

引用格式: 马仪亮, 宋彦亭. 基于大数据空间标记的惯常环境技术定义[J]. 旅游学刊, 2024, 39(5): 161-170. [MA Yiliang, SONG Yanting. Understanding the usual environment in tourism: A technical definition based on big data space marking[J]. *Tourism Tribune*, 2024, 39(5): 161-170.]

# 基于大数据空间标记的惯常环境技术定义

马仪亮<sup>1</sup>, 宋彦亭<sup>2</sup>

(1. 中国旅游研究院, 北京 100005; 2. 农业农村部规划设计研究院, 北京 100125)

**[摘要]** 旅游是在非惯常环境的活动, 但非惯常环境要由惯常环境来定义。无论学术上还是技术上, 惯常环境自身也都缺少相对统一和明确的定义, 使得我国旅游统计执行错乱频发, 各类负面舆情不断。文章探讨了惯常环境技术定义的国际实践、原则和推荐的表达, 认为惯常环境是由以居所为中心和以职学地为中心的两类惯常空间的合集, 一个以不等圆组成的不受行政区划限定的不规则区域。研究还发现: 1) 为减少位置噪声点干扰使得簇质点偏移, 对两个惯常空间进行具有噪声的基于密度的空间聚类时扫描半径宜限定在 1 km 以内; 2) 以居所为中心的惯常空间不超过 1 个, 以职学地为中心的惯常空间个数小于等于 2, 根据位置点衰减情况判断, 前者空间半径以 40 km 为宜, 后者空间半径以 2 km~3 km 更为合理; 3) 无需用全样本位置数据进行标记, 通过代表性用户出游率或抵达率扩样实现总体推算; 4) 不能标记惯常环境的用户, 可假定其出游率或抵达率与能标记的用户相同, 即符合同一性假定。该研究可为大数据的旅游统计规范化应用提供技术参考, 为基于大数据的旅游流研究夯实了基础。

**[关键词]** 惯常环境; 技术定义; 空间标记

**[中图分类号]** F59

**[文献标识码]** A

**[文章编号]** 1002-5006(2024)05-0161-10

DOI: 10.19765/j.cnki.1002-5006.2024.05.015

## 0 引言

旅游学能否成为一门独立的学科, 取决于能否

**[基金项目]** 本研究受科技创新 2030 重大项目“节假日城市旅游客流智能预警决策与疏导关键技术研究”(2021ZD0111403)资助。[This study was supported by a grant from the Scientific and Technological Innovation 2030 Major Project (to MA Yiliang) (No. 2021ZD0111403).]

**[收稿日期]** 2023-05-15; **[修订日期]** 2023-08-25

**[作者简介]** 马仪亮(1979—), 男, 江西永修人, 博士, 副研究员, 研究方向为旅游统计与旅游大数据, E-mail: 381391944@qq.com; 宋彦亭(1989—), 女, 山东莱阳人, 硕士, 工程师, 研究方向为乡村旅游及其客流时空特征, E-mail: 1132343344@qq.com, 通讯作者。

构成一个相对独立的知识体系, 也就是具有特有的概念、原理、命题、规律等所构成的严密的逻辑化的理论体系<sup>[1]</sup>。1999年, 联合国世界旅游组织在定义旅游时首次引入惯常环境概念, 此后各国陆续将惯常环境确立为定义旅游的概念要件。联合国主导的《2008年国际旅游统计建议》, 定义“游客指出于任何主要目的(出于商务、休闲或其他个人目的, 而非在被访问国家或地点受聘于某个居民实体), 在持续时间不足一年的期间内, 出行到其惯常环境之外某个主要目的地的旅行者。”<sup>[2]</sup>实际上, 惯常环境并不指向旅游, 惯常环境之外才指向旅游。张凌云开创性地将惯常环境之外定义为“非惯常环境”, 引发了关于旅游的本质和游客在非惯常环境中游憩行为的深入探讨。一方面, 张凌云提出了旅游就是人们在非惯常环境下的体验和在此环境下的一种短暂的生活方式, 并将旅游的本质定义为“对非惯常环境的体验”<sup>[3]</sup>。一些学者也认同非惯常环境核心概念是旅游学的逻辑起点之一<sup>[4]</sup>。另一方面, 更多学者确实以非惯常环境概念为逻辑起点, 对旅游者非理性行为、旅游欺诈现象、价格感知、生命的意义、体验差异等现象进行了深入研究<sup>[5-7]</sup>, 显著推动了对旅游活动的学术理解。旅游在非惯常环境发生, 学术研究也更多聚焦于此。仔细观察相关研究发现, 广域环境减去惯常环境后就是非惯常环境, 世界上每个人的广域环境在空间上是相等的, 也就是所处发展阶段面向人类生活的全部活动空间。这意味着, 非惯常环境是依靠惯常环境来定义和理解的。从而定义项中的另一个概念“惯常环境”应该要有一个比较准确的已知定义, 才能避免被定义项陷入定义模糊<sup>[8]</sup>。

无论是从地理上还是心理上定义, 惯常环境势必存在理论上的空间边界, 否则惯常环境就与广域环境重合, 也就失去了讨论旅游的逻辑起点。遗憾的是, 惯常环境的理论边界仁者见仁。Govers等定

义惯常环境,与其说是从地方、空间、流动的多视角出发,通过网络、电话、留置问卷等方法进行超大样本的调查,对传统观点的惯常环境进行重置,不如说他们引入游客感知,提出了“让旅游者决定用何种时间和地点的数据来划分惯常环境和旅游、旅行的界限。”<sup>[9]</sup>自此,惯常环境需要从社会地理、环境心理等多视角进行综合考察的观点趋于主流,确定其空间边界的分歧越来越大,而不是相反。甚至出现认为交通的快捷、舒适、方便程度左右人们的心理距离,交通格局的变化可以达到“天涯若比邻”的效果的观点<sup>[10]</sup>,过度强调个体对距离的心理感知,而忽视出行频率。这一逻辑正如韩炳哲所说“电子邮件以此方式造就出即时性,即它完全除掉作为空间间歇的诸多道路。它摆脱了空间本身……每一个远方、每一段距离都要被灭除。要做的是,使一切在此时此地即可得以被支配。”<sup>[11]</sup>随着虚拟现实、元宇宙等数字化技术发展,惯常环境似乎就真的可以与广域环境重合了,届时,继续探讨旅游就没什么实际价值了。学术上百家争鸣是旅游研究之幸,对于旅游统计则可能意味着不知所措和频繁的统计数据质疑。因为,如果人人有自己的惯常环境理解以及在此基础上的旅游活动认知,就难以建立一套各方认可的统计规则。如果技术定义中也依赖游客感知,惯常环境就变成了一个因人而异的主观概念,关于旅游人数的统计就只能经由游客抽样调查这一种方法了,因为能且只能在游客自己感知是否离开惯常环境的基础上进行扩样推算。由于不同经济发展水平、不同文化风俗、不同生长阶段的人对惯常环境大小的感知不同,区域旅游统计数据之间就完全失去可比性了。管婧婧等也认为惯常环境的地理边界应该是开放和发散的,但是从统计操作角度可强行划定物理距离尺度<sup>[8]</sup>。国际上并没有关于惯常环境的统一技术定义,联合国世界旅游组织建议各国根据交通便利情况、人口密度、文化行为、距离行政边界的距离远近等确定一个“唯一的个人惯常环境”<sup>[2]</sup>。一些国家用“出城”“异地”“外地”“离开本地”“离家40 km(或50 km、80 km等)”等表述,含糊其辞。我国自1993年开展国内旅游统计、2000年出台旅游统计制度至今,从来没有对惯常环境下过定义,使得很多学者不理解旅游人次怎么统计得来,也不能准确向学生和基层旅游统计人员传授讲解,特别是各地争相用位置数据监测假日游客接待情况,方法各不相同,且都不是在标记游客惯常环境基础上展开,新一轮旅游数据乱象有愈演愈烈之势。

## 1 基于位置大数据的惯常居住及职学空间标记

### 1.1 惯常环境技术定义的国际隐喻

惯常环境的出现经历了数十年演变,根据张凌云对国际上主流旅游定义的梳理,“非定居地”“离开自己的住地”“离开定居地点较远的地方”“离家”“离开日常生活圈”等表述都出现过(表1)。从表1可以看出,学术上对旅游的定义多从离开居住地展开。各国虽然尽量回避定义惯常环境,但为了统计需要都采用隐喻的方式大致勾勒惯常环境的空间规则。美国对游客出行的时空要求是50 miles(约80 km)<sup>①</sup>,法国则是离开住所超过100 km<sup>②</sup>,日本要求游客离开日常生活环境达到单程80 km<sup>③</sup>,澳大利亚要求过夜游客离家40 km以上,对一日游则为离家往返50 km<sup>④</sup>,西班牙规定游客必须出行至行政边界(市)以外<sup>⑤</sup>。这些国家虽没有明确定义惯常环境,但有两方面的隐喻:1)个体只拥有一个惯常环境;2)住所是多数国家从技术上定义惯常环境的核心。我国官方颁布的旅游统计调查制度使用惯常环境概念,但从来没有对惯常环境进行过界定。2020年颁布的《全国文化文物和旅游统计调查制度》中出现了“惯常环境”“惯常生活环境”“惯常居住地”“惯常居住环境”等多种表述,但没有对惯常环境进行完整定义。从制度的表述中,统计操作上的惯常环境是从居住地展开的。各国技术上对惯常环境的界定可以分为3种:第一种是整圆说,法国、日本和澳大利亚等将惯常环境明确为住所或家周边一定半径距离范围,呈圆圈状;第二种是不规则生活地说,英国限定为游客所在镇或者乡村,西班牙为市政边界内;第三种是“悬置”说,中国、美国和加拿大等都对惯常环境进行模糊处理。

《2008年旅游统计国际建议》关于惯常环境的表述最为全面,但都是抽象性建议,而不是关于时空的具体规定,如“一个人的惯常环境是旅游的一个关键概念,被定义为一个人日常生活的地方(不一定是一个毗邻的地方)”<sup>[2]</sup>“一个人的惯常环境包括其所属住户的惯常居住地,其自身的工作或学习

① 数据来源: <https://apps.bea.gov/scb/2022/02-february/0222-travel-tourism-satellite-account.htm>.

② 数据来源: <https://www.insee.fr/fr/metadonnees/definition/c101>.

③ 数据来源: <https://www.mlit.go.jp/kankocho/>.

④ 数据来源: <https://www.tra.gov.au/en/economic-analysis/tourism-satellite-accounts>.

⑤ 数据来源: [https://www.ine.es/daco/daco42/frontur/frontur\\_egatur\\_metodologia.pdf](https://www.ine.es/daco/daco42/frontur/frontur_egatur_metodologia.pdf).

表 1 “惯常环境”相关概念演变<sup>[10]</sup>  
Tab. 1 Evolution of “usual environment”

学者或机构 Scholars or institutions	年份 Time	所用概念 Concept used
施拉德 H. V. Schllard	1910	非定居地
鲍尔曼 A. Bormann	1931	离开自己的住地
奥格尔维 F. W. Ogilvie	1933	离家
吕克斯曼 R. Glücksmann	1935	旅居地
亨齐克和克拉普夫 W. Hunziker 和 K. Krapf	1942	非定居地
田中喜一 Kiichi Tanaka	1950	原居住地
维也纳经济大学 Vienna University of Economics and Business	-	暂时在异地
让·梅特森 Jean Mateson	1966	定居地点较远的地方
英国旅游协会 Tourism Society of England	1976	常居住和工作之外的地方
贾法利 J. Jafari	1977	常住地
伯卡特和梅特利克 A. J. Burkart 和 S. Medlik	1981	正常居住地
利珀 N. Leiper	1981	离家
“艾斯特”(Aiset)和旅游社团 Aiset 和 Tourism Society	1981	家庭环境之外
马西森和沃尔 A. Mathieson 和 G. Wall	1982	惯常的工作和居住环境
铃木忠义 Tadayoshi Suzuki	1984	日常生活圈
伦德伯格 D. E. Lundlberg	1985	常住地社区
克里斯·瑞安 C. Ryan	1991	离开家
世界旅游组织 World Tourism Organization	1991	惯常居住环境
克里斯·库珀 C. Cooper	1993	自己惯常居住的环境
约翰·特赖布 John Tribe	1997	从一地到另一地
斯蒂芬·威廉斯 Stephen W. Williams	1998	惯常居住地

地以及其定期和经常光顾的任何其他地方,即使该地远离其惯常居住地或在另一地区时也是如此。”<sup>[2]</sup>“每个住户都有一个主要住所(有时称为首要或基本住所),这种住所通常根据逗留的时间确定,其所

在位置用来确定居住国和该住户及其所有成员的惯常居住地。”<sup>[2]</sup>可以知道,惯常环境至少应由住户的惯常居住地和其工作或学习地两部分组成。离开某层级行政区或离家 40 km 是多数国家进行旅游统计的最低空间标准,居住地和工作或学习地往往都在这一空间范围内。我国则不然,没有定义惯常环境,但把离开居住地 10 km 作为游客统计的基本准则,甚至低于一些城市居民的平均通勤距离,很多人工作和学习地与居住地不在以居所为中心的 10 km 范围内,就造成了由居住地去往超过 10 km 的工作地周边多大范围外才计入旅游的技术障碍。把以居所为中心的那部分惯常环境称为以主要居所为中心的惯常空间,半径为  $R^h$ , 根据我国目前的制度规定  $R^h$  暂时等于 10 km。把以职学地为中心的那部分惯常环境称为以职学地为中心惯常空间,半径为  $R^w$ 。如何通过大数据的方法对两个空间进行界定和标记,是本文研究的主要内容。

### 1.2 以居所为中心的惯常空间标记

通过拦访居所所在地住户受访者可能介意隐私避而不答,对于其活动范围又因对距离感知较为含糊难以精确化处理,本文选取位置大数据进行客观标记。随机选取 2022 年 6 月南京电信某用户 M 信令位置点,位置点每半小时提取一次,合计提取到该用户位置点 1188 个(每个点都由经纬度刻画),可以大致看出该用户主要在两个区域活动,位置点总体分布如图 1 所示。

先用  $K$  均值聚类算法 ( $K$ -Means clustering algorithm) 进行空间聚类,算法步骤是预先将数据分为  $K$  组,随机选取  $K$  个点作为初始的聚类中心,然后



审图号:GS(2023)3206号

图 1 中国电信用户 M 2022 年 6 月位置点分布  
Fig. 1 Distribution of location points of China Telecom user M in June 2022

计算每个点与各个种子聚类中心之间的距离,把每个点分配给距离它最近的聚类中心(或簇质心),该聚类中心以及分配给该中心的点集就代表一个聚类(或簇)。一旦全部点都被分配了,每个聚类的聚类中心会根据聚类中现有的点被重新计算,如此反复迭代直至满足误差平方和(各簇内的样本点到所在簇质心的距离平方和)局部最小。简便起见,各簇内的样本点到所在簇质心的距离以两点之间的直线距离表征,即欧几里得距离。设有X点(lat1 long1)和C点(lat2 long2),lat1和lat2分别为两点的纬度,long1和long2分别为两点的经度,则:

$$\begin{cases} d = r \times \arccos[ \sin(\text{lat1}) \times \sin(\text{lat2}) + \cos(\text{lat1}) \times \cos(\text{lat2}) \times \cos(\text{long2} - \text{long1}) ] \\ J(c, u) = \min \sum_{i=1}^n \|X_i - u_c\|^2 \end{cases} \quad (1)$$

式(1)中,d为X点和C点之间的距离(单位为m);r=6371.004 km,为地球半径;i为某个簇的位置点个数,X<sub>i</sub>和C<sub>i</sub>是某个簇位置点的位置属性标记,u<sub>c</sub>为簇质心点;本文聚类中,||X<sub>i</sub> - u<sub>c</sub>||以式(1)中d表征。

对夜晚22:00至次日早晨5:00期间位置点进行

聚类,当K值为2时(图2左上),位置点可以聚类为红色和蓝色2簇。K值为3时(图2右上),可以聚类为蓝色、绿色和红色3簇;K值为4时(图2左下),可以聚类为蓝色、红色、黑色和绿色4簇;K值为5时(图2右下),则可以聚类为蓝色、红色三角、红色圆、绿色和黑色5簇。实际上,簇数越多并不见得越好,K值增多后反而将更多的伪中心标记出来造成更多干扰。如图1所示,该用户在A点(纬度:32.05413°,经度:118.76824°)、B点(纬度:32.053888°,经度:118.765833°)和C点(纬度:32.10333°,经度:118.9408°)3个位置点出现频繁,6月出现次数分别为436次、265次和103次。计算两个经纬度点之间的距离,A点与B点相距228.43 m,A点与C点相距17.15km,B点与C点相距17.38 km,进一步考察表2可以有以下发现。1)加入时间标签,A和B两点为白天活动点,C点为夜晚活动点。2)A点为该用户工作地附近的可能性较大,B点以及周边位置点(纬度:32.043841°,经度:118.78541°;纬度:32.05726°,经度:118.770929°;纬度:32.05404°,经度:118.76583°等)距离A点很近,多数仅相距数百米,大概率为该

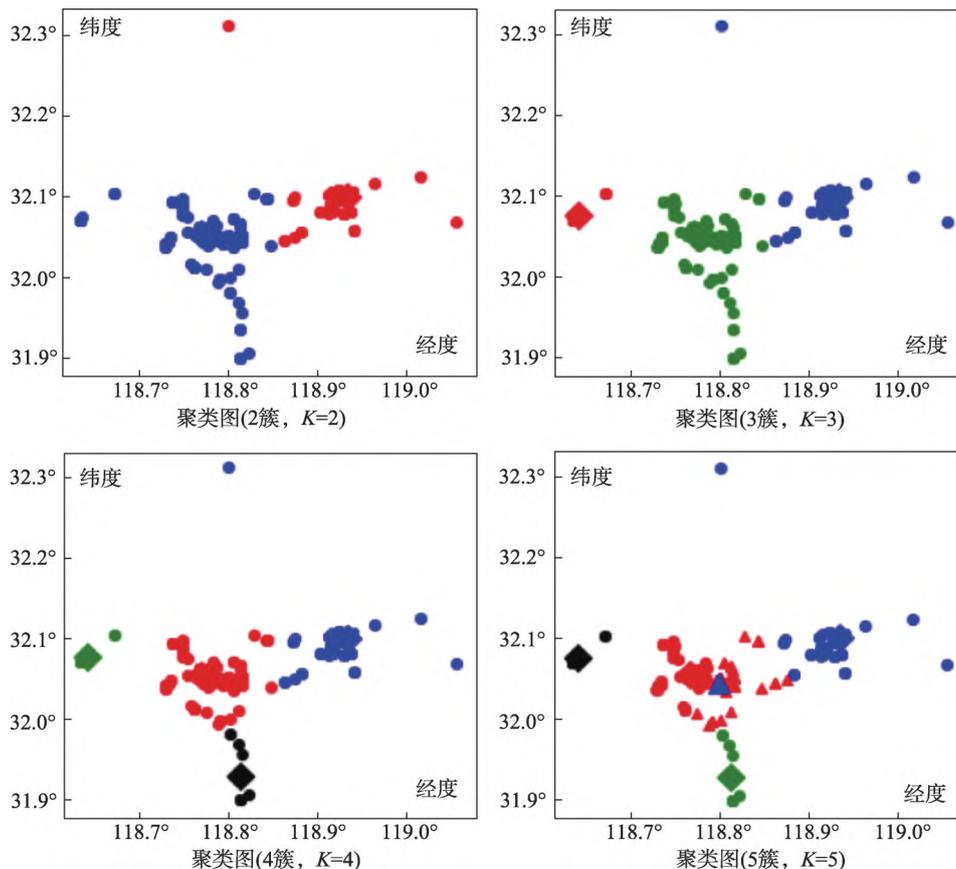


图2 中国电信用户M不同K值K-Means聚类

Fig. 2 K-Means clustering of China Telecom User M with different K-values

表 2 中国电信用户 M 2022 年 6 月聚类时段位置点分布

Tab. 2 Distribution of location points during the clustering period of China Telecom user M in June 2022

位置点纬度 Latitude	位置点经度 Longitude	位置点个数 Number of points	位置点纬度 Latitude	位置点经度 Longitude	位置点个数 Number of points
31.99855°	118.80251°	1	32.05361°	118.77528°	2
31.99593°	118.79186°	1	32.07835°	118.93813°	4
31.99568°	118.79202°	2	32.09859°	118.93440°	28
31.99141°	118.78938°	1	31.97908°	118.80350°	2
32.08432°	118.74894°	1	31.90339°	118.82330°	1
32.30991°	118.80150°	1	31.93208°	118.81513°	2
32.07451°	118.74861°	1	32.03980°	118.79550°	2
32.05413°	118.76824°	366	32.04310°	118.78612°	13
32.05389°	118.76583°	195	32.06973°	118.80672°	2
32.05250°	118.77940°	9	32.06426°	118.81416°	2
32.05231°	118.78220°	9	32.04310°	118.80250°	3
32.04384°	118.78541°	30	32.00722°	118.77528°	1
32.05213°	118.77019°	1	32.03687°	118.77719°	1
32.07840°	118.91661°	4	32.05404°	118.76583°	20
32.10333°	118.94080°	93	32.07333°	118.63639°	1
32.09543°	118.84407°	2	32.06905°	118.63512°	8
32.05726°	118.77093°	14	位置点合计个数		823

注:聚类时段为上午 10:00 至下午 17:00 及晚 22:00 至次日早上 5:00。

用户在工作地周边吃饭、商务谈判等活动点。需要说明的是, A 点是与该用户白天时段手机“握手”最频繁的基站所在位置, 并不是该用户办公室所在位置。城市区域, 两者相距一般不会超过 1 km。3) C 点为该用户居住点附近的可能性较大, 纬度: 32.09859°, 经度: 118.9344° 为该用户在居住地周边主要活动点。也就是当  $K$  值增加时, 将从 B、C 等点附近标记出簇质心, 造成更多干扰。

经过  $K$ -Means 聚类, 发现该用户住所所在地活动点的簇质心为纬度: 32.104218°, 经度: 118.940296°。由于人们居所或职学地个数事先不可知, 在聚类前设定  $K$  值较为不妥。特别是  $K$ -Means 方法容易收到离散点干扰。设  $K=2$ , 需要从图 2 所示红色位置点中聚类出簇质心, 一些离散的位置点容易使得簇质心远离用户真实居所所在地。如果将离散的位置点去除后进行空间聚类, 就能减少簇质心与用户真实居所所在地之间的偏离。为此, 改用具有噪声的基于密度的空间聚类算法 (density-based spatial clustering of applications with noise, DBSCAN) 进行簇质心标记。该方法将簇定义为密度相连的点的最大集合, 能够把具有足够高密度的区域划分为簇, 并可在噪声的空间数据库中发现任意形状的聚类。聚类过程中需要提前设定

两个参数: 扫描半径 (eps) 和最小包含点数 (minPts)。任选一个未被访问的点开始, 找出与其距离在 eps 之内 (包括 eps) 的所有附近点。如果附近点的数量大于等于 minPts, 则当前点与其附近点形成一个簇, 并且出发点被标记为已访问。然后递归, 以相同的方法处理该簇内所有未被标记为已访问的点, 从而对簇进行扩展。如果附近点的数量小于 minPts, 则该点暂时被标记作为噪声点。如果簇充分地扩展, 即簇内的所有点被标记为已访问, 然后用同样的算法去处理未被访问的点。也就是说, DBSCAN 通过参数设定只针对图 2 左上角密集区域红色点位进行聚类, 而将离散点干扰作为噪声点排除在外。对照图 3 中位置点坐标, 通过百度地图坐标反查和测距工具大致观察主要位置点簇心与边界之间的距离和点位数, 设置 eps=1000 m, minPts=100 个进行聚类, 通过 Python 代码迭代聚类出两个簇及其质心<sup>①</sup>。发现该用户居住地在万山庭院小区北门附近楼宇, 簇质点坐标为纬度:

① 笔者比较了不同用户不同 eps 和 minPts 参数聚类结果, 对照地图上簇点与小区、商业楼宇位置发现, eps 为 1 km 时簇点落在楼宇的频率较高, 篇幅所限文中不予赘述。感兴趣的读者可向作者索取: 381391944@qq.com。Python 代码的算法要义是在位置点集合中确定核心点、边界点和噪声点, 并通过密度直达或密度可达关系确定聚类簇。

32.105472°, 经度: 118.94066°, 与用K-Means方法聚类结果直线距离约 143.59 m, 与C点之间相距 238.54 m。

加入用户以其住所为中心, 东南西北各方向活动分布均匀, 则DBSCAN空间聚类后簇质点将非常靠近用户的真实居所。但多数情况下各方向的活动分布是不均匀的, DBSCAN空间聚类得到的簇质点往往与用户居所所在地存在偏差, 但由于基站(包括地下停车场、大型商超等安装的小基站)较为密集, 几百米偏离不会对用户惯常环境的空间划分构成显著影响。而且偏离往往出现在用户活动相对频繁的方向, 也就是簇质点成为居所与周边活动密集点之间的某个位置, 理论上来说, 作为用户在住地周边停留和活动的中心位置更有代表性。找到代表居所的簇质点后, 就需要确定以居所为中心的惯常空间应该多大。我国旅游统计制度规定, 离开惯常居住地 10 km, 也就是  $R^h=10$  km, 与世界各国的规定存在明显差异。比较或探讨哪一个距离标准更合适不是本文研究重点, 主要为了提出用上述空间聚类法研究人们的活动位置点在不同距离尺度的分布情况是今后研究距离标准的可用参考。继续在南京随机选取 500 名电信活跃用户, 占同期南京市活跃用户总数的万分之一, 用户平均年龄 38.1 岁, 男性占比 50.7%。对这些用户夜晚 22:00 至次日早晨 5:00 间轨迹点, 利用 DBSCAN 在不同 eps 参数下进行空间聚类, 发现人们绝大多数位置点在居所所在地簇质点附近数千米以内(表 3), 93.36% 的位置点在 10 km 范围内。笔者曾通过位置大数据监测了国人一年内平均旅游天数约为 7 天, 假设旅游和非旅游时间每天的位置点数相同, 则旅游的位置点占比约 1.92%。对照表 1, 恐怕 40 km 标准对于

消除非旅游活动干扰更加有效。

### 1.3 以职学地为中心的惯常空间标记

考虑到K-Means法存在的适用性弊端, 直接用DBSCAN对电信用户M白天自上午 10:00 至下午 17:00 点之间的位置点进行空间聚类(以下所称白天均指上午 10:00 至下午 17:00 点聚类时段), eps 和 minPts 参数设定同上。聚类出的簇质点在南京师范大学随园校区内(图 3), 具体位置为纬度: 32.059223°, 经度: 118.775472°, 该用户可能为南京师范大学职工或走读的学生。

该用户白天和夜晚只各聚类出一个簇质点, 分别作为该用户的职学地中心位置和主要居所活动区域的中心位置, 两个中心位置点之间的距离为 16.39 km。如前所述, 以居所活动区域的中心位置为圆心半径 10 km(即  $R^h=10$ ), 可以理解为我国目前使用的以居所为中心的惯常空间, 但如果  $R^w = R^h = 10$  km, 则家住五环外在北京前门附近上班的人, 到四环以内的全部区域, 包括去动物园、什刹海、玉渊潭等在内都不能算旅游。人们往往在居所附近的活动半径会大于在职学地的活动半径, 也就是  $R^w < R^h$  可能更加符合实际。中国旅游研究院(文化和旅游部数据中心)研究居民休闲时, 通过连续数年问卷调查发现, 居民多数休闲活动在距家 3 km 范围内开展。浙江省文化和旅游厅相关部门与中国移动合作, 通过观察用户不同时间段停留的栅格数(栅格范围为 250 m × 250 m), 借助 K-Means 和 LSTM(Long short-term memory, 长短记忆网络)相融合的聚类方法, 结合短信调查, 认定将  $R^w=2$  km

表 3 500 名中国电信用户夜晚轨迹点空间聚类  
Tab. 3 Spatial clustering of night trajectory points of 500 China Telecom users

活动半径/km Activity radius	用户位置点占比/% Proportion of location points
2	83.44
3	85.81
5	88.94
10	93.36
20	97.23
30	98.68
40	99.16
50	99.51



审图号:GS(2023)3206号

图 3 中国电信用户 M 2022 年 6 月白天位置点 DBSCAN 聚类  
Fig. 3 DBSCAN clustering of daytime location points for China Telecom user M in June 2022

作为当地大数据旅游客流监测的试用标准。

笔者利用信令数据对 500 名随机电信用户 10:00—17:00 轨迹点,利用 DBSCAN 在不同  $\epsilon$  参数下进行空间聚类,发现用户的白天活动轨迹点中,88.7% 在日常工作或学习中心点 2 km 以内,90.5% 的轨迹点在 3 km 范围内。其中,81.6% 的行为轨迹点在 2 km 范围内的用户达到 7 成(表 4)。由于人们白天行为轨迹更加分散,规律性相对不明显,轨迹点聚类的要求应适当放宽。更为重要的是,如前所述,惯常环境包括两类惯常空间,即以居所为中心的惯常空间和以职学地为中心的惯常空间,前者为主,后者为辅。也就是说,在非惯常环境的旅游活动,从空间移动上而言,首先或者主要是对以居所为中心的惯常空间的否定,从形态上而言,旅游作为异地的短期生活主要是对居住地常态化生活的否定。以职学地为中心的惯常空间,在更多情境下不是作为非惯常环境的对立面存在,而是作为将惯常环境不同区域之间的通勤等出行排除出旅游而存在。也就是说,以职学地为中心的惯常空间首要的作用不是为了否定旅游,而是为了否定某些经常性的空间移动。这也是表 1 中多数旅游的学术定义中指向离开住地,以及前文所述各国在旅游统计中大多规定离开住所的原因。因此,  $R^w$  不宜过大,否则可能将很多旅游活动排除在外。表 4 可以看出,  $R^w$  等于 2 km 或 3 km 轨迹聚类占比差异不大,将  $R^w$  设置为 2 km 或者 3 km 都具有较强的解释力。

#### 1.4 空间标记中的特例处理

按照技术定义,通常每个人都有一个惯常环境,由于有的人没有手机、不同移动运营商底层数据无法融合、有的用户位置点提取较少(较长时间

关机、信号不稳定、手机硬件故障等各种原因)被  $\minPts$  参数过滤等,不是每个用户都可以标记出以居所为中心的惯常空间或以职学地为中心的惯常空间,但并不显著影响位置大数据在旅游统计中的应用,3 种特殊情境需要特别处理。第一种情境,标记不出以居所为中心的惯常空间(无论是用户本就没有主要居所,还是因为技术和数据的原因标记不出,下同),但可以标记出以职学地为中心的惯常空间。这类用户可以直接剔除暂不予标记(后期可标记出再进行标记),按照相似性假定,可认为这些用户的出游率或者到某个目的地的抵达率,与其所在省域、市域或县域活跃用户(可标记明确惯常环境)类似。第二种情境,标记不出以职学地为中心的惯常空间,但可以标记出以居所为中心的惯常空间。这类用户以没有固定工作地点的用户居多,惯常环境等于居所为中心的惯常空间。第三种情境,两个惯常空间都标记不出,处理方法参照第一种情境。

需要解释的是,实践中没有任何一种位置大数据可以覆盖全部人口,往往能够脱敏后接入使用的只有某一家数据商的位置数据,距离全样本覆盖更是相去甚远。此外,就算标记出来惯常环境,依此计算得到的非惯常环境出行也不见得都是旅游,比如出租车司机、货运车司机等每天无规律长距离移动,仅通过位置数据是难以剔除非旅游活动的。为此,需要两方面技术补充。1)通过出游率或游憩抵达率进行反演扩样,比如本文利用中国电信不到 3 亿用户作为样本,标记其惯常环境后计算电信活跃用户的总体出游率作为全国人口的初始出游率。在计算地方旅游接待时,则计算各客源地(根据需要确定不同区域尺度)电信活跃用户到某目的地的不同初始抵达率,并分别作为各客源地全体居民至该目的地的初始抵达率。2)通过短信推送调查剔除非旅游活动。为了认定旅游动机的出行活动,需要针对离开惯常环境最低时间标准以上的用户进行动机调查,计算非旅游动机占比,并以此将初始出游率和抵达率修正为最终出游率和抵达率。未来可以通过机器学习这类人工智能技术判定出游动机,目前来说短信调查的实用性更强。

## 2 惯常环境的技术定义表达

### 2.1 定义原则

对惯常环境进行尽可能简单、最容易理解的技术规定,应该遵循 4 个方面的原则。第一,遵循共性

表 4 500 名中国电信活跃用户白天轨迹点分布  
Tab. 4 Distribution of daytime trajectory points of 500 users of China Telecom

轨迹点占比标准/% Proportion standard of location points	不同活动半径内达标用户占比/% Proportion of users exceeding the standards within different activity radius				
	1 km	1.5 km	2 km	3 km	5 km
95	45.2	50.2	55.4	59.4	65.4
90	53.0	59.0	63.4	67.2	74.2
80	66.2	71.2	75.4	78.0	82.8
70	76.0	79.4	81.6	84.2	87.0
60	82.2	85.4	87.2	88.8	90.6

认知原则,也就是符合尽可能多的人的直觉或经验判断。比如怎样频率的抵达或驻足才能称为惯常,不用坐班的人工作地算不算,一周只有很少几天有课的老师工作地算不算,如果算,那么去单位几天才算,每次去要不要达到一定的时间标准。外卖员、快递员、货车司机、出租车司机等,白天要不要有惯常环境,有的话又需要怎样的标准。对这些问题的规定,要尽可能符合大众的常识性认知。第二,遵循最小理解障碍原则,不能过于繁琐、晦涩。比如将惯常环境界定为居住地、居住环境、常住地、定居地等,读者并不能快速据此确定空间范围。第三,遵循最少特例原则,就是不能针对很多不同的人或情境设置不同的规则,随着社会发展,人群和情境划分越来越多样,无法穷尽。第四,统计手段高适配原则,就是既有的统计抽样或大数据等手段能操作。

## 2.2 定义表达

经过上述关于惯常环境大数据标记的讨论,为明确惯常环境的技术定义打下了基础,虽然仍旧容易招致争议,但对深化相关研究不可或缺。基于以上关于白天和夜间用户活动轨迹的聚类实证,综合考虑《2008年旅游统计国际建议》的概念性表达,以及各国在统计工作中的实际做法和表述上的“隐喻”,以及便于未来大数据旅游统计应用,按照“奥卡姆剃刀”(Occam's Razor)<sup>①</sup>原理尽可能化繁为简,避免陷入无穷无尽的特例考量之中。因此,建议将惯常环境的技术定义阐述为:一个人日常生活和工作的地方,具体为主要住所周边一定范围和职学地周边一定范围构成的空间合集。这一表述需要进行以下说明。1)不再专门强调定期和经常光顾的区域空间。因为“经常”和“定期”无法衡量,每个人都有自己认为的标准,选择任何标准都可能招致非议且不符合“奥卡姆剃刀”原理。为了不把去往定期和经常光顾的地方算作旅游,将在抽样调查中获得此项活动占比并予以剔除。当然,如果光顾确实较为频繁,空间聚类中容易将该地记录为用户的职学惯常空间,不影响旅游统计结果。2)住所和职学空间的空间合集,为两个空间的并集,存在包含、相交和分离3种状况,分离的状况可以跨越行政区,如在天津生活却在北京工作的人,惯常环境涉及天津和北京两地。3)没有惯常居住地的难民、流离失所者或一些行商(驾车巡回表演、维修、游牧、养蜂等群体),其逗留地视为其惯常环境,不计入游客之

列。4)一个人有且只有1个惯常环境,且惯常环境=以居所为中心的惯常空间+以职学地为中心的惯常空间。世界旅游组织在惯常环境的认定中推荐唯一的主要居所,本文的样本测试中,夜间惯常空间识别结果为2的用户数是0,建议以居所为中心的惯常空间不超过1个,1个以上视为度假居所。而有多达25.3%的样本用户识别出两个职学惯常空间,识别出3个职学惯常空间的样本用户不到1%,建议以职学地为中心的惯常空间可以不超过两个。5)住所周边距离标准与职学周边距离标准不要求相等,需要根据人们活动的位置点分布确定。6)惯常环境不受国境线制约,即边境工人的惯常环境可以包括其工作的邻国一定区域。

## 3 结论与展望

第一,惯常环境=以居所为中心的惯常空间+以职学地为中心的惯常空间,两个惯常空间分别是以居所和职学地周边1km(过大容易受噪声干扰使得簇质点发生偏移)范围内位置点的空间聚类簇质点,以居所为中心的惯常空间不超过1个,以职学地为中心的惯常空间个数小于等于2。因此,从空间形态来说,惯常环境是一个以个人不同时段主要活动点簇质点为中心的大小不同、数量不等的圆组成的不规则区域,且可跨越边境。

第二,以居所为中心的惯常空间是惯常环境的核心,以职学地为中心的惯常空间主要不在于表征人们的活动分布,而在于将惯常环境不同区域之间的通勤、求学等出行排除出旅游活动。因此,前者的距离半径要明显大于后者,否则容易将旅游活动过滤掉。《2008年旅游统计国际建议》承认,“各国在人口密度、交通便利情况、文化行为、离国家或行政边界的远近程度等方面通常有差异,这些差异不利于为全球统计确定一个唯一的个人惯常环境。”<sup>[2]</sup>但也“建议领国或属于超国家组织的国家相互磋商,以确保编制的统计数据具有可比性。”<sup>[2]</sup>无论从国际可比还是活动位置点分布来看,采用40 km作为以居所为中心的惯常空间的半径,较10 km更优。如此,在大数据识别中,大多数用户的居所惯常空间与职学惯常空间出现包含关系,只需按照居所惯常空间标记用户的惯常环境,可以大大节省排除通勤

<sup>①</sup> 由英国逻辑学家奥卡姆提出,核心思想是“如无必要,勿增实体”,提倡避繁逐简,当能够用更简单的方式解释事物时,就不要让复杂的理论或假设增添干扰。

活动方面的算力。

第三,对比两种常用的空间聚类方法,发现 K-Means 法操作简单,迭代速度快,可以任意设置聚类数量,但也存在容易受异常点干扰、K 值难以先验设定等缺点。DBSCAN 方法操作性和迭代速度没有优势,但有利于去除异常点干扰。本着质量服从于速度的原则,笔者认为 DBSCAN 方法在居民惯常环境大数据标记中适用性更强。

第四,DBSCAN 聚类中的参数设置对结果影响较大,可以根据城市或农村、城市大小及其交通条件、不同季节人们出行空间分布情况、周末和平时等进行差异化设置,标记的精度将更高。

第五,存在少量夜晚上班白天休息的人群。在进行居所惯常空间聚类的时候,有的人应该聚类白天的空间轨迹,而聚类职学惯常空间的时候也应该聚类夜晚的空间轨迹。有待于利用图神经网络技术,根据聚类出的簇质点属于小区、写字楼等特征进行自动区分标记。

本文关于惯常环境的技术性定义及其设定的大数据标记方法可用于实践,笔者的本职工作也是每月依此生成全国 357 个城市(不含港澳台)矢量旅游流矩阵,支撑全国假日旅游统计、旅游热度预报等工作。未来将进一步考虑城乡、城市体量、季节等因素以及利用误差更小的 GPS 位置数据进行标记,提高惯常环境的标记精度。以惯常环境大数据标记为基础的矢量旅游流监测,凭借计算快、精度高、信息量大、研究的拓展性强等优势,将推动旅游流研究进入新的层次。

#### 参考文献(References)

- [1] 张凌云. 走出混沌: 旅游学科的归属与性质探索[J]. 中大管理研究, 2012, 7(1): 13-34. [ZHANG Lingyun. Out of confusion: An exploration of the attribution and nature of the subject of tourism[J]. *China Management Studies*, 2012, 7(1): 13-34.]
- [2] United Nations. *International Recommendations for Tourism Statistics*

- 2008[M]. New York: United Nations Publication, 2010: 9-34.
- [3] 张凌云. 旅游学研究的新框架: 对非惯常环境下消费者行为和现象的研究[J]. 旅游学刊, 2008, 23(10): 12-16. [ZHANG Lingyun. A study on consumers' behavior and phenomenon under unusual environment[J]. *Tourism Tribune*, 2008, 23(10): 12-16.]
- [4] 张朝枝. 面向旅游学一级学科的核心课程设置探讨[J]. 旅游学刊, 2015, 30(9): 2-4. [ZHANG Chaozhi. A study on the core curriculum design for the first—Level discipline of tourism[J]. *Tourism Tribune*, 2015, 30(9): 2-4.]
- [5] 李庆雷, 吕文艺. 非惯常环境下的历史文化古镇旅游再开发研究——以昆明市官渡古镇为例[J]. 武陵学刊, 2013, 38(1): 61-65. [LI Qinglei, LYU Wenyi. Study on the tourism development of ancient towns in the unusual environment: Taking the Guandu ancient town in Kunming as an example[J]. *Journal of Wuling*, 2013, 38(1): 61-65.]
- [6] 李春晓, 冯浩妍, 吕兴洋, 等. 穷家富路? 非惯常环境下消费者价格感知研究[J]. 旅游学刊, 2020, 35(11): 42-52. [LI Chunxiao, FENG Haoyan, LYU Xingyang, et al. The impact of unusual environmental differences on price perception from the perspective of construal level theory[J]. *Tourism Tribune*, 2020, 35(11): 42-52.]
- [7] 王欣, 邹统钎. 非惯常环境下体验的意义[J]. 旅游学刊, 2011, 26(7): 19-23. [WANG Xin, ZOU Tongqian. On the significance of experience under unusual environment[J]. *Tourism Tribune*, 2011, 26(7): 19-23.]
- [8] 管婧婧, 董雪旺, 鲍碧丽. 非惯常环境及其对旅游者行为影响的逻辑梳理[J]. 旅游学刊, 2018, 33(4): 24-32. [GUAN Jingjing, DONG Xuewang, BAO Bili. A logical disentangling of the concept "unusual environment" and its influence on tourist behavior[J]. *Tourism Tribune*, 2018, 33(4): 24-32.]
- [9] GOVERS R, HECKE E V, CABUS P. Delineating tourism: Defining the usual environment[J]. *Annals of Tourism Research*, 2008, 35(4): 1053-1073.
- [10] 张凌云. 国际上流行的旅游定义和概念综述——兼对旅游质的再认识[J]. 旅游学刊, 2008, 23(1): 86-91. [ZHANG Lingyun. Review on the definitions and concept of tourism currently popular in the world—Recognition of the nature of tourism[J]. *Tourism Tribune*, 2008, 23(1): 86-91.]
- [11] 韩炳哲. 时间的味道[M]. 包向飞, 徐基太, 译. 重庆: 重庆大学出版社, 2017: 82-83. [HAN Byung-Chul. *Scent of Time: A Philosophical Essay about the Art of Lingerin*[M]. HAN

## Understanding the Usual Environment in Tourism: A Technical Definition Based on Big Data Space Marking

MA Yiliang<sup>1</sup>, SONG Yanting<sup>2</sup>

(1. *China Tourism Academy, Beijing 100005, China;*

2. *Academy of Agricultural Planning and Engineering, Ministry of Agriculture and Rural Affairs  
of the People's Republic of China, Beijing 100125, China)*

**Abstract:** The term “tourism” refers to various forms of activities that take place in an unusual environment. This “unusual” environment needs to be defined in terms of its opposite, i.e., the “usual” environment. However, the lack of a relatively homogeneous and unambiguous description of the usual environment, both academically and technically, has led to frequent and repeated mistakes in the implementation of tourism statistics in China. As a result, there has been a continual flood of negative public opinion and a variety of controversies relating to the concept of tourism. This study attempts to address the aforementioned issues through the following procedure. First, we present a review of the international practices, principles, and recommended expressions of the “usual environment” in a technical context. Second, we identify an individual’s usual environment as an ensemble of two distinct types of usual spaces, namely the direct vicinity of a person’s residential address and the region surrounding a person’s place of employment or education. Geographically speaking, the usual environment is an irregular area made up of uneven circles without the limitation of administrative subdivisions. Third, based on the labeling of big data, we employ several spatial clustering algorithms to label the usual environment, and apply the method of inversion and expansion sampling for the monitoring of tourism flows. Finally, we present a preliminary determination of the feasible radius for the two types of usual spaces by comparing the operational parameters of different scanning radii in the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The findings reveal that the scanning radius should be restricted to less than 1 km for the optimal DBSCAN clustering of two usual spaces, as this will minimize the positional noise interference that leads to a mean shift. Moreover, there is typically no more than a single usual residential space, and the number of usual locations relating to a person’s place of employment or education is generally only one or two. Based on the attenuation of location points, the usual environment for a place of residence has a maximum radius of 40 km, whereas that for a place of employment or education has a maximum radius of 2 km~3 km. An inference about the statistical population is reached by expanding the sample space to include representative user travelling rates or arrival rates, rather than labeling with a full sample of location data. In addition, it can be assumed that users who are unable to identify their usual environments have the same travelling or arrival rates as those who are able to specify their usual locations, which is consistent with the assumption of homogeneity. The findings of this study serve as a reference for the standardized and consistent application of big data in tourism statistics, and reinforce the basis for big data-based research on tourism flows. Several significant policy and practical implications can be determined from these findings.

**Keywords:** usual environment; technical definition; space marking

[责任编辑:吴巧红;责任校对:周小芳]