ 上海科技教育出版社



NLIC 2970701414

图书在版编目(CIP)数据

数字缉凶:美剧中的数学破案/(美)德夫林,(美)洛登著;陆继宗译. —上海:上海科技教育出版社, 2011.6

ISBN 978-7-5428-5145-1

I. ①数… II. ①德…②洛…③陆… III. ①数学—应用—刑事侦查—普及读物②电视剧—简介—美国 IV. ①O1-49②J905.712

中国版本图书馆 CIP 数据核字(2010)第 250553 号



引言：英雄是一位数学家？

2005年1月23日，一部名为《数字追凶》的刑侦电视连续剧首播。这部由尼克·法拉奇夫妇和谢里尔·霍顿编写剧本的电视连续剧是派拉蒙电视网络公司(Palamount Network Television)拍摄的，好莱坞资深导演里德利·斯各特(Ridley Scott)和托尼·斯各特(Tony Scott)的加盟使它赢得了广大观众的喝彩声，他们执导的电影有：《异形》、《壮志凌云》、《角斗士》等。自播出以来，《数字追凶》自然而然地击退了竞争对手成了星期五晚上该电视剧播出时段中的收视冠军。

使大家惊奇的是剧中两个英雄之一竟是一位数学家，而且许多行动涉及数学。查利·埃普斯作为一名教授，利用强有力的数学技巧帮助他的哥哥丹(一名联邦调查局特工)找出并抓住罪犯。许多电视剧观众以及有些评论家曾评论说，这些故事很具娱乐性，但基本前提很牵强：你不能简单地利用数学去破案。然而本书却证明，他们错了。你可以利用数学去破案，虽然并非对每一案例都可行，但执法机构常常能使数学成为在对犯罪无休止的战斗中的有力武器。事实上，正如我们将在下一章中讨论的，此连续剧的第一集就是以现实生活中的一个案例为基础的。

我们打算用非技术性的方式来描述目前对警察、中央情报局和联邦调查局有用的一些主要数学技巧。这些方法大部分已在《数

字追凶》的剧本中被提及,在频繁地将我们的解释与播出的剧情相联系的同时,我们的注意力集中于这些数学技巧以及它们如何能被执法者所使用。另外我们也描述一些实际的案例,在这些案例中,数学对破案起了作用,虽然在电视剧中并没有运用过——至少没有直接运用。

在许多方面,《数字追凶》类似于一部优秀的科学小说,以正确的物理和化学知识为基础。每个星期,《数字追凶》播出一集情节跌宕的故事,在这个故事的叙述中实际可行的数学起了关键作用。《数字追凶》的制片人竭尽全力来确保剧中使用的数学知识是正确的,所演示的应用是可能的。虽然有些案例观众看来是虚构的,但它们确实发生了,而有些案例则是非常可能发生的。这部电视连续剧采用了某些戏剧性手法,但本书没有。在本书中,你将发现数学能用于,而且已经用于与现实犯罪作斗争并抓获了真正的罪犯。

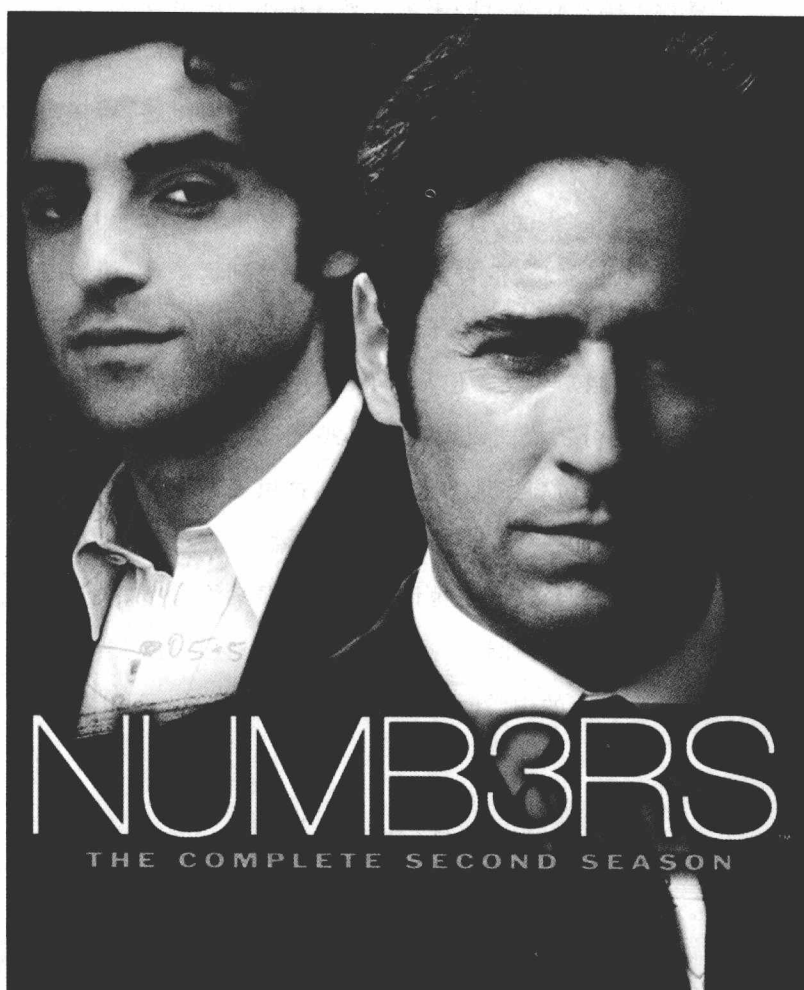
目录

001~209

001	1. 寻找热区——犯罪的地理分析技术
015	2. 使用统计101与犯罪作斗争
029	3. 数据挖掘——在大量的信息中 找出有意义的模式
059	4. 字迹什么时候第一次出现在墙上—— 改变点检测
073	5. 图像的增强和重构
089	6. 预言未来——贝叶斯推理
103	7. DNA图谱分析
123	8. 密码——编码与解码
143	9. 证据有多可靠——对指纹的怀疑
161	10. 点间连接——网络的数学
179	11. 囚徒困境、风险分析和反恐怖主义
203	12. 法庭上的数学
223	13. 赌场中的犯罪——用数学攻击系统
241	附录:《数字追凶》前三季剧情中的数学概要

1

寻找热区——犯罪的 地理分析技术



联邦调查局特工丹·埃普斯(Don Eppes)再看了一眼平摊在他父亲家餐厅桌子上的一张很大的洛杉矶街区图。地图上墨水打的叉号标记了犯罪地点,在过去的几个月内,一个残忍的连环杀手袭击、强奸、然后杀害了数名年轻的女性。丹的任务是在这名杀手再次作案前抓住他。但是调查毫无进展。丹没有线索,也不知道下一步该做什么。

“要帮忙吗?”这是丹的弟弟查利(Charlie)的声音,他是附近加州科学学院^①数学系的一名杰出年轻教授。丹一直对他弟弟不可思议的数学才能佩服得五体投地,真诚地欢迎任何能够得到的帮助。但是……从一个数学家那里得到帮助?

“此案与数字无关,查利。”丹语气尖锐并不是由于生气,更多的是由于遇到了挫折,但查利似乎并未注意到这一点,他的回答完全是就事论事而并非固执:“任何事情都有关数字”。

丹并不信服。当然,他常常听到查利说数学完全是研究模式的:辨认模式、分析模式以及预言模式。但这并不是让一个数学天才去看杂乱无章散布在地图上的叉号。这里面不存在模式,任何人都没有办法预言下一个叉号将打在哪个精确位置上——在哪里下一个年轻的姑娘会被袭击。也许就是今晚罪案将再次发生。如果叉号的排列具有某种规律性,那么可以用一个数学方程来描绘这个模式,就像丹记得的中学时代学到的知识:方程 $x^2 + y^2 = 9$ 描述的是一个圆。

看着地图,即使查利也不得不同意,无法用数学来预言杀手下一次将在哪里作案。他走到窗边,凝视着花园,只有浇灌草坪的自

^① 这是一所虚构的大学,其原型可能是加州理工学院,剧中的不少场景就是在那里拍摄的。——译者

动喷水龙头不断发出的扑哧-扑哧-扑哧-扑哧声音打破了傍晚的宁静。查利的眼睛注视着喷水龙头,但思绪却飞得很远。他不得不承认丹或许是对的。能用数学来做许多的事,要比大多数人认识到的多得多。但是,要利用数学,必须要具有某种模式。

扑哧-扑哧-扑哧-扑哧。喷水龙头连续不停地在工作。纽约有一位著名的数学家用数学来研究心脏工作的方式,帮助医生在病人心脏病发作前找出心跳中细微不规则的症状。

扑哧-扑哧-扑哧-扑哧。银行利用所有那些以数学为基础的计算机程序来跟踪信用卡购物,寻找模式的突然改变,这些模式改变可能标示着身份的盗用,或者卡片被窃。

扑哧-扑哧-扑哧-扑哧。没有聪明的数学算法,查利口袋中的手机将是现在的两倍大,并且重得多。

扑哧-扑哧-扑哧-扑哧。事实上,现代生活中几乎没有哪一个领域可以不依赖数学,而且常常是以一种重要的方式依赖的。但是必须有一种模式,不然数学无法走出第一步。

扑哧-扑哧-扑哧-扑哧。查利第一次注意到了这个喷水龙头,突然他知道该怎么做了。他有了答案。他能够帮助丹破案,这个答案一直在他面前晃来晃去。而他恰恰一直未能认识到这一点。

他把丹拖到了窗前。“我们一直在提一个错误的问题,”他说。“从那些你已经知道的,你根本无法预言凶手下一次将在哪里作案。”他指着喷水龙头说。“就像不管你对溅落草坪上的每滴水滴的位置研究得如何透,你都无法预言下一滴水滴将溅落在何处。有太多的不确定性。”他看了丹一眼,确信他的哥哥正在听。“但是设想你看不到这个喷水龙头,你所能得到的是溅落到地上的所有水滴的分布模式。然后利用数学你就能精确推算出喷水龙头在什么地

方。你不能利用水滴的模式去预言下一滴水滴将落在哪里,但它能够用它反推出源头在哪里。这对你寻找凶手来说也是一样的。”

丹觉得很难接受他弟弟正在给出的建议。“查利,你是否是在告诉我,你可以推出凶手的落脚之处?”

查利的回答很简单:“对。”

丹仍然怀疑查利的想法是否真的管用,不过他弟弟的信心和热情鼓舞了他,他同意让查利协助调查。

查利第一步是学一些犯罪学的基本科学常识:首先,连环杀手的行为是怎样的?在这一点上,他作为一个数学家的多年经验教会他如何辨认哪些是关键因素,而忽略所有的其他因素,以使一个看似复杂的问题能够简化成一个仅包含少数几个关键变量的问题。例如,在与丹以及他哥哥工作的联邦调查局办公室里的其他特工交谈后,他知道了暴力连环案的罪犯在地点的挑选上会呈现出某种倾向性。他们倾向于在靠近他们的住所的地方实施犯罪,但不会挨得太近;他们会环绕他们的居住所在地设立一个“缓冲区”,在此区域内他们不实施犯罪,这个区域太靠近住处故要保持舒适状态;而在这舒适区外,犯罪点出现的频率是随着离住所距离的增加而减少的。

然后,查利回到他在加州科学学院数学系的办公室,兴奋地认真工作了起来,黑板上布满数学方程和公式。他的目标:找到数学的钥匙,确定一个“热区”——地图上的一个区域,由作案地点反推出来,是罪犯最可能居住的地区。

就像他在着手解一个艰难的数学问题时经常发生的那样,查利尝试种种方法却都没有成功时,时间在飞快地流逝。最后,他有了一个他认为可以管用的想法。他擦掉了以前写在黑板上的潦草

字迹,写下了这个看起来很复杂的公式:^①

$$P_{ij} = k \sum_{n=1}^g \left(\frac{\phi}{(|x_i - x_n| + |y_j - y_n|)^f} + \frac{(1-\phi)(B^{gf})}{(2B - |x_i - x_n| - |y_j - y_n|)^g} \right).$$

“要用一个小窍门,”他喃喃自语。

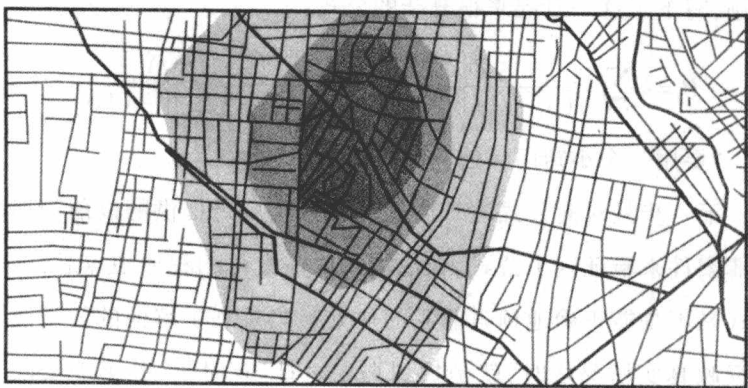
下一步是对此公式进行微调,就是用由丹提供的以前的连环犯罪案件来验证这个公式。当他输入以前案件的犯罪地点后,这个公式是不是能正确地预言罪犯的住处?当查利了解他的数学能不能反映真实情况之时,就是找到真相之日。有时候,并不能反映真实情况,他就知道当初决定的哪些是考虑在内的因素,哪些是可忽略的因素,一定有误。不过,这时候,在查利作了某些微小的调整后,这个公式似乎就管用了。

第二天,查利的活力和自信大爆发,他在联邦调查局办公室展示了一份夺目地标明了“热区”的犯罪地点打印地图。正如丹回忆起在中学时代学到的用 $x^2 + y^2 = 9$ 描述一个圆那样,把这个方程输入一台有合适程序的计算机时它就画出一个圆周,同样查利把他的新方程输到他的计算机中时,计算机也会产生一个图形。不过这次不是一个圆周——查利的方程要复杂得多。计算机在丹的洛杉矶犯罪地点图上给出了一系列着色的同心区域,颜色逐步向中心加深,这些区域就是凶手居住在其中的热区。

虽然有了这张图也仍然还有许多工作留给丹和他的同事们去做,但是此时寻找杀手,已不再像是在干草堆中捞针了。多亏了查利的数学,一堆干草顷刻间变为仅仅一袋干草。

查利向丹和调查此案的其他联邦调查局特工解释,虽然这个连环杀手在他认为随机挑选的犯罪地点寻找被害人,试图不暴露

^① 我们马上会仔细地考察一下这个公式。



他的住处,但是数学公式揭露了真相:以很高的概率确定了罪犯居住的热区。丹和他的小组决定侦查居住在这个热区中的某一年龄段的男子,并使用监视以及隐蔽的手法,从嫌疑人抛弃的烟头、吸管或者其他类似物上获得 DNA 证据,看看哪个与从犯罪现场侦查获得的 DNA 样本相符。

在短短几天之内——经历了几次让人心跳停止的紧张时刻后——他们抓到他们正在寻找的罪犯。案件破了。丹告诉他的弟弟:“查利,这要归功于你得到的公式。”

事实,抑或虚构?

除少数戏剧性情节外,上面所叙述的就是电视观众看到的2005年1月23日首播的《数字追凶》第一集中的情节。许多观众不相信数学能以这种方式帮助捉拿罪犯。事实上,整个第一集中的情节是基于一个非常相似的真实案例,在此案中用了一个数学方程来确定罪犯居住地的热区。它就是观众看到的,查利写在黑板上的那个方程。

现实生活中得到这个公式的数学家,名叫金·罗斯莫(Kim

Rossmo)。罗斯莫用来预言连环犯罪案凶犯居住在哪里的数学技巧叫做地理分析技术^①。

20世纪80年代,罗斯莫是加拿大温哥华警察局的一名年轻的警察。使他成为一名不同寻常的警官的是他的数学天赋。在校期间,他一直是一名“数学能手”,是那种使得其他同学(常常还有教师)神经有点紧张的学生。有个故事说,他在十二年级开学后不久,嫌数学课进度太慢,因此在这个学期的第二周就要求进行期终考试。在得了一百分后,他被准予免修这门课程的其余部分。

同样也是抱怨涉及连环暴力犯罪的侦查太过缓慢,罗斯莫决定回到学校去,他获得了西蒙·弗拉塞大学(Simon Fraser University)犯罪学的博士学位,是加拿大第一个得此学位的警察。他博士论文的导师保罗·布兰廷汉和帕特里夏·布兰廷汉(Paul & Patricia Brantingham)是开发犯罪行为数学模型(实际上就是描写某种情况的方程组)的先驱,特别是基于罪犯的生活、工作和娱乐来推断罪案最有可能在哪里发生的那种模型。(正是布兰廷汉夫妇注意到了连环犯罪的地域分布模式,即观众看到的查利从丹和他的联邦调查局同事处得知的那种分布模式。)

罗斯莫的兴趣与布兰廷汉夫妇的稍有不同。他不想研究犯罪行为的地域分布。作为一名警官,他想用与未知的单个行凶者的犯罪地点相联系的真实数据作为一种侦查工具来帮助警察发现罪犯。

罗斯莫已在重新分析老案件上取得了一些初步的成功,在获得了他的博士学位并被提升为探长后,他把兴趣倾注到了开发更好的数学方法来进行他所称的犯罪地理定位。其他人则称此方法为“地理分析技术”,因为它与著名的“心理分析技术”相辅相成,后

^① 一种由系列犯罪地点的地理关系来推断犯罪嫌疑人可能落脚点的方法。——译注

者被侦查人员用来根据犯罪分子的行为和心理特征来寻找罪犯。地理分析技术企图由分析犯罪分子的犯罪地点来确定一个罪犯可能的基本行动规律。

罗斯莫那个外观神奇的公式后面的关键思想,是1991年某一天他在日本的一列子弹列车上获得的。他发现身边没有可供记录的笔记本,就把它写在一张餐巾纸上。经过后来的一些改进,这个公式成了罗斯莫编写的一个计算机程序的基本组成部分,这个程序叫做 Rigel (读作 Rye-gel,是以猎户座中的一颗恒星命名的^①)。如今罗斯莫把 Rigel 销售给了遍及全球的警察以及其他侦查机构,帮助他们寻找罪犯,并且提供培训和咨询。

在罗斯莫向对此程序感兴趣的执法机构描述 Rigel 是如何工作的时候,他提供了一个他所喜欢的比喻——就是由分析草坪上的一个旋转喷水龙头喷洒在地面上的水滴的分布来确定龙头的位置。《数字追凶》的编剧谢里尔·霍顿和尼克·法拉奇在编写他们试播集的剧本时,就采用了罗斯莫的比喻作为查利偶然想到这个方程和向他哥哥解释这个思想的途径。

罗斯莫早期在加拿大从事系列犯罪案件侦查时,已有一些成绩,但真正使他在整个北美地区的执法机构中家喻户晓的是发生在路易斯安那州拉斐特 (Lafayette) 的那桩城南强奸案。^②

十多年来,一名像强盗一样用头巾包住脸的不明身份袭击者一直在跟踪小镇上的妇女,并且袭击她们。1998年,被数以千计的

^① Rigel 是猎户座中的 β 星,一颗蓝白色的超巨星,中文名为参宿七。它是猎户座中最亮的星,全天第七亮星。其质量是太阳的 17 倍,光度约为太阳的 25000 倍。Rigel 本意为巨人的左腿。——译者

^② 1985 年到 1995 年间发生在美国路易斯安那州小城拉斐特南部地区的一起系列强奸案件,十多年一直未破案,后来在罗斯莫的帮助下案件才告破。罗斯莫以及拉斐特警察局都因此案而闻名全美。——译者

举报信息以及相应数目的嫌疑人困扰着的当地警察局邀请罗斯莫来帮忙。罗斯莫使用 Rigel 程序分析犯罪地点数据,并制成了一幅酷似《数字追凶》中查利所示的地图,上面用着色的地带来标明热区以及它的愈加热的内圈。这幅地图使得警察能够把搜索范围缩小到半平方英里^①,约十多名嫌疑犯中间。使用和《数字追凶》中相同的手法,秘密侦查人员遍查热区,以得到区域中适合年龄段中所有男性的 DNA 样本。

热区中的每一名嫌疑人都被有关 DNA 的证据刷清了嫌疑,侦查遭到了挫折。但是随即他们运气来了。侦查负责人麦克(Mac,全名 McCullan Gallien)收到了一条匿名举报的信息,检举一名未必可能的犯罪嫌疑人——邻近机构的代理司法长官。由于这只是已获得的堆积如山的举报信息中的又一条,麦克倾向于仅把它归档了事,但是一时兴起的他决定查一下这个代理司法长官的地址。结果甚至连靠近热区都谈不上。不过他仍然感到有什么在牵引着他,他又深入挖掘了一下。这次他中了“头奖”。这个代理司法长官以前居住在另一个地方——就在热区中!从一个香烟头收集来的 DNA 证据与从犯罪现场采集来的相符合。此人被逮捕了,罗斯莫则迅速成为犯罪侦查圈内的著名人物。

有趣的是,当霍顿和法拉奇基于真实案例编写《数字追凶》试播集的剧本时,他们不可抗拒地在结尾处把同样戏剧性的曲折编写了进去。当查利第一次使用他的公式时在热区的嫌疑人中也没有发现相符的 DNA,就像罗斯莫的公式在拉斐特发生的那样。查利坚定地相信他的数学分析,所以当丹告诉他搜索的结果为零时,一开始他拒绝接受这一结果。“你们一定把他漏掉了。”他说。

① 1 英里相当于 1 609.344 米。——译者

遭到挫折心情不佳，查理和丹聚在他们父亲艾伦的屋子里交谈，艾伦说，“我知道不是数学的问题，查利。一定是其他问题。”这一提醒促使丹认识到寻找杀手的住所可能是个错误的目标。“如果试图在我住的地方找到我，你可能会失败，因为我几乎从不住在那里，”他提示说。“我一直在工作。”查利抓住这条提示设计了一条不同的进攻路线，通过修改他的计算来寻找两个热区，一个可能包含这个杀手的住所，而另一个是他的工作场所。这次查利的数学管用了。丹成功地确定了杀手的身份，并恰好在这个罪犯正要杀害另一个被害人之前抓住了他。

现在，罗斯莫的环境犯罪学研究公司提供专利产品计算机软件包 Rigel，同时包括教他们如何使用它来有效地破案。罗斯莫自己则奔波于全世界，去亚洲、非洲、欧洲以及中东，协助刑事调查，给警察和犯罪学家上课。要熟练掌握这个程序以适用于一个特定罪犯的行为癖性，需要受罗斯莫或他的一名助手培训两年。

Rigel并不是每次都赢得很大的胜利。例如，罗斯莫也被请来参加侦破臭名昭著的环形路狙击手案^①，这是一起发生在2002年10月的3个星期里的案件，共有10人被杀，3人重伤，2名连环杀手在华盛顿特区和周边地区作下此案。罗斯莫得出的结论是狙击手的基地是在华盛顿北郊的某个地区，但后来证明这两个杀手并不居住在这一地区，他们迁移得太频繁了，因而用地理分析技术无法确定地点。

Rigel并不总是管用的事实，对于任何了解把数学应用于杂乱无章的人类现实世界时会发生些什么的人来说并不突然。许多人

^① 环形路狙击手案是一起轰动世界的大案，从10月2日发生第一起案件到24日两名凶犯被捕历时3个星期。环形路是指环绕大华盛顿地区的州际高速公路I-495，由于两名凶犯都是手持狙击步枪在I-495的周围频繁地滥杀无辜，故称此名。——译者

