

城市公共自行车租赁数据的可视分析方法^①

陈晓燕¹, 毛佳媚², 李泽界²

¹(绍兴职业技术学院 信息工程学院, 绍兴 312000)

²(浙江工业大学 计算机科学与技术学院, 杭州 310023)

摘要: 自行车已经成为广大民众公共出行的一种常见交通工具, 杭州市公共自行车系统分布范围已覆盖了整个市区. 研究公共自行车系统的可视化及分析方法, 需要查询每个站点的实时租赁数据及历史统计, 包含上周、昨天和今天的统计数据, 用户可以选取一些站点, 在同一个视图内显示它们的统计数据, 通过交互分析, 预测每个站点的借用趋势, 帮助管理人员充分理解和分析这些数据, 并了解这个区域内的整体趋势.

关键词: 公共自行车; 电子地图; 网页抓取; 交互分析系统

Visualization Method of City Public Bicycle Rental Data

CHEN Xiao-Yan¹, MAO Jia-Mei², LI Ze-Jie²

¹(College of Information Engineering, Shaoxing Vocational & Technology College, Shaoxing 312000, China)

²(College of Computer and Science Technology, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract: Bicycle has become a common transport means of the general public in public travel. The distribution range of Hangzhou public bicycle system has covered the whole Hangzhou. Visualization and analysis method of study of public bicycle system includes gaining each station's rental data by query, and historical statistic data including that of last week, yesterday and today. Furthermore, the users can select some stations and this system will show their statistic data in one graph in order to realize the trend of an area by interactive analysis to help people fully understand and analyze data and forecast each station's rental trend.

Key words: public bicycle system; electronic map; web crawl; interactive analysis system

在全世界范围内, 公共交通系统发展得如火如荼. 自行车是我们出行使用最广泛的交通工具. 三十年前, 北欧国家就倡导建立健康、经济和环保的自行车系统. 荷兰、比利时、西班牙、奥地利和德国这些年也开始效仿. 杭州市公共自行车系统的发展在全国范围内处于领先地位. 近些年来, 杭州市公共自行车系统从最初的 61 个服务站、2800 辆自行车和过 5000 的日租用量, 发展到 2 千多个服务站、6 万多辆自行车和过 30 多万日租用量. 2011 年, 在英国广播公司(BBC)旅游频道的投票中, 杭州市位列“8 个最佳公共自行车服务城市”之一.

作为城市短程非机动车载人设备, 自行车具有廉价、

便捷、健身、环保节能的优点. 它已经成为人们生活中重要的出行方式之一. 在中国, 它已被很多城市纳入到公共交通体系之中, 它可以解决“最后一公里”的问题, 能够实现“低碳旅行, 美化城市”. 在 21 世纪, 自行车应该成为公共交通的重要工具之一^[1].

信息可视化研究是一门抽象数据视觉化表示以加强人类认知的研究. 抽象的数据包括数值和非数值数据, 如文本和地理信息^[2]. 通过信息可视化研究, 我们可以深入了解, 找出问题的答案, 发现各种各样的关系, 也可能让我们发现在其他形式的情况下不容易找到的模式.

因此, 借助信息可视化^[3]和数据挖掘的优势, 使用

① 基金项目:浙江省高等学校访问学者教师专业发展项目(FX2012111)

收稿时间:2015-04-21;收到修改稿时间:2015-06-11

可视化分析技术^[4], 识别、描述和解释一个大规模、多属性的数据集内的自行车行为, 将有利于促进公共交通的发展. 本文研究旨在实现杭州市公共自行车系统的可视化, 帮助人们了解站点的分布情况和它们的实时状态, 更直观、更方便地获取各个自行车站点的相关信息. 本文研究的系统不仅可以查询和浏览每个站点, 还可以了解它们的历史统计数据来知晓其租借趋势, 实现公交自行车的有效管理, 具有重要的现实意义.

1 概述

1.1 数据来源

本系统的数据通过网页抓取获得, 来源是杭州市公交出行实时信息服务系统(<http://www.hzbus.cn>)和巴士地图(<http://wap.busditu.com>). 其中, 在杭州市公交出行实时信息服务系统中, 获取 2122 个站点的基本信息, 整理成一个单独的文件, 用于存储系统整体信息. 本文研究的系统所提供的实时查询功能, 是通过向巴士地图网站发送请求, 然后获取帮助. 每个站点的统计数据也是通过定时、定间隔、顺环地访问巴士地图, 解析 Html 页面后整理获得的.

每个站点的基本信息来自杭州市公交出行实时信息服务系统, 基本信息包括: 编号、名称、精度、维度、地址、总容量、服务时间、联系电话、服务项目和区域标号.

实时数据则来自巴士地图, 包括: 每个站点的可借量和可还量. 这部分的数据是定频采样集合, 采样时间间隔是 15 分钟, 数据空间是离散的.

数据本身具有时空特性, 包括空间维度、时间维度、属性维度三个方面. 但本次数据属性维度相对稳定, 只有容量(可借量和可还量)是变化的, 其余则是固定的. 由于在所抓取的杭州市公共自行车数据中, 缺少每辆自行车的轨迹数据, 所以无法可视化其运动轨迹和探索其行为模式.

1.2 系统概述

本文研究所开发的系统被设计成有两个部分: 网页抓取部分和交互式可视化系统部分.

网页抓取部分旨在收集每个公共自行车站的基本信息及其统计数据, 它也有两个部分: 基本信息抓取和实时数据抓取. 基本信息抓取是获得每个公共自行车服务站的基本信息. 实时数据抓取是实时获取每个

公共自行车服务站的数据--可借量. 累积存储在文件中来获得每个站点的统计数据.

交互式可视化系统部分由三个视图组成: 地图视图、统计视图和区域视图. 本系统的用户可以是普通公民或者是公共自行车系统的研究者. 本系统希望帮助人们更好地了解杭州市公共自行车系统, 为他们的日常出行提供便利, 实现的功能如下: (1)在地图上实现对每个公共自行车服务站的展现, 提供一个查询功能, 查看某个站点的实时数据; (2)实现对某个站点统计数据的展现, 提供一个浏览功能; (3)实现对某一区域内所有站点统计数据的展现.

2 网页抓取

2.1 基本信息抓取的设计

考虑到复用性, 该部分程序被设计为 WPF 窗口程序, 使用 .NET 的 WebBrowser 控件访问 web 页面. 它可以被用来复制 IE 浏览器的功能; 在加载完网页后, 让程序来解析 html 文本; 然后执行内嵌的 javascript 事件来进行页面跳转. 这样不仅可以利用超链接进行跳转, 还可以实现更为复杂的 HTTP 请求, 并且直观可控. 整个抓取过程利用计时器控制. 因此, 该部分基本实现了 IE 浏览器的自定义及网页抓取等功能, 并实现利用独立的配置文件保存页面访问、解析和存储的相关行为. 这样, 程序就可以根据杭州市公共自行车系统中站点变化, 更改抓取的规则来更新基本信息文件.

获取数据后, 它将被解析转换成为一种在数据每部分标识其使用目的的格式. 这个部分使用逗号分隔文件(CSV)保存其结果. 文件的每一行分别表示一个站点的信息. 每行的数据字段分别表示各个站点的属性.

2.2 实时数据抓取的设计

这个部分将被一直运行, 并被设计为使用 java.net.URL 类直接访问在互联网上的 URL 资源; 利用多线程加快获取数据的过程; 利用计时器控制整个程序及其执行过程. 该程序根据上面的站点基本信息文件, 每 15 分钟执行一次, 其抓取规则也独立保存在配置文件中. 同样地, 每次抓取后进行数据整理, 并且带上时间标签, 以逗号分隔文件(CSV)保存其结果. 数据存储文件中, 每一行以一个时间标签开始, 依次记录了所有站点在某一时刻的容量.

3 交互式可视化系统

这部分的主要工作是实现公共自行车服务系统的视觉展示. 它被设计为三个视图: 地图视图, 统计视图和区域视图. 这三个视图分别实现不同功能需求. 而且有一个提示框来解释如何使用本系统. 整个系统是单视图的, 需要在不同视图间切换, 此外给使用者提供丰富的交互与反馈.

3.1 整体设计

对于地理空间数据, 利用地图投影及采用点地图(dot map)的方式来表示公共自行车系统; 对于时间序列数据, 利用二维坐标轴分别来表示每个站点数据信息的时间维度和特征属性.

这个部分的设计基于 MVC 框架模式实现业务逻辑和数据显示分离的方法.

首先设计站点模型类来抽象和封装了每个公共自行车站点, 其拥有每个站点的基本信息、与应用程序相关的状态和一些相关行为(响应状态查询、应用程序功能和通知视图改变). 然后, 同理设计统计数据模型类来读取和存储来自网页抓取部分的数据, 分别表示今天的、昨天的和上周的各个站点的统计数据. 在系统运行时, 实例化每个自行车站点模型和三个统计模型.

然后, 我们设计一个控制类来负责定义应用程序行为, 将用户动作映射成模型更新, 并选择相应的视图. 接着, 我们设计了三个视图类负责解释模型, 通过模型更新请求绘制界面, 将提供丰富的事件响应, 发送用户请求给控制类, 以提供强大的交互功能. 整个结构如下图:

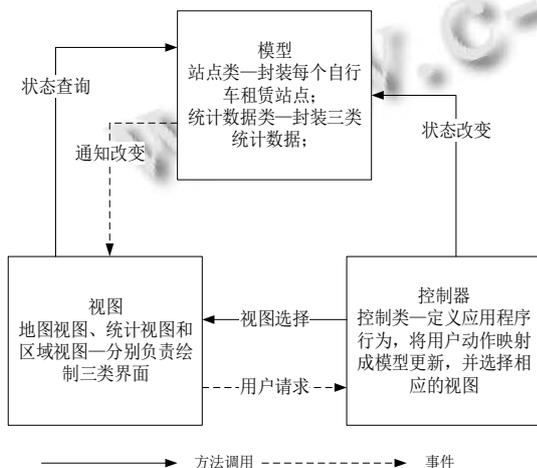


图1 交互式可视化系统的MVC框架

当更新数据或者产生视觉变化的时候, 不断地为用户显示数据和界面的转变是很重要的. 在值与值之间插入值能够帮助用户跟踪发生的变化以及产生的变化提供背景情况, 从而让“变成”改为“过度”到另外一个值. 为了这样做, 我们设计了一个插值器来实现插值, 并使用简单的物理方法. 利用过度前的原始值作为弹簧变形后的长度, 目标值是弹簧的原始长度, 并根据弹簧的力学物理公式(又称为胡克定律, $F = k * x$)和牛顿第二运动定律($F = m * a$)来模拟一个弹簧以实现插值, 其公式如下:

$$a = \frac{k * (t - v)}{m}$$

$$v_j = v_{j-1} + (Vel + a) * d,$$

其中, 将物体的质量 m 单位化为 1, 设弹簧的劲度系数为 k , 阻尼系数为 d , 长度的原始值为 v , 目标值为 t , 长度变化速度为 Vel , 并假设每次刷新时计算插值, 直到 v 近似于 t , v_0 等于 v , v_j 是插值结果, 表示弹簧当前的长度.

3.2 地图视图的设计

我们使用地图视图来可视化整个杭州市自行车系统的分布及使用情况.

3.2.1 界面设计

地图视图中, 我们将各个站点绘制在地图上, 每个站点用地图上的一个圆点来表示, 并利用圆点的颜色来表示每个站点的拥有量(当前可借量); 而其大小会随着地图缩放比例进行适宜地缩放, 以优化用户体验不表示任何属性(见图2). 所选的配色方案中有6个颜色值: 左边的颜色代表了该站点可借量小于50%; 右边的颜色代表了其可借量超过50%(见图3).

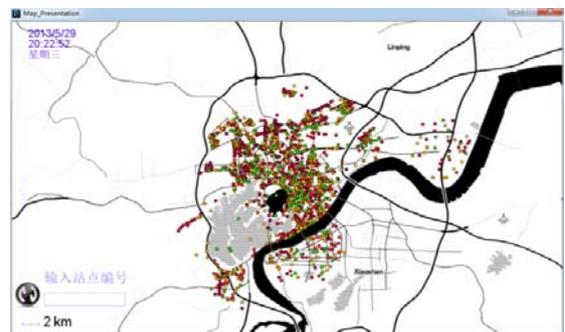


图2 利用颜色反应当前每个站点的拥有量情况



图3 配色方案

该视图允许用户自行选择是否利用圆点的颜色来表示每个站点的当前可借量, 避免视觉疲劳, 以方便用户选取每个站点及其他操作. 并且本视图允许用户切换地图图层(平面图、卫星图和墨粉地图), 来方便用户以不同视角进行观察, 例如平面图方便用户从平面角度了解站点分布路线, 卫星地图可以帮助用户了解真实的地理信息, 墨粉地图可帮助用户了解站点集中区域(见图 4).

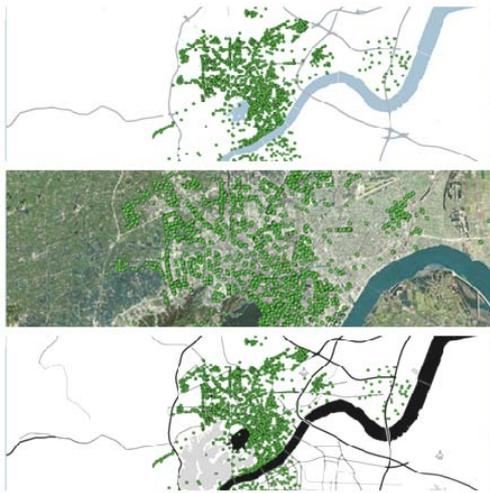


图 4 三种图层分别表示平面图(上)、卫星图(中)和墨粉地图(下)

地图视图的左上角用于显示时间和日期, 左下方有四个部分: 提示信息(用来提供交互反馈的), 输入框和比例尺和图片按钮(用于切换查询方式).

3.2.2 交互设计

地图视图提供了基本的地图操作: 缩放、拖拽. 地图视图的左下角有四个部分: 提示信息, 输入框和比例尺和图片按钮(用于切换查询方式), 用来实现站点查询的功能. 关闭利用颜色表示站点的当前可借量后, 用户可以通过站点的编号、名称或地址以进行查询(见图 5).

为了让用户意识到变化, 我们同样利用插值来实现, 当用户通过编号进行查询, 并键入一个数字后, 那些符合查询条件的每个站点的点将会变亮; 而那些不符合的点将会变暗. 我们根据上面提到的插值器, 重新设计颜色插值器. 为了避免遮挡, 那些符合的点覆盖在上面; 当符合的站点数量只有一个时, 系统将显示该站点实时查询的结果, 其包括所有基本信息和

当前可借数量. 如果用户通过名称或地址进行查询, 将会在视图上显示匹配的各个站点的名称.

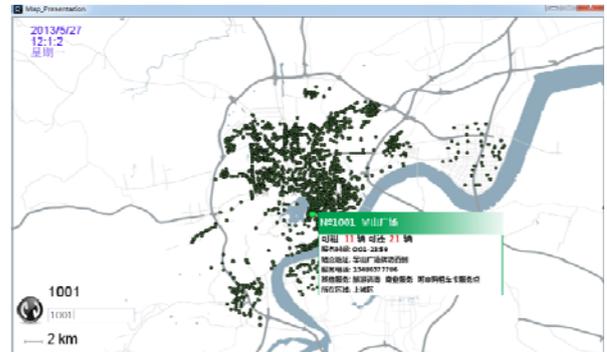


图 5 通过编号查询的效果

在地图视图中, 单击某个站点后, 将在视图上显示其实时查询结果. 双击某个站点后, 将切换到其统计视图(见图 7). 为了方便用户观察在同一区域内的站点集合, 用户可以选择一些站点后, 切换到区域视图(见图 8).

由于地图的比例尺无法给用户更为直观的距离感, 我们设计了地图距离计算工具. 用户将鼠标移到某站点上, 本系统将以该点显示一组同心圆, 其直径分别为 100 米, 300 米或 500 米(见图 6). 这可以显示该站点周围的距离范围, 增强用户在地图上的距离感.



图 6 移动鼠标的效果

利用 unfolding 地图的 API, 在当前地图缩放级别下, 设计了获取像素距离与实际距离之间关系的算法, 如下:

- ①根据地图 API 的缩放级别, 定义一个数组来表示各个缩放级别下比例尺的单位距离;
- ②通过当前缩放级别, 计算出最接近的比例尺单位距离 l ;
- ③设一个地图上起始点 s , 利用地图 API 计算获得其在水平正向相距 l 的目标点 d , 计算得出两点在屏幕

上的像素坐标距离 L ;

④根据像素坐标距离 L 和当前缩放级别下的比例尺单位距离 l , 返回它们的比值, 即为单位实际距离对应像素距离;

3.3 统计视图的设计

我们使用统计视图来可视化自行车系统中每个站点的基本信息及使用情况.

3.3.1 界面设计

统计视图中, 将折线图和柱形图相结合, 来显示每个站点的租借统计数据. 其中, 我们首先绘制柱形条来表示从今天开始的过去几个小时内的统计情况, 然后在其上先后绘制两条折线分别表示昨天的和上周的统计情况. x 轴表示一天中的 24 小时, y 轴表示每个站点的自行车拥有量(见图 7). 用户可以结合柱形条和折线, 来比较这些统计数据, 预测在接下来的几个小时内各个站点可借量的变化趋势. 统计视图的左上角用于显示时间和日期; 中央是绘图区, 其顶部是该站点的编号和名称, 右侧是图例; 中间利用柱形图和折线图来表示数据.

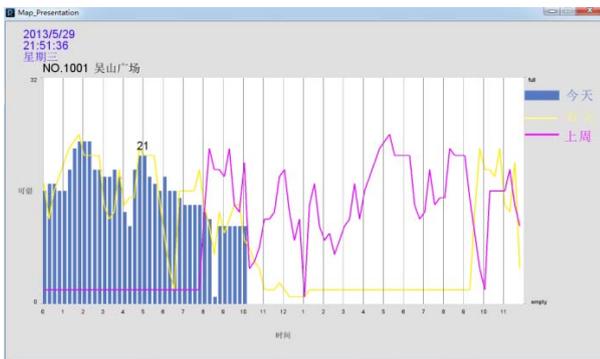


图 7 统计视图

3.3.2 交互设计

用户可以通过键盘操作浏览每个公共自行车服务站的统计数据及返回到地图视图. 由于切换各个站点的统计视图会产生视觉疲劳, 为了增强用户对于变化的感知, 我们利用上面提到的插值器, 在值与值之间插入值帮助用户跟踪发生的变化以及产生的变化提供背景情况, 根据弹簧的力学物理公式和牛顿第二运动定律来模拟一个弹簧以实现插值(具体公式见本文 3.1 节), 为柱形条和折线附加弹簧渐变效果, 使得在浏览各个站点的时候, 产生动态变化效果. 在统计视图中, 用户将鼠标移到柱形条上, 系统会显示该柱形条所代

表的数值(见图 7). 用户可以根据站点的统计视图预测分析站点的借用趋势.

3.4 区域视图的设计

我们使用区域视图来可视化自行车系统中某一块区域内所有站点今天的使用情况.

3.4.1 界面设计

区域视图是用来显示同一区域内一些站点从今天开始的过去几个小时内的统计数据(见图 8); 它利用堆叠柱形图来显示该选中区域内所有站点的可借数量, 并用不同颜色来区分不同的站点, 其中 x 轴表示一天中的 24 小时, y 轴表示这些站点的自行车拥有总量. 区域视图的左上角用于显示时间和日期; 中央是绘图区, 其顶部是该站点的编号和名称, 右侧是图例; 中间利用堆叠柱形图来表示数据.

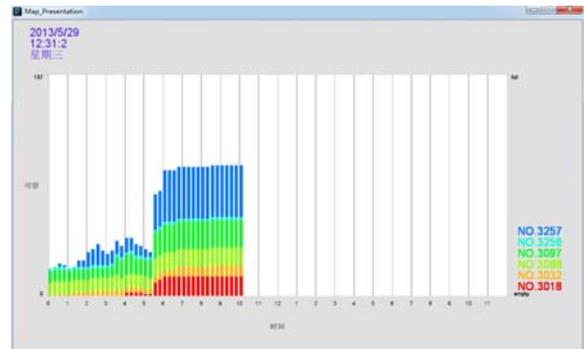


图 8 区域视图

3.4.2 交互设计

用户可以通过键盘操作浏览每个公共自行车服务站的统计数据及返回到地图视图. 区域视图中, 用户将鼠标移到柱形条上, 系统会显示该柱形条所代表的数值.

3.5 系统实现

交互式可视化系统部分的操作流程图见图 9.

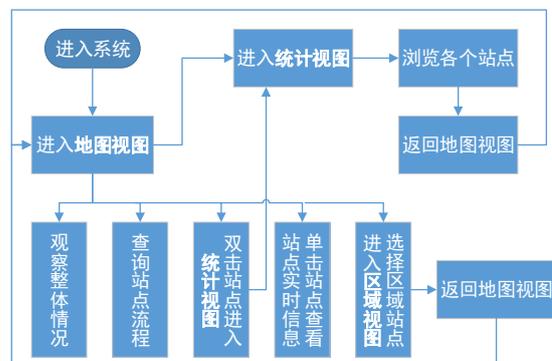


图 9 交互式可视化系统部分流程图

它是利用 Processing 的 java 库以及 Unfolding 库。Processing 是一个开源的编程语言和编程环境,建立在 Java 语言的基础之上,但使用简化的语法和图形编程模型。Processing 一开始是面向数字艺术家而开发的绘图软件,它通过封装底层的图形操作,使得语法简单易用。Processing 支持许多现有的 Java 语言架构,因此其完成的作品可在个人本地端运行,也可以以 Java Applet 的模式发布到网上。我们使用 Processing 的 java 库,并在 eclipse 平台结合 Swing 进行开发。而 Unfolding 是一个用来在 Java 和 Processing 上创建交互式地图和地理可视化的库。它像我们提供了丰富的地图类型以及交互事件。这部分系统还提供一个提示框来向用户解释系统的操作细节(见图 10)。

杭州公交自行车系统可视化展示

1. 最初的地图视图显示了所有租赁站点,并以圆形的边框颜色深浅代表了该站点的容量。颜色越深表示满的,越浅表示空的。通过鼠标可以控制地图,如滑动滚轮放大,拖动地图。
2. 在地图视图中,点击左下方的按钮切换查询方式,直接敲击键盘输入编号、站名和地址进行搜索。其他将鼠标滑到某站点上,将显示其编号和站名;单击显示其基本信息和实时租赁数据;双击显示其统计视图。用鼠标右键选取一些点后按下“z”键,将显示区域统计视图;按下“<”和“>”键可以切换地图,按下“f”键也可以进入视图。
3. 在统计视图中,通过“[”键和“]”键向前向后浏览各个站点;其中显示每一个站点今天、昨天和上周的数据;通过鼠标移动到柱状条上可以显示数值;按下“m”键返回到地图视图。
4. 区域统计视图中叠加了各个站点今天的数据;按下“m”键返回到地图视图。
5. 按下“h”键将弹出使用说明的提示框。

图 10 提示框的显示效果

4 结语

随着可视化技术的发展,信息可视化(或数据可视化)在现实生活中扮演着越来越重要的角色。本次研究的目的是可视化杭州市公共自行车租赁数据,帮助人们更好地了解和获取关于杭州市公共自行车系统的相关信息,甚至为规划其建设提供帮助。在接下来的研究中,将进一步采集自行车的租赁轨迹数据,可视化自行车流量轨迹,探索其行为模式,为公共交通发展提供参考。

参考文献

- 1 Demaio PJ. Smart bikes: Public transportation for the 21st century. *Transportation Quarterly*, 2003, 57(1).

- 2 Munzner T. Process and pitfalls in writing information visualization research papers. *Information Visualization*. Springer Berlin Heidelberg, 2008: 134–153.
- 3 Wood J, Slingsby A, Dykes J. Visualizing the dynamics of London's bicycle-hire scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 2011, 46(4): 239–251.
- 4 Beecham R, Wood J, Bowerman A. A visual analytics approach to understanding cycling behaviour. 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE. 2012. 207–208.
- 5 Aldred R. On the outside': constructing cycling citizenship. *Social & Cultural Geography*, 2010, 11(1): 35–52.
- 6 石晓凤,崔东旭,魏薇.杭州公共自行车系统规划建设与使用调查研究.城市发展研究,2011,18(10):105–114.
- 7 张丽珺.公共自行车交通发展研究[硕士学位论文].西安:长安大学,2011.
- 8 Lathia N, Ahmed S, Capra L. Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, 2012, 22: 88–102.
- 9 von Landesberger T, Bremm S, Andrienko N, et al. Visual analytics methods for categoric spatio-temporal data. 2012 IEEE Conference on Visual Analytics Science and Technology (VAST). IEEE. 2012. 183–192.
- 10 Fry B. *Visualizing data: exploring and explaining data with the processing environment*. O'Reilly Media, Inc., 2008.
- 11 Yau N. *Visualize this: the FlowingData guide to design, visualization, and statistics*. Wiley.com, 2011.
- 12 Lliinsky N. *Beautiful Visualization: Looking at data through the eyes of experts*. O'Reilly Media, Inc., 2010.
- 13 孙国道,梁荣华,何贤国,等.高维时空房地产数据的可视分析.计算机辅助设计与图形学学报, 013, 25(8):1169–1176.
- 14 蒲剑苏,屈华民,倪明选.移动轨迹数据的可视化.计算机辅助设计与图形学学报,2012,24(10):1273–1282.