

城市旅游交通优化决策支持平台：基于杭州访客大数据的城市出行优化分析

【论文摘要】随着经济的发展和人们生活质量的提升，旅游出行逐渐成为人们生活的重要活动。杭州作为中国最大旅游城市之一，每年会迎来大量访客，访问目的包括休闲旅游、公务出差、探亲访友等。在其中，旅游服务的质量不仅关系到访客满意度，也与杭州市整体形象息息相关。大量的游客在旅游旺季涌入杭州，然而，杭州市在旅游交通线路规划等方面目前还存在着诸多问题，如专线巴士少、基础设施欠缺或分布不均等。本调研报告从不同类型访客时空分布角度切入，通过百万级数据的可视化并结合文献查阅、问卷发放、实地走访等调研结果寻找访客出行规律、探索内在原因、建立相关优化模型并进行验证，为城市旅游交通优化决策提供相关方案。

【关键词】旅游者，空间分布，可视化，路线优化

Urban tourism traffic optimization decision support platform: Urban travel optimization analysis based on bigdata of visitors in Hangzhou

Abstract: Coincident with the advancement of economy and the improvement of people's lives qualities, tourism becomes an important part in daily activities. Hangzhou, being the biggest tourist city in China, attracts many tourists each year. Visitors come for leisure travel, business travel, visiting family and friends and so on. The quality of tourism service not only concern to the visitor satisfaction but also to the overall image of the city. A mass of visitors crowded into the city , nevertheless, defects exist in many aspects such as traffic route planning, including the lack of shuttle buses and the maldistribution of urban infrastructure will undoubtedly affect tourist experience. This report try to find out the regularity and inner reason of visitors' getting around from the angle of different kind of visitors' spatial and temporal distribution by visualizing mega data combined with literatures, questionnaire and field visit. After analyzing the results, we will build relative

optimization model and offer advice for urban tourism traffic optimization decision.

Key words: tourist, spatial distribution, visualization, route optimization.

1. 问题的提出与研究方法

1.1 研究的背景

1.1.1 当今的环境

随着时代的发展和社会的进步，人们逐渐不再满足于物质层面的需求，越来越多的人开始追求精神层面的消费，而旅游便是其中举足轻重的一部分。一方面，经济上的宽裕刺激了旅游的需求，另一方面，科技的快速发展又助力了人们对于旅游的选择：交通工具的创造与更新，如高铁，解开了空间上的束缚；GPS 的普及应用与地图 app 的发展，方便了人们智能化出行等等。

与此同时，为跟进旅游业发展的脚步，开发商不断开发具有潜在商业价值的旅游地，各级政府也制定相关政策以适应与引导相关旅游产业，更有不少城市打造旅游品牌，以旅游产业作为当地主要的经济引擎^[1]，杭州便是其中的典型代表。杭州景点众多，现代化建设又日趋完善，然而由于往来旅客数量庞大，旅游业的飞速发展也对旅游管理行业带来了更大的挑战。

1.1.2 问题的提出

旅游的大众化以及杭城旅游热度的居高不下，不可避免地引起人们对于旅游相关部门所存在瑕疵的关注。其中，最为直观的便是部分景点（如西湖）游客流量的过度膨胀，这从历年的新闻报告中便可见一斑。随之而来的是过重的交通压力。调查结果显示，交通拥堵，尤其是连结热门景点的部分站点的拥堵，成为大部分人积极反应的问题。迅速增长的游客数量也无疑给景区的基建带来了压力，而不完善的基础设施建设则对游客及旅游业本身造成了损害。再者，杭州旅游辅助设施的开发应用并未跟上旅游业本身的发展，如旅游专线的缺乏，这也反过来加重了上述压力。针对这些问题，杭城旅游交通优化决策支持平台的搭建亟需提上日程，如何优化游客旅行路线、如何改善相关部门质量也成为亟待解决的问题。

1.1.3 该问题解决的意义与价值

游客旅游的交通路线优化不仅关系到游客自身的体验，更对杭州乃至全国的旅游业和相关产业有重大影响。科学的路线规划将会缓解城市交通压力，平衡各景点人流量，提升游客满意度，从而美化杭州的旅游城市形象。同时，城市交通

路线的优化也成为相对不热门的景点发展的契机，帮助缓解热门景区爆满的情况，使杭州游客分布全面化、均衡化。该问题的出现亦要求了更科学的交通体系设计、更完善的基础设施和配套服务建设以及更加优化的整体城市规划。

杭州作为我国旅游城市的典型代表，对杭州交通问题的分析将对国内其他城市旅游规划起到借鉴和指导作用。通过大数据可视化技术、建模等方式找到问题的突破点并加以解决，将会成为旅游产业革新的催化剂。

1.2 文献综述

对于此课题国内外已有一些研究成果，经典的旅游线路设计利用网络规划或组合优化理论，对于这方面的研究发展已较为成熟，比较有代表性的便是 Gham 的基于树模型的推荐路径算法^[2]、Ahuja 等人的网络流算法^[3]、Dantzig 等提出的 TSP 问题解决方法^[4]等；而当今在旅游线路设计方面，许多研究者对于景点的算法进行了改进，并结合景区拥堵的实际情况：吕琼艺等利用改进的 Dijkstra 算法等研究旅游线路规划^[5]，通常针对最短路问题；Ahulre 等通过推荐系统原理研究旅游线路的规划与安排^[6]；同时，在提升游客满意度方面，Oscar Claveria 等人提出了利用机器学习原理根据访客需要来进行线路推荐^[7]；再加上实际景区拥堵状况的考虑，有龙涛等通过空间聚类以及 ArcGIS 辅助可视化研究游客兴趣点以及景区密度^[8]从而做出决策，李攀等设计了数据可视化的城市路网拥堵特性描述分析方法^[9]等。另外，在城市大数据研究方面，龙瀛利用城市大数据进行城市旅游相关的定量研究^[10]，杨正洪等考量了大数据在智慧城市中的应用^[11]，秦萧等通过城市大数据时定性分析空间行为并讨论访客行为方式的不同^[12]，党安荣等基于大数据对于城市理性规划与城乡治理提出了一些新的解决方案^[13]。

1.3 数据收集方式

1.3.1 定性数据收集

为了深入了解访客的看法，从游客本身的角度出发研究旅游出行优化问题，我们制作了网络调查问卷，进行了实地走访调研并收集了定性数据，深化了对杭州旅游的认识。

问卷调查我们采用自填式问卷，包括访客出行时间地点选择、活动类型、满意度等方面，通过问卷调研，我们能够从游客角度了解城市交通，探寻景点

热度以及导致这种分布差异的内部原因，从而探索旅游交通线路及基础设施规划中存在的具体问题；访谈我们使用结构式访问，通过线下或电话咨询城市交通及旅游管理决策部门的相关人员，通过访谈了解城市中整体存在的问题，从较为宏观的角度深化我们对城市旅游线路规划的理解；实地调研中，我们访问杭州市较大景区，详实记录了景区人流、分布及基础设施相关数据，同时，在较大的商城、地铁站以及杭州火车站、机场，我们进行访客随机调查以及交通状况的观察记录，得到了相当全面与充分的定性数据。

1.3.2 定量数据收集

在定量数据收集方面，我们采用线上线下结合的方式。一个城市的访客信息往往来源于以下方面：搜索引擎、网站分析系统和现代追踪设备。随着大数据技术的发展，“调研之中样本=总体”便有实现的可能。借鉴大数据思维，我们希望利用这些数据来分析旅游者的喜好、动机以及旅游过程中的体验。因此，我们通过线下抽样调查与大规模线上数据爬取获得了大量数据，线上数据主要来源于个推公司向我们提供的经用户授权的路径点位信息，数据量为百万级别，包括访客年龄等特征、来源地、访问时间以及访问地点轨迹等；线下调研数据来源于我们在实地调查中通过相关部门以及问卷收集到的定量数据。我们按照数据的产生方式、获取方式进行分类以及相应的预处理，百万级个推轨迹数据通过游客类别的划分进行相应的处理，方便后续模型建立与研究。

1.4 研究可能的创新之处

1.4.1 研究方法的创新

除了采用观察和实验、模型演绎、计算机模拟仿真的传统研究范式外，我们引入了当下兴起的科学研究的第四范式——基于大数据的科学研究，通过对用户行为轨迹的数据分析挖掘得到有效的信息，并将其应用到旅游管理等实际运作层面。与普通的数据分析不同的是：我们采用第四范式去发现寻求隐藏在数据背后的规律，并且在关系规律的基础上，结合实际情况，如交通设施，地标建筑等，分析结论中变量之间的先关联系，将数据发现的结果运用逻辑去实证。

1.4.2 研究对象与角度的创新

查阅以往文献，我们发现过去也有不少人做过对于杭州的旅游调查以及路径优化的研究。不过这些研究一般从访客行为入手，或者聚焦于某个景点在一

定时间段的拥挤程度。我们则从另一个角度出发，直接对访客时空分布进行研究，有助于把握整体的特征和趋势。

同时，随着互联网时代的到来，杭州已渐渐从“旅游胜地”变成“互联网之城”。此次研究将旧的命题放在新的时代背景下，运用数据科学的思维对杭州城市旅游交通进行优化，并且对于以下新的方面进行了创新性思考：

1. 归纳个人行为的类似性，给出最人性化，旅游体验最佳的道路优化方案。
2. 利用访客分布图，发掘潜在服务需求
3. 改进旅游公共资源的配置以达到资源高效运用等。

1.4.3 研究技术的创新应用

在研究过程中，我们运用一些计算机技术及数理模型来解决实际问题，交叉学科的研究有助于对于实际问题更好的把握。具体技术上，对于每个人及其所在群体的数据，我们运用统计学方法，通过相关性分析将人群进行分类，并对杭州市访客时空分布结构进行分析，挖掘其中的规律，使用数据可视化技术，将分布结构清晰呈现，同时基于大数据进行建模与验证，利用机器学习等手段进行线路的预测。

除此之外，我们将可视化分析得出结论辅以实地考察、调查问卷数据进行验证，通过模型及其验证结果，提出可行方案并为杭州市旅游规划决策提供更为科学的参考，有效辅助城市出行线路管理。

2. 杭州市旅游设施与访客分布特征的统计分析

2.1 杭州市景点调研分析

2.1.1 热门景区与主题线路分析

根据实地走访、调研以及网上检索杭州市范围内的主要景点，可以将杭州的景区分为名胜古迹、自然风光、主题公园等。名胜古迹无疑是杭州的主打招牌。西湖及环西湖的众多大大小小的景点是杭州游客的首选目的地，这其中包括岳王庙、苏堤、白堤、断桥、三潭印月等著名景点。自然风光主要包含西溪国家湿地公园、千岛湖景区等自然风光优美，但人文气息相对淡薄的景区，主题公园则以还原了宋代都市风貌的宋城最为熟知。

同时我们对马蜂窝、携程等线上旅游网站进行了考察，从单个景点热度来看，西湖、千岛湖、宋城、西溪湿地等位列前茅；而从各大网站上推荐的旅游

线路来看，一般线路首先都会包含西湖景区，接着根据旅游时长的差异，再加入自然风光以及主题公园类别中的一个或两个景区，如果旅游时间足够长，还会包括博物馆、大学、森林公园等相对冷门的景区。进一步分析可以总结出，类似于“第一日西湖+第二日宋城+第三日西溪湿地及购物”或“第一日西湖+第二日千岛湖”的时长为两三天的短期旅游将会是热门线路中的热门，因为它们不仅囊括了最主要景点，还在较短时间内包含了较多类别，满足了游客“到此一游”的打卡心理。

2.1.2 各景区吸引力与游客满意度

为更深入研究游客对景区的选择，可以借助 Leiper 的思路从社会学框架角度出发对旅游吸引力进行理解：人类的需求和吸引人的地方的存在，都是由相关的信息推介形成的^[14]，它使潜在旅游需求变为真正的旅游行为，才是有“吸引力”的系统。在互联网时代，旅游吸引力在一定程度上可由互联网媒介上的相关信息所决定^[15]。

以“大众点评”网站上的数据为基础进行线上数据收集统计（表 2-1），可以总结出：传统的“热门景区”——西湖风景名胜区及其周围诸景点（包括灵隐飞来峰景区、南宋御街、清河坊、胡雪岩故居等）在评论人数和环境方面都相对较高，这些景区一般为部分开放式景区，人均费用相对较低（如果不选择乘坐游船西湖基本是免费的），性价比指数较高。“自然风光”类景区则差异较大，如：对比西溪湿地和千岛湖风景区，前者吸引力较大且人均花费更低；而“主题公园”类景区中，具有代表性的宋城，评分较高，但需要的花销也较大。

表 2-1 杭州市各景区评分数据统计

景点名	评论数	总分	环境	服务	人均花销
西湖风景名胜区	11579	8.2	7.9	7.4	88
胡雪岩故居	1619	9.1	9.0	9.0	20
灵隐飞来峰景区	5634	7.8	7.6	6.7	65
西溪国家湿地公园	6033	7.9	7.9	7.2	117
岳王庙	848	7.0	6.9	6.2	27
南宋御街	1430	7.3	7.4	7.0	61

杭州宋城景区	7027	8.8	8.4	8.0	363
清河坊	1484	6.9	6.6	6.0	90
京杭大运河	647	8.6	8.6	8.6	71
钱塘江	228	8.4	8.4	8.0	45
大明山	781	8.2	7.6	7.2	152
长乔极地海洋公园	1434	8.2	7.8	7.3	438
千岛湖风景区	923	7.9	7.5	7.1	382
浙西大峡谷	393	6.6	6.7	5.8	248

通过数据统计还可以发现，西湖、灵隐寺等“热门景区”与大运河、大明山等“冷门景区”在各项评分及花销方面较为接近，然而因为知名度差异导致来访人数差异巨大，如果有旅游线路能同时囊括满意度较高、主题相近的“热门”和“冷门”景区，将具有高质量与丰富文化特色的双重优势^[16]。

2.2 访客时空分布特征

2.2.1 时间分布特征

根据所得数据，能够得到三个维度的访客时间分布，分别为访客抵杭时间、访客离杭时间、访客在杭出行时间。

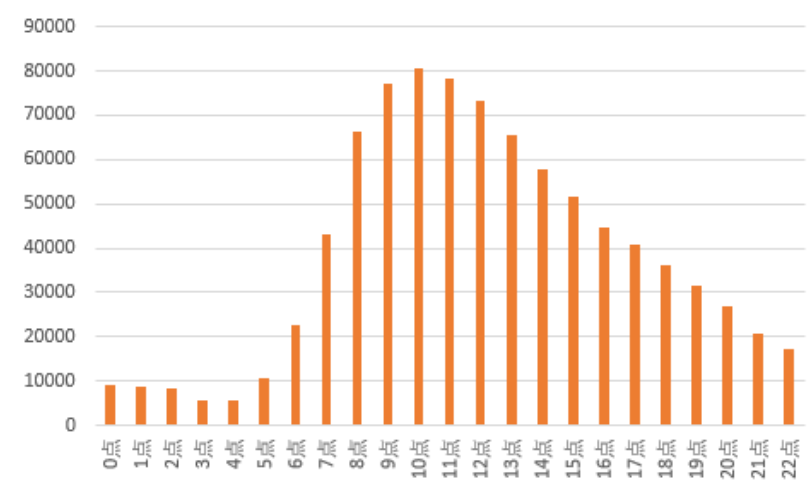


图 2-1 旅客一天中抵杭时间分布

从抵杭时间来看，访客分布最集中的是早上八点至下午三点，下午三点后的访客数量逐渐减少，趋势较为平缓。而凌晨抵杭的访客数量最少，其中凌晨三点

达到了极小值，从凌晨五点后访客数量又急剧上升。因此，可以推测 8 点至 15 点机场、高铁站及火车站的游客疏导压力较大，容易造成站内拥挤，也会连带导致站点周围的市内交通出现超负荷运转的情况。如何处理这一时间段过多的人流量便是需要注意的问题。

相对地来看离杭时间访客分布，可以发现此时人流量集中在下午一点至晚上七点。同样，凌晨时段离杭人数是最少的，从凌晨六点开始，访客数量开始上升，且增长速度逐渐加大。

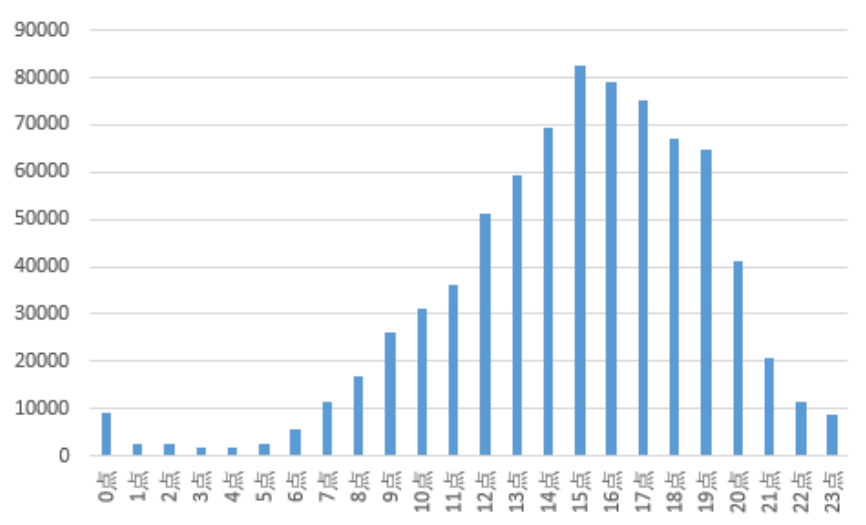


图 2-1 旅客一天中离杭时间分布

西湖每日访客人数

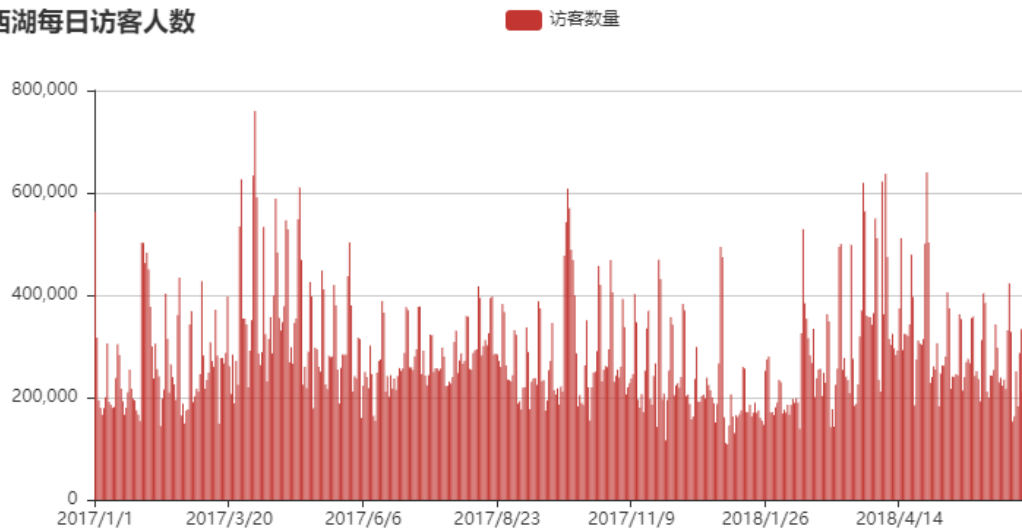


图 2-3 西湖景区访客人数变化

另外，特定景区的访客分布也有一定的规律可循，以西湖为例，可以看到，大概每七天有一个峰值，一周为一个周期。现实情况中，周末西湖人数一般都会

明显高于工作日。同时，清明、国庆假期也同样出现了峰值，这说明西湖等杭州市热门景区的拥挤现象在公众假期期间尤为明显。

2.2.2 空间分布特征

在分析空间分布特征的过程中，我们使用 Echarts 处理访客数据点，这里我们从一百万条个推访客数据中随机抽出一万五千条进行可视化处理，图 2-3 中圆点的大小代表着访问该地点的人数。

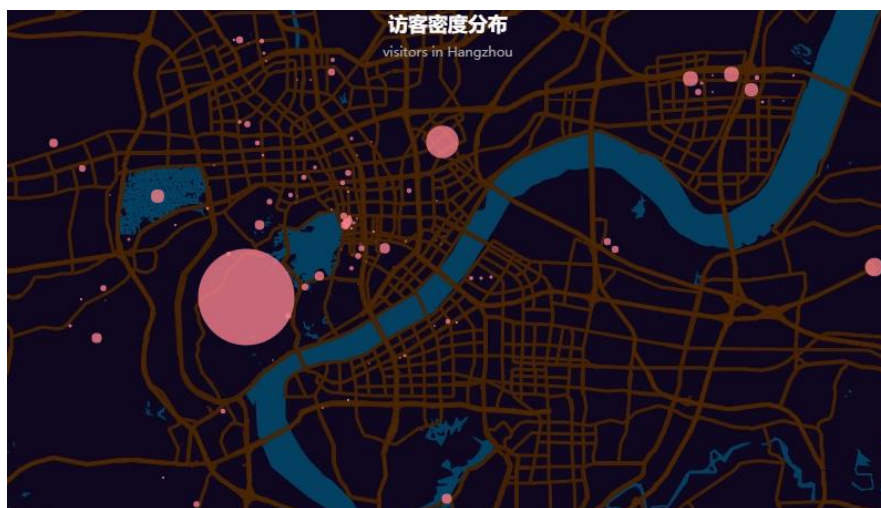


图 2-3 外地访客最常去的地点

可以看到，西湖及其周边景区、火车东站等地点聚集了大部分游客。来访人员对于地点的选择呈现出明显的集聚特点，绝大多数游客都会涌入少数几个地点，这容易造成周边道路以及景区的严重拥堵。结合地理学第一定律 (Tobler's First Law) 的观点，访客在杭州市内的分布总体上呈集聚分布，在西湖周边地区有很多空间距离相近的聚落，根据游客更倾向于选择相近的地点组合游览的特点^[19]，这些人群集聚地区正是节假日出行压力较大的地区^[20]，改善杭州出行交通等设施应该先从该类地区入手。

2.3 旅游基础设施分布分析

2.3.1 吃住购娱设施的分布

食宿一般是旅客到访后最先确定的事项，因此餐馆与酒店等基础设施的分布必然会对旅客的分布产生一定的影响。在对杭州餐馆与酒店的分布做了可视化处理之后，可以发现，杭州的餐馆与酒店基本呈块状分布，即有明显酒店数量密集且相距较远的区域，而在密集区域之外，餐馆酒店较多地倾向于沿地铁线及省道

线分布。

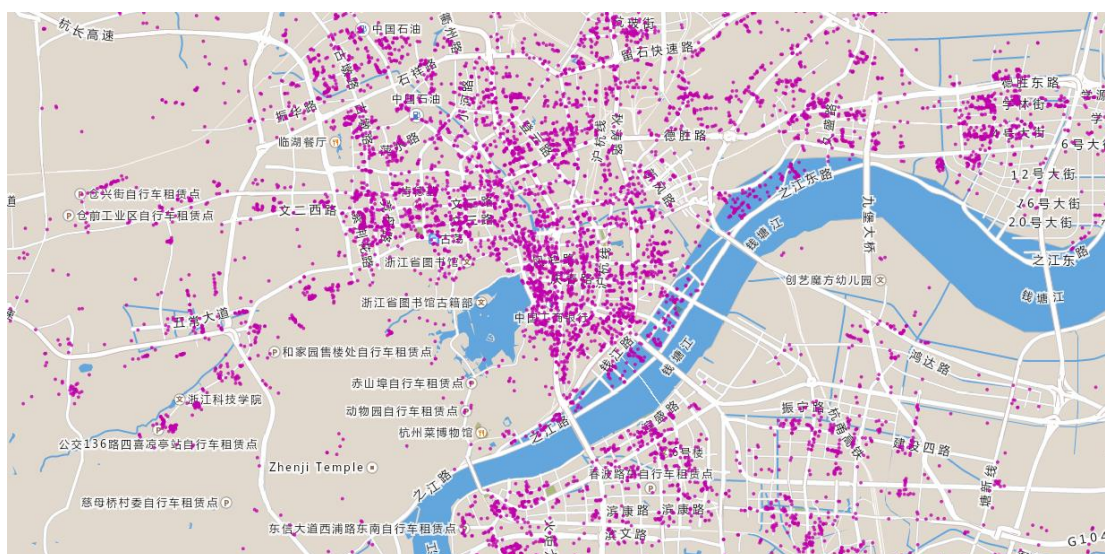


图 2-4 杭州酒店的分布

我们对于酒店数据进行了可视化显示，不难发现杭州酒店分布最为集中的区域是西湖及其周边一带，其次还有临安市太湖源西部，造成此现象的重要原因是附近有西湖、灵山、青山湖国家森林公园等景点；萧山区萧山体育中心周边也是酒楼分布的密集区域，该区域与湘湖、长乔极地海洋公园、杭州乐园相邻；地铁一号线末（下沙大学城方向）和余杭地铁战附近等交通较为便捷的地方也是酒店较为集中的区域。

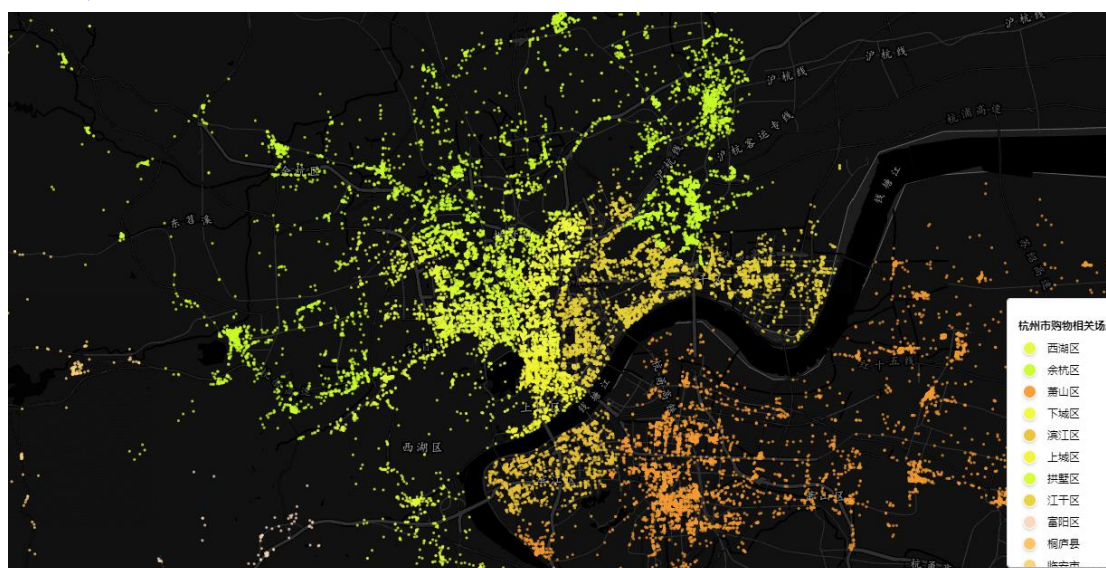


图 2-5 杭州市内购物场所分布

购物也是旅游中的重要环节之一，购物场所的分布也会影响访客线路的选

择，单位区域内购物场所数量将会影响访客密度。购物相关场所可视化分布图较明显地显示了其分布特征：钱塘江两岸、主干路附近及各区交界处是购物场所集中分布的地带。购物设施最密集的是西湖区、上城区、下城区，其中区交界处——古墩路、申花路一带尤为显著，西湖周边购物场所数量众多，其中延安路与解放路、庆春路、西湖大道的交汇处较为突出。拱墅区、余杭区、江干区、滨江区购物设施密集程度有所降低，而萧山区、临安、建德、桐庐最为稀疏，只在相关景点附近有稍多的分布。

此外，娱乐活动也是旅客选择时的考虑因素之一，在此以公园和城市广场为例进行分析。杭州公园分布最大的特征便是呈现辐射状，辐射中心依然是钱塘江尾两岸区域，其中最为密集的是西湖沿岸、上塘路高架以及钱塘江畔的白塔附近。由此为中心发散开去的有一条较为清晰的射线：上述区域与临安的连线，即省道S25、S23、S31沿线。另外，中心区域外的萧山体育中心、富春江沿岸、太湖源西面、G25S14交汇处（附近有良渚）也是公园分布较多的地方。公园相对来说自然属性较强，而城市广场类的娱乐场所社会属性占主导，因此，城市广场的分布有自己的特性：不再倾向于分布在紧靠景区的地带，而是倾向于靠近景区附近的主干道路。

从上述满足旅客一般需求的基础设施的分布描述中可以看出，基础设施的密集程度与景区数量及热门程度呈现明显正相关关系，某区域景区数量越多，越热门，往往其基础设施数量也越多。这一点，根据数据库，我们利用最小二乘回归进行了实证分析。将位于杭州拱墅区、余杭区、江干区、滨江区、上城区、下沙区的297个样本点按照城市规划的道路进行分区，统计每个区域内的景点个数和人数，将景区大致分为9个区域，令其为 $a_1, a_2, a_3 \dots a_9$ ，对每个区域的景点个数统计为 x_i ，人数总数为 $y_i, i = 1, 2, 3 \dots 9$ 。再进行回归分析，由于样本量较小，我们估计，若得到的 β 系数为正，且拟合优度 $R^2 = 1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}$ 大于0.75，则说明基础设施个数和地区热门程度呈正相关。

根据结果（代码间附录G-1）， R^2 达到了0.993， β 的值为1884.8，为正值，即在所取的样本内，景区数量和人数是有明显的正相关关系的。

但是旅游地只是影响基础设施建设的最主要因素之一，却非唯一影响因子。但完善的基础设施能够对旅客产生吸引力，拉动旅游业发展。

2.3.2 基础设施与访客分布不平衡性

从上文中的分析中可以很明显看出，旅客的分布和基础设施的分布虽然有很高的相关性，然而它们的空间分布特征仍有巨大的区别，杭州市内与旅游者息息相关的设施，如饭店、购物中心等，大多数沿着主要道路或是景区；然而游人分布聚集化程度更高，呈现出分团分块的特征。

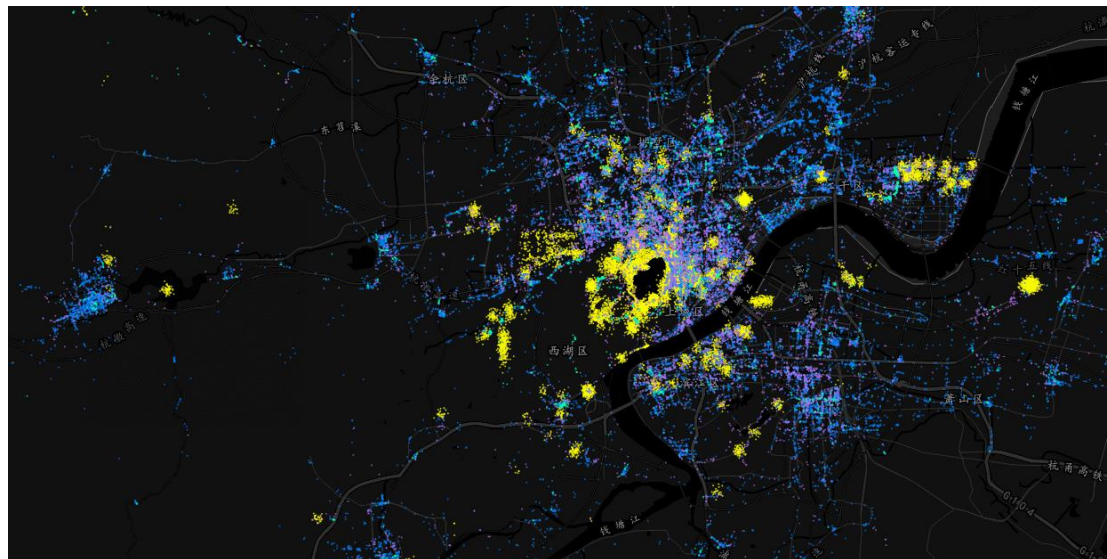


图 2-6 杭州购物场所分布

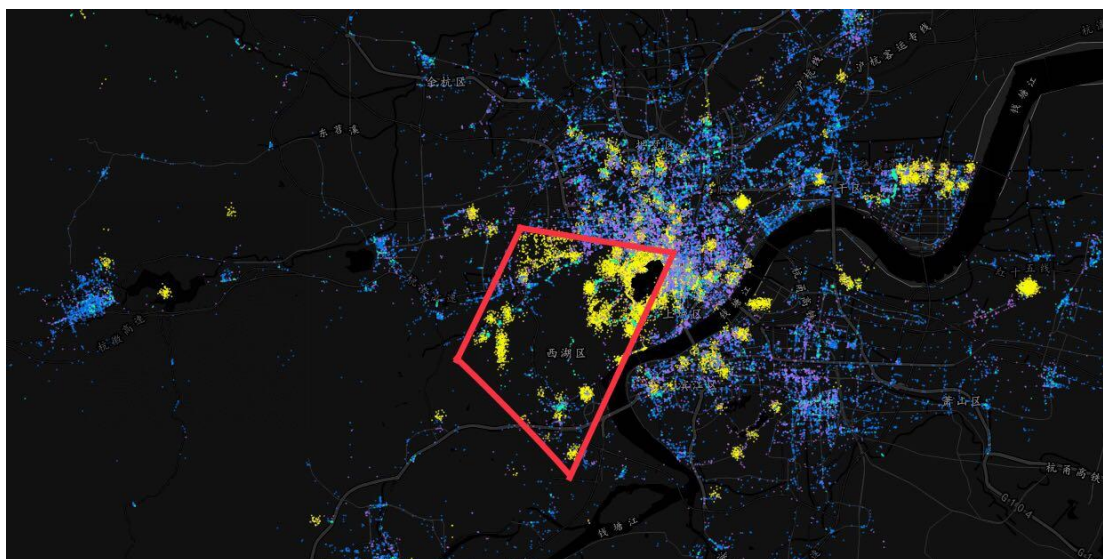


图 2-7 基础设施明显不足区域

将游人数与基础设施数据整合并在一张图上可视化后可以明显看出这种特征。图 2-6 中，黄色散点为访客，浅蓝色、深蓝色、紫色散点分别为为杭州市旅馆、购物中心和停车场所的分布，在西湖周边、凤起路等大型地铁中转站附近，游人密度远高于周边设施的的密度。这就不难从总体上解释节假日西湖

附近饭店、旅店爆满的现象了。

从图中我们可以划分出旅游设施供不应求的区域，如图 2-7，在红色多边形区域内，旅游基础设施明显不足，而这一块区域恰恰包括了西湖、西溪湿地、宋城等杭州大多数热门景区。除去中间无数据点区域(山区)，在四周游客聚集区域，对于出行交通等设施的优化势在必行。然而大多数景区已经规划成熟，很难进行基础设施的再建设，因此，优化的基本思路是从游人体验角度出发，进行新线路的设计以及开发新景点进行分流。

3. 访客出行线路优化分析及主要结论

3.1 交通规划对访客线路选择的影响

杭州市区范围内的城市快速路总体上呈现“三横二纵”的格局，“三横”连接了城西与城东地区，并与“二纵”相接驳。“二纵”向南联络城站、滨江，并直穿杭州中心城区，与西湖景区以及滨湖商圈为邻。“三横”与“二纵”作为城市长距离快速通行的主干道，对在杭游客远距离景区间的转运具有重要意义。

公交车、出租车等传统的出行方式承载了很大一部分客流量^[21]，也是多数旅客点对点出行的主要方式，主干线路快速路省事省时，然而倘若主干道拥堵，可能会迫使旅客放弃预先设计好的游玩线路，从而影响游客的满意度^[22]。

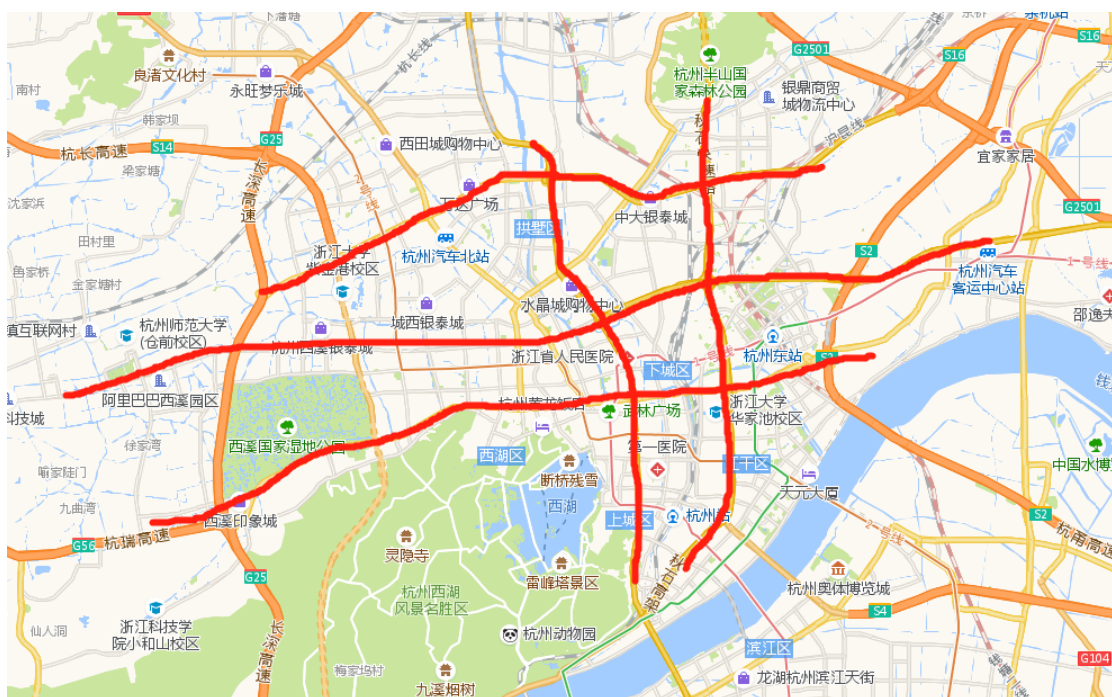


图 3-1 杭州市区主干道路

地铁也是出行的重要交通工具，一般为外来人员首选，杭州已建成 1、2、4 号三条地铁线路。三条地铁中，以连接了东站、城站以及滨湖商圈的 1 号线最为繁忙。凤起路地铁站紧邻滨湖商圈与武林商圈，同时又是两条地铁线的交汇点，其繁忙程度可想而知。然而，经过走访与调研，我们发现，对于游客来说，地铁的拥挤问题与其带来的便捷相比之下会显得相对不那么重要，许多外地游客为了能享受长时间的观光娱乐也愿意承担较短时间的拥挤的成本。不过，大多数旅客认为，地铁未连通机场是一个大问题，这对外来访客或杭州本地人出行造成了诸多不便。

共享单车、共享电单车也是访客常用的交通工具。它们可以应用在短途出行之上，如西湖某侧两公里之内的景区接驳，可以通过共享单车的方式来实现。不过，共享单车停放问题为人所诟病^[23]，节假日景区附近堆积如山的共享单车也会很大程度上影响访客体验。

在优化的基础上，这些传统出行方式也可以实现更高的运输效率。同时，可以通过设置旅游专线、旅游专车等方式拓宽游客出行方式的选择范围^[24]；也可以开发冷门景区线路来解决上述问题，具体的优化模型将在下一节提出。

3.2 出行线路优化的模型设定

3.2.1 景区热点分析

杭州市景区多为开放式景区（如西湖风景区），景区道路主要面临着两大压力：一是过境交通压力，即景区内道路承担着城市道路的功能；二是旅游交通压力，每年旅游旺季，景区旅游交通繁忙，对游客的游览感受及旅游质量造成负面影响^[25]。

为了更加细致反应不同景点的热度，我们借助核密度分析的思想对于杭州市地点人流量进行热点分析：除了关注一个点本身数值大小，还会关注周围点数值大小，将局部数值进行加总后与所有点数据总和比较，利用 Z 得分判断某地数值是否具有统计学上的显著性差异（即验证是否为随机产生的结果）。

若地点 i 的总人数为 x_i ，则地点 i 的热度为

$$H_i = \frac{\sum_j^n w_j x_j}{\sum_j^n x_j} \quad (3.2.1)$$

其中 w_j 为空间权重，将 H_i 标准化后可得到 Z 分数：

$$Z_i = \frac{H_i - E(H_i)}{\sqrt{Var(H_i)}} \quad (3.2.2)$$

若 Z 分数为数值较高，说明该地点及其附近的访客密度相对较高，该地区访客集聚分布，高值聚类紧密，为一访问热点；反之，若 Z 分数为负值且 $|Z|$ 数值较大，说明该地访客分布稀疏，即访客空间集聚现象较弱或此地较少访客来访。

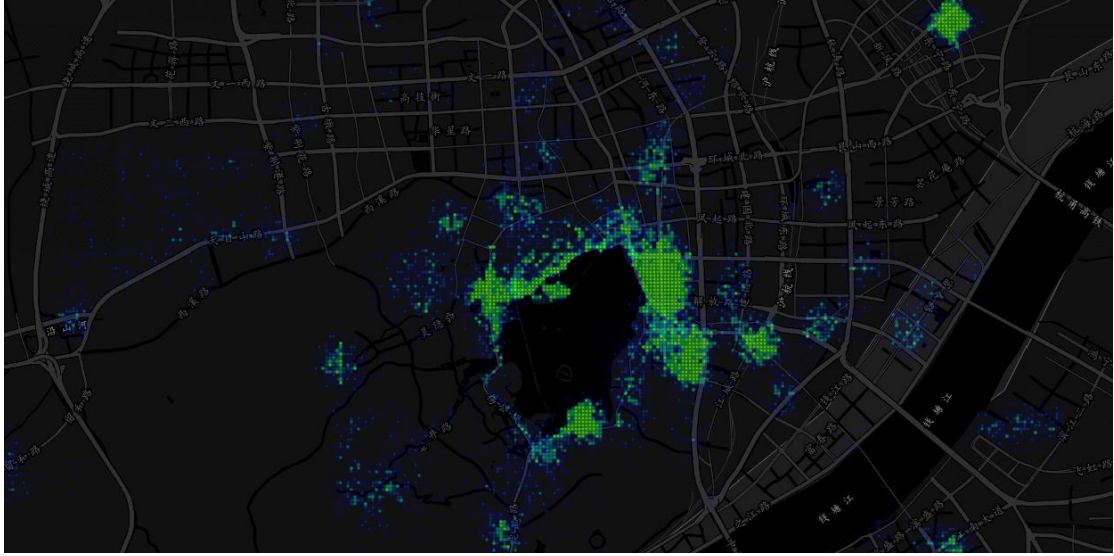


图 3-2 西湖及其周边访客热点图

仅以西湖周边地区为例，各个地区的 Z 分数进行可视化后（如图 3-2）可以发现核心热点区高度集中于西湖环线以及延安路两侧商业区，西湖一面与城市核心商业区接壤，而另一面由于地理原因只有少数道路能够疏通人流，因此有非常大的可能造成节假日拥堵，而杭州市中类似情况的景点不在少数。改造已成熟景区或限制人流并不是解决此问题较为可行的方案，因此，我们将重点放在游客疏导上，主要向非集聚区域进行疏导，区域需满足：

$$Z_i \leq f(z_1, z_2, \dots) \quad (3.2.3)$$

其中 $f(z_1, z_2, \dots)$ 为该区域道路最大通行能力，是关于 z_1, z_2, \dots 的多元函数， f 的值取决于道路车道数、周边道路连接点数量等自变量，涉及因素多且较为复杂，这里不详细展开讨论。

3.2.2 城市个性化旅游专线改造模拟

个性化旅游线路的设计出发点为缓解景区拥堵，然而新增一条旅游线路，还需要对访客满意度进行考量，为了能够量化访客选择某地点的倾向，可以使用相似数据的推荐思路，该模型基于以下假设：绝大多数游客出行时首选自己

未去过的景区，且评分高者优先，旅游专线的推荐算法基于游客-景点评分矩阵：

$$G = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{pmatrix} \quad (3.2.4)$$

其中 r_{ij} 为访客 i 对于景点 j 的评分，然而由于景点数量众多而访客所去过的景区相当少，因此 $G_{m \times n}$ 为一个非常稀疏的矩阵。为了能根据矩阵中已知的访客评分来预测访客对于未去过地点的评分，利用 SVD（奇异值分解）的方法处理评分矩阵 G ，SVD 的基本思路是：访客对于某景点的评分，与和该访客景点选择较为相的似访客评分存在一定相关性。比如，甲乙两个人同时去过很多相同的地方且对这些地方评分相似，甲去过 A 地且给 A 地较高的评分，乙未去过 A 地，但是可以预测他对 A 地的评分很可能也比较高。当访客数据量足够多时，这种预测准确率会大幅提升。

对评分矩阵 G 进行 SVD 操作：

$$G_{m \times n} = U_{m \times m} \times S_{m \times n} \times (V_{n \times n})' \quad (3.2.5)$$

其中 $S_{m \times n}$ 对角线元素为 G 的奇异值且降序排列，其余元素均为零，由于前几个奇异值的平方和占据了所有奇异值平方和的绝大部分比例（该比例又称奇异值平方和占比的阈值，设为 p ， p 为 0-1 之间的数），对于海量游客，为了提升推荐算法的计算效率，可以将 G 降至 k 维处理，选取 k ：

$$\frac{\sum_{i=1}^k s_i^2}{\sum_{i=1}^m s_i^2} \geq p \quad (3.2.6)$$

s_i 为 $G_{m \times n}$ 的奇异值，通常 p 取 0.9，且当 m 、 n 较大时，(3.2.6)式能够保证 k 远小于 m 、 n ，从而达到减少计算时间的目的并能够处理大量游客评分数据，此时 $G_{m \times n}$ 可近似表示为

$$G_{m \times n} \approx U_{m \times k} \times S_{k \times k} \times (V_{n \times k})' \quad (3.2.7)$$

此时可以得到景点数据在 k 维向量空间变换后的数据矩阵 $V_{n \times k}$

$$V_{n \times k} = (G_{m \times n})' \times U_{m \times k} \times S_{k \times k}^{-1} \quad (3.2.8)$$

$V_{n \times k}$ 中每一行为一个景点向量 \vec{x}_i ($i=1, \dots, n$)，对于用户未去过的景点 u ，使用余弦相似度衡量 \vec{x}_u 和 $V_{n \times k}$ 中第 i 行所代表景点 \vec{x}_i 的相似度：

$$\theta_{ui} = \frac{\overrightarrow{x_u} \cdot \overrightarrow{x_i}}{||\overrightarrow{x_u}|| \cdot ||\overrightarrow{x_i}||} \tag{3.2.9}$$

计算 $\overrightarrow{x_u}$ 与 $V_{n \times k}$ 所有行的相似度并加权求和，得到该景区 u 的预测评分 g_u

$$g_u = \frac{\sum_{i=1}^n w_i \theta_{ui}}{\sum_{i=1}^n \theta_{ui}} \tag{3.2.10}$$

通过这种方式，我们可以将每一位访客未去过的地点进行评分预测，而旅游线路可以根据预测评分的高低，同时在交通状况允许的条件下，为访客推荐分数最高的景区。

3.3 模型结果检验与分析

为了验证旅游专线规划模型的有效性，从杭州访客评分数据中选取 20 名访客、15 个景区进行模拟验证分析，表 3-1 中每一格为访客对相应景区的评分：

表 3-1 游客对不同景区评分

SpotId	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	西	西	宋	大	黄	岳	京	清	西	杭	钱	杭	吴	玉	宝
	湖	溪	城	明	龙	王	杭	河	湖	州	塘	州	山	泉	寿
		湿		山		庙	大	坊	乐	剧	江	植		校	山
VisitorI	地						运		园	院		物		区	
d							河					园			
0	8.5	0	7	0	0	8.5	0	0	0	8	0	0	0	0	0
1	7	0	0	0	9	0	0	0	0	0	0	9.5	0	0	0
2	9	8	0	0	7.5	7.5	9	0	0	7	0	0	0	0	7
3	9.5	0	7	0	0	0	7.5	0	0	0	0	0	0	0	0
4	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	7	0	9
6	0	0	0	0	0	0	0	0	8	9	8	0	0	8.5	0
7	0	8	8	0	7.5	0	0	0	6.5	0	0	0	0	0	0
8	0	0	0	0	0	8.5	0	0	0	0	0	0	9.5	0	7
9	0	0	0	7.5	0	0	0	9	0	9	0	10	0	0	0
10	0	0	0	0	0	8	0	0	0	7.5	0	0	0	0	0
11	0	8	8.5	0	0	0	0	0	0	0	7	0	0	7	0

12	0	0	0	8.5	0	0	0	6.5	0	0	0	0	8	0	0
13	0	9	0	0	8.5	0	0	0	8	0	0	0	0	0	8.5
14	0	0	9.5	0	0	0	0	0	0	0	0	0	0	8.5	0
15	0	10	0	0	0	0	9	0	7	0	0	0	0	0	0
16	0	9.5	0	0	0	0	0	0	0	7	0	0	7	0	0
17	0	0	0	8.5	0	0	0	7.5	0	0	0	0	7	0	0
18	0	8	8	0	0	0	7	0	0	0	8.5	0	0	0	8
19	0	0	7.5	0	0	6.5	0	0	0	0	8	0	0	0	0

其中前五人曾去过西湖，使用 SVD 对该矩阵进行处理后，以 Id=10 的访客为例(具体代码参见附录(G-2))，输出结果如下：

For visitor No.10, we recommend:		
	spot Id	spot score
0	2	7.803317
1	14	7.789637
2	12	7.776284
3	4	7.768304
4	0	7.746036

图 3-3 推荐旅游线路输出结果

据此可以看出，西湖（Id=0）仅在推荐列表中排名第五，对于该访客而言，如果有一条 2-14-12 的旅游专线，将会使其满意度得到较大提升；而对于城市而言，若较多旅客推荐线路与此相似，则可以分流部分访客，达到缓解景区拥堵的作用。

对上万的访客评分数据使用 SVD 算法，从而生成访客-景点推荐矩阵，能够为相关部门提供建设新旅游专线的决策建议^[26]，同时，访客分流也将大幅减少西湖等热门景区的压力，同时对于开发其他相对冷门的景区具有一定推动作用。

3.4 模型修正与进一步改良

在为游客推荐相关景区的同时，还需要关注所推荐景区的人口密度以及交通距离^[27]，同时，以西湖为例，由于部分访客慕西湖之名而来，此部分访客分流难度较大，而若访客曾来过杭州，为其推荐冷门景区较为合适。按照这种方式，以 2018 年 10 月 5 日西湖访客为例，对于访客进行筛选后推荐相关景区，可以将西

湖人数减少约 13%:



图 3-3 2018.10.5 西湖景区游人(上图)及线路设计后对于访客的分流(下图)

另外, 基于大数据思维, 可以将访客去过所有地点及其评分记录作为游客-景点评分矩阵的一部分, 地点不必局限杭州, 可以扩展到全国景区^[28], 利用 SVD 方法进行评分预测的精确性将会有更大的提升, 而综合所有人的推荐结果, 可以选取推荐重叠度最高的线路开设旅游专线, 为城市优化访客出行提供决策支持平

台。

4. 基于大数据分析的改进方案

4.1 增设旅游线路

4.1.1 必要性及可行性

杭州作为世界闻名的旅游城市，其游客接待数量逐年递增，然而，其道路规划和建设还无法满足大量的旅游需求。而若能根据旅客需求增设旅游专线，则可以一定程度上缓解访客出行问题。杭州历史文化悠久，景区类型丰富多样，设置一定的主题线路既可以分流一部分访客，为访客出行提供便利；又可以促进杭州旅游业的发展。

在互联网时代，基于旅游大数据规划旅游线路以成为可能。杭州作为互联网产业蓬勃发展的城市，可以根据访客数据进行分类，个性化推荐线路，提升访客出行的满意度^[29]。

4.1.2 具体决策方案

从机场和车站出发增设旅游专线，这里以机场为例，杭州市通向市中心的旅游大巴大多需在武林门中转，而此处正是交通严重拥堵的区域，因此，可以考虑为旅游者增设直达西湖景区、西溪湿地等地的旅游专线。

新增专线的同时，也可以将旅游专线进行分类，根据所去景区的特征为专线设置主题，如“自然风光专线”、“文化之旅专线”等，专线巴士可以点对点传送游客。另外，专线巴士可以同城市交通数据平台相连，根据城市交通实时数据躲避拥堵，而不一定需要经过固定线路，只需将游客送至目的地，保证出行效率能够最大化。

4.1.3 可能面临的问题及影响

设计新的旅游专线往往会面临资源不足的问题，比如巴士数量的不足或人员不足等，另外，旅游专线巴士停泊也会限制专线的拓展，比如在太子湾、杨公堤等地，开拓一个区域让专线巴士停泊较为困难，在这种情况下，专线巴士为了避免拥堵可能需要停在离景区相对较远的位置，这在一定程度上增加了游客行走的成本。

4.2 将冷门景区加入旅游专线

4.2.1 冷门景区对访客满意度提升的作用

近年来，随着热门景区在黄金时段人数激增，部分访客也会选择相对冷门但性价比较高的一些景区。然而根据调研，冷门景区与热门景区游客评分相差不大，在拥堵时段，舒服地在一些相对还比较有特色的冷门景区游览反而更能增加游客的满意度。同时，二次来访者也会更倾向于去一些冷门景区，冷门景区对于消费者回头率也有一定的提升作用。

4.2.2 冷门景区对城市交通疏导作用

就杭州而言，杭州景区分布广泛，然而热门地点较为集中，将游客分散到冷门景区明显能够减轻城市及景区的交通压力。例如，拱墅区拥有京杭大运河、浙江艺术博物馆、江南小镇等景点，将西湖区、下城区游客分散到拱墅区可以助力西湖沿线区域拥堵的缓解，同时也可以挖掘杭州市旅游潜力，使更多的线路可被访客进行选择。

4.2.3 具体改进方案

在新旅游主题专线建设的基础上，可以将具有相近主题的冷门景点加入，如经过宋城、西湖的旅游线路同时加入京杭大运河、历史文化古镇等地点，同时，由于访客活动通常具有区域性，并且倾向于选择距离较近的景点，专线的设计可以基于杭州市景点大数据，通过推荐模型进行新景点的选取与路径规划。

4.2.4 可能面临的问题及影响

冷门景区主题线路开发将成为杭州市旅游业发展的趋势之一，然而根据产品周期理论，冷门景区的受众面扩大以及相关设施的完善仍需要较长的一段时间，简单推荐冷门景区对于提升访客满意度可能难以达到预想的效果。最有效的发展方式需要旅游部门间的沟通，确保专线与冷门景区建设的同步性。

4.3 增加或改造景区设施

4.3.1 基础设施改造的必要性

基础设施完善程度对于游客体验有很大的影响，并且直接影响到城市的旅游品牌形象。调查结果显示，近半数游客认为杭州景区基础设施建设尚不完善，尤其体现在道路设施、住宿设施及公卫等方面。其中热门景区尽管有比较好的开发基础，但由于开发较早、承压较大、维护与建设边际成本较高，基建情况同样没能满足多数旅客的要求。

4.3.2 改造具体方案及可行性分析

可行的改造方案可分为现有基础设施的扩展翻新以及新类型的开发。新类型的开发方面，需根据针对基础设施的调查，结合基础设施地理信息大数据来确定需要新开发的基础设施种类，同时也可根据景区人流量及有该需求的特定人群的分布确定新基建的数量及可考虑的选址。而对于老设施而言，要先评估它的价值和游客满意度的影响，进行循序渐进的翻新，在增建一类基础设施时要注意保留原先的设施避免出现供不应求的状况。

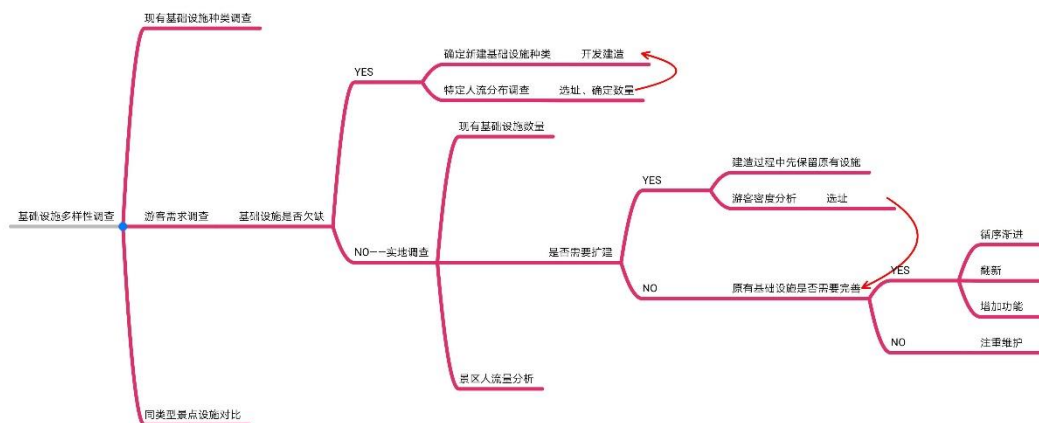


图 4-1 基础设施改进思路

基于该思路可以对杭州许多景点的基础设施进行改造，如增加连结西湖沿岸道路的通道，优化自然景观公共卫生间及周边酒楼建设，改善人文景区（如宋城）中的商铺外观及分布等。

整个改造流程基于景区现状和旅客反馈，具有现实基础，且调查成本不高，具有可行性；同时改造的过程将建造施工所造成的负面影响最小化，确保了其能够持续进行。从大的方面来看，由于旅游业是杭州重要的产业支柱，对于景区基础设施的改造能够得到政府的支持，技术的进步则使得基础设施改造的时间成本降低，并且使其功能更加完善。

4.4 优化景区等旅游项目的定价

效用价值论的角度来看，旅游产品能够满足人们对文化价值、美学价值的欲望，而定价会对消费者旅游体验主观感受产生相当大的影响，通过优化旅游产品定价也会为访客出行提供一些指导^[33]。在分散热门景区人流以提升游客体验方面，提供差别定价方案是一种解决方式，在冷门景区和线路的发展阶段，可以提供一些优惠，如特色线路团购价，或者是制定不同线路组合的不同优惠措施，激励游

玩者优先选择性价比较高的冷门景区^[34]。同时可采取了线上订票优惠、线下加价的差别定价方式，以达到分流旅客，人数管控等目的^[35]。

4.5 城市交通设施循序渐进改进

结合杭州市旅游业现状以及旅游业的蓬勃发展趋势，改造城市交通设施应该成为一个循序渐进的过程。城市可以借助实时访客数据平台来进行决策，设立特色旅游专线或主题游览线路，同时考量不同时间、不同地域、不断变化的旅游需求，进行实时的交通规划与调度；可以建立旅游景点评价反馈机制，紧随消费升级过程与趋势，在保持景点特色与风貌的前提下进行改造设计，提高效率；可以加入更多的旅游产品供访客进行选择，设置开放的交通数据可视化平台，为访客展现景点实时数据，方便访客进行决策。

5. 总结与展望

随着人们生活水平的提高，旅游出行越来越成为节假日人们消遣的主要方式之一，然而，旅游城市——以杭州为代表，必然会面临景区拥堵等状况。国庆等公众假期旅游俨然成为花钱找罪受。城市的形象往往受到当地交通设施的影响，同时，出行者的体验也与景区规划、交通便捷程度和拥挤程度等指标息息相关。杭州市在假期景区的拥堵现象已经持续多年，并且有不断加重的趋势。

然而，随着大数据技术的应用与发展，城市可以用一种新思路来解决此问题：利用大数据可视化发现规律，包括不同时间、地理位置的访客行为模式，分析出访客出行线路规划的共同规律，并运用于景区疏通，特色主题线路设计等，从而提升访客出行体验，提高来杭人员的满意度。

通过调研以及海量的数据分析，能够较为直观看出杭州市来访人员的时空分布特征，而城市亦可以根据这些特征来进行旅游线路的个性化推荐。从短期来看，个性化旅游线路的设计能够缓解热门景区的拥堵状况，同时为访客提供一个契机访问杭州其他有吸引力但是相对没那么热门的景区；从长期来看，特色旅游线路可以提升相对较为冷门景区的知名度，带动当地的经济发展，促进杭州不同区域的协同发展。总而言之，基于杭州访客大数据的出行线路优化将成为杭州市提升城市整体建设质量的推动力，智能旅游与交通疏导也必定是城市发展的趋势。

参考文献

- [1] 张立明.敖荣军等编著[M]旅游学概论：武汉大学出版社，2003.7
- [2] Graham R L, Hell P. On the history of the minimum spanning tree problem[J].Annals of the History of Computing, 1985,7:43-57
- [3] Ahuja R K, Magnanti T L, Orlin J B. Network Flows: Theory Algorithms and Applications[M].New Jersey: Prentice-Hall, 1993
- [4] Dantzig G B, Fulkerson D R, Johnson S M. Solution of a large scale traveling-salesman problem[J].Operations Research, 1954,2:393-410
- [5] 吕琼艺.基于改进的 Dijkstra 算法的旅游规划线路研究与实践——以鼓浪屿景区为例 [J] Journal of Liuzhou Vocational & Technical College.2017,17(2)
- [6] Ahulre Frank.Tourist route design based on the principle of recommendation system [J/OL] arxiv.org/ftp/arxiv/papers
- [7] Oscar Claveria, Enric Monte, Salvador Torra.Modelling tourism demand to Spain with machine learning techniques [J/OL] arxiv.org/ftp/arxiv/papers
- [8] 龙涛.上海中心城区游客兴趣点(POI)时空分布特征研究[D].上海师范大学.2016
- [9] 李攀. 基于数据可视化的城市路网拥堵特性描述分析方法[P]. CN201610806496.0
- [10] 龙瀛. 城市大数据与定量城市研究[J]. 上海城市规划, 2014 (2014 年 05): 13-15, 71.
- [11] 杨正洪. 智慧城市：大数据，物联网和云计算之应用[M]. 清华大学出版社, 2017.
- [12] 秦萧，甄峰，熊丽芳，等. 大数据时代城市时空行为研究方法[J]. 地理科学进展, 2013, 32(9): 1352-1361.
- [13] 党安荣，袁牧，沈振江，等. 基于智慧城市和大数据的理性规划与城乡治理思考[J]. 建设科技, 2015 (5): 64-66.
- [14] 元媛,刘向前,梁留科.基于 IPA 法的特殊兴趣旅游目的地游客满意度评价——以开封专业美食夜市为例[J].洛阳师范学院学报,2018,37(09):24-29.
- [15] 陈秀玲.基于 ACSI 模型的主题公园节庆活动游客满意度指数研究[D].上海师范大学,2018.
- [16] 刘超,胡宝贵.基于 probit 模型和智慧旅游满意度影响因素实证研究——以黄果树景区为例[J].青岛农业大学学报(社会科学版),2018,30(03):45-51.
- [17] 潘禄生.基于 LBS 的个性化旅游线路推荐算法研究[J].电子测试,2018(21):57-59.
- [18] 吴念.开放式景区交通治理模式研究[D].复旦大学,2014.

- [19] 余杰.城景相依型景区交通组织优化研究[D].东南大学,2016
- [20] 卞显红.城市旅游空间结构研究[J]. 地理与地理信息科学, 2003, 19(1): 105-108.
- [21] 关宏志, 任军, 刘兰辉. 旅游交通规划的基础框架[J]. 北京规划建设, 2001 (6): 32-35.
- [22] 黄柯, 祝建军, 蒲素. 我国旅游交通发展现状及研究述评[D]., 2007.
- [23] 卞显红. 城市旅游空间规划布局的影响因素分析[J]. 地域研究与开发, 2003, 22(3): 93-96.
- [24] 陈传康. 城市旅游开发规划研究提纲[J]. 旅游学刊, 1996 (5): 31-34.
- [25] 李德仁, 姚远, 邵振峰. 智慧城市中的大数据[J]. 武汉大学学报信息科学版, 2014, 39(6): 631-640.
- [26] 刘春琳, 冷红. 基于大数据挖掘的城市关注平台的构建与应用[J]. 城乡治理与规划改革——2014 中国城市规划年会论文集 (04 城市规划新技术应用), 2014.
- [27] 徐蓉艳. 旅游大数据与挖掘及其在旅游行业的应用方向[J]. 中国市场, 2014 (51): 204-205.
- [28] 郭鑫. 旅游大数据与挖掘分析研究[J]. 电脑知识与技术, 2013 (5X): 3215-3216.
- [29] 梁昌勇, 马银超, 路彩红. 大数据挖掘: 智慧旅游的核心[J]. 开发研究, 2015 (5): 134-139.
- [30] 王显飞, 陈梅, 李小天. 基于约束的旅游推荐系统的研究与设计[J]. 计算机技术与发展, 2012, 22(2): 141-145.
- [31] 杨璇. 海量旅游统计数据可视化的研究与应用[D]. 武汉邮电科学研究院, 2018.
- [32] 朱桂祥, 曹杰. 基于主题序列模式的旅游产品推荐引擎[J]. 计算机研究与发展, 2018, 55(5): 920-932.
- [33] 李雪梅. 公共资源型景区门票价格及其影响因素实证研究[D].浙江财经大学,2015.
- [34] 贺颖. 公共景区免费开放的经济学研究[D].浙江大学,2012.
- [35] 滕玮峰.对景区价格体制创新的探索与思考——杭州西湖景区免费准人体制的效益分析及启示[J].价格理论与实践,2008(03):23-24.

致谢

经过两个月的努力，从零开始入手一个项目非常不容易，每一位队员为此付出了巨大的努力，查资料、整理材料、写作论文。论文得以完成，需要感

谢非常多人的支持。首先要感谢为本文提出宝贵指导意见的陈熹、林姗姗两位导师。他们在我们的论文写作遇到瓶颈的时候，指出我们的不足，给我们提了许多有价值的修改建议；在我们寻找数据遇到困难时，两位导师主动为我们联系市场上为数不多的有价值的数据公司，帮我们解决了燃眉之急；同时，在团队成员遇到障碍，情绪低落时，两位导师会和我们谈心，为我们做心理辅导，减轻了我们精神上的压力。

接着我们要感谢为我们提供大量珍贵数据或者接受我们线上采访的组织和公司，包括个推大数据公司、极海数据有限公司、携程网杭州站、大众点评、杭州市旅游委员会、杭州公共交通运营管理部门、杭州市西湖区区政府等有关部门。他们提供的数据量大、质高、类全，为我们的建模、测算、检验等工作提供了充足的保障，同时感谢他们能够耐心回答我们的疑问，为我们解决课题上的问题提供了巨大的帮助。

论文的顺利完成，也离不开其它各位老师、同学和朋友的关心和帮助。在整个的论文写作中，各位老师、同学和朋友积极的帮助我查资料 and 提供有利于论文写作的建议和意见，在他们的帮助下，论文得以不断的完善，最终帮助我完整的写完了整个论文。另外，要感谢在大学期间所有授课的老师，他们的悉心教导使得团队成员能够较好地掌握专业课知识，这也是论文得以完成的基础。本文是在团队成员的通力合作，导师的悉心指导以及数据公司与相关部门的大力支持下才得以完成的，少了其中任何一方的参与都难以有最终成果，由于篇幅有限难一一列举完，也一再次并表示我们最衷心的感谢。

附录

A、项目团队介绍

B、研究的日程安排

C、线下访谈整理

D、问卷备份

E、问卷结果报告

F、实地调研总结反思

G、具体模型的代码实现

H、数据可视化报表

I、数据来源汇总

附录 A 项目团队介绍

本团队由来自竺可桢学院 17 级的学生组成，团队成员来自数学与应用数学、金融学、信息与计算科学、信息管理与信息系统、计算机等专业, 专业互补性较强, 有利于互相沟通与开阔眼界; 项目具体成员为: 队长, 武子越, 竺可桢学院大二年级本科生, 交叉 1702 行政班, 统计学(信息管理与信息系统交叉创新平台); 队员刘沛林, 竺可桢学院, 统计学(信息管理与信息系统交叉创新平台); 马万腾, 竺可桢学院, 统计学(信息管理与信息系统交叉创新平台); 刘洵孜, 竺可桢学院, 数学与应用数学(金融学交叉创新平台); 翟泳皓, 竺可桢学院, 数学与应用数学(金融学交叉创新平台); 余林恩, 竺可桢学院, 信息与计算科学; 阮诗贇, 经济学院, 金融学; 周瑞祺, 竺可桢学院, 计算机专业。

同时, 本项目团队导师由陈熹教授和林珊珊教授担任。陈熹老师是来自浙江大学数据科学研究中心的教授与博士生导师, 信息管理与信息系统本科项目主任, 为信息系统和大数据分析领域的专家; 林珊珊老师来自浙江大学旅游管理专业研究生导师, 为该领域的较有影响力的专家。

附录 B 研究的日程安排

第一阶段 选题与初步开展

2018.11.3	提出构想与团队组建
2018.11.4-8	初步小范围社会意见征集
2018.11.10	汇总意见讨论研究方向
2018.11.9-11	联络并选定指导老师
2018.11.11	拟定研究课题
2018.11.11	制定调研方案, 完成前期调研资料整理汇总

第二阶段 数据探索与文献研究

2018.11.12-25	联络大数据公司和旅游相关部门
2018.11.13-15	首批数据收集与预处理
2018.11.16-18	基于 Web 的交互式可视化实现

2018.11.19	课题组拆分，细分研究方向
2018.11.21	线上问卷调研开始
2018.11.20-29	收集相关文献，整理已有资料和研究成果
2018.11.20-29	成员培训，现学新知识
第三阶段 实地调查与结果验证	
2018.12.1-7	群众意见的再次收集
2018.12.9\16\31	景区走访调查
2018.12.9	文章撰写开始
2018.12.10	数学模型初步建立
2018.12.11	导师组会讨论意见修改
2018.12.12-14	数据可视化完善，研究着手研究城市规划方案
2018.12.15-17	整理研究数据和成果，形成分组研究报告
2018.12.16	模型代码实现
2018.12.15-18	模型修正与代码复现
2018.12.15-18	线下访谈进行
第四阶段 成果整理与完善	
2018.12.17-21	线上讨论各小组成果
2018.12.17	提出景区定价策略改进方案
2018.12.20-22	组织内部评估，模拟评审，互相提出修改意见
2018.12.24-31	根据评估结果，修改研究报告
2019.1.1	总结小组成果，积极准备接受评审

附录 C 线下访谈报告

访谈对象：经常在杭州市出行游玩的同学。

1. 你觉得杭州旅游交通有什么问题？

同学：一个是景点之间的公共交通不是太方便，而且重要交通组与景点间的联系较弱，就拿机场为例，萧山机场离杭州市中心和其他景点都比较远，但却没有机场到旅游景点的直达快捷的交通方式。这就让人很头疼，初到杭州时，为了到相应的景点玩玩，还需要对杭州的交通路线好好研究一下，特别麻烦。

2. 能够具体谈一下对杭州公交和地铁交通的看法吗？

同学：杭州城市发展很快，但和老城区相比，新城区的交通好像还不是那么方便。而且公交站点相对较少，有些站点看着不远，但斑马线设置不太合理，走过去要绕一大圈。而且这几年感觉来杭州旅游的人越来越多了，特别是西湖景区，人满为患，让杭州不断上“热搜”。但景点的公交线路少，让整个假期期间的交通拥堵成为大家旅游的一个巨大痛点。至于地铁线路，现在感觉地铁出行是挺方便的，但是感觉从地铁站去往周边还是不够方便。而且有些地铁施工造成的一些主干道的拥堵，没有得到好的缓解和治理。

3. 你可以谈谈关于像大商场，医院，学校等人口密集的重点场所的交通的看法吗？

同学：至于大商场，医院这些地方嘛……最典型的问题当然还是找不到停车位了，而且由于某些知名医院设计施工早，并没有考虑到太多的车辆停滞问题，导致不光是医院门口堵，周边的几条主干路都会受其影响，而商场也是类似的，所以现在出行不管是打车还是开车，感觉还是没有自己走着快。学校嘛，就拿我们学校来说，感觉周边的公共交通真的太少了，就只有几个公交站点，着实让出行有点麻烦。

附录 D 问卷设计备份

问卷内容：

1. 您的年龄是

- a. 15 岁及以下
- b. 16-25 岁
- c. 26-35 岁
- d. 36-50 岁
- e. 50 岁以上

2. 您是杭州本地人吗

- a. 是（转第 4 题）
- b. 不是

3. 来杭州旅游，您通常选择哪种交通工具到达杭州

- a. 高铁/动车
- b. 火车
- c. 长途汽车
- d. 飞机
- e. 自驾
- f. 其他

4. 倘若您要进行一次杭州旅游，您会选择去哪些地方【多选】

- a. 西湖风景区（包括雷峰塔、三潭映月、太子湾公园等）
- b. 灵隐寺&六和塔
- c. 宋城
- d. 杭州动物园&野生动物世界
- e. 清河坊街
- f. 长乔极地海洋公园
- g. 西溪湿地公园
- h. 九溪烟村
- i. 杭州乐园&杭州烂苹果乐园
- j. 千岛湖
- k. 大明山景区
- l. 岳王庙景区
- m. 浙西大峡谷
- n. 钱塘江大桥

o.京杭大运河

p.名人故居（如胡雪岩故居）

q.其他

5. 影响您选择的因素有哪些【多选】

a. 对景点类型的偏好（如对自然景观有特殊偏好）

b. 经济因素（如门票定价、酒店费用等）

c. 交通便利程度

d. 景区周围基础设施完善程度（如医院、餐馆、公共卫生间等）

e. 景点知名度

f. 影响体验感的客观因素（如人流量、天气等）

g. 安全因素

h. 陪同他人

i. 其他

6. 您认为哪些出行方式适合在杭州旅游【多选】

a. 地铁

b. 公共汽车

c. 出租车

d. 共享单车

e. 自驾

f. 步行

g. 其他

7. 站在旅游者视角，您认为杭州交通有哪些不足之处【多选】

a. 线路单一，缺少灵活性

b. 交通网络覆盖不全面，触角不够长

c. 某些站点拥堵（如景区附近站点）

d. 整体道路拥堵

e. 可供选择的交通方式不够多样

f. 缺少专业化的旅游专线

g. 乘坐公共交通体验感差

h. 其他

i. 没有不足之处

8. 您有了解过杭州的旅游专线吗

a. 有

b. 没有

9. 您认为是否有必要普及旅游专线

a. 有必要

b. 没有必要

c. 无所谓

10. 您认为杭州景点拥堵状况如何

a. 几乎都很拥堵

b. 只有某些景区总是拥堵

c. 只有在节假日某些景区会拥堵

d. 不怎么拥堵

11. 景点拥堵会影响您的出游心情吗

a. 会，很影响心情

b. 有一点，不影响整体体验

c. 没有影响

12. 您认为杭州旅游地带的基础设施完善吗

a. 完善，不需要改进（转第 14 题）

b. 不够完善，需要改进

13. 您认为杭州旅游地带哪些基础设施建设不够完善【多选】

a. 酒店或民宿

b. 医院

c. 商场或商店

d. 餐馆

e. 良好的通信网络

f. 公共卫生间

g. 宽阔的道路、桥梁等方便出入和内部交通的设施

h. 文化娱乐设施（如游乐中心、游泳池）

i. 体育或疗养设施

j. 水、电、气系统

k. 其他

14. 作为旅游者，您通常会选择什么时段出行

a. 上午

b. 中午

c. 下午

d. 晚上

e. 随机

15. 您的出行时间会受哪些因素影响【多选】

a. 个人喜好

b. 目的地人流量估计

c. 规避交通拥堵心理

d. 天气等自然环境

e. 其他

16. 您认为杭州旅游景点的空间分布合理吗

a. 合理

b. 不合理，存在改进余地

c. 不合理，但难以改进

17. 您认为杭州旅游景点分布存在哪些问题【多选】

- a. 热门景点过于集中，导致人流量严重失衡
- b. 没有形成旅游区块，相似景点较分散
- c. 某些景点过于偏僻
- d. 景点开发不完全或存在较多潜在开发景点
- e. 其他
- f. 不存在问题

18.您希望有某种工具或机制能针对您的旅行地目标对您的旅行线路进行优化吗

- a. 非常希望
- b. 不希望
- c. 无所谓

附录 E 问卷结果报告

此问卷主要针对杭州旅客进行了有关旅游交通、基建、景点选择等方面的调查，通过对收集到的 264 份问卷的数据归纳及对比分析，能够得出作为旅游城市的杭州存在的几个显著特征：

1. 交通枢纽单一

问卷显示大部分游客（70.97%）选择动车或高铁作为抵杭的交通工具，另也有旅客选择飞机与火车这两种方式，相对而言动车或高铁结合了速度快与成本低两个优点，因此成为旅行首选，杭州东站成为往来杭州最主要的交通枢纽。

2. 地铁和共享单车为主要出行方式

82.64%受访者都认为“地铁”是适合在杭州旅行的出行方式，同时超过一半的人都会选择在出行时用共享单车代步，这说明地铁和共享单车是杭州旅游交通的主力，这两方面的建设需予以重视。

3. 交通网络存在明显问题

通过对杭州交通可能存在的问题的调查，可见“交通网覆盖不全面”这一问题较为明显（有 34.5%的旅客赞同）。而地铁作为杭州城内主要的出行交通工具，其建设不完全正是杭州城市交通网络覆盖不全面的主要原因。从结果上看，该问题所导致的不必要的转乘、打的等行为极大破坏了访客的旅游体验。此外，热门景区附近站点拥堵问题更是引起了广泛的共鸣（77.8%），需要用线路多样化、站点扩建等方法进行治理。

4. 基础设施不够完善

杭州市新型城区建设与规划并未普及，部分景区附近基础设施也未与时俱进，如街道在设计时没有考虑较大的人流量，市内的天桥、地道等大型出入口类基础设计十分少，未能充分利用城市的空间层次来分流人群以缓解路面出入口交通压力。同时随着经济发展，旅客数量剧增，对景区周边的民宿、酒店等生活必需设施需求增大，而供给并未跟上需求，从而将降低旅客的体验度。

5. 西湖景区景点过于集中

这是杭州旅游最明显也是最棘手的问题。西湖景区作为一个综合性的景区，其内除西湖本身外，还包括了雷峰塔、三潭映月，太子湾公园等景点，又

与附近的灵隐寺、杭州动物园等形成了旅游大区块，成为旅客的聚集点。这种情况直接导致了交通压力过大，行人拥挤等一系列旅游问题。

横向对比分析：

不只是杭州，世界上许多著名旅游城市都存在着交通问题，与这些城市治理方式的对比是必要环节，利于当下旅游治理的反思与经验的汲取。先观察其他城市：日本箱根景区主要通过公共交通来连接各景点、发挥了水上交通的功能；法国巴黎圣母院景点依托强大的公共交通配套保障来进行交通治理；北京什刹海景区实施区域性交通管制、推出景区内的交通代步服务、开辟水上交通服务、完善景区周边的停车设施；上海南京路步行街强化区域交通组织，挖潜社会停车资源，拓展车辆观光服务，设置路障、信号灯及标志，加快地铁人流疏散；武汉东湖景区则采取加强交通疏导，采取交通管制，实行公交优先政策、落实衔接管理等措施。

反观杭州，杭州同样采取了诸多措施，如西湖风景区采取“区域单双号”限行措施并改善苏堤南口交通节点，组织实施单行线，确保公交优先和换乘方便等，但这些措施仍然不能完全解决杭州市旅游出行问题。

借鉴其他城市管理方式以及通过实地调研，能够为当前杭州市的交通治理提出如下建议：1. 限制私家车进入景区内部，通过增设巴士、环保观光车、电瓶游览车等设施有效加强景区与停车场的联系，引导游客换乘进出旅游景区；2. 结合杭州自然环境发挥水上交通运输功能，水上交通既可以发展为旅游产品，又能够发挥交通功能；3. 考虑建设机械式立体停车库，加大景区的停车设施供给规模，或者考虑利用景区外的空旷场地建设路外停车场，进一步缓解车辆过挤及泊车困难的情况。

附录 F 实地调研总结与反思

短短两个月的时间内，在实地考察阶段，我们走访了西湖、灵隐寺、西溪湿地等主要风景区。采访时间为 2018 年 11-12 月的周末。采取的交通方式为最具代表性的公共交通（地铁，公交车）和单车、步行。

在走访西湖的过程中我们发现，从凤起路下地铁后向西湖行进的道路人流量明显增多。尤其是位于北山街的景区入口，其人流量长时间居高不下。由于是周末，早晨 9 点之前景区内更多是晨练的老年人，而到了近中午外来游客明显增

多。特别是中午到下午的时段，游客数量增加，以至于在关键入口使用单车骑行出现困难。

走访灵隐寺和西溪湿地时，由于天气转寒，灵隐寺和西溪湿地的游客明显减少。但灵隐寺仍有不少游客慕名而来，经询问发现许多都是杭州本地居民前来拜佛祈福。

由于调研时段不是类似十一黄金周这样的旅游佳节，客流量并没有达到景区的承载极限，所以走访体验较好，但从中我们仍可以窥见杭州景区在旅游高峰时出现的一些问题。例如相关基础设施较少，关键出入口人流量过高等。

同时我们也有以下收获与反思：

1. 实地调研的时间选取一定要有代表性。由于课题时间原因，天气变冷后，许多游客不愿上山或参观湿地，导致我们调研时人流量较小。如果能选区旅游高峰期进行调研，获得的结果会更有参考性。

2. 要重视本地游客对整体旅游的影响。以灵隐寺为例，其游客很大比例是杭地信佛的老年人们，且这部分人前往灵隐寺拜佛的频率叫高。在旅游高峰期，这部分人的出行和移动会和旅客有极大部分的重叠。而这也是旅游优化中应当考虑的问题。

3. 大型景区入口处的引流设施十分重要。走访调查时，西湖的东北入口出长期滞涩，很大程度上是因为人行道与机动车道相交，必须人车交替通行。而这大大放慢了人流同行速度。杭州市内景区入口规划并没有满足日益增长的旅游需求，入口附近的街道并没有相应数量的天桥、地道等大型出入口，未能充分利用城市的空间层次来分流人群，缓解路面出入口交通压力。可以考虑在市内景区的入口添加天桥等人行设施，达到人车分流或部分分流的目的。

附录 G-1 基础设施密度与景区热度相关性验证代码

```
data=xlsread('Location_new.xlsx');
loc=data(1:297,1:2);
st=[mean(loc(:,1)),mean(loc(:,2))];
loc=loc-ones(length(loc),1)*st;
data2= xlsread('lines.xlsx')-ones(4,1)*[st,st];
linep=[];
for i=1:4
    temp=data2(i,:);
    b=(temp(1,1)-temp(1,3))/(temp(1,2)-temp(1,4));
```

```

a=temp(4)-temp(2)*b;
t=-1:1;
xh=t;
yh=a+b*xh;
plot(xh,yh);
hold on
linep=[linep;a,b];
end
area1=[];area2=[];area3=[];area4=[];area5=[];area6=[];area7=[];area8=[];area9=[];
for i=1:length(loc)
    temp=loc(i,:);
    if temp(1)<temp(2)*linep(2,2)+linep(2,1) &&
temp(1)<temp(2)*linep(3,2)+linep(3,1)
        area1=[area1,data(i,5)];
    end
    if temp(1)<temp(2)*linep(2,2)+linep(2,1) &&
temp(1)>temp(2)*linep(3,2)+linep(3,1) && temp(1)<temp(2)*linep(4,2)+linep(4,1)
        area2=[area2,data(i,5)];
    end
    if temp(1)>temp(2)*linep(2,2)+linep(2,1) &&
temp(1)<temp(2)*linep(3,2)+linep(3,1) && temp(1)<temp(2)*linep(1,2)+linep(1,1)
        area3=[area3,data(i,5)];
    end
    if temp(1)<temp(2)*linep(2,2)+linep(2,1) &&
temp(1)>temp(2)*linep(4,2)+linep(4,1)
        area4=[area4,data(i,5)];
    end
    if temp(1)>temp(2)*linep(2,2)+linep(2,1) &&
temp(1)>temp(2)*linep(3,2)+linep(3,1) && temp(1)<temp(2)*linep(1,2)+linep(1,1) &&
temp(2)<temp(2)*linep(4,2)+linep(4,1)
        area5=[area5,data(i,5)];
    end
    if temp(1)<temp(2)*linep(3,2)+linep(3,1) &&
temp(1)>temp(2)*linep(1,2)+linep(1,1)
        area6=[area6,data(i,5)];
    end
    if temp(1)>temp(2)*linep(2,2)+linep(2,1) &&
temp(1)>temp(2)*linep(4,2)+linep(4,1) && temp(1)<temp(2)*linep(1,2)+linep(1,1)
        area7=[area7,data(i,5)];
    end
    if temp(1)>temp(2)*linep(3,2)+linep(3,1) &&
temp(1)>temp(2)*linep(1,2)+linep(1,1) && temp(1)<temp(2)*linep(4,2)+linep(4,1)
        area8=[area8,data(i,5)];
    end
end

```

```

        if temp(1)>temp(2)*linep(1,2)+linep(1,1) &&
temp(1)>temp(2)*linep(4,2)+linep(4,1)
            area9=[area9,data(i,5)];
        end
    end
end
x=[length(area1),length(area2),length(area3),length(area4),length(area6),length(area9)
];
y=[sum(area1),sum(area2),sum(area3),sum(area4),sum(area6),sum(area9)];
reglm(y,x);

```

附录 G-2 基于 SVD 个性化线路推荐的 Python 代码实现

```

import numpy as np
import pandas as pd
#Using liner algebra module in numpy
from numpy import linalg as la

#Using test data for validation
def DataInput():
    f=open('C:/Users/Tinky/Desktop/data.csv')
    M0=pd.read_csv(f)
    M=M0.iloc[:,1:]
    return np.mat(M)

#Calculate the cosine similarity between vector X and vector Y
def cosSim(X,Y):
    vec_product=float(X.T*Y)
    norm_product=la.norm(X)*la.norm(Y)
    normalize=0.5*(vec_product/norm_product) +0.5 #Normalization
    return normalize

#Calculate the threshold value of singular value proportion(K)
def sig_percentage(sig,percentage):
    tol_sig_Squaresum=sum(sig**2)
    initK_sig_Squaresum=K=0
    for i in sig:
        initK_sig_Squaresum+=i**2
        K+=1
        if initK_sig_Squaresum>=tol_sig_Squaresum*percentage:
            break
    return K

#Dimension reduction
def Reduction(data,percentage):

```

```

U,S,Vt=la.svd(data)
k=sig_percentage(S,percentage)
sigK=np.mat(np.eye(k)*S[:k]) #The singular value matrix of SVD
Vk=data.T*U[:, :k]*sigK.I #Converted data of SPOT in k dimension linear space
return Vk

#Evaluate the spots that visitors haven't been to by similarities
def Predict_score(data,visitor_Id,spot,percentage):
    n=np.shape(data)[1]#Read the column(size of the matrix's Second Dimension)
    tol_similarity=Gtol_similarity=0
    SVDreduction=Reduction(data,percentage)
    for j in range(n):
        visitorgrades=data[visitor_Id,j]
        if visitorgrades==0 or j==spot:
            continue
        similarity=cosSim(SVDreduction[spot,:].T,SVDreduction[j,:].T) #Calculate the
similarity
        tol_similarity+=similarity
        Gtol_similarity+=similarity*visitorgrades
    if tol_similarity==0:
        return 0
    else:
        predict_score=Gtol_similarity/tol_similarity #Calculate the predict score
        return predict_score

#Recommend the first k highest-score spots fot visitors
def recommend(data,visitor_Id,highest_k,percentage=0.9):
    unvisited_spot=np.nonzero(data[visitor_Id,:].A==0)[1]
    if len(unvisited_spot)==0:
        return 0
    Scores=[]
    # Predict the score
    for spot in unvisited_spot:
        Score=Predict_score(data,visitor_Id,spot,percentage)
        Scores.append((spot,Score))
    # Descending order
    Scores=sorted(Scores,key=lambda x:x[1],reverse=True)
    return Scores[:highest_k]

visitor_Id=10
result=pd.DataFrame(recommend(DataInput(),visitor_Id,highest_k=5))
result.rename(columns={0:'spot Id',1:'spot score'}, inplace = True)
print("For visitor No.{}, we recommend:".format(visitor_Id))
print(result)

```

附录 H 数据可视化报表

论文中由于分析的必要加入了部分数据可视化图表，其余的地理数据可视化图表将汇总在此，由于篇幅原因，此处将具有重要信息的图表呈现：

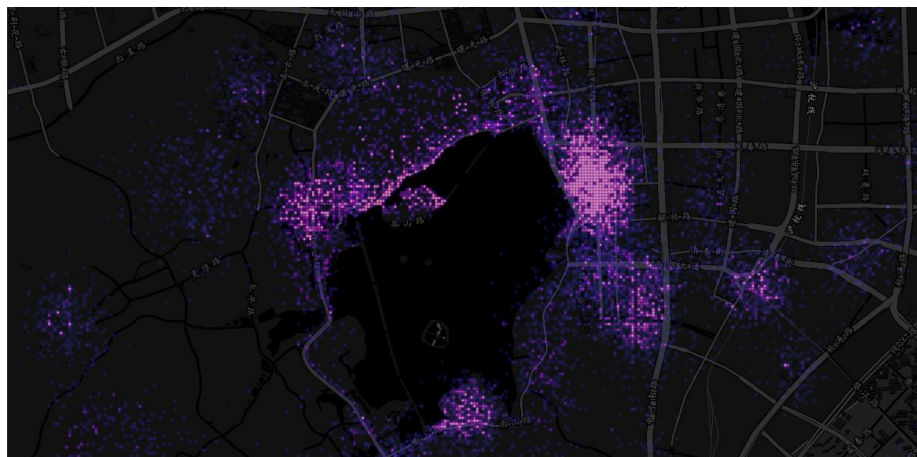


图 H-1 西湖景区的人口密度 (2018. 9)

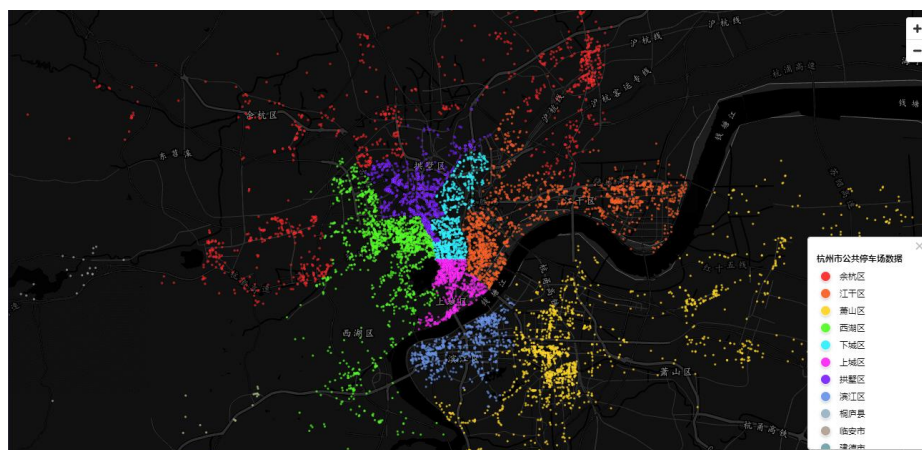


图 H-2 杭州市公共停车场区域数据

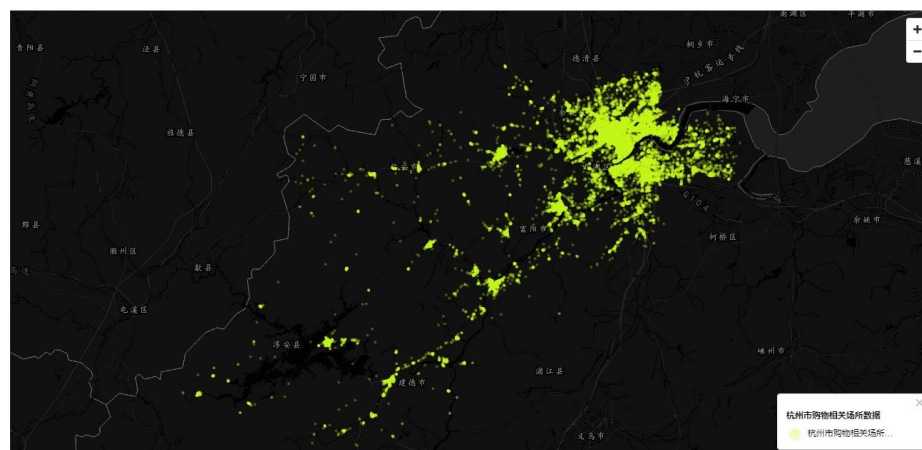


图 H-3 杭州市购物相关场所数据



图 H-4 西湖周边访客分布散点图(2018.9-10)

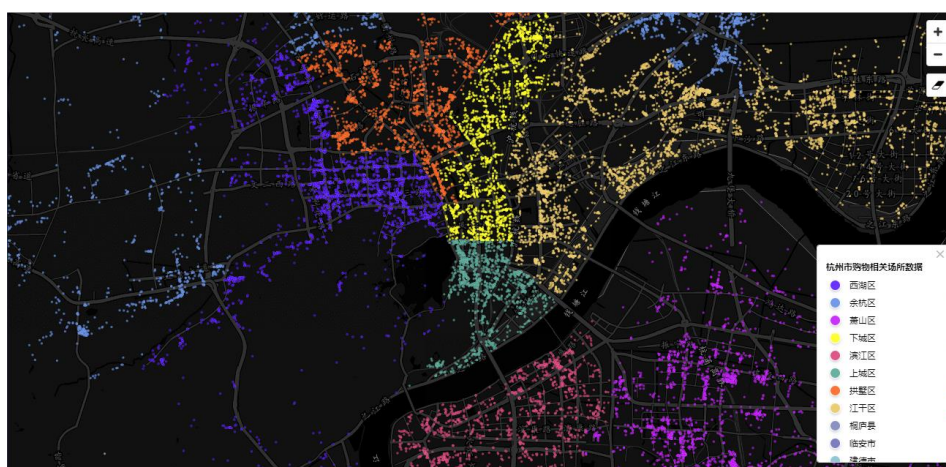


图 H-5 杭州市购物场所与购物中心(20km 范围)

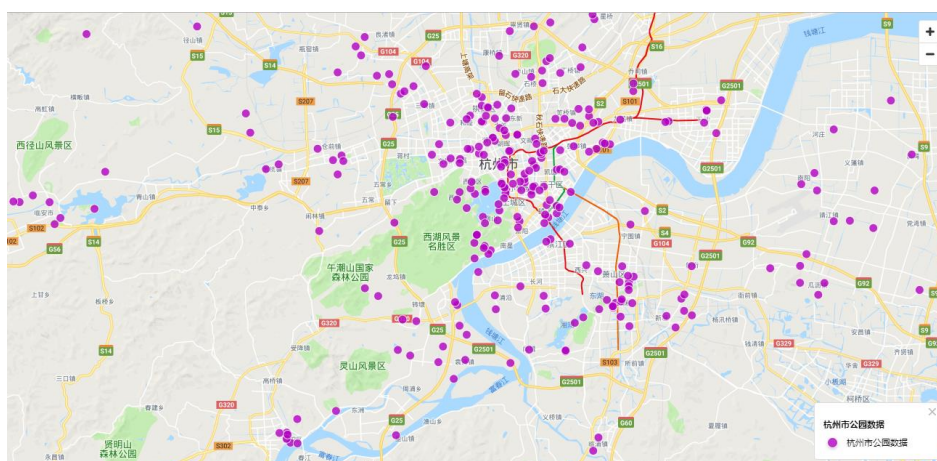


图 H-6 杭州市公园分布

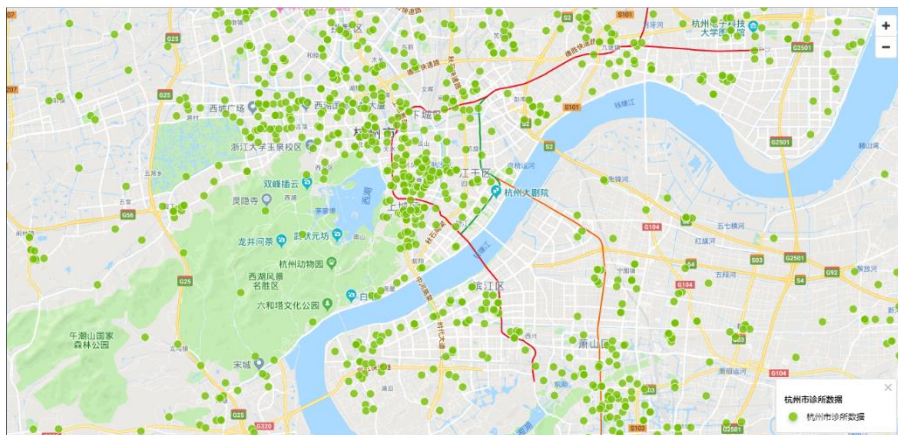


图 H-7 杭州市诊所数据

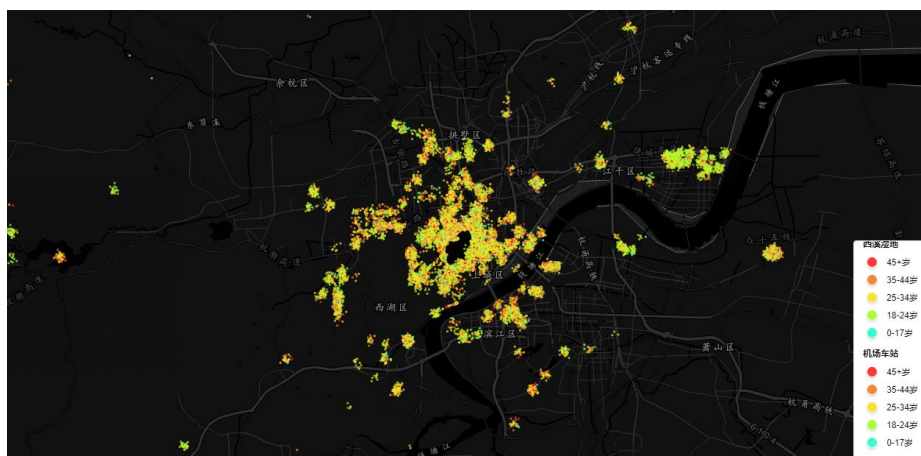


图 H-8 各大景区访客年龄散点图

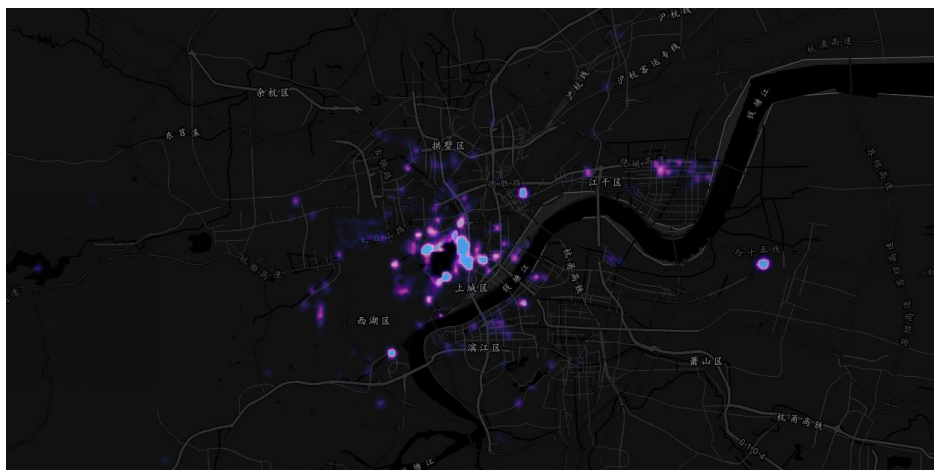


图 H-9 景区热度 (35km 范围)

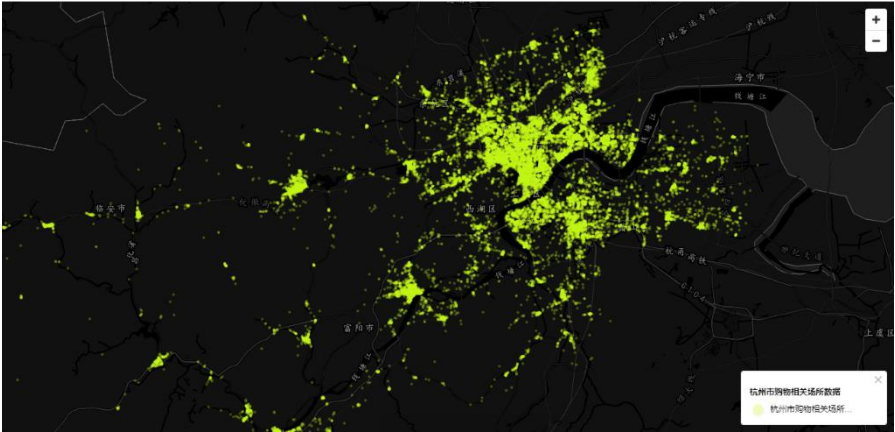


图 H-10 节假日开放购物中心分布

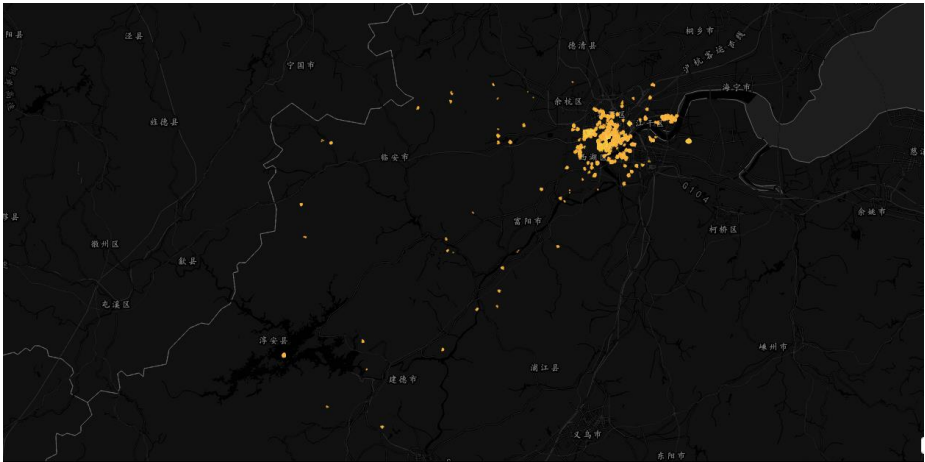


图 H-11 杭州市热门景点分布

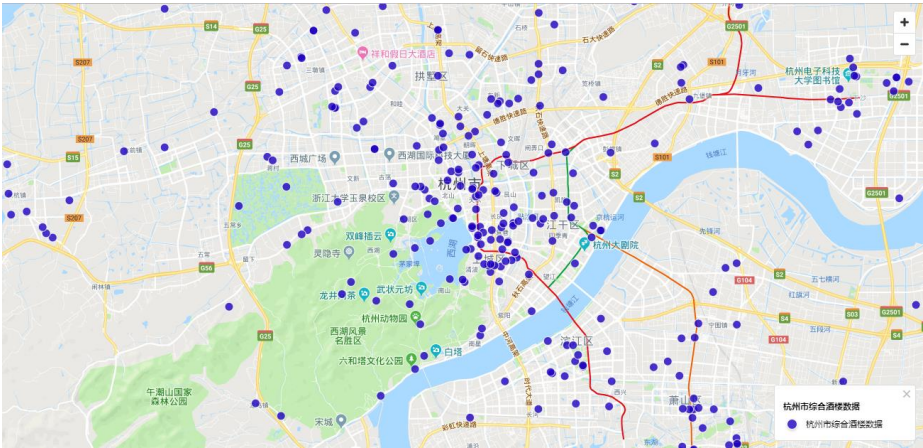


图 H-12 杭州市大型酒楼分布

附录 I 数据来源汇总

编号	数据量	描述	来源
1	939939	2017-2018 杭州市访客轨迹数据	个推移动大数据服务商

2	45224	杭州市基础设施分布数据	极海数据有限公司
3	547	2017.1-2018.6 西湖区人数(天级)	杭州市旅游局
4	21220	杭州市地铁实时数据	杭州市地铁公安分局
5	231	杭州知名景区评价与评分	大众点评网
6	2659	杭州全市实时游客数	杭州市旅游局
7	297	杭州市主要景点地理位置数据	百度地图
8	2075	杭州市购物场所热度数据	极海数据有限公司
9	107	景点综合评分数据	携程网
10	34315	杭州东站访客进出量	个推移动大数据服务商
11	34	访客来源地数据	个推移动大数据服务商
12	135	杭州行政区人口数据	中国统计网