

高密度群体轨迹数据的微观可视分析

朱素佳, 孙国道^{*}, 江槑, 夏旺, 梁荣华

(浙江工业大学计算机科学与技术学院 杭州 310023)
(guodao@zjut.edu.cn)

摘要: 人类活动会持续产生大量时空轨迹数据, 探索和理解数据中的隐藏模式可帮助理解人类生产生活的规律、提高管理效率。前人工作大多关注于轨迹数据的宏观可视分析, 缺乏对高密度小区域内的个体行为展开深入分析。针对高密度群体轨迹数据, 提出多视图结合的可视分析方法以探索和分析其微观演化情况。该方法主要包括 4 个部分: 基于 small-multiples 的特征统计视图, 展示轨迹数据在时间和空间维度上的特征分布; 采用多分辨率河流视图支持从多层次聚合角度展示轨迹数据的变化; 采用轨迹投影视图揭示高密度群体轨迹间的相关性; 采用多层次空间地图以展示轨迹数据的具体分布情况。上述方法形成的可视分析系统已用于 2 个高密度群体轨迹数据集的分析中, 为后续高密度群体轨迹数据的可视分析提供一种解决思路。

关键词: 时空轨迹数据; 高密度群体; 微观演化; 河流图聚合

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2020.18206

Microscopic Visual Analysis of High-Density Group Trajectory Data

Zhu Sujia, Sun Guodao^{*}, Jiang Qi, Xia Wang, and Liang Ronghua

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

Abstract: People's daily life is surrounded by Spatio-temporal trajectory datasets. Exploring and understanding the hidden patterns can help us grasp the characteristics of human productivity and improve management efficiency. Most previous works focused on the macroscopic analysis of trajectory datasets and lacked in-depth analysis of individual behaviors within high-density trajectory datasets in a small region. In this paper, we proposed a visual analysis system to analyze the high-density trajectory datasets. The method consists of four components, a small-multiples based feature statistical view for showing the features of various dimensions within dataset, a glyph-embedded multi-stream view for displaying overall evolution patterns and highlight critical modes, a projection view for illustrating the relative correlation among trajectories, and a multi-mode hierarchical spatial map for presenting detailed characteristics in the dataset. The interactive system has been used in two trajectory datasets to provide a solution for visually analyzing high-density trajectory datasets.

Key words: trajectory dataset; high-density population; micro-evolution; aggregated stream graph

人类的生产活动会持续产生大量的时空轨迹数据, 充分挖掘时空轨迹数据中隐藏的模式, 能够

帮助了解人类生产和生活规律。许多学科的研究都涉及了时空轨迹数据, 例如, 经济学家基于移民

收稿日期: 2020-01-10; 修回日期: 2020-07-01. 基金项目: 国家自然科学基金(61972356); 浙江省重点研发计划(2019C01009). 朱素佳(1992—), 女, 硕士研究生, 主要研究方向为信息可视化; 孙国道(1988—), 男, 博士, 副教授, 论文通讯作者, 主要研究方向为信息可视化; 江槑(1997—), 男, 硕士研究生, 主要研究方向为信息可视化; 夏旺(1997—), 女, 硕士研究生, 主要研究方向为信息可视化; 梁荣华(1974—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为可视化和计算机图形图像处理。

流量数据研究了地域经济变化，疾病学家基于人口流动数据研究了疾病在地区和国家之间的传播。

轨迹数据的可视化工作一直以来都是可视化领域的一个重要的挑战，它不仅需要从宏观角度展示时空上的情境，还需要从微观的角度展示某时刻的多维信息。分析轨迹数据的微观变化对资源精细化调整、评估资源精准配置的合理性以及细微模式发现至关重要。许多研究人员对时空轨迹数据的研究重点是从宏观的角度进行探索^[1-3]，然而缺乏对高密度小区域内的个体行为展开深入分析^[4-5]。例如，Chu 等^[6]将轨迹记录点都映射为最近的街道名称，使用文本分析方法提取出租车数据的话题，可视化道路中话题的时序演化。Furnas^[7]提出 TripVista 来从多维度探索道路交叉口的微观交通模式和异常行为，然而该方法不能同时分析多个交叉口的交通数据的异常行为，并且它不能展示频繁移动的高密度群体轨迹之间的潜在关系。

因此，本文提出一种多视图结合的可视分析方法，以探索和分析高密度群体的微观演化情况。本文方法主要分为 4 个部分：基于 small-multiples 的特征统计视图展示轨迹数据在时间和空间维度上的特征分布；轨迹投影视图采用多维缩放 (multiple dimensional scaling, MDS) 算法，将轨迹数据投影至二维平面，然后利用基于密度的噪声应用空间聚类 (density-based spatial clustering of applications with noise, DBSCAN) 算法，结合字形表达生成的簇内特征，以及套索和弹出框等交互式技术展示群体间的差异性和相似性；多分辨率河流视图结合内嵌图元帮助用户快速发现重点信息以及上下文模式；空间层次地图包含 3 种模式：空间层次分布模式展示空间划分的等级，轨迹图模式展示包括具体的轨迹路径、方向、速度、逗留时间等微观数据，热力图模式展示数据在特定语境下的空间分布情况。

本文首先根据相关文献总结设计需求，然后设计可视化组件以及交互式工具，从不同的角度进行细粒度的时空轨迹数据分析，并以场馆内参会者真实的轨迹数据为例，帮助用户探索时空轨迹数据中的潜在模式。

1 相关工作

本文采用多个视图来展示高密度轨迹数据的微观轨迹演化情况。

1.1 宏观轨迹数据可视化

在轨迹数据的可视化领域，文献[8-9]提出 3 种通用的轨迹数据可视分析方法：直接绘制、模式提取和视觉聚合。直接绘制方法能够展示大量的数据信息，但是会造成视觉混乱，并给用户带来认知负担；模式提取方法采用自动分析来提取数据^[10]；视觉聚合^[11]是将运动轨迹进行聚合，以揭示更高水平的动态图。另一些工作对区域间的行为变化进行处理和可视分析，Liu 等^[12]研究位置之间的路径多样性，并提出形似时钟的径向布局来显示时间统计分布。Guo^[1]和 Adrienko 等^[2]对地理区域进行分区，进而将数据处理为区域间的移动，在精简了轨迹数据的同时保留了数据的主要特征。Lu 等^[3]设计了 ODWheel 以分析起始点到终止点的动态变化。Zeng 等^[13]从公共交通系统中起始点开始将移动的路线进行可视化，并对路线之间进行比较。基于以上方法，本文将空间区域进行多层次划分以支持用户不同聚合层次的探索，通过路线之间的比较发现个体间的相似性和差异性。

Wang 等^[5]提出一种基于道路的查询模型，其关注单条道路的状态，查询经过该道路的出租车的轨迹，并显示其速度、流量、密度的时空分布。然而当所关注路段的车辆速度和密度等呈现双峰分布时，该路段的统计属性会给用户带来误导。另外，该方法不能检测离群点数据，而交通事故的发生经常发生在离群点之间。所以从微观的角度探索轨迹数据，对资源精准配置以及精细化的调整意义深远(如定位具体的事故和拥堵的根源)。

1.2 微观模式可视化

少数可视分析方法研究轨迹的微观模式，例如，Anwar 等^[14]提出 Traffic Origins，以分析从微观层面上观察交通事故和由此引起的交通拥堵之间的视觉关系。其用一个扩大的圆圈标记事件的位置，以发现潜在的交通流图，并在结束时，该圆圈消失。这不仅可以将注意力转移到即将发生的事件上，还可以观察交通事故对附近车辆流量的影响，以及多起事故对道路网络的连锁影响。然而该方法用于较为密集且频繁移动的群体轨迹时，会由于密度较大导致自动识别群体为异常事件，这时标记事件位置的圆圈也会布满整个界面，导致可读性较低。Chu 等^[6]使用了隐含狄利克雷分布(latent Dirichlet allocation, LDA)的文本分析方法，提取了车辆数据的“话题”，每个话题以一定的概率出现在道路中，每条车辆轨迹以一定的概率出现在每一个话题上。其中该工作通过火花云

(spark cloud)方式显示了话题在轨迹中的分布。然而该方法用于较为密集且频繁移动的群体轨迹时, 会由于话题转化频繁和轨迹密集而导致视觉杂乱。

Furnas^[7]提出 TripVista 来分析一条道路交叉口的交通数据, 从时间、空间等单一维度或从多个维度的角度分析了微观交通模式和异常行为; 然而该方法不能同时分析多个交叉口的交通数据, 而且当个体间较密集且在小范围内有规律地频繁移动时, 该工作不能展示轨迹之间的潜在关系。本文提出的对高密度群体轨迹数据的微观可视化, 不仅能够在地图视图中展示轨迹数据的位置、速度、逗留时间等微观信息, 从而推测人员类型, 还能在投影视图中展示轨迹之间的相似度, 从而推测轨迹之间的潜在关系。

1.3 时序流可视化

时态数据中的模式通常出现于不同的聚合度上, 例如, 几天、几周和几个月等, 这使得基于时间聚合数据的可视化具有挑战性。通常使用矩阵图和流图表示特定时间或者间隔时间段内的流量, 并通过滑块的功能帮助用户探索每一个时间段内的数据。Flowstrates^[15]和 MapTrix^[16]将时空轨迹处理为 OD(origin destination)数据来描述时间序列上的流量变化, 并将轨迹起点和终点嵌入不同的地图中, 使用热力图映射时间序列上的流量, 以关联热力图中像素与对应的地图区域。在流的控制层

中, Flowstrates 使用矩阵的一行代表流; MapTrix 集合矩阵每个单元格中的流, 可以更好地聚合像素。采用颜色映射流量信息, 可给人直观的感受, 但是这种抽象的可视编码很难精确地表达每个像素的具体流量, 可视化长时间的流量会降低可读性。Byron 等^[17]使用堆积图表达了数据的时序变化, 时序堆叠表示定量值的演变, 而随着时序流量的增加, 视觉效果会降低。因此, Morrow 等^[18]基于时间的图表提供焦点和上下文, 以实现跨时间刻度的模式解释。Xu 等^[19]设计了 EvoRiver 以可视化社交媒体中主题复杂的动态演化, 进而分析主题之间合作与竞争关系。而传统的堆栈图不能展示具体条带间的变迁, 所以孙国道等^[20]提出嵌入式线条堆栈图可视化城市群在时间上的演变, 以展示一个城市如何转换其所属的城市群。

当时间跨度较大时, 传统的堆栈河流图不能同时展示高密度群体轨迹的概览和细微变化^[7]。本文提出从不同空间聚合角度探索高密度群体的演化, 拓展了 MultiStream^[21], 以替代传统堆栈图, 并结合嵌入式图元来解决此问题。同时, 本文采用焦点+上下文交互技术^[7,22], 以展示轨迹数据的动态聚合, 并结合视图之间的协同交互来帮助用户探索高密度微观群体演变。本文结合空间层次地图展示详细轨迹信息, 解决了可伸缩性问题, 能够直观地表达高密度群体在时空上不同层次的聚合信息。图 1 所示为本文系统概览图。

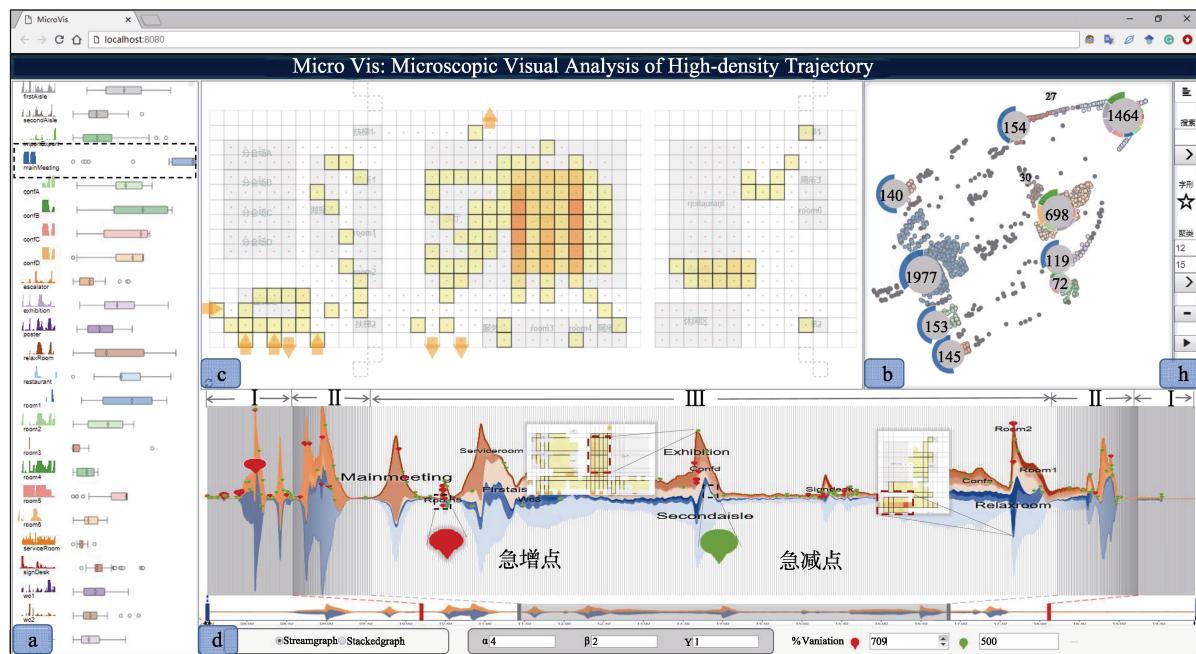


图 1 系统概览图

2 需求分析

本文首先对相关高密度群体轨迹数据可视分析文献进行调查分析，并总结出设计需求：

R1. 展示轨迹数据在不同维度上的分布情况。高密度群体的轨迹数据具有相似度较高的特征，因此分别从时间和空间维度进行可视分析，能够整体掌握轨迹数据的分布特征。

R2. 展示各个群体之间的时序差异性以及群体的演化趋势。高密度群体之间不可避免地会产生集群，因此探索群间的差异性对把握轨迹数据之间的相对关系至关重要。

R3. 探究高密度群体内部个体之间的差异。了解群体内部个体的特点有助于发现潜在的紧急事件，如分析移动速度、方向和位置等能够帮助相关人员确定潜在的紧急情况，以便更好地应对。

R4. 关注个体的详细轨迹信息。由于异常多发于少数的个体之间，因此展示个体的原始轨迹数据有助于用户从细节上保护个体的安全。

R5. 探索潜在的异常事件。用户对大量的轨迹数据进行探索，观察其大量的微观信息会使得工作繁琐。因此，本文定义异常事件类型来预处理潜在的异常事件，缩短用户探索事件。

针对 R1~R5 的用户需求分析，提出 S1~S4 的解决方法来指导可视化视图的设计。

S1. 基于 small-multiples 的特征统计视图。该视图对轨迹数据在时间和空间维度上的特征分布进行可视化，直方图可视化轨迹数据的时序分布特征，盒图可视化轨迹数据的空间分布特征；所以用户能够快速地概览轨迹数据的特征分布。

S2. 基于 MDS 的轨迹投影视图。该视图对个体之间的相对关系展开可视分析，帮助用户理解集群的特征。

S3. 多分辨率河流图。该视图对时间和空间进行聚合，同时在多层次聚合度下展示数据的时序演化趋势，在一定程度上解释群体的移动规律。

S4. 多层次空间分布地图。该视图设计了多种模式分别对群体的分布和个体移动的具体信息(速度、方向、位置等)进行可视化。

3 系统框架

基于以上设计需求，本文设计了一个可视分析框架来探索高密度群体轨迹数据的演化趋势。该框架包括数据处理、可视化组件设计和交互设计。

(1) 数据处理模块。该模块主要实现数据处理和分层。因为本文任务包括将轨迹数据进行统计分析以及概览轨迹数据之间的关系，所以本文通过计算空间的时序流量来揭示数据的时空特征，以及计算轨迹之间的相似度来揭示数据之间隐藏的相互关系。由于对大规模轨迹数据的可视化存在存储量较大和查询方式较慢，导致交互式可视分析无法进行，因此对轨迹数据进行空间分层能够降低系统计算成本、提高系统工作效率。

(2) 可视化组件设计模块。该模块包括了基于 small-multiples 的特征统计视图，基于 MDS 的轨迹投影视图，空间层次地图、具有分层空间结构模式、热力图模式、原始轨迹路线模式，多分辨率河流图和参数设置面板。通过组件之间的协同工作来有效地辅助用户理解数据分布，概览轨迹数据关系，探索微观动态演变，观察详细轨迹数据。

(3) 交互设计模块。该模块设计了丰富的交互，用户可以通过这些设计探索系统中的微观信息。轨迹投影视图设计了聚类参数可调、套索、弹出框等。多分辨率河流图基于 MultiStream 的交互，设计了阈值可调的嵌入式图元，视图间筛选交互等。

4 可视分析与设计

基于以上的分析，本文提出对高密度轨迹数据的可视分析系统，本节将对视图进行详细描述。

4.1 基于 small-multiples 的特征统计视图

为了引导用户，系统对轨迹数据根据空间区域进行四分位统计处理，将每个区域的数据分为最小值、中位数、最大值以及离群值等。采用如图 2 中 a 区域所示直方图可视化每个特征的时序流量，采用如图 2 中 b 区域所示盒图可视化数据的空间分布特征，将两者结合能够直观地展示轨迹数据在时间和空间维度上的分布，帮助用户理解数据。对于没有先验知识的用户可以从统计图中获得数据的时空特征分布，从而选择感兴趣的特征。

4.2 空间层次分布地图

空间层次分布地图系统中设计了 3 种模式，结合视图间的交互，帮助用户对群体轨迹数据从概览其分布到具体理解其详细信息。

(1) 轨迹空间层次分布模式。层次结构是根据空间的属性进行划分(如结构属性、功能属性)。以本文处理的数据为例，从原始数据的空间维度出发，最小单元为第 1 层(如图 3 中 c 区域传感器检测到

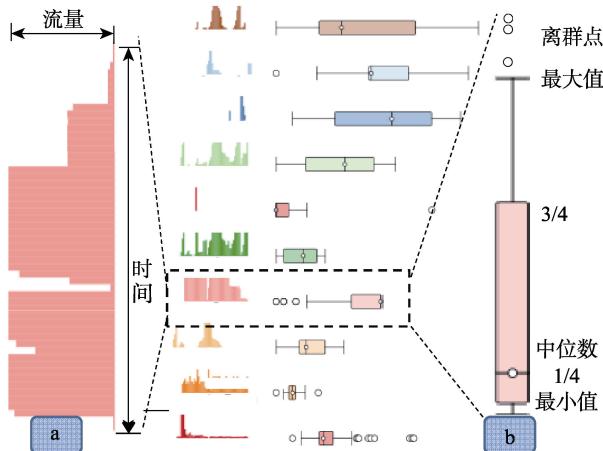


图 2 特征分布统计视图

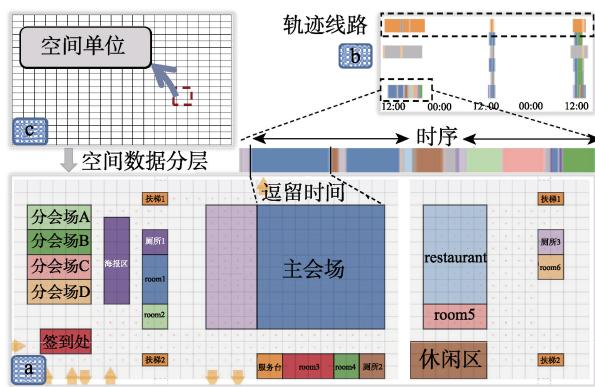


图 3 空间属性可视化映射

的空间区域所示), 按照功能属性划分为第 2 层(如将案例 1 中将场馆划分为功能不同的 24 个区域), 按照结构属性划分为第 3 层(如将功能不同的 24 个区域分为 1 楼和 2 楼). 用户根据空间结构信息, 在多分辨率河流图中指导其探索感兴趣的区域.

(2) 轨迹热力图模式. 空间层次分布地图能够较直观地展示轨迹的分布, 但展示所有轨迹数据会造成视觉杂乱的现象. 所以本文设计了如图 1 中 c 区域所示的热力图模式, 从空间的角度显示高密度群体整体分布情况; 采用渐进式颜色映射连续时间段内的数据量, 颜色越深, 表示数据量越大. 点击该区域查看相关轨迹在投影视图中的分布.

(3) 轨迹图模式. 该模式将轨迹数据的原始数据进行可视编码, 包括路径、方向、速度和停留时间等. 如图 4 虚线所示不同的颜色编码轨迹不同的日期, 线条重叠表示 2 天及以上具有相同的轨迹路线. 本文采用圆映射人员在相应的位置停留的时间, 采用三角形映射人员的移动速度, 如图 4 中的 b 区域所示, 半径越大, 表示停留时间越长; 三角

形越大, 表示移动速度越快. 提示栏显示个体的状态, 采用动画的方式显示相关个体的移动方向和速度.

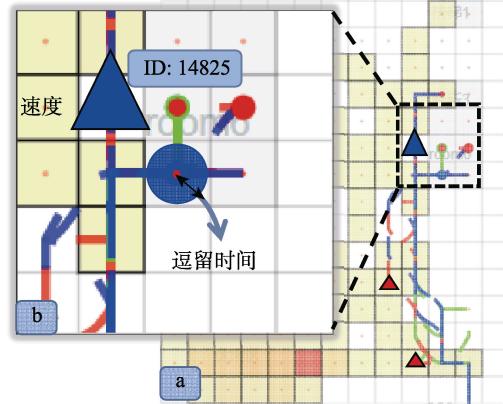


图 4 个体的微观轨迹信息可视化

4.3 基于 MDS 的轨迹投影视图

由于本文任务包括识别集群及群体特征, 因此使用了 Levenshtein 距离^[23]计算轨迹之间的相似度. 此度量标准是将 2 个序列之间的距离定义为将一个序列转换为另一个序列所需的插入、编辑和删除操作的次数. 若 a, b 的轨迹序列分别为 $S_a = \{a_1, a_2, a_3, \dots, a_i, \dots\}$ 和 $S_b = \{b_1, b_2, b_3, \dots, b_i, \dots\}$, 那么 a 和 b 之间的 Levenshtein 距离为

$$l_{a,b}(i,j) = \begin{cases} \text{if } (\min(i,j) = 0) : \max(i,j) \\ \text{else } \min \begin{cases} l_{a,b}(i-1, j) + 1 \\ l_{a,b}(i, j-1) + 1 \\ l_{a,b}(i-1, j-1) + [a_i \neq b_j] \end{cases} \end{cases} \quad (1)$$

其中, i 和 j 分别代表时间段内的事件序列下标. 如果 2 个事件相同, $a_i = b_j$, 则 $[a_i \neq b_j] = 0$; 否则 $[a_i \neq b_j] = 1$.

本文基于式(1)的相似度距离计算, 采用 MDS 算法, 将轨迹数据降维并投影到二维平面, 如图 5a 所示, 散点图中的点代表一条轨迹, 2 点之间的距离代表个体之间的相似度, 距离越小, 代表相似度越高, 关系越密切. 投影视图帮助用户从概览的角度了解群体之间的相似性与差异性, 用户可以通过点击对应的点查看该轨迹的具体路线. 但是在探索集群特征时, 该操作方式不仅需要遍历所有的集群, 还需要记忆操作过程, 效率较低, 给用户带来负担. 因此, 本文设计字形来展示集群特征. 本文利用 DBSCAN 聚类算法发现集群, 如图 5b 所示. 定位集群的中心点, 如图 5c 所示. 复杂的字形能够展示较为丰富的信息, 但在视觉上给用户带

来认知负担, 因此本文基于图 5d 所示南丁格尔玫瑰图设计字形。内圆半径代表簇的大小; 扇形的颜色代表特征, 扇形的大小代表空间特征的贡献度。用户通过字形概览数据的整体分布, 获得集群的特征。考虑用户不同的需求, 系统支持用户交互式设置面板的参数, 以获得适合的集群模式。

进行个体之间的比较, 为了缓解视觉上的混

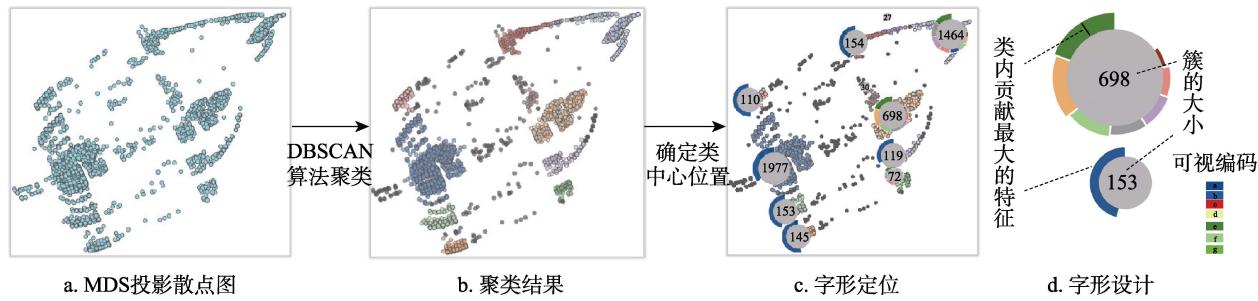


图 5 字形位置确定过程

4.4 多分辨率河流图

本文任务包括在不同聚合层次上展示数据在时间和空间维度上的信息, 突显特殊模式(如峰值), 所以本文使用多分辨率河流图展示时空数据在不同聚合层次上的演变趋势。它的层次结构数目是由轨迹数据的空间分层结构决定的, 可以由多个层次结构组成。使用较低层次且拥有较高分辨率的结构可查看局部详细信息, 采用较高层次且分辨率较低的结构可概览整体。为了使得层次结构的转变过程缓和, 在高、低数据结构中间加入缓和层。在多分辨率河流视图中显示了 2 个层次的结构, 最高和最低层次; 而实际中的层次结构则是多层。因此为了处理多层次的结构数据, 提高系统的可扩展性, 本视图在上下文区域中显示高层次结构, 在详细层中显示最低层次结构。

多分辨率河流视图展示了不同时空聚合度的时序演化情况。通过拖动如图 1 中滑动条的操作方式, 观察最低层次结构下的数据演变的细节信息, 效率较低, 当时间序列较大时给用户较大的负担, 不能精确且快速地筛选到特殊目标信息。因此本文基于 MultiStream 设计如图 1 黑色框所示嵌入式图元, 来表示急剧增加/减少的特定时刻。

若当前时刻数据为 C_X , 前一时刻 X 区域的数据为 P_X , X 区域数据变化为

$$P(X) = \frac{(C_X - P_X)}{P_X} \times 100\% \quad (2)$$

根据用户设定的急增/急减阈值(分别为 T_{up} 和

T_{down}), 像素轨迹图展示了不同轨迹的时序信息, 将空间属性用颜色进行可视化映射, 如图 3 中的 a 区域所示色块的颜色代表空间区域。如图 3 中的 b 区域的像素轨迹图所示, 横坐标为时间, 纵坐标为个体编号, 并用色块的颜色和宽度分别映射区域和逗留时间; 其中的黑色框区域展示了轨迹个体的高层次轨迹。

T_{down}), 确定图元的位置为

$$\begin{cases} L_T = T_{up}, & P(i) > T_{up} \\ L_T = T_{down}, & |P(i)| < -T_{down} \end{cases} \quad (3)$$

系统根据获取的急增/急减点的信息, 在河流视图的对应区域嵌入图元。用户通过图元的位置快速且准确地捕捉到关键信息节点, 缩减探索时间。多分辨率河流图中具有以下功能:

(1) 层次聚合。为了探索高密度群体轨迹数据在不同的聚合层次上的时序演变趋势, 系统支持用户浏览层次结构。如图 1 中, 高聚合层 I 聚合无关数据; 过渡层 II 提供上下文信息; 详细层 III 聚焦重要数据, 探索关键点的详细信息。

(2) 高亮图元。为了提高用户探索效率, 快速发现数据中隐藏的模式, 系统自动突显数据的关键点, 支持用户自定义调整阈值并定位出急剧增加/减少的区域和时间段。其中红色图元表示高于设定的上升阈值, 绿色图元表示高于设定的下降阈值。由分析可知, 图元常出现于峰值或低谷处前。所以图元不仅具有提高探索效率的作用, 而且具有预测的作用, 可帮助用户提前做好应对策略。

5 案例分析

本节介绍了系统的案例分析。

5.1 数据处理

本文采用 ChinaVis 2019 挑战赛^①参会人员轨迹数据集和 VAST 2015 挑战赛^②公园来访者轨迹数据

① <http://chinavis.org/2019/challenge.html>

② <http://vacommunity.org/VAST+Challenge+2015>

集。如在图 3 中 c 区域所示人员聚集的区域内, 分布了较为密集的传感器记录个体的移动轨迹数据。图 6 中 a 区域为个体的轨迹信息 $S_i = \{p_1, p_2, p_3, \dots, p_i, \dots\}$ 。其中, $p_i = [x, y, t]$ 表示个体 i 在 t 时刻的坐标; m 表示个体轨迹中定位到的坐标数量。

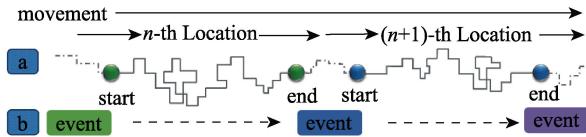


图 6 移动轨迹序列示意图

(1) 数据集分析。考虑空间属性位置序列表征了轨迹的特征具有实际的意义, 为了提高轨迹数据的可扩展性以及提高运行速度, 在预处理阶段提取的事件序列上, 本文根据空间属性特征将空间进行多层次划分, 如图 3a 中多个网格组合成一个高层次区域, 从而将轨迹数据转化为如图 6b 所示具有实际意义的空间事件序列 $S_i = \{e_1, e_2, e_3, \dots, e_i, \dots, e_m\}$ 。

(2) 活跃度计算。人流量信息能反映区域间的流动量, 但无法描述群体内部人员的移动信息。所以本文提出活跃度 H_{X_θ} 来表征区域内部人员的运动状态, 进而分析不同区域的活跃度差异。定义 θ 为当前时刻, X_i^θ 为 θ 时刻在区域 X 内的第 i 个成员所在的区域编号, n 为 θ 时刻 X 区域的人员总数, 那么区域 X 在 θ 时刻的空间活跃度为

$$H_{X_\theta} = \sum_{i=1, i \in X}^n f(X_i^\theta) \quad (4)$$

其中,

$$f(X_i^\theta) = \begin{cases} 0, & X_i^{\theta-1} = X_i^\theta \\ 1, & X_i^{\theta-1} \neq X_i^\theta \end{cases} \quad (5)$$

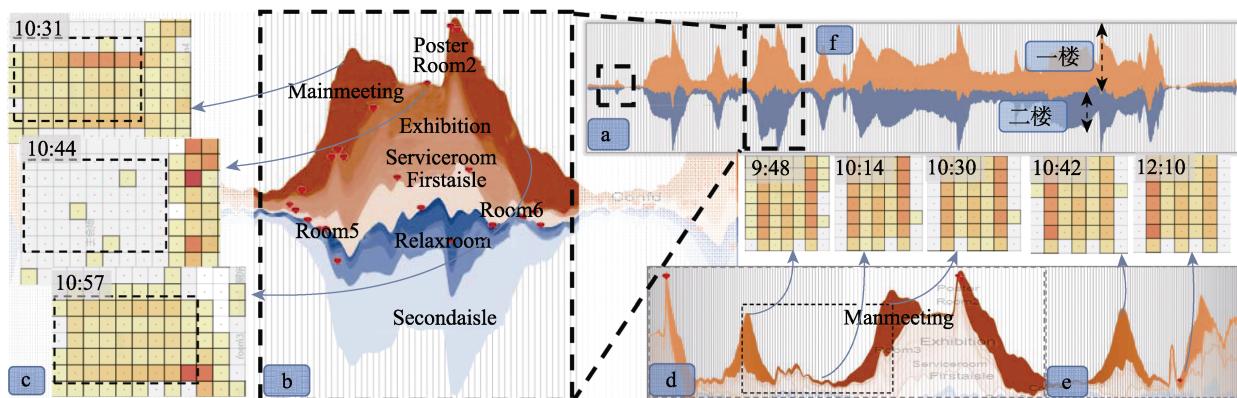


图 7 各层次活跃度演变视图

5.2 群体活跃度探索

探索轨迹中坐标转变频率, 对高密度群体内部活跃度的分析具有重要意义。考虑数据 1 中的会议规模和实际需求, 本文设置急增点、急减点阈值分别为 100%, 来快速锁定活跃度突变节点。

从图 7 中 a 区域所示参会群体活跃度演化趋势的概览图发现, 黑色框的区域具有 2 个较高的峰值, 通过滑动 MultiStream 的控制器来选择多分辨率河流图的聚焦区域。如图 7 中的 b 区域所示 f 区域的聚焦信息, a 区域展示了 24 个人流的时序演化趋势。从图 1 中 a 区域所示各区域的统计特征发现, 2 个黑色框区域人流量较大, 且出现较多的离群点, 所以点击河流图对应区域的图元和峰值时刻, 得到图 7 中 c 区域所示一系列对应的热力图。对比发现, 在 10:44 左右黑色框区域热度较小、活跃度较低, 且前后时刻维持热度大、活跃度较高的状态。因此, 推测参会者从出入口(图 7 中 c 区域的 10:57 红色网格)涌入、涌出、再涌入, 导致在较短时间内群体活跃度出现 2 个峰值。

为了进一步探究该区域的活跃度在时序上的演变趋势, 本文结合过渡层和详细层, 将对应的条带如图 7 中 d 区域进行高亮显示。可以看出, 在黑色框区域内的活跃度变化较大, 呈 U 字形, 而该时间段内的图 7 中 d 区域所示黑色框内蓝色箭头所示热力图分布变化较小, 尤其是 10:14 左右, 人员密度较大而活跃度较小。因此, 推测 2 个峰值之间的时间段内, 橙色代表的区域举办了大型的会议活动; 同时在图 7 中 e 区域所示多分辨率河流图的过渡层中, 该区域在 2 个峰值时间段内出现人员聚集。

由于从图 7 中 a 区域观察到会议活动开始出现人员聚集的现象, 因此为了探索该群体细节特征, 点击对应时间段的热力图发现, 其在轨迹投影视图 8 中 a 区域。通过套索选中投影中轨迹相似的人

员, 显示其像素轨迹图。点击出现详细轨迹路线, 如图 9 中 a 区域所示, 签到处具有半径较大的圆, 因此该参会者为本次会议负责签到的工作人员。通过多个工作人员的微观轨迹信息总结得出, 工作人员在工作期间在各自的岗位长时间停留(如图 8 中 b 区域所示的红色色块为签到处), 非工作期间在 room6 区域逗留, 推测 room6 区域为工作室。例如, 图 9 中 a 区域和 b 区域中的工作人员到达工作区之前在 room6 处做准备工作。图 9 中 a 区域所示为负责签到处的工作人员(ID15367)的轨迹, 会议期间在签到处工作。图 9 中 b 区域所示为负责服务台的工作人员(ID15800)的轨迹, 在服务台处工作直至会议结束。

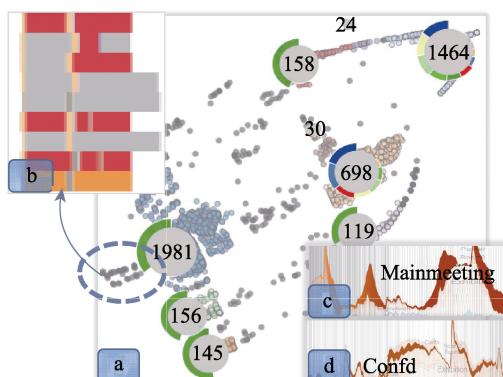


图 8 轨迹投影视图

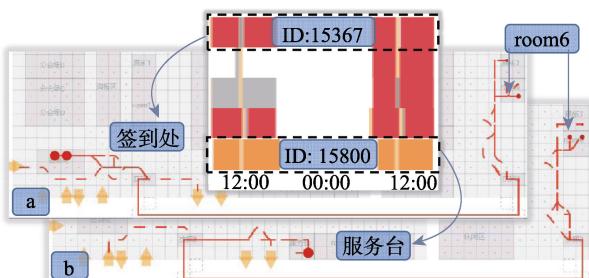


图 9 签到处和服务台人员轨迹移动的微观信息

为了从整体上掌握参会者的轨迹特征, 本文探索轨迹投影视图, 发现参会者的轨迹特征主要为蓝色(代表主会场), 有少部分的簇的特征为绿色(代表分会场)。其区域活跃度的时序变化分别如图 8 中的河流图所示, 图 8 中 c 区域的活跃度出现多个峰值, 且条带宽度大于图 8 中 d 区域, 说明了蓝色区域的人员访问量大于绿色区域的人员访问量; 同时图 8 中 c 区域的条带宽度变化较大, 所以活跃度不稳定。

5.3 基于流量分析的异常事件探索

考虑人流量密度大于一定阈值时容易发生踩踏事件, 本文通过改变阈值来快速锁定人流突变

节点。考虑数据 2 中的公园面积和数据的实际需求, 设置急增点、急减点阈值分别为 200%。

如图 10 中 a 区域所示, 在多分辨率河流图的高聚合层区域, 橙色表示出入口的人流量时序变化, 紫色和红色分别是 2 类游戏区域的人流量变化, 蓝色和黄色分别表示餐厅和休息区。可观察到, 紫色和红色的高度大于蓝色和黄色的高度, 所以人员进入公园后更倾向于参加一些游戏项目。同时发现在出入口出现突变点, 点击对应时刻的河流图可得到该时刻的热力图; 发现图 10 中 b 区域所示出入口游客较多, 图 10 中 c 区域所示出入口游客较少, 2 个区域的人流量差别较大。推测此情况是由于大量游客在早高峰时期进入公园, 并且前者比后者距离娱乐项目近(黑色线段的长度表示距离)。因此, 议公园应该在图 10 中 c 区域所示偏远的出入口附近增加一些娱乐项吸引游客, 以缓解大量游客在图 10 中 b 区域所示出入口打卡造成的拥挤压力。

如图 10 中 d 区域所示, 在多分辨率河流图的详细层, 展示了 68 个区域的人流量时序演化趋势。从中发现, 图 10 中蓝色 e 区域和紫色 f 区域条带代表的区域出现阶段性的游客容纳量较大的现象。首先, 分别点击蓝色和紫色条带嵌入的急增点、急减点, 得到一系列对应时刻的热力图。然后, 通过对比 e 区域的出入口(如图 10 中 d 区域所示红色框)热力图, 发现在人员持续密集的 2 个阶段出入口的人员密度变化频繁, 由此推测该区域的游客出入比较频繁。之后, 通过对比图 10 中 f 区域的热力图发现, 14:30—15:16 时间段的人流量较大, 故推测该时间段内 f 区域的活动较受欢迎。因此, 在详细层点击条带的急增/减点能快速锁定人员密度异常的现象。例如, 点击图 10 中 g 区域条带的急增点, 得到该区域的热力图, 结合前后时间段的热力图分布发现, 16:00 出现大量游客涌入该区域的异常现象, 建议举办方加强对热门区域的管理。因此, 热力图分布特征能够帮助举办方合理且精准地分配人力资源, 如在人流密度较高区域分配较多的工作人员。

为了进一步发现异常现象, 本文探索人员轨迹投影视图, 如图 11 中 a 区域所示。点击轨迹投影视图中的离群点, 得到一系列个体微观移动轨迹视图, 如图 11 中 b 区域和 c 区域所示。首先, 观察图 11 中 b 区域的轨迹发现, 该人员在出现较大的三角形、代表间隔时间内速度较大; 网格边长为 5 m, 该人员 12 s 内在横纵坐标移动网格数目分别为 23 和 4, 考虑高密度区域易引起踩踏事件, 故判定该人员移动速度异常。然后, 观察图 11 中 c

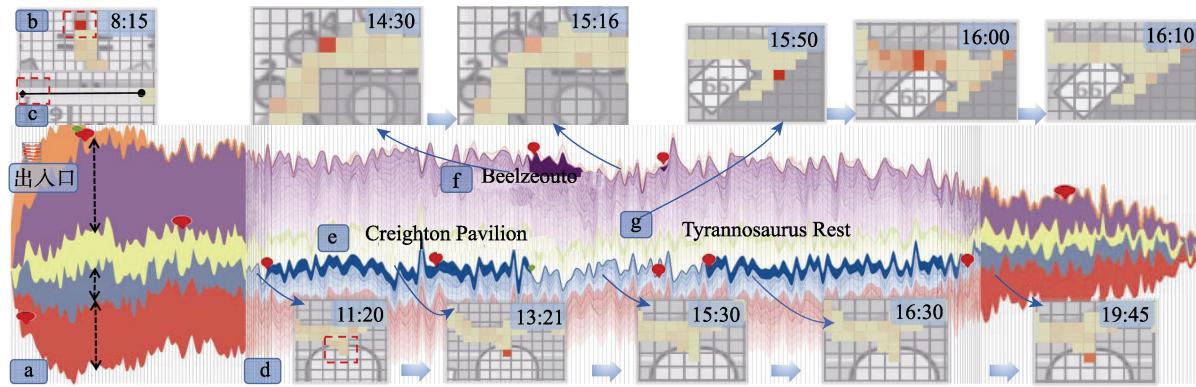


图10 不同空间聚合度下群体流量的时序演化

区域发现, 红色、蓝色和绿色的线条重合度高, 代表该人员3天的轨迹相似。其中红色的圈内出现红色和蓝色的圆, 代表该人员2天在该区域停留时间较长。因此推测该人员为公园工作者, 而在绿色代表的日期内, 该人员没在此处工作, 本文判定该轨迹为异常现象。

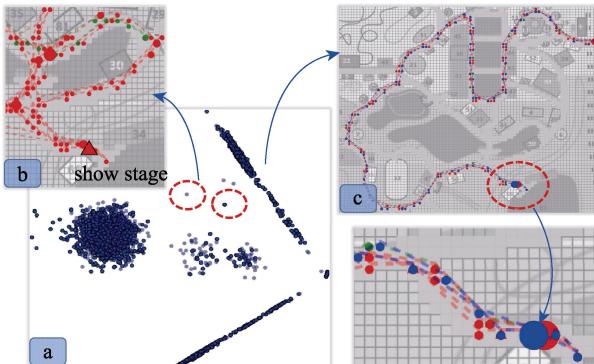


图11 轨迹投影视图

综上, 本文能够对高密度群体轨迹数据进行异常检测。图元标识符通常出现在区域流量较高之前, 用户能够根据图元出现的时间和地点, 提前做好应对策略。例如, 在红色图元时刻通常出现大量人员涌人, 相关人员应该在对应区域分配较多的服务人员; 在绿色图元时刻通常出现大量人员涌出, 则相关人员应该在对应的区域做好疏散等工作。同时, 本文根据个体微观移动轨迹推断异常现象, 如移动速度异常, 则相关人员通过轨迹信息判断人员轨迹异常, 从而加强对人员的管理。

6 讨论

本文设计一种可视分析系统, 以探索和分析高密度群体轨迹数据的微观演化情况。通过多种可视编码的方式展示数据的特点, 结合 small-multiples 的特征分布统计视图、轨迹投影视

图、多分辨率河流视图以及空间层次地图等多视图多模式展示轨迹数据的微观变化。随着研究人员对轨迹数据的深入研究, 不断提出了新的问题和挑战。同时也存在一些局限性:

(1) 可拓展性问题。本文探索轨迹数据的微观变化, 对大规模高密度群体轨迹数据的应用会由于存储量较大和查询方式较慢导致交互式的可视分析无法进行。针对此问题, 需要构建特殊的数据结构来处理这些差异性较小的轨迹数据。

(2) 轨迹数据可视化问题。本文在轨迹图模式中, 同时展示大量轨迹路线的具体信息会造成视觉杂乱的效果。针对此问题的解决方案是使用相似路线绑定算法对数据进行可视化。

(3) 异常状态检测。本文在多分辨率河流图中使用嵌入式图元识别潜在的拥堵事件, 在轨迹投影视图中具有识别离群点的功能, 但本文还需要考虑更多个体轨迹异常行为的检测。

参考文献(References):

- [1] Guo D S. Flow mapping and multivariate visualization of large spatial interaction data[J]. IEEE Transactions on Visualization and Computer Graphics, 2009, 15(6): 1041-1048
- [2] Adrienko N, Adrienko G. Spatial generalization and aggregation of massive movement data[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(2): 205-219
- [3] Lu M, Liang J, Wang Z C, et al. Exploring OD patterns of interested region based on taxi trajectories[J]. Journal of Visualization, 2016, 19(4): 811-821
- [4] Guo H Q, Wang Z C, Yu B W, et al. TripVista: Triple Perspective Visual Trajectory Analytics and its application on microscopic traffic data at a road intersection[C] //Proceedings of the IEEE Pacific Visualization Symposium. Los Alamitos: IEEE Computer Society Press, 2011: 163-170
- [5] Wang F, Chen W, Wu F R, et al. A visual reasoning approach for data-driven transport assessment on urban roads[C] //Proceedings of the IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2015: 103-112

- [6] Chu D, Sheets D A, Zhao Y, et al. Visualizing hidden themes of taxi movement with semantic transformation[C] //Proceedings of the IEEE Pacific Visualization Symposium. Los Alamitos:IEEE Computer Society Press, 2014: 137-144
- [7] Furnas G W. Generalized fisheye views[J]. ACM SIGCHI Bulletin, 1986, 17(4): 16-23
- [8] Krüger R, Thom D, Wörner M, et al. TrajectoryLenses-a set-based filtering and exploration technique for long-term trajectory data[J]. Computer Graphics Forum, 2013, 32(3): 451-460
- [9] Andrienko G, Andrienko N, Dykes J, et al. Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research[J]. Information Visualization, 2008, 7(3/4): 173-180
- [10] Dosman J A, Cotton D J, Graham B L, et al. Chronic bronchitis and decreased forced expiratory flow rates in lifetime nonsmoking grain workers[J]. American Review of Respiratory Disease, 1980, 121(1): 11-16
- [11] Wang Z C, Lu M, Yuan X R, et al. Visual traffic jam analysis based on trajectory data[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2159-2168
- [12] Liu H, Gao Y, Lu L, et al. Visual analysis of route diversity[C] //Proceedings of the IEEE Visual Analytics Science and Technology. Los Alamitos:IEEE Computer Society Press, 2011: 171-180
- [13] Zeng W, Fu C W, Arisona S M, et al. Visualizing mobility of public transportation system[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1833-1842
- [14] Anwar A, Nagel T, Ratti C. Traffic origins: a simple visualization technique to support traffic incident analysis[C] //Proceedings of the IEEE Pacific Visualization Symposium. Los Alamitos: IEEE Computer Society Press, 2014: 316-319
- [15] Boyandin I, Bertini E, Bak P, et al. Flowstrates: an approach for visual exploration of temporal origin-destination data[J]. Computer Graphics Forum, 2011, 30(3): 971-980
- [16] Yang Y L, Dwyer T, Goodwin S, et al. Many-to-many geographically-embedded flow visualisation: an evaluation[J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(1): 411-420
- [17] Byron L, Wattenberg M. Stacked graphs - geometry & aesthetics[J]. IEEE Transactions on Visualization and Computer Graphics, 2008, 14(6): 1245-1252
- [18] Morrow B, Manz T, Chung A E, et al. Periphery plots for contextualizing heterogeneous time-based charts[C] //Proceedings of the IEEE Visualization Conference. Los Alamitos: IEEE Computer Society Press, 2019: 1-5
- [19] Xu P P, Wu Y C, Wei E X, et al. Visual analysis of topic competition on social media[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2012-2021
- [20] Sun Guodao, Hu Yajuan, Jiang Li, et al. Urban agglomerations-based visual analysis of air quality data[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(1): 17-26(in Chinese)
(孙国道, 胡亚娟, 蒋莉, 等. 基于城市群的空气质量数据的可视分析方法研究[J]. 计算机辅助设计与图形学学报, 2017, 29(1): 17-26)
- [21] Cuena E, Sallaberry A, Wang F Y, et al. MultiStream: a multiresolution streamgraph approach to explore hierarchical time series[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 24(12): 3160-3173
- [22] Cockburn A, Karlson A, Bederson B B. A review of overview+ detail, zooming, and focus+context interfaces[J]. ACM Computing Surveys, 2009, 41(1): 1-31
- [23] Li Y J, Liu B. A normalized levenshtein distance metric[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1091-1095