

基于位置大数据的青藏高原人类活动时空模式

许珺¹,徐阳^{1,2},胡蕾^{1,2},王振波³

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室,北京 100101;
2. 中国科学院大学资源与环境学院,北京 100049; 3. 中国科学院地理科学与资源研究所
中国科学院区域可持续发展分析与模拟院重点实验室,北京 100101)

摘要: 随着青藏高原城镇化和旅游的发展,人类活动强度不断加剧,有必要对青藏高原不同类型的人类活动时空模式进行研究,以便了解不同区域人类活动对自然环境的影响以及更好地采取生态环境保护措施。利用腾讯位置大数据,分析青藏高原在旅游淡季和旅游旺季的人类活动时空模式,研究旅游人口对青藏高原人类活动的影响。本文分别选取2018年1月和7月的腾讯定位请求数据,构建多时空维度的高阶张量,采用张量分解的方法对数据进行降维,获取旅游淡季和旺季不同时空维度的主要特征,分析活动模式,发现青藏高原在旅游淡季和旺季的活动模式在时间和空间上都有很大不同,体现了旅游人口造成的影响。在旅游淡季,青藏高原居民的活动在凌晨有个高峰,在午间有个低谷,并且表现有特殊的节日模式,工作日的人类活动主要分布在城镇等人口聚集区,休息日向城镇周边发散;旺季凌晨活动分为两个高峰,一个在时间上较淡季提前,与午夜活动相连,另一个较淡季的时间推迟,旺季活动没有午间的低谷和特殊的节日模式,且工作日和休息日的日间活动都分布较广。

关键词: 青藏高原;人类活动;淡旺季;大数据;张量分解

DOI: 10.11821/dlxb202007006

1 引言

随着青藏高原城镇化的发展和人口增长,人类活动强度不断加剧。据统计,1990—2010年青藏高原的人类活动强度增加了约30%^[1]。快速的城镇化过程和人口集聚带动了当地经济发展,但也为脆弱的高原生态环境带来巨大的压力,并对当地文化系统造成冲击^[2-4]。不同的人类活动强度和类型作用于不同脆弱性等级的高原生态系统之中,将会带来不同的生态环境效应,因此为了揭示不同时空维度人类活动对青藏高原的影响,亟需研究青藏高原地区人类活动的时空模式和分布规律。此外,旅游业是青藏高原地区经济支柱型产业,2000年以来,青藏高原旅游人口年均增长率高达25.31%,2017年共有5100万游客登上青藏高原。因此,除了当地人口的活动,旅游人口的活动不容小觑。旅游人口的活动和当地居民的活动不同,导致的环境影响也不同,将旅游人口的活动与当地人口的活动区分开,对全面评价人类活动的生态环境影响具有正面意义。

多种类型位置大数据的获取,例如公交车智慧卡、手机数据、社交媒体数据、出租车轨迹数据、POI数据等,为研究人类活动提供了多样的数据源,在很大程度上改善了

收稿日期: 2019-03-29; 修订日期: 2020-04-20

基金项目: 中国科学院战略性先导科技专项(XDA20040401); 国家自然科学基金项目(41771477) [Foundation: Strategic Priority Research Program of Chinese Academy of Sciences, No.XDA20040401; National Natural Science Foundation of China, No.41771477]

作者简介: 许珺(1972-),女,博士,副研究员,中国地理学会会员(S110007304M),主要从事地理空间认知、知识表达、空间数据挖掘研究。E-mail: xujun@lreis.ac.cn

人类活动研究的手段^[5-6]。Sorokin等^[7]曾提出人类的社会活动具有一定的韵律, 近期利用大数据的研究也揭示了城市人类活动空间规律和时间律动^[8-10]。正是这种时空律动为发现人的行为和城市空间功能提供了可能, 为分析人类活动模式提供了依据, 使人们可以从时空大数据中挖掘出城市空间功能和土地利用类型^[11-12]。为了从高维度大数据中发现模式, Sun等^[13]使用张量分解模型对数据进行降维, 从新加坡城市公交车数据中发现城市区域模式; Wang等^[14]利用张量分解模型对北京市出租车轨迹数据进行分析, 揭示城市区域模式。此外, Zhi等^[15]提出一种对高维数据低秩近似的方法, 从社交媒体数据中提取潜在时空活动模式。这些研究说明利用合理的降维手段能够对大数据进行有效分析。

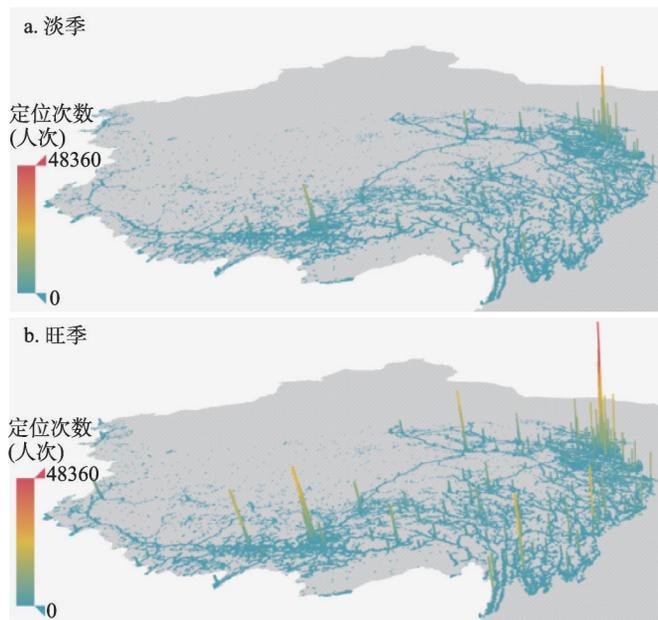
鉴于大数据在行为模式挖掘中的有效应用, 本文采用从腾讯位置大数据 (<https://heat.qq.com>) 抓取的定位请求数据分析青藏高原人类活动的时空模式。一定区域内的定位请求次数与区域内的人数有相关关系, 能够在一定程度上反映区域内人群聚集程度^[16], 而人群聚集程度的变化体现了人类活动的时空模式。本文在对西藏的拉萨、日喀则、那曲、昌都以及青海的西宁、海西、海东、海南、黄南、玉树等地的调研中发现青藏高原居民普遍使用智能手机, 手机定位数据能够反映青藏高原的人群聚集特征, 用大数据研究青藏高原人类活动切实可行。因此本文采用腾讯位置大数据作为人群时空分布的依据, 通过分析人类活动时空模式了解青藏高原人类活动。为了发掘旅游群体对青藏高原总体人类活动时空模式的改变和不同时间段人群的时空活动特征, 本文选择旅游淡季和旺季的数据进行对比分析, 以淡季的活动作为青藏高原人类活动的本底, 旺季的活动作为被旅游人口改变的活动模式, 从而区分当地居民和旅游人口活动的时空模式。张量能以高阶形式表达数据的多方面特征, 并且能通过张量分解对数据进行有效降维^[17], 因此本文对定位数据构建张量, 通过张量分解的方法获取以日、小时和空间格网为单位的特征模式。

2 研究区和数据

研究区为中国境内西南部的青藏高原区域, 西起帕米尔高原, 东至横断山脉, 南自喜马拉雅山脉南缘, 北迄昆仑山—祁连山北侧, 东西长约2945 km, 南北宽约1532 km, 包括青海省和西藏自治区大部分以及新疆、甘肃、四川和云南的小部分地区^[18]。作为世界第三极, 青藏高原的隆起造成了地域上的隔绝, 既形成了特殊的气候和地理环境, 也造就了独特的民族文化风情。另外, 青藏高原地广人稀, 城镇化水平低^[2], 人类活动与中东部地区有较大差异。

腾讯位置大数据实时记录全球每 $0.01^{\circ} \times 0.01^{\circ}$ 经纬网内通过腾讯软件定位的次数。我们选择1月作为旅游淡季, 7月作为旅游旺季。从定位数据中提取出青藏高原研究区范围内的数据, 选择包含青藏高原淡旺季各7天的完整定位数据, 其中淡季为2018年1月16日—22日, 旺季为2018年7月10日—16日, 分别对应于周二至下周一。本文将研究区内的定位数据按 $0.01^{\circ} \times 0.01^{\circ}$ 为单元建立栅格格网, 获取的定位数据即为每个格网的定位数据, 通过计算得到每个格网的日平均定位数(图1)。淡季和旺季的定位数量有较大差距, 但是空间分布趋势大体一致, 主要集中在青藏高原东部人口密集的城镇和道路附近, 西部广大的无人区定位数据非常稀疏。

图2显示了淡季和旺季青藏高原范围内每天定位总人次的对比, 旺季每天的定位请求次数大致为淡季的两倍。图3为淡季和旺季一天中24小时的平均定位请求数量, 图中显示淡季和旺季的定位数量在一日内的分布有非常大的差异, 淡季在清晨5点、6点和中



注：地图调用百度地图js API用Echarts三维显示。

图1 青藏高原淡季和旺季日平均定位数空间差异的三维显示

Fig. 1 3D display of average daily positioning requests' distribution in off-season and peak season of tourism

午12点分别有个高峰，而旺季的定位数在清晨是低谷，日间不断攀升，在夜间达到最高值。鉴于淡季和旺季数据分布的差异，本文对旺季和淡季数据分别构建张量进行分析。

3 研究方法和流程

3.1 张量分解

本文采用张量分解的方法挖掘青藏高原人类活动的时空模式。 X 张量是一种多维数组，不同维度可以具有不同的坐标系。一阶张量就是向量，二阶张量是矩阵，三阶及三阶以上的张量统称为高阶张量。张量分解可以看作奇异值分解和主成分分析的高阶推广，常见的为Tucker分解^[17]。Tucker分解是一种高阶主成分分析，它把一个张量分解为一个核心张量沿每一个模乘上一个因子矩阵，每个模上的因子矩阵称为张量在每个模上的基矩阵或主成分。以三阶张量 $\tilde{X} \in R^{I \times J \times K}$ 为例，图4表示对其进行Tucker分解的结果，分解为一个核心张量和3个因子矩阵，分解模数为 $P \times Q \times U$ ，其中核心张量 $G \in R^{P \times Q \times U}$ ，记录各模式之间的关系； $A \in R^{I \times P}$ 、 $B \in R^{J \times Q}$ 、 $C \in R^{K \times U}$ 为3个因子矩阵。压缩之后核心张量的存储空间明显小于原始张量的存储空间，从而达到对高维数据降维的目的，起到数据压缩与特征提取的作用。核心张量表示因子矩阵之间的相互作用或联系强度， \tilde{X} 是由核心张量与3个因子矩阵重新构建的张量（公式(1)），是原始张量 X 的近似张量， X 与 \tilde{X} 的差异可用来衡量近似张量 \tilde{X} 的质量：

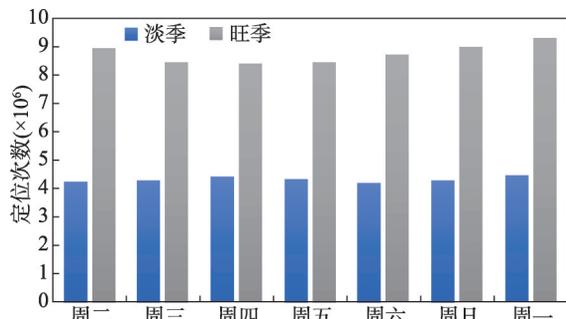


图2 青藏高原淡季和旺季每日定位总人次

Fig. 2 Daily positioning request numbers in off-season and peak season of tourism

$$X \approx \tilde{X} = G \times A \times B \times C \quad (1)$$

给定一个张量 X , 张量分解解决的问题就是求解其核心张量 G 及对应的因子矩阵, 目的是用核心张量和因子矩阵乘积组成的近似张量 \tilde{X} 替代原始张量 X 来降维。本文采用 Mørup 等^[19] 提出的稀疏非负 Tucker 分解模型 SN-TUCKER 求解。

3.2 方法流程

方法流程如图 5 所示。首先, 对数据进行预处理, 根据矢量边界提取数据、建立格网区域、剔除异常数据; 接着以日、小时和空间格网为单位划分数据, 构建三阶张量, 然后运用 Tucker 模型, 选择合适的参数和规则, 对张量分解求解, 获取数据在日、小时和空间维度上的因子矩阵; 最后通过因子矩阵和核心张量分析人类活动的时空特征, 并对空间因子矩阵进行聚类得到活动模式相似的空间格网集合, 属于同一区域模式的格网中的定位数在时间上具有相似的变化规律。以下详细介绍张量的构建方法和张量分解的参数选择过程。

3.2.1 张量构建 本实验使用淡季和旺季各 7 d 的定位数据, 分别构造数据张量 $X=(\text{定位点、时段、日})$, 每天划分为 24 h, 青藏高原有定位点的格网为 90694 个, 分别构建淡季和旺季 $90694 \times 24 \times 7$ 的张量, 即 7 张日切面, 每张切面共有 90694 行 24 列, 其中每一张垂直切面表示一天; 垂直切面的每一行表示一个定位点, 每一列表示 24 h 中的一个时段, 其中 1 时段对应 0—1 点, 依次类推 24 时段对应 23—24 点; 垂直切面中格网元素值表示该定位点在对应时刻的定位数量。

3.2.2 参数选择 张量分解中的重要问题是选择合适的参数求解其核心张量 G 及因子矩阵。因子矩阵包括区域模式矩阵 A 、时段模式矩阵 B 和模式矩阵 C 。核心张量中的值显示不同模式之间的相关程度, 一个模数较高的核心张量也许能很好地拟合样本数据, 但同时也会带来过拟合的问题, 而核心张量模数过小则会造成各类别模式冗杂在一起, 无法区分有差异的类别模式。为了使区

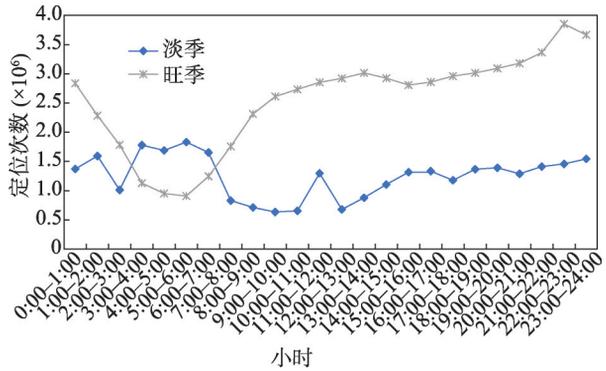


图 3 青藏高原淡季和旺季分时段定位人次
Fig. 3 Hourly positioning request numbers in off-season and peak season of tourism

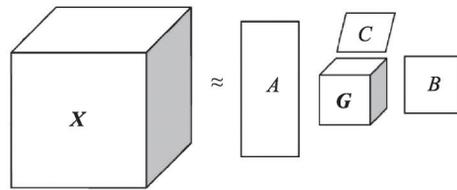


图 4 张量的 Tucker 分解模型
Fig. 4 Tucker model of tensor decomposition

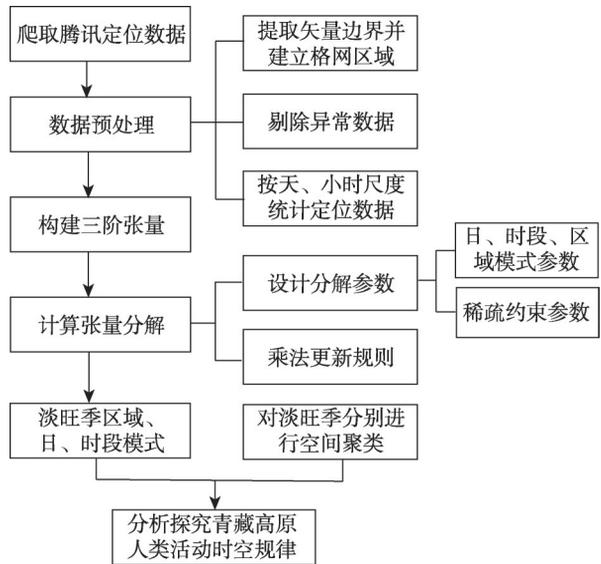


图 5 研究流程
Fig. 5 Flowchart of the research

域模式划分清晰、时段模式分布合理且日模式明显，本文对模数和稀疏约束参数进行多组实验，分别设置区域模式模数 $S \in [3, 15]$ ，时段模数 $T \in [3, 6]$ ，日模数 $D \in [2, 3]$ ，共计24组实验，模数选择范围如表1所示。由于实验变量较多，采取控制变量的方法进行分析，根据目标函数和模式组合表现来确定3个因子矩阵模数的最终取值。

表1 张量分解模数选择范围
Tab. 1 Parameter selection for tucker decomposition

参数	参数含义	参数范围
S	区域模式特征模数	[3, 15]
T	时段模式特征模数	[3, 6]
D	日模式特征模数	[2, 3]

图6和7分别是淡季和旺季数据张量分解在不同模式组合下的损失目标函数。由图6可知，淡季在日模数为2、时段模数取3和4时和日模数为3、时段模数取3时，目标函数能够较快趋于平稳，同时各模式组合对应有清晰的分解结果。通过比较目标函数大小，选取淡季的分解模数为 $9 \times 3 \times 3$ ，但由于分解结果中有一类空间模式几乎为空，所以最终确定为 $8 \times 3 \times 3$ 。由图7可知，旺季在日模数为2、时段模数取3和4时和日模数为3、时段模数取3时，目标函数能够较快趋于平稳，同时各模式组合对应有清晰的分解结果，通过对这几组模数分解结果的可视化，旺季的分解模数最终确定为 $8 \times 4 \times 2$ 。

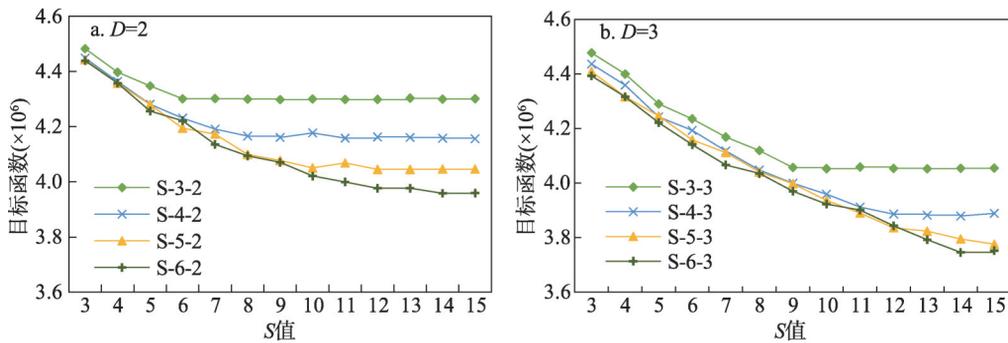


图6 淡季数据张量分解目标函数与S值

Fig. 6 Cost function values of different S-T-D compositions for dataset in off-season of tourism

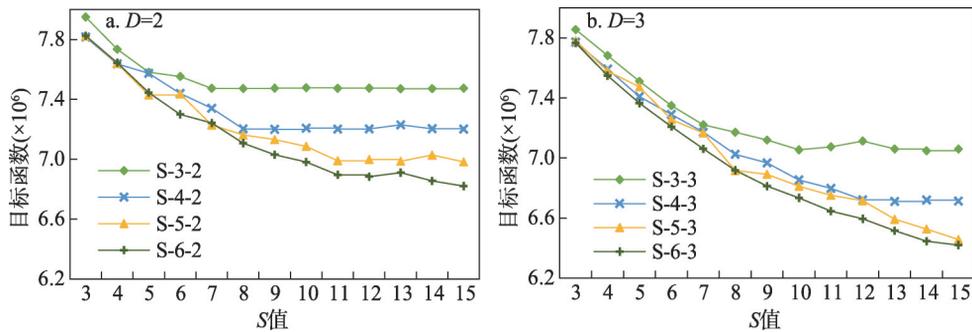


图7 旺季数据张量分解目标函数与S值

Fig. 7 Cost function values of different S-T-D compositions for dataset in peak season of tourism

4 结果与分析

通过Tucker分解，将淡季定位数据的张量分解为一个核心张量 $G \in R^{8 \times 3 \times 3}$ 和3个因子矩阵 $A \in R^{90694 \times 8}$ 、 $B \in R^{24 \times 3}$ 、 $C \in R^{7 \times 3}$ 的乘积形式，将旺季定位数据的张量分解为一个核

心张量 $G \in R^{8 \times 4 \times 2}$ 和 3 个因子矩阵 $A \in R^{90694 \times 8}$ 、 $B \in R^{24 \times 4}$ 、 $C \in R^{7 \times 2}$ 的乘积形式。 A 、 B 、 C 可被认为是每一特征维度上的主成分, 即空间维度、时段维度和日维度的主成分, 对应的模式值越大意味着相关程度越高, 下面分别对各特征模式的结果进行分析。

4.1 日模式

淡季和旺季的日模式因子矩阵如图 8 所示。图 8a 为淡季日模式, 对应的日因子矩阵为 $C \in R^{7 \times 3}$ 。淡季的日模式有 3 个, D1 模式对应周三至周五, D2 模式对应周六至周一, 1 月 16 日这天单独对应一个特殊的模式 D3。藏历显示, 2018 年 1 月 16 日 (阴历 30 日) 为释迦牟尼佛节日, 藏民多去寺院朝拜、转古拉 (绕寺、绕山、绕湖)、煨桑等户外活动, 因此活动模式与平常不同。图 8b 为旺季日模式, 对应的日因子矩阵为 $C \in R^{7 \times 2}$ 。旺季的日模式有 2 个, 分别为周二至周五对应的 D1 模式和周六至周一对应的 D2 模式。与通常所了解的工作日和周末的活动模式不同, 青藏高原的人类活动呈现周一与周末相同的模式, 而周二至周五为另一种模式。本文将淡、旺季的 D1 模式称为周中部模式, D2 称为周头尾模式, 称 D3 为特殊日模式。旺季的 7 月 13 日也是释迦牟尼佛节日, 但是没有表现出特殊的活动模式, 与周中部模式相似。

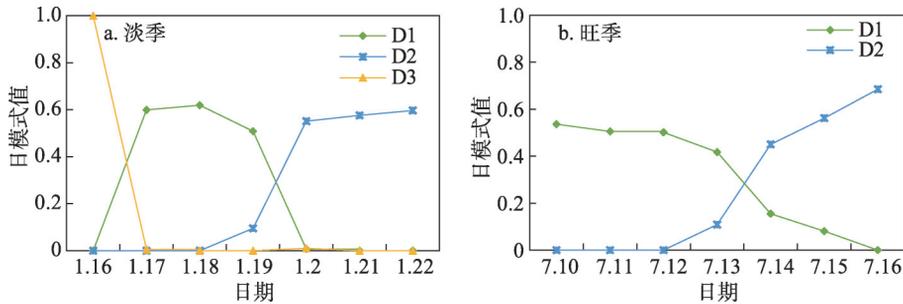


图 8 淡季和旺季日模式分布

Fig. 8 Distributions of daily patterns in off-season and peak season of tourism

4.2 时段模式

淡季和旺季的时段模式分布如图 9 所示。图 9a 为淡季的时段模式分布, 对应的时段因子矩阵为 $B \in R^{24 \times 3}$ 。淡季的时段模式有 3 个, T1 在凌晨阶段较为活跃; T2 从 8:00—16:00 期间处于高值状态, 午间有个低谷, 对应为日间的活动, 而且日间活动中间有午休; T3 在 17:00—24:00 为高值, 并且在正午有个小高峰, 填补了午休时间的活动, 这些时间通常对应的是饭点和休闲娱乐时间, 可归并为休闲活动。T1 凌晨模式是青藏高原旅游淡季一个非常显著的活动模式。中国城市的人类活动都显示在 0:00—6:00 是个活动低谷, 从 6:00 点以后逐渐增高并在日间维持高值^[20-21]。而青藏高原旅游淡季在凌晨有个活动高峰, 甚至比旅游旺季的活动强度还大 (图 3), 代表了青藏高原居民的一种特殊活动方式。图 9b 为旺季时段模式分布, 对应的时段因子矩阵为 $B \in R^{24 \times 4}$ 。旺季的时段模式有 4 个, T1 在 0:00—03:00 较为活跃, 为凌晨活动; T2 在 07:00—10:00 处于高值状态, 对应为清晨的活动; T3 在 10:00—20:00 处于高值, 对应为日间活动; T4 在 20:00—24:00 达到高值, 对应为晚间活动。相比于淡季, 旺季凌晨活动的峰值提前, 并随后在 07:00—10:00 有另一个高峰模式, 旺季日间的模式没有正午的低谷, 分布比较连续, 且时间上比淡季推迟, 旺季的晚间活动时间也比淡季的晚间活动推迟, 但是峰值较高。

4.3 空间模式

分解结果中淡季空间因子矩阵 $A \in R^{90694 \times 8}$ 和旺季空间因子矩阵 $A \in R^{90694 \times 8}$ 分别表示在

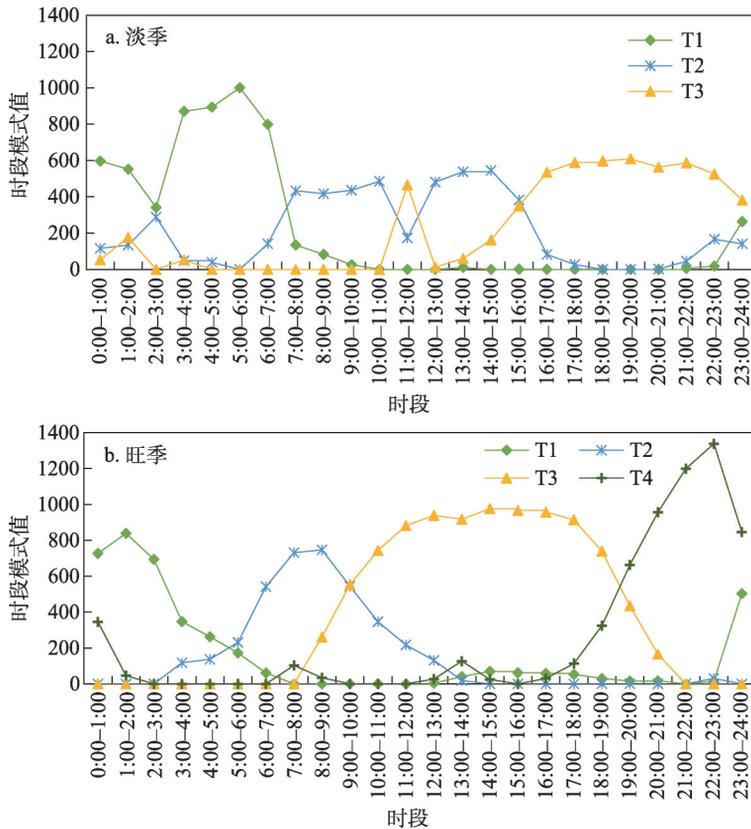


图9 淡季和旺季时段模式分布

Fig. 9 Distributions of hourly patterns in off-season and peak season of tourism

淡季和旺季研究区域的空间格网分别对应有8种定位模式，因子矩阵中每一行的8个模式值代表每个格网与这8种定位模式的相关程度。格网对应的每个模式值表示该格网在模式中的贡献，模式值越大则贡献越大，说明该模式在格网中越突出。由于每个格网不是只有一种空间模式，而是多种空间模式的混合，不同活动类型区域的空间模式混合方式不同。根据归一化后每个格网中8种空间模式的贡献率，将空间模式组合相似的格网聚类，得到人类活动相似的空间聚类。本文采用k-means聚类方法，根据Calinski-Harabasz (CH) 指标将淡季的空间模式聚为11类，旺季的空间模式聚为10类（图10）。从图10中各类的聚类中心可以看出，除了8种单一空间模式的类别S1-8，淡季还有分别以S2和S5为主且包括其它空间模式的类别S2+和S5+，以及多种空间模式混合的类别，旺季有以S5为主的类别S5+和多种空间模式混合的类别。图11显示了淡季和旺季空间模式聚类的分布，可见无论是淡季还是旺季，大部分区域都是多种空间模式混合的类别。

4.4 青藏高原人类活动分析

张量分解的结果显示青藏高原人类活动具有一定的时空规律，而核心张量反映了各个维度不同模式之间的联系强度。通过不同维度的人类活动模式和核心张量所反映的不同维度之间的关系，可以进一步推断活动的内容。

表2显示淡季8种定位模式、3个时段模式和3个日模式之间的联系强度，例如S1空间模式与D1日模式的T1时段和T2时段有较强联系，即工作日的凌晨和日间活动较强，休息日活动较弱，则S1很可能是与工作有关的场所，但如前文所述，T1是青藏高原凌晨

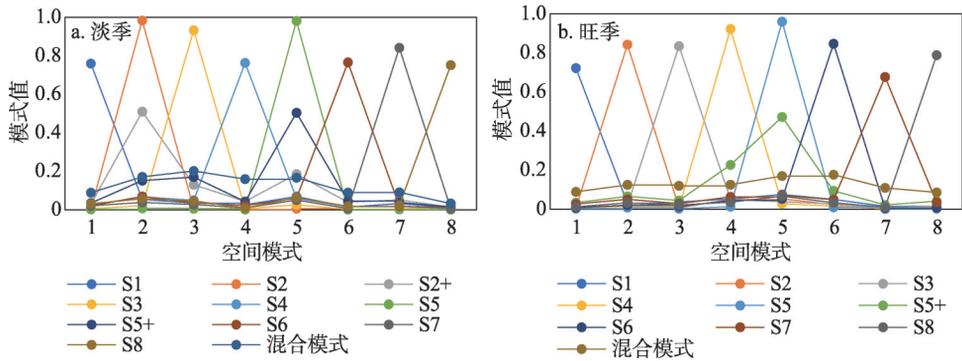


图 10 淡季和旺季空间模式聚类中心

Fig. 10 Cluster centroids of spatial patterns in off-season and peak season of tourism

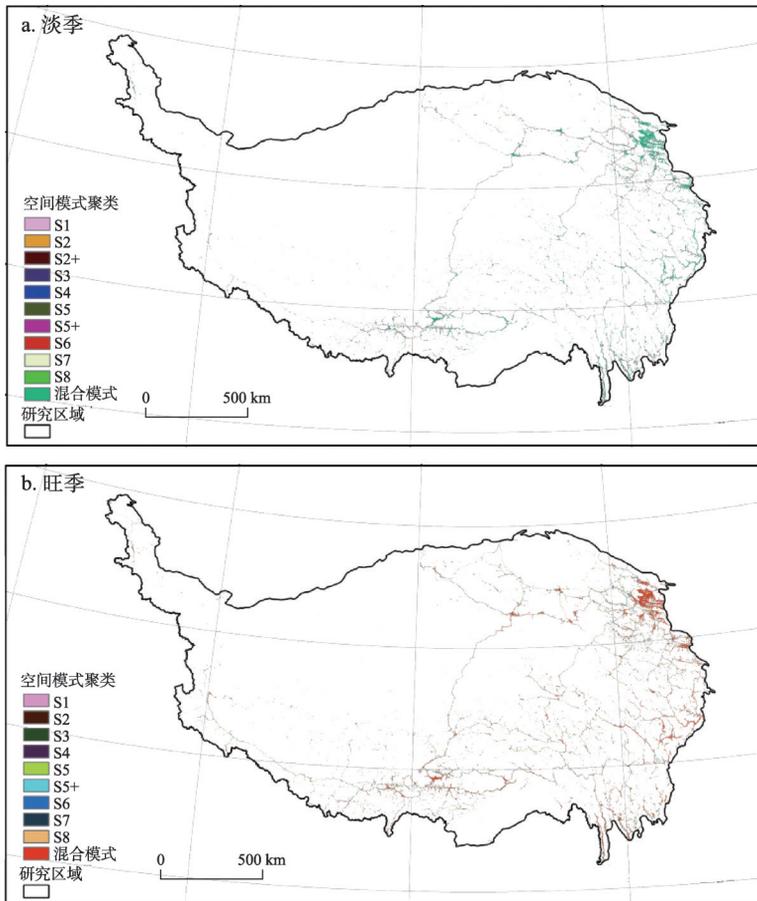


图 11 淡季和旺季空间模式聚类

Fig. 11 Clusters of spatial patterns in off-season and peak season of tourism

时段的一种特殊活动,是与宗教转经活动有关还是与工作有关,值得进一步研究;S5与D2日模式的T3时段有较强联系,即周头尾的休闲时段活动较强,S5是休闲场所的可能性很大;S1和S3与T1时段有较强联系,则在清晨活动较强,S6、S4、S8则分别与D3的T1、T2、T3时段有较强联系,它们是与特殊日活动有关的场所。

从表2可以计算出特定时段模式在不同日子的发生概率。清晨T1时段的活动在D1和D2的概率较大,而在特殊节日D3的概率较小。但是从对应的空间模式上看,D1和D2的清晨时段活动分布范围小且大多散布在居民点内部或附近,但是D2的清晨时段活动更多沿旅游路线分布;D3的清晨活动比较密集,在居民点内部大面积连片分布,并且沿交通路线附近也较多,说明特殊日的活动对居民生活影响很大。日间T2时段的活动在D3的概率较高,且在分布面积较大,主要在大居民点附近,说明特殊日的活动主要是日间活动,因此S4是特殊日佛教活动的主要场所,活动多聚集在大居民点附近。D1的日间活动分布范围小,较为零散,且与D1清晨的活动分布范围接近,而D2的日间活动比清晨的活动分布更向城市周边发散或沿旅游路线分布,说明D2的日间活动偏向出游,D1的日间活动偏向工作。晚间T3时段的活动在不同日子的概率相差不大,但是从分布上来说,D1和D2的晚间活动分布较广,且多有重合,而D3的晚间活动分布比较零散,且与D1和D2的晚间活动区域不重合,说明D3的晚间活动较弱,且与平日的晚间活动不同。

表3显示旺季8种定位模式、4个时段模式和2个日模式之间的联系强度。凌晨T1时段的活动在D1的概率比D2大,但是T1时段活动在D2的分布范围略大,且与D1多有重合。T1和T2时段的活动分布有地区差异,拉萨的T1时段活动分布较均匀,T2时段的活动分布非常分散,而西宁T1时段活动分布较分散,T2时段的活动分布较广。日间和晚间的活动在D2的概率比D1大。西宁人口总量和建成区面积分别是拉萨的2.6倍和1.5倍,西宁基本形成了“两个中心,八个片区”的复杂带状组团式城市结构,而拉萨只有“两岸三区”的简单组团结构。因此,通过腾讯位置大数据显示的人类活动模式,表现出

表2 淡季张量分解的核心张量

Tab. 2 The core tensor of dataset in off-season of tourism

	T1			T2			T3		
	D1	D2	D3	D1	D2	D3	D1	D2	D3
S1	0.1231	0.0000	0.0000	0.2751	0.0000	0.0000	0.0000	0.0000	0.0000
S2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3983	0.0000	0.0000
S3	0.2224	0.2617	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
S4	0.0000	0.0000	0.0000	0.0000	0.0000	0.3775	0.0000	0.0000	0.0000
S5	0.0000	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000	0.4192	0.0000
S6	0.0005	0.0000	0.2564	0.0000	0.0000	0.0000	0.0003	0.0012	0.0000
S7	0.0000	0.0958	0.0000	0.0000	0.2938	0.0429	0.0000	0.0000	0.0000
S8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3913

注:有底色数值表示模式与时段之间有较强联系。

表3 旺季张量分解的核心张量

Tab. 3 The core tensor of dataset in peak season of tourism

	T1		T2		T3		T4	
	D1	D2	D1	D2	D1	D2	D1	D2
S1	0.4005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
S2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3904
S3	0.0000	0.0000	0.3422	0.0000	0.0000	0.0000	0.0000	0.0000
S4	0.0000	0.0000	0.0000	0.0000	0.0000	0.3618	0.0000	0.0000
S5	0.0000	0.0000	0.0000	0.0000	0.2748	0.0000	0.0000	0.0000
S6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.3493	0.0000
S7	0.0000	0.3427	0.0000	0.0000	0.0030	0.0001	0.0062	0.0000
S8	0.0000	0.0000	0.0000	0.3524	0.0000	0.0000	0.0000	0.0000

注:有底色数值表示模式与时段之间有较强联系。

同时段差异性的人类活动空间分布规律。与淡季工作日日间活动的零星分布相比,旺季无论是工作日还是休息日,都有大面积的人类活动,说明旅游人口对青藏高原旺季的人类活动空间分布影响很大。与淡季晚间活动主要分布在城市内部不同,旺季晚间时段的活动更多分布在城市周边,反映了旅游人口与本地人口晚间休闲活动的不同。

5 结论与讨论

本文利用腾讯位置大数据,研究青藏高原旅游淡季和旺季的人类活动时空模式。定位数量表现出青藏高原的人类活动强度在旺季大于淡季,在空间上淡季和旺季定位相差较大的区域也是原本人类活动密集的地区。总体来说,青藏高原在旅游淡季和旺季的活动模式在时间和空间上都有很大不同,反映了旅游人口造成的影响。通过构建张量和运用张量分解的方法,得到青藏高原旅游淡季和旺季的人类活动时空模式:

(1) 青藏高原的人类活动模式与通常所了解的工作日和周末模式不同,呈现出周末尾模式和周中部模式,并且具有特殊节日模式。这一方面反映了青藏高原生活节奏较慢,周一的人类活动很大程度上延续了周末的休闲状态;同时也说明宗教文化对青藏高原居民的活动内容和活动轨迹具有深刻的影响。

(2) 在人类活动在时段划分上,青藏高原凌晨出现人类活动高峰模式,主要源于时差因素影响。调研发现,受时差影响,拉萨和西宁等大城市市民睡眠普遍较晚,且夜生活丰富,居民一般通过唱歌、喝酒、打麻将等方式缓解一整天的工作劳累。

(3) 淡季和旺季的空间模式都是8种,但是单一模式的区域很少,大部分区域都是多种空间模式的混合,说明多种活动类型在空间上交互混杂。该结论表明,青藏高原人类活动受城市人口规模、空间规模和功能组团的影响较大。

(4) 通过对核心张量的分析,推断活动类型,发现旅游淡季的凌晨比较活跃,这主要是因为藏族同胞在清晨朝拜寺院、转古拉(绕寺、绕山、绕湖)已经形成了较为稳定的模式。另外,淡季工作日的人类活动主要分布在城镇等人口聚集区,休息日分布向城镇周边和旅游区发散,主要表现出当地居民工作日在城镇活动,周末在城市周边休闲的活动规律。而旅游旺季受大量游客的行为模式影响很大,日间定位活动分布较广,且工作日和休息日差别不大。

大数据获取迅速,包含大量人群的活动信息,有助于人类活动的挖掘。但是大数据的数据量大,处理难度大,需要高效的挖掘算法,并且与调查数据相比,大数据语义信息不足,需要结合土地利用、文本语义等多源数据解读更多类型的人类活动。本文的研究表明利用大数据研究人类活动的方法可行,但是由于青藏高原面积广阔,还有广大的无人区,人类活动多样,因此本文利用张量分解在大部分区域得到的都是混合模式。应当降低空间尺度,选择人类活动频繁的区域深入研究,才能得到进一步细化的人类活动类型。

参考文献(References)

- [1] Li S, Zhang Y, Wang Z, et al. Mapping human influence intensity in the Tibetan Plateau for conservation of ecological service functions. *Ecosystem Services*, 2018, 30: 276-286.
- [2] Fang Chuanglin, Li Guangdong. Particularities, gradual patterns and countermeasures of new-type urbanization in Tibet, China. *Bulletin of Chinese Academy of Sciences*, 2015, 30(3): 294-305. [方创琳,李广东. 西藏新型城镇化发展的特殊性与渐进模式及对策建议. *中国科学院院刊*, 2015, 30(3): 294-305.]
- [3] Li Guangdong, Wang Zhenbo, Liu Shenghe. Strategies and countermeasures for science and technology supporting

- Tibet and regional scientific and technological cooperation in Tibet. *Bulletin of Chinese Academy of Sciences*, 2015, 30(3): 333-341. [李广东, 王振波, 刘盛和. 西藏区域科技合作与科技援藏的战略思路及对策. 中国科学院院刊, 2015, 30(3): 333-341.]
- [4] Zhang Huiyuan. Problems and progresses in protecting ecological environment on the Qinghai-Tibetan Plateau. *Environmental Protection*, 2011(17): 20-22. [张惠远. 青藏高原区域生态环境面临的问题与保护进展. 环境保护, 2011(17): 20-22.]
- [5] Degrossi L C, de Albuquerque J P, Rocha R S, et al. A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. *Transactions in GIS*, 2018, 22(2): 542-560.
- [6] Pei Tao, Liu Yaxi, Guo Sihui, et al. Principle of big geodata mining. *Acta Geographic Sinica*, 2019, 74(3): 586-598. [裴韬, 刘亚溪, 郭思慧, 等. 地理大数据挖掘的本质. 地理学报, 2019, 74(3): 586-598.]
- [7] Sorokin P A, Merton R K. Social time: A methodological and functional analysis. *American Journal of Sociology*, 1937, 42(5): 615-629.
- [8] Demissie M G, Correia G A, Bento C. Analysis of the pattern and intensity of urban activities through aggregate cellphone usage. *Transportmetrica*, 2015, 11(6): 502-524.
- [9] Sagl G, Loidl M, Beinat E. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. *ISPRS International Journal of Geo-Information*, 2012, 1(3): 256-271.
- [10] Xu Y, Shaw S L, Zhao Z, et al. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation*, 2015, 42(4): 625-646.
- [11] Pei T, Sobolevsky S, Ratti C, et al. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 2014, 28(9): 1988-2007.
- [12] Tu W, Cao J, Yue Y, et al. Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, 2017, 31(12): 2331-2358
- [13] Sun L, Axhausen K W. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, 2016, 91: 511-524.
- [14] Wang J, Gao F, Cui P, et al. Discovering urban spatio-temporal structure from time-evolving traffic networks//Chen L, Jia Y, Sellis T, et al. *Web Technologies and Applications: 16th Asia-Pacific Web Conference*. Switzerland: Springer International Publishing, 2014: 93-104.
- [15] Zhi Y, Li H, Wang D, et al. Latent spatio-temporal activity structures: A new approach to inferring intra-urban functional regions via social media check-in data. *Geo-spatial Information Science*, 2016, 19(2): 94-105.
- [16] Ma T, Lu R, Zhao N, et al. An estimate of rural exodus in China using location-aware data. *PLoS ONE*, 2018, 13(7): e0201458. Doi: 10.1371/journal.pone.0201458.
- [17] Kolda T G, Bader B W. Tensor decompositions and applications. *SIAM Review*, 2009, 51(3): 455-500.
- [18] Zhang Yili, Li Bingyuan, Zheng Du. A discussion on the boundary and area of the Tibetan Plateau in China. *Geographical Research*, 2002, 21(1): 1-8. [张懿理, 李炳元, 郑度. 论青藏高原范围与面积. 地理研究, 2002, 21(1): 1-8.]
- [19] Mørup M, Hansen L K, Arnfred S M. Algorithms for sparse non-negative Tucker decompositions. *Neural Computation*, 2008, 28(8): 2112-2131.
- [20] Ma T, Pei T, Song C, et al. Understanding geographical patterns of a city's diurnal rhythm from aggregate data of location-aware services. *Transactions in GIS*, 2019, 23(1): 104-117.
- [21] Cai L, Xu J, Liu J, et al. Sensing multiple semantics of urban space from crowdsourcing positioning data. *Cities*, 2019, 93: 31-42.

Discovering spatio-temporal patterns of human activity on the Qinghai-Tibet Plateau based on crowdsourcing positioning data

XU Jun¹, XU Yang^{1,2}, HU Lei^{1,2}, WANG Zhenbo³

(1. State Key laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

Abstract: The activities of local people and tourists have great effects on the ecological environment on the Qinghai-Tibet Plateau. Different kinds of activities may cause different impacts on ecology and environment. To effectively protect the ecological environment, it is necessary to study the spatiotemporal patterns of different kinds of human activities. In this paper, two Tencent positioning datasets which record one-week location requests in January and July of 2018, respectively, are used to explore the human activities in off-season and peak season of tourism on the plateau. A Tucker tensor decomposition method is employed to reduce the dimension of massive data and obtain the principle modes of human activities. The data in off-season are decomposed into 3 daily patterns, 3 hourly patterns and 8 spatial patterns, and the data in peak season are decomposed into 2 daily patterns, 4 hourly patterns and 8 spatial patterns. By analyzing the core tensor, different kinds of activities are inferred through the relations among different dimensions of data, and the human activities in off-season and peak season of tourism are analyzed. The human activities on the Qinghai-Tibet Plateau are found to be different from those in other places. Different from ordinary weekday and weekend patterns, there is a mid-week pattern (Tuesday through Friday) and an inter-week pattern (Saturday, Sunday and Monday) on the Qinghai-Tibet Plateau, and there is a special holiday pattern in off-season of tourism. It is also found that the human activities in off-season and peak season of tourism are different, which indicates different activities of the local residents and the tourists. In off-season of tourism, the positioning activities are very active in the morning, however, the activities are less active during the daytime of mid-week days than during the daytime of inter-week days, and the activities are mostly found in the cities in the mid-week days but mostly in the outskirts of the cities or on the way to scenic spots in the inter-week days. In off-season, there exist the activities of local residents. In peak season, there are less activities in the morning, but the activities during the day are more broadly distributed both in the mid-week days and in the inter-week days. It is indicated that the activities of tourists are significant in the peak season. After clustering spatial grids with similar patterns, we find that there are mixed spatial patterns in most parts of the study area, which discloses that there are usually multiple kinds of human activities in a region.

Keywords: Qinghai-Tibet Plateau; human activity; off-season and peak season; big data; tensor decomposition