

数据偏误、行为选择与政策助推 ——行为科学视角下城市信息治理的陷阱和机遇

Data Bias, Behavioral Choices, and Policy Nudge:
Pitfalls and Opportunities in Urban Information Governance from a Behavioral Science Perspective

许立言
Xu Liyan

摘要：城市信息学的长足进步为更加精细化、智能化的城市治理提供了新的机遇。通过社会感知途径获取的大量新数据包含了丰富的城市态势和市民行为信息，具有数据挖掘和知识发现上的可观潜力，但其有偏性是值得高度重视的问题。本文以社会感知数据偏误为研究对象，首先枚举了城市治理中的各类常见新数据及其应用场景，并扼要讨论了其中选择偏误、信息偏误和反事实偏误等三种基本偏误的构成和主要性质。继而，本文从行为科学角度，基于“暴露—认知—行动”框架分析了导致数据偏误问题的行为根源，并提出了一种基于自然实验的因果推断以实现各类偏误纯净效应分解的方法。最后，本文以城市信息化治理中常见的网格员巡查记录、执法监控视频等典型数据资源所反映的违章停车、违章游商、噪声扰民等问题为例，展示了在准确认识数据偏误的基础上，对其进行校正、归因和“助推”式利用，以服务于城市态势感知、问题发现与溯源、针对性政策与法规制定等城市治理目标的应用途径。

Abstract: Recent progress of urban informatics provides new opportunities for more refined and intelligent urban governance. The large amount of new data acquired through Social Sensing contains rich information on urban dynamics and citizens' behaviors, and implies potential for data mining and knowledge discovery. However, observational bias in Social Sensing data is worthy of deep concerns. This paper, in investigating the data bias problem, firstly enumerates various kinds of common new data and their application scenarios in urban governance, and briefly discusses the main properties of the three basic bias types, namely selection bias, information bias, and counterfactual bias. Next, this paper analyzes the behavioral roots of the biases from the perspective of behavioral science with an "exposure-cognition-action" framework, and then proposes a methodological framework which utilizes natural experiment design-based causal inference to realize the decomposition of the pure effects of the three bias types. Finally, this paper takes the typical data resources such as urban management inspection records and law enforcement surveillance videos as examples, and demonstrates ways of correction, attribution, and "nudge"-style utilization of data bias with the proposed methodological framework which helps improve people's well-being.

关键词：城市信息治理；选择偏误；信息偏误；反事实偏误；行为科学；自然实验

Keywords: Urban Information Governance; Selection Bias; Information Bias; Counterfactual Bias; Behavioral Science; Natural Experiment

作者：许立言，博士，北京大学建筑与景观设计学院，北京大学建筑与景观设计学院，研究员，博士生导师；时空智能与地理设计实验室主任。
xuliyan@pku.edu.cn

引言

党的十九届四中全会作出了推进国家治理体系和治理能力现代化的重要决议，其中城市治理（广义上包括国土空间规划、建设、管理、体检评估等环节的体系和能力的现代化）是这一顶层设计的关键组成部分之一^[1-2]。在城市治理新范式的构建中，信息作为一种新的驱动力是近年来一个引人瞩目的方向^[3]。随着信息技术发展，特别是互联网的发展而涌现的各类新数据资源以其样本覆盖广泛^[4]及天然蕴含的“自下而上”特性^[5]，恰与通过监测、调查等传统“自上而下”渠道所获取的数据互补，成为观测城市运行态势、诊断城市问题、评估工作绩效的良好工具；其时空信息显式呈现^[6]的特征亦直接回应了城市治理的空间面向^[7]。依托于新数据资源、以诸如“城市大脑”等称谓呈现的数字化平台已广泛应用于城市规划、建设、日常管理、应急响应、体检评估等诸多环节^[8]，相关基础和应用研究也雨后春笋般开展起来。这些进展切实提高了城市治理的精细化水平，有助于提升人民的幸福感和获得感^[9]，从而契合了新时期我国城市高质量发展的整体目标^[10]。

城市数据应用的普及凸显了通过信息治理确保数据可靠性的重要意义。然而与数据应用的繁荣局面形成鲜明对照的是，城市信息治理的理论和方法体系尚未系统建立，无法充分满足城市精细化治理中深层次问题发现、归因和体系化治理的需求^[11]。概括而言，当前城市治理工作中的信息可靠性问题，特别是新数据源的此类问题集中体现于“稀”“薄”“偏”三个方面。

其一，时空稀疏性。城市信息通常具有时空属性，

因此不可避免地具有时空数据典型的、在特定分析尺度和语境中的稀疏性问题，致使表面看起来的“大数据”在特定的应用场合往往并不足够“大”。这不仅局限了城市信息应用的舞台，更因可用性限制导致数据分析的尺度受限，从而产生因经典的地理分析可变量元问题（MAUP: the Modifiable Areal Unit Problem）^[12]而引发的谬误。

其二，语义单薄性。城市信息，特别是通过新的传感和“社会感知”^[5]途径获取的数据通常依赖单一的采集渠道，因此除必要的时空标签之外，往往仅具有极为有限的语义信息^[13]。相反，传统的监测、调查等数据获取方式虽然样本容量有限，但允许广泛语义信息的采集。二者的区别可类比人口普查中的“短表”与“长表”——从信息论角度看，后者虽样本量较小，但包含较为丰富的信息。理论和实践均表明，多种语义信息的关联分析是挖掘数据价值的关键，而新数据的语义单薄性阻碍了这一潜力的发挥。

其三，观测有偏性。传统的监测、调查数据一般基于严格的抽样设计，而新数据，特别是社会感知数据通常不具备这一条件。因而，其对城市态势的反映依赖于特定的数据生成主体；作为对城市“总体”的一个抽样，这些群体样本往往是相当偏颇的^[14]。这种抽样有偏性又进而受到数据应用中的算法“信息茧房”和自我强化等效应，甚至出于种种动机的数据伪造行为的进一步扭曲^[15]，导致对客观事实的严重偏离。

由于上述三个问题的存在，现有新数据对城市态势的反映往往不准确、不全面、不充分。早在“计量革命”时代，人们即已意识到基于不可靠数据源进行的分析难免产生“不可靠结果”^[16]，乃至最终奏响了“大模型的安魂曲”^[17]。殷鉴不远，若对当今的各类城市新数据不加分辨地盲目使用，难免导致对城市问题的感知失真、分析失准、对策失当^[18]。因此，在信息治理层面深入认识新数据的稀疏性、单薄性和有偏性，是新时代城市治理事业中亟待解决的重要问题。

在三个问题中，有偏性尤其重要。这一方面是由于其关乎基本的正误问题，没有妥协余地；且一旦有偏性问题得以解决，处理稀疏性、单薄性问题有着相对成熟的方法，如稀疏数据的扩样或生成式补全^[19-20]、基于多源数据融合的语义增强^[13]等。另一方面，有偏性问题涉及原始数据生成过程中的诸多认知、行为因素，其并不能单纯通过统计技术解决。这就使得理解数据偏误的性质成为解决城市信息治理问题的关键。

本文以城市数据中的偏误为对象，分析常见城市数据中各类偏误的构成及其基本性质，指出新数据中的偏误应从行为科学的视角加以认识。进而，本文基于行为科学中的“暴露—认知—行动”框架，提出了一种基于自然实验的因果推

断以实现各类偏误纯净效应分解的方法，并结合实际案例讨论了在此基础上对数据偏误进行校正、归因和“助推”式利用的方法路径。

1 城市数据及其偏误的性质

1.1 行为科学视角下的新兴城市感知数据

一般而言，人们可通过三种途径感知城市态势，即遥感（remote sensing）、传感（infrastructure sensing）和社会感知（social sensing）^[5]。其中，遥感和传感被认为是相对“传统”的观测途径，其获取的数据往往较为结构化，人们对其偏误等性质的认识也较为充分。“社会感知”则是一种新兴的城市态势观测途径。随着近年来信息和通信技术的快速发展，市民个体能够通过硬件设备随时随地产生带有个体标记和时空语义信息的地理数据^[21-22]，由此在探测并揭示人类社会环境方面扮演了与遥感、传感设备类似的“传感器”角色，所生成的具有时空标记的地理大数据也因此被称为社会感知数据，构成了城市新数据的主体。下文如无特别指出，“城市数据”这一概念均特指源自社会感知渠道的新数据。

常见的社会感知数据包括来源于用户个体的社交媒体、移动运营商、私人 and 公共交通系统数据^[23-24]，以及来源于互联网平台的街景照片和智慧城市平台数据等^[25]。依其记录方式，可以将社会感知数据分为被动、主动记录两类。

被动记录的数据是在被观测对象无意识或不知情的情况下进行的。常见的被动记录社会感知数据包括手机信令数据、交通轨迹数据、视频监控数据以及街景照片数据等。其中，由于对当代社会的高渗透率，移动通信运营商记录的手机位置信息已成为最流行的时空大数据来源之一^[26-27]；随着GPS追踪技术的发展，许多车辆被位置感知设备追踪，产生了大量的单个车辆运动数据，反映了接近实际运动路线的交通轨迹^[28-29]；街景影像给出了较宽时空覆盖范围内的建成环境特征^[30-32]，使人们能够从人类（水平）的视角来检验视觉特征，这是传统的建成环境感知方式，如航空或卫星遥感所不能提供的^[33-34]。

与被动记录相反，主动记录数据是指用户在有意识的情况下自发“制造”的数据。其中，应用最广泛的可能是在线社交媒体数据。此类数据具有数量大、空间覆盖范围广、生成速度近乎实时的特点，且通常可以自由访问，因而构成了城市研究中一种特别有价值的数据来源^[35-36]。此外，对城市研究而言，随着近年来智慧城市平台建设的陆续到位而产生的另一类“新政务数据”^[37-38]同样值得关注。此类平台的城市管理模块往往同时记录了城管执法人员、网格员和社区志愿者报告的各类城市事件^[37]，并收集了来自城市问题投诉热线、微信公众平台等自下而上信息采集渠

道的城市事件^[39]，如中国的“12345”热线和美国的“311”热线^[40]等。

主被动两种视角下的数据相互补充，构成了对大时空范围内高频城市事件的重要观测途径^[41-42]。在应用上，社会感知数据蕴含了丰富的城市语义和空间交互信息，对其中有效信息的深度挖掘极大地促进了人们对人口分布和动态、土地利用、各类出行行为，以及从城市、国家到全球层面的空间社会网络性质的理解^[24]，为认知人类社会提供了一个前所未有的机会。这方面的例子包括但不限于人类移动^[43]、人口分布动态^[27]、空间社会网络^[44]和城市治理^[39]等。

1.2 城市数据中的基本偏误类型及其性质

虽然社会感知数据具有种种优势，但其获取方式的固有限制可能导致观测结果的潜在偏误问题。偏误又称为系统误差 (systematic error) 或偏倚、偏差^[45-48]，是指由于观测的设计或执行缺陷而产生的系统性偏向，导致观测结果向某一方向偏离真值^[49]，从而破坏了相关研究的内部和外部有效性^[50]。与随机误差 (random error) 不同，随着数据规模的增加，系统误差并不会降低^[51]，因此大数据之“大”对消除数据偏误并无帮助，其问题必须从数据获取的源头解决。

根据上述定义，如果以“客观现实”作为概念性的“真值”，则社会感知数据中的偏误可分为两种基本类型，即选择偏误 (selection bias) 和信息偏误 (information bias，又称为测量偏误 [measurement bias])^[52,54]。前者是指所观测的样本并不代表目标观测对象的全貌^[53,55-56]，后者则指数据收集或测量过程中产生的错误^[52,54]。进一步，倘若将上述定义中的“真值”概念延伸为一种“反事实” (counter-facts，后详)，则可进一步定义第三类偏误，即反事实偏误。可见，以上三种偏误的形成机制完全不同，故不同社会感知数据的偏误构成也各异。

(1) 选择偏误

选择偏误的本质即抽样偏差，被动与主动记录数据中选择偏误的构成具有不同的特点。对于前者而言，选择偏误由特定信息采集设备的渗透率、使用率等因素决定，这些因素即反映了相应观测渠道对全体人群的抽样方法。问题在于，在多数城市研究情形中，其抽样方法都未必符合预定研究对象的总体特征。例如：手机信令数据的生成主体是“拥有并日常使用手机的人”，这一群体通常仅涉及部分老人和更少的儿童。由于其样本对全部人群这一总体来说是一个有偏抽样，因而并不能直接适用于以全部人群为对象的城市研究^[26]。又如：出租车轨迹和智能公交卡刷卡记录都是方便易得的城市交通大数据，但二者的生成主体即出行人群明显不同，且均不是“城市出行行为”这一总体的无偏抽样，其反映的

交通态势也因此具有相当差异^[57-60]，均仅为局部的交通态势刻画。

与被动记录的社会感知数据相比，主动记录社会感知数据同样依赖用户所持设备上的传感器以获取部分信息。区别在于，人作为一种“传感器”，一般无法被“部署”。因此，主动记录社会感知数据仍然存在因用户群体差异导致的、相对于全体人群的有偏抽样，但与被动记录数据的情况不同，这里的抽样逻辑是内生的。以社交媒体数据为例，大量研究表明此类数据集并非人群的一个随机样本^[61]，且不同社交媒体在用户年龄、性别、种族、社会经济状况等属性的分布上具有很大差异^[62,64]。内生的抽样机制，意味着此时的选择偏误包含“自我选择” (self-selection) 效应，这也是主被动记录社会感知数据选择偏误的主要区别之处。

(2) 信息偏误

信息偏误是由不准确的测量导致的。被动记录社会感知数据通常由用户所持设备上的传感器自动记录，因此其潜在的信息偏误完全取决于特定传感器的性能。以光学传感器为例，可能存在由光学畸变、色差、元件阈值等原因造成的测量偏误。一般而言，这种偏误的物理机制是清晰的，通常可以预知特定传感器的激活曲线，即测量值与真值之间的转换函数，并通过算法实现纠偏^[65]。由于被动记录社会感知数据的信息偏误产生机制与传统的传感途径没有实质性区别，且常用定位、摄像等传感器多在技术上相当成熟，可以认为被动记录的社会感知数据并不存在难以解决的信息偏误问题。

主动记录社会感知数据的信息偏误形成机制则截然不同。信息偏误固然同样由传感器的特征决定，但此时的“传感器”是人本身。相应地，其“激活曲线”由用户在“主动上传”这一环节中展现出的多样化感知和行为模式决定，其机制远比物理传感器的测量误差复杂。例如：有研究比对了蓝牙 (被动记录) 和签到 (主动记录) 两个渠道下同一城市事件的发生时间，发现后者通常较前者有所延迟。这是因为人们通常不会在第一时间完成“签到”的动作，而被动记录则不存在这一问题^[23]。出于类似原因，社交媒体上带有地理标签的发帖记录可能同样无法反映用户的确切实时位置，从而导致信息偏误^[66,68]。又如：对比手机信令数据和社交媒体数据，前者会自动上传，用户在此缺乏有意义的自主选择权；而后者是用户选择何时、何地、在哪个社交媒体上传或不上传何种内容。本质上，由“上传”这一动作导致的偏误同样源于自我选择，只是此时的自我选择效应并非作用于抽样阶段，而是作用于测量阶段，因而产生了信息偏误。

(3) 反事实偏误

除选择偏误和信息偏误外，社会感知数据中还普遍存在

另一种特殊形式的偏误,其来源是反馈效应,有时被称为“观察者效应”或“霍桑效应”^[69],即当人们意识到自己的行为(发帖、签到或某种线下行为)正在被人观察时,可能作出与“常态”不同的行为选择。这种偏误在主被动记录的社会感知数据中都可能存在。前者仍以社交媒体数据为例,大量研究发现社交媒体舆论在公共事务上所反映的情绪比一般人群更加极端^[70],且对个人生活有关的话题更加坦率^[71]。这里的比较基线是用户在较为中立条件下的言论,而二者的差别当然源自用户根据其言论的预期受众反应而产生的行为扭曲。后者的常见例子则是当人们发现附近存在视频监控设备时,常会收敛各种违规行为。也就是说,“观测”这一行动产生了某种威慑作用,从而改变了人们的行动。

反馈效应所导致的偏误的特殊之处在于,此时的观测本身可能准确地反映了“客观真实”,但这种“客观真实”偏离了在无观测条件下的“自然状态”。由于观测确实已经发生了,因而此时的“无观测条件”是一个反事实假设(counter-factual hypothesis),故不妨称这种偏误为“反事实偏误”。反事实偏误一定程度上并不符合以“客观真实”为比较对象的经典偏误概念,但无论是在理论上还是在城市治理实践中,这种偏误同样是不可忽视的。例如:如果我们希望认识某种违规行为的时空规律,并基于这种认识指导城市建设和管理,则必须意识到现实条件下的观测可能受到已存在的执法观测系统(如视频监控)的影响,导致其结论并不能准确反映被观测对象在缺乏此类系统时、或系统的形态有差异时的行为模式。换句话说,此时的“客观真实”本身是扭曲的、具有误导性的,而反事实状态才是人们真正期望获得的有效知识,这就凸显了排除反事实偏误的重要性。

2 理解城市数据偏误的行为根源:基于“暴露—认知—行动”框架

虽然各类城市数据中偏误的类型和表现形式相当多样,但不难看出,上述各类偏误均以某种方式与人类行为相关:选择偏误中的感知偏差、信息偏误中的行动偏好,以及反事实偏误中的观察反馈。与此同时,可以注意到数据生成过程实际上包含了选择、观察、行动等多个环节,而上述各类效应发生在不同环节。这启发我们从行为科学视角,建立理解城市数据偏误的一般概念框架。

一般而言,我们可以将社会感知数据的生成过程视为行为科学基本范式下的一个“暴露—认知—行动”过程^[72]。如图1所示,首先,被观测对象的客观事实构成了一种“暴露”(图1-II)。其次,暴露作用于观察者的过程是一个感知过程,图1中从II到III的过程A即为感知过程,这里的观察者对被动记录数据来说可能是物理传感器(如手机上的定位装置),对主动记录数据来说则是人。需要注意的是,在人作为“传感器”的条件下,此时的感知结果(图1-III)并没有直接形成数据,而是一个人脑中的状态。接下来,观察者将采取某些行动,从而将其认知结果转化为数据记录,图1中从III到IV的过程B即为行动过程。对于被动记录数据来说,其“行动”即是“存贮”这一操作;而对于主动记录数据来说,这里的行动意味着“上传”“签到”之类动作。无论哪种情况下,被如此记录的信息即为最终得到的数据(图1-IV)。最后,我们考虑观察者的存在所引发的反馈效应,该效应将导致被观测对象即暴露本身偏离无观测条件下应具有的状态,即反事实状态(图1-I),图1中从I到II的过程C即为反馈过程。

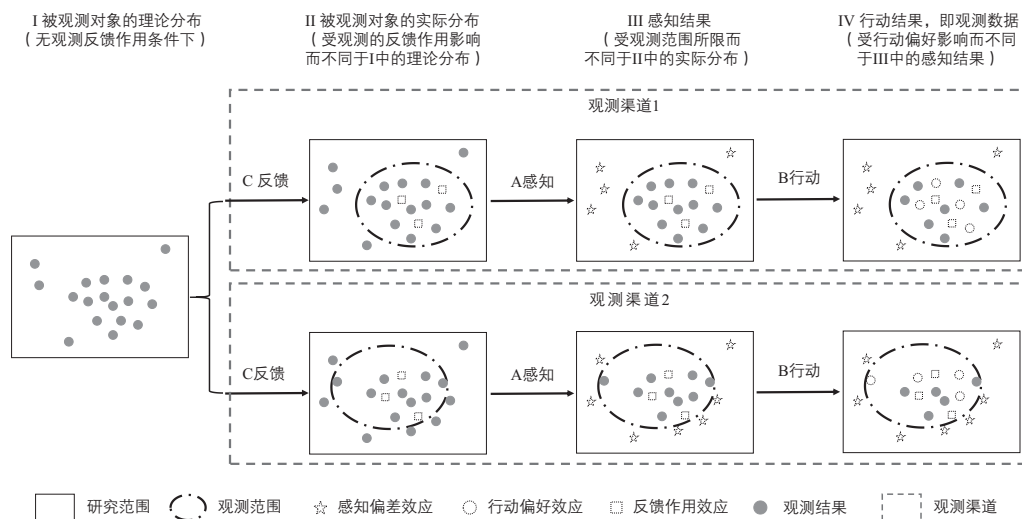


图1 行为科学视角下城市信息生成过程的概念模型及三种偏误效应

在上述图景下，不难厘清前文所分析的选择、信息和反事实三种偏误效应的行为根源。以城市管理中的违章停车事件为例，假定存在视频监控（被动记录）、网格员巡查（主动记录）两种观测渠道，仍如图 1 所示。

(1) 感知偏差 (perception deviation) 效应导致的选择偏误，即因观测能力不足以完整覆盖被观测对象的时空或属性范围而产生的、相对于被观测对象全貌的偏差。在图 1 中，可见由于部分研究范围无法被特定观测渠道覆盖（例如：城管网格员有一定的巡逻路线，视频监控有特定的部署点位，二者通常都不能形成对整个城市时空范围的全覆盖），其覆盖范围之外的事件将不可能体现在图 1-III 的感知结果中，其造成的选择偏误如图 1-III 中的空心星形所示。

(2) 行动偏好 (behavior preference) 效应导致的信息偏误，即人们在根据感知结果采取行动时，因其主观选择的差异而导致的行动效果差异。如前所述，感知结果是一个仅存在于人脑中的状态，通常不可直接观测，而观测者实际观测到的是某种行动的效果，该行动是人们基于其感知结果而采取的。这里，额外的一步行为“过滤”，可能导致观测结果与感知结果有所差异，该差异即为行动偏好效应的体现。例如：巡查人员可能出于某种原因，对其发现的违停事件（感知）采取视而不见的态度而不予记录（行动）。如此，其行动效果与感知结果便有所差异，因而产生了信息偏误，如图 1-IV 中空心的圆形所示。

(3) 反馈效应导致的反事实偏误，即“观测”这一行为本身会对被观测对象产生反作用，使之偏离无观测条件下的面貌。例如：违停监控摄像头或执法巡查的存在对违停行为具有一定威慑力，使得部分本来可能发生的违停行为实际上没有出现。其效果如图 1-II 中空心的方形所示。倘若以此时的观测结果（即图 1-II）为依据预测无监控或巡查覆盖地区的违停情况（即图 1-I），则将低估违停行为的程度。

经过三种效应的累计，最终体现在观测数据中的结果即图 1-IV 中的实心圆形，可见其无论相对被观测对象的“客观真实”（图 1-II 中的实心圆形）还是“理论”状态（图 1-I 中的实心圆形）均有很大偏差。

3 应对城市数据偏误：借助自然试验设计实现各类偏误纯净效应的分解、归因和校正

在城市规划和治理实践中，数据偏误通常是需要避免的，由此产生了偏误校正的要求，该要求是城市数据治理的基本任务。根据上述概念框架，社会感知数据中往往混杂着各类性质完全不同的偏误，这些偏误各有针对性的校正方法。因此，校正数据偏误的先决条件应为对各类型偏误纯净效应的分解，以便对症下药，分别进行校正处理。

一般而言，欲分解出某影响因素的纯净效应，可采用受控实验方法。然而，社会感知的对象是社会现象，通常缺乏进行受控实验的条件。鉴于这一现实，并受上述概念图景启发，本文提出一种基于自然实验设计而实现三种偏误纯净效应分解的方法。

自然实验设计的基本思路是，由于任一观测渠道均有其特定的偏误构成和性质，如果对同一城市事件进行多渠道且各渠道有部分重复的观测，则可借助比较多个观测结果中各类偏误构成的不对称性来实现其纯净效应的分解，不妨称之为“多渠道部分重复观测”方法。

仍以上文违章停车事件为例。被动记录渠道（视频监控）至少存在选择偏误（摄像头仅能覆盖部分城区）和反事实偏误（人们由于摄像头的威慑而收敛了违章行为），同时亦可能存在一定信息偏误（智能算法甚至人工判读均未必能识别出全部违章行为）；主动记录渠道则同时存在选择偏误（巡查路线仅能覆盖部分城区）、信息偏误（即使发现了违章事件，巡查员可能出于种种原因而不上报，从而无法在数据中留下记录），以及反事实偏误（人们鉴于巡查人员的活动而收敛了违章行为）。

基于多渠道部分重复观测的偏误分解原理如图 2 所示。假定待观测对象的总体为 P——注意 P 为一个反事实状态，其为一个时空分布。一般情况下，假定存在对该观测对象的两种感知渠道 A 与 B，只要这两种感知渠道在时空上不完全吻合，二者即构成了对观测对象的多渠道部分重复观测。在这一图景中，可依两种感知渠道的覆盖和记录范围，将对总体 P 的观测结果划分为 SI-SV 5 个样本子集。不难看出，SI-SV 具有不同的偏误构成。例如：样本 SI 被感知渠道 A 所覆盖，但并未被感知渠道 B 覆盖，且仅通过渠道 A 被记录，因此其观测结果中包括了感知渠道 B 的感知偏差和行动偏好（没有反馈效应，因为样本集不在 B 的感知范围内），以及感知渠道 A 的反馈效应（没有感知偏差和行动偏好，因为这里二者具体指的是“看不见”和“视而不见”，显然对于已经被记录的事件，不存在这两种问题）。又如：样本 SV 中包括了同时被两种感知渠道覆盖且同时被两个渠道记录的观测结果，因此其中不包含任何感知偏差和行动偏好，但同时反映了两种感知渠道的反馈效应。以此类推。

我们可进而通过对各样本子集的简单线性组合计算分解各类偏误的纯净效应。具体来说，在上述图景中，感知偏差效应可通过比对各观测渠道的覆盖范围之差而直接给出，即 P-SI、P-SIII 分别给出了 A 和 B 的感知偏差。在控制感知偏差效应之后，SI、SIII 中分别仅包含 A 或 B 感知渠道下的反馈效应，因而通过 SIV-(SI+SIII)、SII-(SI+SIII) 即可分别分解出 A 或 B 渠道下的行动偏好效应。最后，SV-(SI+SIII) 则为消除了

偏误效应的无偏样本。由此即实现了三种偏误效应的分解。

最后，基于所分离出来的净偏误效应，可通过构建统计推断模型实现各类偏误的归因和校正。假定待观测对象 Y (注意其为一个时空分布) 具有特定的时空发生机制。一般而言，表示为广义函数：

$$Y = f(X)$$

其中 X 为解释变量。以该模型为基线，基于图 2 的图景，有如下模型。

模型一： $f_1(X|X \in S_1)$ ——在基线模型的基础上，仅包含来自感知渠道 A 的反馈作用。

模型二： $f_2(X|X \in S_1) - f_1(X|X \in S_1) - f_3(X|X \in S_{III})$ ——在基线模型的基础上，仅包含来自感知渠道 B 的行动偏好。

模型三： $f_3(X|X \in S_{III})$ ——在基线模型的基础上，仅包含来自感知渠道 B 的反馈作用。

模型四： $f_4(X|X \in S_V) - f_1(X|X \in S_1) - f_3(X|X \in S_{III})$ ——在基线模型的基础上，仅包含来自感知渠道 A 的行动偏好。

模型五： $f_5(X|X \in S_1) - f_1(X|X \in S_1) - f_3(X|X \in S_{III})$ ——反事实模型，不含任何行为因素干扰的违停“净”发生机制模型。可见，此即为基线模型。

以上各模型直接给出了各类偏误效应的生成机制。换句话说，其实现了对各类偏误的归因。最后，基于所拟合的模型参数进行反演，即可实现偏误的校正，其意义不再赘述。

上述偏误分解的方法具有潜在的广泛适用性。首先，所描述自然实验的实现条件并不苛刻。事实上，以各种彼此具有一定重叠的渠道对同类城市事件进行同步观测，正是当今城市数据信息获取的主要方式，因此其保障了相当普遍的自然实验机会。其次，上述例子仅包含两种感知渠道，但明显容易推广到两种以上观测渠道的情形。事实上，更多的观测渠道提供了以组合数递增的线性组合机会，原则上甚至可能通过对各种不同组合分析结果的一致性检验来改善估计的可靠性。即使考虑多感知渠道导致的操作复杂性，研究者总是

可以采用两两比对的方式进行实际操作。

原则上，在城市规划和治理实践中，通过对上述各子集中城市事件时空分布模式及其影响因素的分析，即可得知任一种偏误的时空规律和形成机制，并实现相应偏误效应的校正。然而，在应用上述方法时，有两个值得一提的注意事项。首先，图 2 仅给出了一个抽象的一般图景，在具体的问题语境中，各感知渠道及其特定的感知偏差、行动偏好和反馈效应的表现形式是因语境而异的。因此，在自然实验设计中，首先应在概念上明确上述表现形式的社会—物理实质。其次，由于偏误归因是以统计推断的方式给出，该推断函数的具体形式和自变量选择应针对其问题语境，基于可靠的基础理论和先验知识给出。总之，虽然方法在形式上是普适的，但以上两点强调了具体问题语境和相关领域知识的重要性。接下来，我们通过两个案例，从概念上展示将该方法应用于各类城市规划和治理场景的一般途径。

3.1 城市违章游商治理

违章游商是与人民生活息息相关的日常性城市问题。由于其重要性，在当前的数字化城市管理系统中，通常存在执法巡查、网格上报、视频监控、居民投诉等多种感知渠道同时对违章游商现象实施观测，由此满足了基本的“多渠道部分重复观测”自然实验条件。在下文的讨论中，我们为了简化而仅考虑居民投诉和执法巡查两种感知渠道，仍分别对应图 2 中的 A 和 B；同样，所有观测结果仍可划分为图 2 中的 SI-SV 5 个样本子集。

我们首先界定此场景下两种感知渠道的三类行为效应。两种感知渠道的感知偏差和反馈效应均十分直观，即分别为“看不到”和“威慑作用”。至于行动偏好效应，对于居民投诉而言，其意味着居民虽然感知到了游商现象，但选择不去投诉；对于执法巡查而言，其意味着执法人员虽然感知到了游商现象，但选择不采取执法行动。也就是说，这里的行

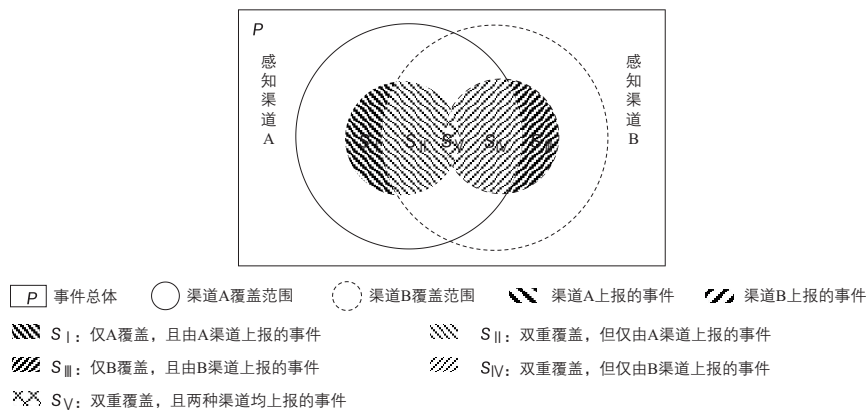


图 2 通过多渠道部分重复观测实现三种偏误效应分解的自然实验设计

动偏好效应指的是“视而不见”。

接下来，我们构建偏误分解的机制模型。根据犯罪地理学理论^[73-75]——虽然违章游商并非“犯罪”，但事件时空发生机制意义上的基本地理学原理是一致的：游商事件的发生应具有一般的、稳定的时空行为规律。因此，以游商事件为因变量，以催生该事件的内外因素为自变量，按前述方法分别对SI-SV 5个样本子集构建推断模型。

囿于篇幅，分析过程不再赘述。结果显示，首先，居民投诉的感知偏差与城市土地利用的容积率负相关（即容积率越高，越容易出现“视而不见”），而执法巡查的感知偏差与人口密度、容积率、道路密度负相关。其次，两种感知渠道的反馈效应均不明显。最后，有趣的是二者的行动偏好：居民的行动偏好与场所的“居民区”属性，其中的儿童数量、“休息时段”的时间属性，以及“郊区”的城市区位特征强烈负相关。这意味着在儿童较多的居民区，居民会倾向于积极投诉违章游商事件，而不是选择对其视而不见。作为对比，巡查人员的行动偏好则与场所的“政府和商业区”属性、其中的工作人口数量、“工作时段”的时间属性，以及“市中心”的城市区位特征强烈负相关。这意味着他们倾向于对政府单位、核心商业区周边的违章摊贩行为严格执法，在其他城市市场则展现了执法的灵活性^[37]。

总的来说，以上发现暗示了两类感知主体不同的行为逻辑。对于居民来说，有小孩的家庭会更加关注摊贩活动导致的烟尘、噪声、食品安全、卫生等扰民问题，特别是在休息时间，因为这个时候家长和小孩都倾向于在家，对以上问题的反应也最为强烈。至于“郊区”这个区位特征，可能是由于研究区的城郊存在大量新开发居住区，且这些地区的城管力量相对薄弱所致。对于城管执法人员来说，其对工作时段中政府单位、核心商业区周边违章摊贩问题的格外重视自然地反映了城市管理者对城市形象维护的诉求。在上述情况之外，两类人群都倾向于对违章摊贩视而不见。这些发现揭示了居民的邻避心理和一线执法人员的选择性执法倾向对违章摊贩问题数据图景的扭曲。对于中立的观察者或关注全局的城市管理人员来说，认识到这些扭曲并进行校正，有利于在规划方案和治理政策形成阶段推演不同执法资源配置场景下的违章摊贩问题图景，作为方案和政策制定的重要参考；更进一步，亦可针对上述行为机制采取补救措施，从而在城市治理的精细化提升中有的放矢地指导执法资源配置，以收获全局意义上满意的治理绩效。

3.2 城市环境噪声治理

城市环境噪声是城市治理中最常见的日常性问题之一^[76]。对此通常存在基于物理传感（噪声等级监测）和社会感知

（居民投诉等）等不同原理的感知渠道。我们即以这两种渠道为例进行讨论，二者的感知偏差和反馈效应仍分别为“听不到”和“威慑作用”；对于行动偏好效应，由于噪声监测依赖于物理传感器，原则上其除了受到传感器自身激活曲线特征的影响之外，不存在人类行为意义上的偏好效应，居民投诉的行动偏好则仍为“选择不投诉”这一行动。观测结果的样本子集拆分和推断模型构建方法同上文，不再赘述。

分析结果显示^[40]，研究区域的昼间环境噪声水平显著高于夜间，但夜间的噪声投诉远多于白天。可见，若以环境噪声水平为待观测的客观事实，两种感知渠道中至少一种必定存在偏误。在这里，对于噪声监测器，忽略由外生的部署策略造成的感知偏差，且考虑到这种监测器非常隐蔽，不太可能存在“威慑”效果，则可认为噪声监测的结果在其感知范围内是无偏的。相比之下，居民投诉视角下的环境噪声图景则大不相同。其感知偏差主要受到“居住区”这一城市市场语义的影响，行动偏好则同时受到居住区、老人和儿童以及夜间因素的影响，但未观察到明显的反馈作用。这就是说，居民大体上仅对居住区的环境噪声敏感，并更倾向于在夜间而非白天时段、由于噪声对老人和儿童的干扰而拨打投诉电话，但这种投诉几乎不会自动产生对噪声源的威慑作用。就观测结果而言，上述行为因素在相当程度上扭曲了噪声强度和投诉数量的简单正相关关系，导致了社会感知结果相对“客观事实”的偏误。

以上发现启发了对城市环境噪声管理的深入思考——到底是“客观”的噪声水平更重要，还是人们对噪声的主观感知更重要？对比《声环境质量标准》（GB 3096-2008）中对各类声环境功能区噪声极限水平的规定和12345热线中的投诉分布，发现噪声水平达标的区域未必没有投诉，反之，超标区域也不一定就会有投诉，且噪声投诉最密集的地方并不是超标最严重的区域。这些事实说明，按照客观分贝数制定的声环境质量标准与人们对噪声的容忍度之间存在相当的偏差——即使其标准完全得以落实，依然无法避免群众的噪声投诉。因此，从城市精细化治理的角度，在未来的声环境标准修订中引入社会感知视角，以人们的感受为规范目标，可望更好地回应市民的心理需求。

当然，这一思路与现行标准是互补而非替代关系。这里的逻辑与违章摊贩之类的城市问题有微妙的不同。违章摊贩可以认为是城市治理中的灰色地带，它在一定程度上以其非正规性构成了对正规商业部门的补充，因而有其存在的价值。因此，城管对违章摊贩的选择性执法反映了二者的某种默契——城市精细化治理的目标并不是完全消除这一现象，而是在维持其必要功能的同时，尽量消除其负外部性。此时，通过居民投诉渠道观察到的违章摊贩现象即构成了对那些具

有显著负外部性摊贩行为的良好指示。相反，对城市声环境治理而言，物理与心理噪声指标是同样重要的，因为“投诉”意味着人们对噪声忍无可忍，即使人们主观上能够忍受，噪声在客观上也有害健康。这种危害通常是慢性的，并不能因为人们能一时容忍就忽略^[77]。换言之，对噪声而言，并非“被投诉了的问题才是真问题”，不同感知渠道的信息都是重要的。对比上述两例，可以看出深入理解城市数据偏误的性质及其行为根源的现实价值。

4 “利用”城市数据偏误：“助推”思想及其城市规划和治理应用

最后，对于城市数据偏误，除归因和校正之外，人们甚至可以尝试在某些特定情况下“利用”偏误，使其成为解决城市问题的直接抓手。这种潜力是通过对偏误生成机制中反馈效应的主动强化实现的。反馈效应的特殊之处在于，它具有影响甚至塑造人类行为的潜力，由此暗示了城市治理主体通过主动利用反馈效应从而达成特定治理目标的可能性。这种可能性的理论基础即为2017年诺贝尔经济学奖得主理查德·泰勒的“助推”（Nudge）理论^[78]。助推，在我国语境中可理解为“四两拨千斤”式的政策设计思维。该理论指出，由于人的直觉性和理性两种思维方式^[79]的存在，人的行动往往以某种“可预测的非理性”^[80]的方式表现出特有的“惯性”，或普遍的行为倾向——如“观测”这一行动对被观测者行为的影响即为一例。倘若可以在城市治理中利用此类行为倾向，因势利导地设计政策，则可望低成本地实现治理目标。在这里，特别值得注意的是，作为助推基础的行为机制，正可以通过导致数据偏误的行动偏好、反馈作用等效应体现出来。从这个意义上说，数据偏误并不总是坏事。通过对各类偏误纯净效应的分解，厘清导致城市问题的人类行为机制，提供了一个化偏误根源为助推机遇，巧妙达成高效治理绩效的机会。

助推思维有着多种具体实现途径，可视情况应用于城市规划和治理的各个环节。仍以上述违章停车管理场景为例。首先，在日常行政和执法实践环节，可采用行动效果提示、创造同侪压力等助推方法。例如：通过在所发现的违停高发地广泛张贴“此区域有视频监控”的提示（不一定真的部署监控设备），甚至设置醒目的假监控探头，或提供公开违停举报平台并警告将严格执法等方式，可强化执法监控的吓阻作用，避免违停事件。这些方法在国际实践中展现了屡试不爽的效果^[78]。

其次，在城市规划和治理政策制定环节，可采用默认规则设计、选择体系优化等助推方法。例如：违停行为的发生有时是由于人们不知道该区域禁止停车，或不清楚附近哪里可以停车，对此可通过改善停车引导系统等方式进行正面引

导；同时，大量违停行为的持续时间并不长，且在某些时间地点的违停并不真正构成危害。对此，可考虑在不同地点、不同时段设置短至数分钟、长至一两个小时的允许停车时段，且区分临停（人不离车）和泊车（人离车）行为，以此丰富停车选择体系，在维持基本街面秩序的前提下提供适当的弹性，从而提升城市治理的精细化水平。这些措施均基于对人们行动偏好的主动适应，因此同样依赖于数据偏误分析的行动偏好认识。

更深层次的助推应用则可深入法律法规甚至标准规范制定层面。例如：违停事件本质上是邻避的——当违停车辆并不妨碍自身时，多数人并不介意其存在。反过来说，哪些违停事件是真正有害的？与前面的违章摊贩案例类似，该问题在很大程度上可以通过居民投诉得到直接回答。因此，在上述案例中，倘若能够进一步引入12345热线、城管微信平台等反映自下而上声音的观测渠道，城市管理部门则可通过一段时间的数据积累，对违停事件的实际影响进行时空差异化的评估，并将其结果运用于城市管理条例甚至相关标准规范的修订中。本质上，该方法反映了一种响应式的、自组织的城市治理思维——城市生活自有其逻辑，原则上，法律法规、标准规范等治理工具应适应之，而非尝试塑造之；而对数据偏误性质的挖掘，则为城市管理者提供了深入认识这种规律的窗口。

最后，值得指出的是，助推策略不仅是低成本的，而且是公正的，因为常用的助推策略的效用均为因势利导的结果，并不损害人的隐私和自由选择权，因而体现了“数据正义”——这也是新时代城市治理精细化所追求的深层逻辑之一。

5 结语

本文力图从城市信息学本体出发认识城市数据偏误。通过对常见城市数据的分析，本文指出各类数据中普遍存在选择偏误、信息偏误和反事实偏误这三种性质截然不同的偏误类型，它们各有其独特的行为根源。本文提出，基于行为科学中的“暴露—认知—行动”框架，通过引入基于多渠道部分重复观测的自然实验方法，可实现对各类数据中三类偏误纯净效应的分解，由此实现对偏误根源的归因推断和偏误校正，从而为城市规划和治理提供对症下药的方法工具；更进一步，通过深刻认识数据偏误中所反映的市民行为模式，可基于“助推”思维进行政策设计，从而高效、低成本、公正地改善城市治理的精细化水平。

作为一个概念框架，本文提出的数据偏误处理方法路径当然仍有局限性。如在技术层面，偏误分解的统计推断模型可能拟合优度不高。这固然无碍于定性层面的偏误归因，

但若以定量的偏误校正（而非归因）为主要目的，则尚待开发更优的预测方法，诸如基于人工智能的方法等。总之，对数据偏误行为根源的认识不仅有助于揭示城市问题本质，更具有直接服务于人民福祉的意涵，可望助力于新时代的城市治理体系建设。笔者愿与学界同侪共同探索这一具有前景的方向。UPI

致谢：作者感谢东湖高新区国家智能社会治理实验综合基地项目对本研究的支持。

参考文献

- [1] 姜晓萍, 董家鸣. 市域社会治理现代化的理论认知与实现途径[J]. 社会政策研究, 2019, 4(4): 24-31.
- [2] 伍江, 程进, 张京祥. “城市精细化治理与高质量发展”学术笔谈[J]. 城市规划学刊, 2020, 67(2): 1-11.
- [3] 孙轩, 孙涛. 大数据计算环境下的城市动态治理: 概念内涵与应用框架[J]. 电子政务, 2020, 17(1): 20-28.
- [4] VIKTOR M S, KENNETH C. Big data: a revolution that will transform how we live, work, and think[M]. Reprint Edition. New York: Harper Business, 2014.
- [5] LIU Y, LIU X, GAO S, et al. Social sensing: a new approach to understanding our socioeconomic environments[J]. Annals of the Association of American Geographers, 2015, 105(3): 512-530.
- [6] 孙轩, 孙涛. 大数据时代公共管理应用决策4M思维: 理论思考与实践探索[J]. 上海行政学院学报, 2019, 20(1): 56-65.
- [7] 张勇, 何艳玲. 论城市社区治理的空间面向[J]. 新视野, 2017, 40(4): 84-91.
- [8] 马亮. 数字政府建设: 文献述评与研究展望[J]. 党政研究, 2021, 36(3): 99-111.
- [9] 叶裕民, 王晨跃. 城市治理研究范式转移与一般分析框架创新[J]. 城市规划, 2022, 46(2): 42-52, 99.
- [10] 王佃利, 王玉龙. 情境变迁、理论沿革与价值承继: 城市管理的发展逻辑[J]. 陕西师范大学学报(哲学社会科学版), 2020, 49(4): 156-164.
- [11] 锁利铭. “城市大脑”建设何以更加有序[J]. 国家治理, 2021, 8(17): 11-15.
- [12] OPENSHAW S. The modifiable areal unit problem[M]. Norwich: Geo Books, 1984.
- [13] WU L, YANG L, HUANG Z, et al. Inferring demographics from human trajectories and geographical context[J]. Computers, environment and urban systems, 2019, 77: 101368.
- [14] YUAN Y, LU Y, CHOW T E, et al. The missing parts from social media-enabled smart cities: who, where, when, and what? [J]. Annals of the American Association of Geographers, 2020, 110(2): 462-475.
- [15] 邓凯, 吴灏文. 数字化助推: 一种智慧城市的公共政策视角[J]. 决策探索(下), 2020, 36(8): 11-15.
- [16] BOYNE G A. The determinants of variations in local service contracting: garbage in, garbage out? [J]. Urban affairs review, 1998, 34(1): 150-163.
- [17] LEE D B. Requiem for large-scale models[J]. Journal of the American Institute of Planners, 1973, 39(3): 163-178.
- [18] 彭兰. 假象、算法囚徒与权利让渡: 数据与算法时代的新风险[J]. 西北师大学报(社会科学版), 2018, 55(5): 20-29.
- [19] MOECKEL R, SPIEKERMANN K, WEGENER M. Creating a synthetic population[C] // Proceedings of the 8th international conference on computers in urban planning and urban management (CUPUM). Sendai, Japan, 2003: 1-18.
- [20] 钮心毅, 林诗佳, 秦思娴, 等. 移动定位大数据支持城市人口规模监测的技术途径[J]. 当代建筑, 2020, 1(12): 39-43.
- [21] GOODCHILD M F. Citizens as sensors: the world of volunteered geography[J]. GeoJournal, 2007, 69(4): 211-221.
- [22] MILLER H J, GOODCHILD M F. Data-driven geography[J]. GeoJournal, 2015, 80(4): 449-461.
- [23] KWAN M P. Algorithmic geographies: big data, algorithmic uncertainty, and the production of geographic knowledge[J]. Annals of the American Association of Geographers, 2016, 106(2): 274-282.
- [24] LIAO C, BROWN D, FEI D, et al. Big data-enabled social sensing in spatial analysis: potentials and pitfalls[J]. Transactions in GIS, 2018, 22(6): 1351-1371.
- [25] PANDHARIPANDE A. Social sensing in IoT applications: a review[J]. IEEE sensors journal, 2021, 21(11): 12523-12530.
- [26] BLONDEL V D, DECUYPER A, KRINGS G. A survey of results on mobile phone datasets analysis[J]. EPJ data science, 2015, 4(1): 10.
- [27] DEVILLE P, LINARD C, MARTIN S, et al. Dynamic population mapping using mobile phone data[J]. Proceedings of the National Academy of Sciences, 2014, 111(45): 15888-15893.
- [28] GONG L, LIU X, WU L, et al. Inferring trip purposes and uncovering travel patterns from taxi trajectory data[J]. Cartography and geographic information science, 2016, 43(2): 103-114.
- [29] SUN L, AXHAUSEN K W, LEE D H, et al. Understanding metropolitan patterns of daily encounters[J]. Proceedings of the National Academy of Sciences, 2013, 110(34): 13774-13779.
- [30] ZITOUNI M S, SLUZEK A, BHASKAR H. Visual analysis of socio-cognitive crowd behaviors for surveillance: a survey and categorization of trends and methods[J]. Engineering applications of artificial intelligence, 2019, 82: 294-312.
- [31] AL-SMADI M, ABDULRAHIM K, SALAM R A. Traffic surveillance: a review of vision based vehicle detection, recognition and tracking[J]. International journal of applied engineering research, 2016, 11(1): 713-726.
- [32] ZHANG Z, ZHUO K, WEI W, et al. Emotional responses to the visual patterns of urban streets: evidence from physiological and subjective indicators[J]. International journal of environmental research and public health, 2021, 18(18): 9677.
- [33] ZHANG F, WU L, ZHU D, et al. Social sensing from street-level imagery: a case study in learning spatio-temporal urban mobility patterns[J]. ISPRS journal of photogrammetry and remote sensing, 2019, 153: 48-58.
- [34] BILJECKI F, ITO K. Street view imagery in urban analytics and GIS: a review[J]. Landscape and urban planning, 2021, 215: 104217.
- [35] GOODCHILD M F. The quality of big (geo)data[J]. Dialogues in human geography, 2013, 3(3): 280-284.
- [36] DIEZ ROUX A V, MAIR C. Neighborhoods and health[J]. Annals of the New York Academy of Sciences, 2010, 1186(1): 125-145.
- [37] LI C, HUANG Y, SHEN Y, et al. Spatiotemporal patterns and mechanisms of street vending from the social sensing perspective: a comparison between law-enforcement reported and residents complain events[J]. Cities, 2022, 124: 103597.
- [38] WU L, LI J, HU X, et al. Gleaning Wisdom from the past: early detection of emerging rumors in social media[C] // Proceedings of the 2017 SIAM International Conference on Data Mining (SDM). Philadelphia: Society for Industrial and Applied Mathematics, 2017: 99-107.
- [39] PENG X, LI Y, SI Y, et al. A social sensing approach for everyday urban problem-handling with the 12345-complaint hotline data[J]. Computers, environment and urban systems, 2022, 94: 101790.
- [40] SI Y, XU L, PENG X, et al. Comparative diagnosis of the urban noise problem from infrastructural and social sensing approaches: a case study in Ningbo, China[J]. International journal of environmental research and public health, 2022, 19(5): 2809.
- [41] HARTMANN S, MAINKA A, STOCK W G. Citizen relationship management in local governments: the potential of 311 for public service

- delivery[M] // PAULIN A A, ANTHOPOULOS L G, REDDICK C G. Beyond bureaucracy: towards sustainable governance informatisation. Cham: Springer International Publishing, 2017: 337-353.
- [42] SOGA K, SCHOOLING J. Infrastructure sensing[J]. *Interface focus*, 2016, 6(4): 20160023.
- [43] GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns[J]. *Nature*, 2008, 453(7196): 779-782.
- [44] ZHONG C, ARISONA S M, HUANG X, et al. Detecting the dynamics of urban structure through spatial network analysis[J]. *International journal of geographical information science*, 2014, 28(11): 2178-2199.
- [45] SHAO J. *Mathematical statistics*[M]. New York, NY: Springer New York, 2003.
- [46] FISHER R A. *Statistical methods for research workers*[M] // KOTZ S, JOHNSON N L. *Breakthroughs in statistics*. New York, NY: Springer New York, 1992: 66-70.
- [47] MARRIOTT F H C. *A dictionary of statistical terms*[M]. 5th ed. New York: Longman Scientific & Technical, 1990.
- [48] HENDERSON M, PAGE L. Appraising the evidence: what is selection bias? [J]. *BMJ ment health*, 2007, 10(3): 67-68.
- [49] COGGON D, BARKER D, ROSE G. *Epidemiology for the uninitiated*[M]. 5th edition. New Jersey: BMJ Books, 2003.
- [50] BARRAZA F, ARANCIBIA M, MADRID E, et al. General concepts in biostatistics and clinical epidemiology: random error and systematic error[J]. *Medwave*, 2019, 19(7): e7687-e7687.
- [51] LASH T L, FOX M P, COONEY D, et al. Quantitative bias analysis in regulatory settings[J]. *American journal of public health*, 2016, 106(7): 1227-1230.
- [52] ALTHUBAITI A. Information bias in health research: definition, pitfalls, and adjustment methods[J]. *Journal of multidisciplinary healthcare*, 2016, 9: 211-217.
- [53] TRIPEPI G, JAGER K J, DEKKER F W, et al. Selection bias and information bias in clinical research[J]. *Nephron clinical practice*, 2010, 115(2): e94-e99.
- [54] DELGADO-RODRÍGUEZ M, LLORCA J. Bias[J]. *Journal of epidemiology & community health*, 2004, 58(8): 635-641.
- [55] ELLENBERG J H. Selection bias in observational and experimental studies[J]. *Statistics in medicine*, 1994, 13(5-7): 557-567.
- [56] KLEINBAUM D G, MORGENSTERN H, KUPPER L L. Selection bias in epidemiologic studies[J]. *American journal of epidemiology*, 1981, 113(4): 452-463.
- [57] YUE M, KANG C, ANDRIS C, et al. Understanding the interplay between bus, metro, and cab ridership dynamics in Shenzhen, China[J]. *Transactions in GIS*, 2018, 22(3): 855-871.
- [58] LIU L, HOU A, BIDERMAN A, et al. Understanding individual and collective mobility patterns from smart card records: a case study in Shenzhen[C] // 2009 12th International IEEE Conference on Intelligent Transportation Systems. 2009: 1-6.
- [59] LI B, ZHANG D, SUN L, et al. Hunting or waiting? discovering passenger-finding strategies from a large-scale real-world taxi dataset[C] // 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). 2011: 63-68.
- [60] ZHANG X, XU Y, TU W, et al. Do different datasets tell the same story about urban mobility - a comparative study of public transit and taxi usage[J]. *Journal of transport geography*, 2018, 70: 78-90.
- [61] GOLUB B, JACKSON M O. Naïve learning in social networks and the wisdom of crowds[J]. *American economic journal: microeconomics*, 2010, 2(1): 112-149.
- [62] ZICKUHR K. *Location-based services*[R]. Washington DC: Pew Research Center, 2013.
- [63] HARGITTAI E. Potential biases in big data: omitted voices on social media[J]. *Social science computer review*, 2020, 38(1): 10-24.
- [64] BOYD D, CRAWFORD K. Critical questions for big data[J]. *Information, communication & society*, 2012, 15(5): 662-679.
- [65] NAMUDURI S, NARAYANAN B N, DAVULURU V S P, et al. Review-deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors[J]. *Journal of the electrochemical society*, 2020, 167(3): 037552.
- [66] GRAHAM M, HALE S A, GAFFNEY D. Where in the world are you? geolocation and language identification in Twitter[J]. *The professional geographer*, 2014, 66(4): 568-578.
- [67] DAVIS JR. C A, PAPPAS G L, DE OLIVEIRA D R R, et al. Inferring the location of Twitter messages based on user relationships[J]. *Transactions in GIS*, 2011, 15(6): 735-751.
- [68] ELWOOD S, GOODCHILD M F, SUI D Z. Researching volunteered geographic information: spatial data, geographic research, and new social practice[J]. *Annals of the Association of American Geographers*, 2012, 102(3): 571-590.
- [69] MCCARNEY R, WARNER J, ILIFFE S, et al. The Hawthorne Effect: a randomised, controlled trial[J]. *BMC medical research methodology*, 2007, 7(1): 30.
- [70] MORSTATTER F, PFEFFER J, LIU H, et al. Is the sample good enough? comparing data from Twitter's streaming API with Twitter's firehose[J]. *Proceedings of the International AAAI Conference on Web and Social Media*, 2013, 7(1): 400-408.
- [71] LANSLEY G, LONGLEY P A. The geography of Twitter topics in London[J]. *Computers, environment and urban systems*, 2016, 58: 85-96.
- [72] BELL P A, GREENE T C, FISHER J D, et al. *Environmental psychology*[M]. 5th edition. Fort Worth, TX: Psychology Press, 2005.
- [73] 孙峰华, 李世泰, 黄丽萍. 中外犯罪地理规律实证研究 [J]. *人文地理*, 2006, 21(5): 14-18.
- [74] TELEP C W, WEISBURD D. What is known about the effectiveness of police practices in reducing crime and disorder? [J]. *Police quarterly*, 2012, 15(4): 331-357.
- [75] CULLEN F T. Criminal circumstance: a dynamic multicontextual criminal opportunity theory[J]. *Contemporary sociology*, 2004, 33(3): 359-361.
- [76] 生态环境部. 中国噪声污染防治公报 [R]. 北京: 中华人民共和国生态环境部, 2023.
- [77] BASNER M, BABISCH W, DAVIS A, et al. Auditory and non-auditory effects of noise on health[J]. *The lancet*, 2014, 383(9925): 1325-1332.
- [78] RICHARD H. THALER, CASS R. SUNSTEIN. *Nudge: improving decisions about health, wealth, and happiness*[M]. New Haven: Yale University Press, 2008.
- [79] KAHNEMAN D. *Thinking, fast and slow*[M]. 1st edition. New York: Farrar, Straus and Giroux, 2013.
- [80] KAHNEMAN D, TVERSKY A. Subjective probability: a judgment of representativeness[J]. *Cognitive psychology*, 1972, 3(3): 430-454.

(本文编辑: 王枫)



本文更多增强内容扫码进入