

论文引用格式:

平行坐标轴动态排列的地理空间多维数据可视分析

周志光^{1, 2}, 余佳珺¹, 郭智勇¹, 刘玉华¹

1. 浙江财经大学 信息管理与工程学院, 杭州市, 310018;

2. 浙江大学 CAD&CG 国家重点实验室, 杭州市, 310058

摘要: **目的** 平行坐标是经典的多维数据可视化方法, 在其用于地理空间多维数据分析过程中, 往往存在空间位置信息缺失, 空间关联分析不确定等问题。因此, 本文拟设计一种有效关联平行坐标和地图的地理空间多维数据可视分析方法。**方法** 根据多维属性信息对地理空间位置进行聚类分析, 引入 Voronoi 图和颜色明暗映射对地理空间各类区域进行显著标识。利用平行坐标呈现地理空间多维属性信息, 引入互信息度量地理空间聚类与属性类别的相关性, 动态地确定平行坐标轴排列顺序。进一步计算属性轴与地图之间数据线的绑定位置, 对数据线的布局进行优化处理, 降低地图与平行坐标系间数据线分布的紊乱程度。**结果** 有效集成上述可视化设计及数据分析方法, 设计与实现一种基于平行坐标轴动态排列的地理空间多维数据可视化分析系统, 提供便捷的用户交互模式, 通过 2 组具有明显地理空间多维属性特征的数据进行测试, 验证了本文可视分析方法的有效性和实用性。**结论** 所提出的可视分析方法和工具可以帮助用户快速地分析地理空间多维属性存在的空间分布特征及其关联模式, 为地理空间多维数据的探索提供了有效手段。

关键词: 可视分析; 地理空间; 多维数据; 平行坐标; 互信息

Visual analysis of geospatial multi-dimensional data via dynamically arrangement of parallel coordinates

Zhou ZhiGuang^{1, 2}, Yu JiaJun¹, Guo ZhiYong¹, Liu YuHua¹

1. School of Information, Zhejiang University of Finance and Economics, Hangzhou 310018;

2. State Key Laboratory of CAD & CG, Zhejiang University, Hangzhou 310058

Abstract: Geospatial multi-dimensional data is mainly composed of spatial location and attribute information, which can effectively record and describe events and phenomena such as social and economic development, natural environment changes, and human social activities. As a commonly-used method for multi-dimensional data visualization, parallel coordinates do not work well for the visual exploration of geospatial multi-dimensional data, because of the lack of spatial information and uncertainty of spatial correlation. Therefore, it is of great significance for the analysis and understanding of geospatial multidimensional data to establish an effective association between spatial locations and multiple attributes. In this paper, we propose a novel geospatial multi-dimensional data visualization method, which uses geographical map to display spatial locations, visualizes multi-dimensional attributes by means of parallel coordinates, and associates map and parallel coordinates through data lines. We design a corresponding visual analysis system, allowing users to interactively explore and analyze the spatial distribution of geospatial multi-dimensional data and its associated feature patterns, based on the initial geospatial multi-dimensional data, including different spatial locations and their corresponding multi-dimensional attribute information. Spatial areas are classified into different clusters according to multi-dimensional attributes and spatial distance, and the Voronoi diagrams and color mappings are designed to visually represent different clusters. The attribute information of the geospatial multi-dimensional data is represented by parallel coordinates, and the data on different attribute axes is clustered and analyzed. The mutual information is used to dynamically calculate the correlation between the geospatial clustering and the attribute categories, and the ordering of the parallel coordinate plot is adaptively determined. Then, the map is embedded into the

parallel coordinates based on the above axis alignment results, and the map view and parallel coordinate systems are effectively correlated through data lines. Furthermore, the binding position of the data line between the attribute axis and the map is dynamically calculated according to the geospatial clustering, and the layout of the data line is optimized to reduce the disorder of the data line distribution between the map and the parallel coordinate system. We design and implement a geospatial multi-dimensional data visualization analysis system with the above visual designs and data analysis methods integrated. A convenient user interaction mode is provided, and two case studies are conducted based on the datasets with multi-dimensional geospatial attributes, to demonstrate the validity and practicability of the proposed visual analysis system. The GDP data containing 11 attributes and 32 spatial locations is visualized based on our visual analysis system. Comparing the geospatial clustering and the actual urban development in the map view, it is proved that the geospatial clustering algorithm which comprehensively considers the data attribute information and spatial locations information proposed in this paper is useful. By observing the arrangement of parallel axes, it is verified that there is certain rationality in the dynamic arrangement of parallel axes based on mutual information. In the second case based on geospatial multi-dimensional data, we mainly explore the spatial distribution of the attribute information in a certain spatial cluster. When the user clicks on a geospatial clustering of interest, the system will rearrange the parallel coordinate axes. By comparing the distribution of attributes of the same geospatial clustering on different times, we find that the proposed method is more sensitive to data. When the data changes slightly, the order of parallel axis will change, making the map embedded parallel coordinates better match the spatial distribution of multi-dimensional attribute information. We further invite experts from different fields such as geography and economy to use and evaluate the system. The validity and practicability of the geospatial multidimensional data visual analysis system is further verified by one-on-one interviews. Through a set of case studies and expert feedbacks, it is shown that the visual analysis methods and tools proposed in this paper can help users to quickly analyze the spatial distribution characteristics and associated patterns of geospatial multi-dimensional attributes, and provide domain experts with an effective means for the exploration of geospatial multi-dimensional data.

Key words : visual analysis, geographic space, multi-dimensional data, parallel coordinates, mutual information

0 引言

地理空间数据通常由地理位置和属性信息构成。地理位置用于描述对象的空间几何信息，一般由经纬度坐标表达；属性信息用于描述空间对象的状态特征，一般由多种专题属性数据表达^[1]。可见，地理空间数据是具有明确空间位置标识的多维属性数据集合，能够有效记录社会经济发展，自然环境变化，人类社交活动等现象。典型的地理空间数据如宏观经济统计数据，空气质量监测指标数据，以智能手机为代表的移动终端数据等。因此，有效地探索和解析地理空间多维属性数据，对经济社会发展形式，人类生活模式进行全面而细致的分析论证，进而为相关部门提供决策支持，是当前数据分析以及可视化领域的研究热点，具有重要的理论价值和现实意义。

可视分析是近年来快速发展的交叉学科，有效融合了数据挖掘，模型分析等理论和方法，利用交互可视化界面对复杂的数据进行探索和分析，引导

用户全面而细致地分析探索地理空间多维属性数据中潜在的关联特征模式^[2]。平行坐标是一种经典的多维属性数据可视化技术，将多维属性抽象为平行的垂直或者水平坐标线，将数据对象按照多维属性信息映射为跨越平行坐标轴的组合线段，进而直观地呈现多维属性空间中数据的分布及关联特征^[3]。大量研究工作将平行坐标引入地理空间多维属性数据可视分析，主要可以分为将空间位置信息直接映射坐标轴以及多视图交互两类方法。然而，由于空间位置分布不均，且不具备顺序属性，将其直接布局于平行坐标轴上会导致空间信息丢失，不利于空间关联特征的分析与挖掘；多视图交互的可视化效果借助用户意境创建地理空间概念视图，多维属性信息及地理空间位置的分离，不利于多维属性地理空间分布特征及关联模式的直观呈现和深入挖掘。

因此，本文提出一种有效融合地图和平行坐标的地理空间多维属性数据可视分析方法。首先，综合考虑空间位置信息，从专题属性角度出发，对地理空间对象进行聚类分析，并且在地图视图中利用

Voronoi 图和颜色明暗程度对地理空间对象进行划分和显著标识。进而,在地图空间嵌入平行坐标系,根据多维属性分布对数据对象进行聚类分析,利用互信息动态地计算获取地理空间聚类与属性类别之间的相关性,自适应地确定平行坐标轴排列顺序。为了降低大量数据线造成的分布紊乱情况,进一步计算各个地理空间聚类的数据线绑定位置,对数据线的布局进行优化处理。最后,提供便捷的用户交互模式,有效关联上述可视化设计及数据分析方法,设计与实现基于平行坐标轴动态排列的地理空间多维数据可视分析系统,为用户快速地分析和探索地理空间数据中多维属性空间分布特征及其关联模式提供交互识别手段和量化分析依据。

本文的主要贡献如下:

1. 提出一种有效融合地图和平行坐标的地理空间多维属性可视分析方法,克服传统平行坐标用于地理空间多维属性可视化过程中存在的问题,帮助用户快速分析和探索地理空间数据中多维属性空间分布特征及其关联模式;

2. 综合考虑地理空间对象专题属性和地理位置属性,设计地理空间多维属性聚类分析策略,有效识别地理空间聚类特征,Voronoi 图和颜色明暗程度映射的设计和引入,进一步增强地理空间类别特征表达;

3. 引入互信息有效地度量地理空间聚类与专题属性类别的相关性,动态地优化平行坐标轴排列顺序及数据线空间分布,支持用户交互式地聚焦与探索感兴趣空间类别特征,增强多维属性空间分布及关联特征的视觉感知。

本文的组织结构如下:第 1 部分将简要阐述相关工作,第 2 部分介绍本文的系统流程及算法要点,第 3 部分详细论述本文数据挖掘算法及可视化设计,第 4 部分结合案例分析和专家反馈验证本文所提可视分析系统的有效性和实用性,并对不足之处进行了分析和讨论,第 5 部分对全文工作进行总结。

1 相关工作

本节将从多维属性可视化,平行坐标,地理空间数据可视化三个角度出发,对相关工作加以论述。

1.1 多维属性可视化

大量经典而有效的可视化方法将多维属性数据映射至人眼易于感知的二维空间中,帮助用户快速分析与探索其在原始多维空间中的分布特征和关联模式,如散点图矩阵^[3],降维算法^[4],平行坐标^[5]

等。散点图矩阵是将原始数据的多维属性排列式地布局到二维直角坐标系中,能够有效呈现任意两个属性间的相关关系。为了减少散点图矩阵中大量数据重叠所引起的视觉混淆,Matejka 等人^[6]通过定义平均有效不透明度变量,自适应地优化散点图矩阵的渲染透明度,增强多维属性分布特征的理解与认知。Goodwin 等人^[7]通过设计具有特定含义的图标(Glyph),有效表示散点图矩阵中的不同单元,有助于感知和识别多变量数据在不同尺度、不同地理空间区域的分布差异。然而,散点图矩阵局无法展示三维及三维以上属性之间的相关关系,且受空间资源局限导致其可拓展性不强。降维算法可以将多维数据以线性或非线性的方式投影至低维空间中。经典的降维算法如主成分分析^[8](Principal Component Analysis, PCA),多维标度^[9](Multidimensional Scaling, MDS),自组织映射^[10](Self-organizing Maps, SOM)等。Turkay 等人^[11]有效融合 PCA 和 MDS 降维算法,实现多维数据的结构感知及特征分析。Yuan 等人^[12]设计并提出有效融合 PCA 和 MDS 算法的维度投影矩阵/树,支持用户交互式地探索和优化多维属性数据可视化结果。Sacha 等人^[13]利用 SOM 算法对多维数据进行聚类分析和数据库概览,支持筛选具有代表性的属性,以达到降维可视化的目标。然而,降维投影的坐标不代表具体的数据含义,为多维属性数据的理解带来困扰。

1.2 平行坐标

平行坐标是将多维属性抽象为二维空间中平行排列的坐标轴,进而连接不同属性轴上的坐标,获得多维属性数据的直观分布^[14]。随着数据规模的增加,大量重叠,交叉的折线容易造成的数据分布混乱,导致平行坐标系中多维属性关联特征难以识别与发现。Wu 等人^[15]设计面向手机基站数据的交互式可视分析系统 TolCoVis,利用 binning 技术对平行坐标属性数据进行聚类。Nguyen 等人^[16]利用平行坐标的点线对称性,设计并提出 DSPCP 方法,增强平行坐标表达复杂多维数据关系的效率与精度。Wang 等人^[17]构建嵌套式的平行坐标系,有效探索集合模拟数据多分辨率之间以及单个分辨率内部的数据分布特征。Qin 等人^[18]对维度间不同类别的连接关系进行重新规划与排列,解决平行坐标系中边捆绑方法存在的数据量表达不准确以及维度间视觉连续性丢失等问题。Wu 等人^[19]提出一种基于力引导分段式骨骼布局的边绑定方法,增强平行坐标系中数据分布特征表达的准确性。

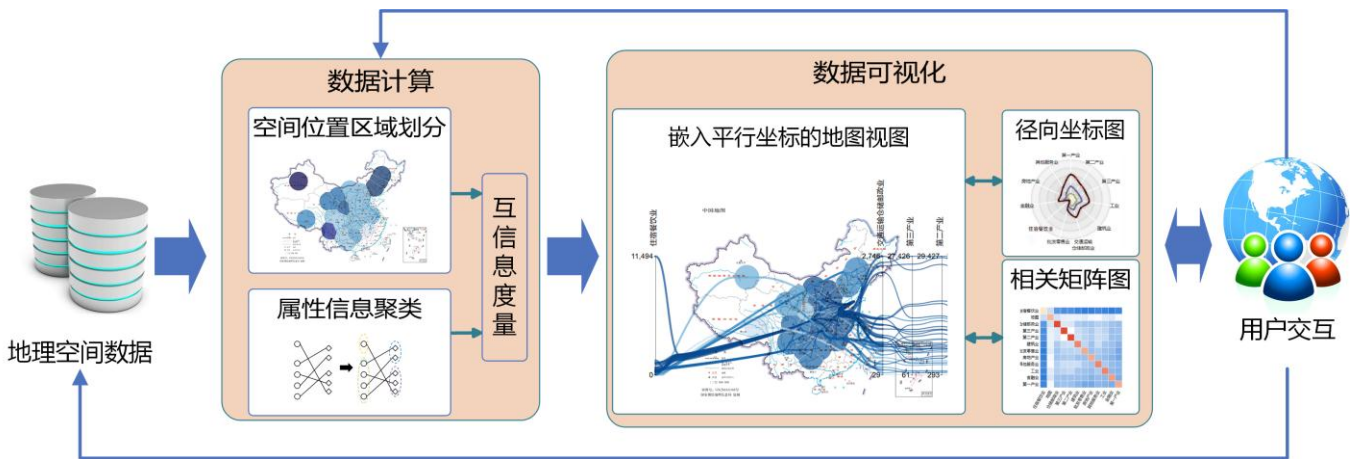


图1 系统流程图

Fig.1 System overview

平行坐标轴排列顺序对于多维属性数据的探索和感知具有重要的意义。Wegman 等人^[20]将平行坐标轴排列的解空间看作一个图模型，利用欧拉路径和哈密顿分解确定最优的平行坐标轴排列顺序。Qu 等人^[21]利用加权完全图可视化呈现维度间的相关性，根据相关系数编码边的权重，帮助用户交互式地确定平行坐标轴排列顺序。Zhang 等人^[22]将平行坐标系中的每个属性轴抽象为网络图中的结点，引入相关系数度量边的权重，进而利用旅行商问题 (Traveling Salesman Problem, TSP) 求解算法查找最优平行坐标轴排列顺序。Zhen 等人^[23]根据 Spearman 秩相关系数衡量多维属性之间的相关度，进而利用贪心算法优化平行坐标轴排列顺序。Zhou 等人^[24]提出一种基于聚类相关性的平行坐标轴排列算法，对各个专题属性维度进行层次聚类，利用信息熵度量多维属性相关性，结合图遍历最小代价路径的方式，对平行坐标轴排列进行优化。然而，由于空间位置不具备顺序属性，采用平行坐标展示地理空间多维属性信息会导致空间信息丢失，不利于空间关联特征的分析与挖掘。

1.3 地理空间数据可视化

地理空间对象通常具有多维属性信息描述，有效融合地理空间位置和多维属性信息，为用户提供全面而完整的可视化结果，具有重要的意义和研究价值。Chen 等人^[25]设计 Glyph 编码地理空间对象的多维属性及概率分布等信息，并且按照其对应的空间位置映射在地图中，帮助用户快速浏览地理空间多维属性及其概率分布特征。Turkay 等人^[26]从交互可视的角度出发，面向英国人口普查数据，设计地图和散点图矩阵关联的可视分析系统，支持用户交互定义分析路径，利用 Small multiples 展示多维经

济指标的相关性及变化规律。Wang 等人^[27]面向多分辨率气象集合模拟数据，设计内嵌式平行坐标，实现多维参数关联关系的协同可视，并且利用 Small multiples 描述模拟数值与真实观测数据之间的差异。Lu 等人^[28]针对出租车轨迹数据，设计 OD-Wheel 可视分析方法，将环形视图与线性树图结合，帮助用户分析和对比各个区域间流量的时序分布特征。Chen 等人^[29]设计基于热力图和放射环的可视分析系统 SunMap，在地图上叠加热力图呈现数据分布，利用基于节点排序的放射环对具有显著层次结构特征的数据进行探索与可视化。Yang 等人^[30]构建能够有效融合 OD 矩阵和流图的可视分析工具 Mapatrix，利用二次规划算法查找视图的最佳连接位置，对地理空间信息和 OD 分布进行关联可视分析。可以看出，Glyph 设计，Small Multiples，交互展示，地图嵌入式等手段能够有效融合空间位置与属性信息，是地理空间数据可视化的有效途径。然而，上述地理空间数据可视化方法，容易造成多维属性信息及地理空间位置的感知分离，不利于地理空间多维属性数据的空间分布特征及关联模式的直观呈现和深入挖掘。

2 系统概述

本文基于平行坐标研究一种有效关联地图的地理空间多维数据可视分析方法，帮助用户交互式地探索和分析地理空间多维属性的空间分布及其关联特征模式。图 1 出示了本文系统的总体流程。

首先，导入初始的地理空间多维属性数据，包括不同地理空间对象及其对应的多维属性信息；进而对空间位置进行空间位置区域划分，借助地图可视化展示空间位置属性，并对属性信息进行聚类分

析。采用颜色明暗程度的映射方案对不同的地理空间聚类进行标识, 结合互信息度量地理空间聚类与属性类别的相关性, 自适应确定轴排列顺序, 进而通过数据线有效关联两视图; 最后提供相关矩阵视图, 径向坐标视图等可视化视图, 以及便捷的用户交互模式, 并有效关联上述可视化设计及数据分析方法, 为用户快速分析地理空间多维属性数据存在的空间分布特征及其关联模式提供有效手段。

3 地图关联平行坐标多维属性可视化

本节将从地理空间聚类分析、地图关联平行坐标以及数据线布局优化设计等角度出发, 对本文系统涉及的数据计算算法及可视化设计进行详细地解析和阐述。

3.1 地理空间聚类分析

为有效提取地理空间多维属性数据的地理空间分布特征, Li 等人^[31]提出一种兼顾空间位置和属性信息的地理空间聚类算法, 能有效构建空间位置邻近且属性信息相似的地理空间聚类分析。在此基础上, 本文设计一种地理空间聚类分析及可视化策略。首先, 读入初始的地理空间多维属性数据集 $F = \{f_1, f_2, \dots, f_n\} (n \geq 2)$, 其中 f_i 是该地理空间数据集 F 中第 i 个样本, 包含空间位置, 即经纬度坐标, 以及属性信息。根据公式 (1) 计算样本 f_i 和 f_j 之间的空间位置距离。

$$D_{\text{geo}}(f_i, f_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (1)$$

结合闵可夫斯基距离计算样本 f_i 和 f_j 之间的专题属性距离, 公式如下所示:

$$D_{\text{attr}}(f_i, f_j) = \left(\sum_{a=1}^n \frac{|A_{ia} - A_{ja}|}{D_{a\text{max}}} \right)^{1/n} \quad (2)$$

其中, A_k 表示样本 f_i 的第 k 维专题属性, $D_{k\text{max}}$ 为第 k 维专题属性上最大最小属性信息的差值 $D_{k\text{max}} = A_{k\text{max}} - A_{k\text{min}}$, 并根据实际案例需求交互定义空间位置距离和专题属性距离的最小阈值 ϵ_{dis} 、 ϵ_{attr} , 以此构建基于专题属性的地理空间聚类算法, 过程如下:

Step1: 从当前未标记的样本中随机选取某一样本 f_i , 迭代计算其余样本与样本 f_i 之间的空间距离, 并构建集合 $F_{\text{geo}} = \{f_x | D_{\text{geo}}(f_i, f_x) \leq \epsilon_{\text{geo}}\} (x \leq n)$ 。

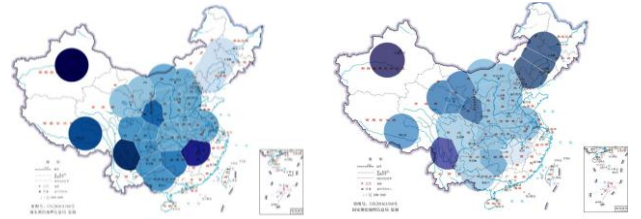
Step2: 计算集合 F_{geo} 中所有样本与 f_i 之间的属性距离, 构建集合 $F_D = \{f_y | D_{\text{attr}}(f_i, f_y) \leq \epsilon_{\text{attr}}\} (y \leq x)$ 。

Step3: 对 F_D 中的每一样本进行循环递归搜

索, 寻找所有与 f_i 相连的集合, 构建集合 F_L , 并对 f_i , F_D 和 F_L 进行相同颜色编码, 视为同一类别。

Step4: 当全部样本均被颜色编码, 则视为算法结束, 否则返回 **Step1** 进行下一轮递归搜索。

为有效呈现地理空间聚类分析结果及其分布, 本文利用 Voronoi 对地图进行多边形划分。在多边形划分的基础上, 结合地理空间聚类分析结果, 对属于同一聚类的多边形进行相同颜色映射, 用户可以直观地发现属性相似且空间邻近的地理空间位置。图 2 展示了不同参数下, 对应空间位置聚类结果对比。在 $\epsilon_{\text{attr}}=4.7$, $\epsilon_{\text{dis}}=2.72$ 时, 地理空间位置聚为 7 类, 如图 2 (a) 所示, 大量位于华北、华东、华中的城市被聚为一类。我们降低专题属性最小阈值, 并调高空间位置距离最小阈值, 聚类结果如图 2 (b) 所示, 在 $\epsilon_{\text{attr}}=4.247$, $\epsilon_{\text{dis}}=3.219$ 时, 地理空间位置被聚为 8 类, 华北、华东、华中所属城市在一定程度上被区分。



(a) $\epsilon_{\text{attr}}=4.7$; $\epsilon_{\text{dis}}=2.72$ (b) $\epsilon_{\text{attr}}=4.247$; $\epsilon_{\text{dis}}=3.214$

图2 不同参数下对应的地理空间位置聚类可视化结果

Fig. 2 Spatial clusters based on different parameters

(a) $\epsilon_{\text{attr}}=4.7$, $\epsilon_{\text{dis}}=2.72$ (b) $\epsilon_{\text{attr}}=4.247$, $\epsilon_{\text{dis}}=3.214$

3.2 地图关联平行坐标

本文利用平行坐标展示地理空间数据的多维属性信息, 其中数据线有效关联地图视图和平行坐标, 能够帮助用户直观地探索和理解地理空间多维属性的空间分布及其关联特征。然而, 具体的平行坐标与地图进行关联的过程存在如下挑战: 首先, 地理空间对象不具有顺序特点, 直接映射至坐标轴上容易导致空间信息丢失; 其次, 不同属性信息在地理空间的分布情况不尽相同, 若随机排列属性轴, 通过数据线关联地图和平行坐标视图, 容易导致数据线分布紊乱, 妨碍地理空间多维属性的空间分布特征探索与感知。

为此, 本文引入互信息度量属性信息的空间分布相关性, 为有效关联地图视图以及平行坐标轴排列提供依据。首先, 基于核密度估计对地理空间多维属性数据集中每个属性信息进行聚类分析。本文采用高斯核密度估计, 计算如公式 (4) 所示。

$$\hat{f}_h(x) = \frac{1}{\sqrt{2\pi nh}} \sum_{i=1}^n \exp\left(-\frac{(x-x_i)^2}{h^2}\right) \quad (4)$$

其中， h 表示不同距离样本点对密度的影响程度。迭代比较核密度曲线中不同样本点之间对应密度的大小，得到 $l+1$ 个波谷，每两个相邻波谷间对应属性信息数据视为一个类，以此构建 l 个聚类。属性信息的聚类结果与带宽的选取有关，如图 3 所示。当 h 选择过大时，核密度曲线过于光滑，导致所有的数据被聚为一类，当 h 选择过小时，核密度曲线波谷较多所得聚类较多。因此本文设置参数接口，交互式的定义带宽，进而得到合适的属性聚类结果。

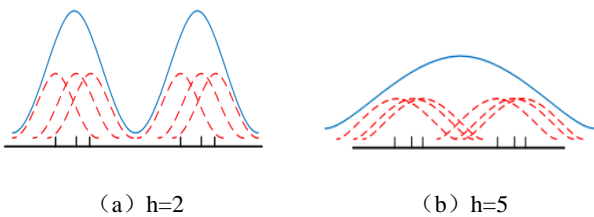


图 3 核密度曲线

Fig. 3 Kernel density curves (a) $h=2$ (b) $h=5$

在获得每个属性信息聚类特征的基础上，根据公式 (5) 结合互信息计算属性信息与空间位置的相关性大小。

$$VI(C; C_k) = 2H(C, C_k) - H(C) - H(C_k) \quad (5)$$

其中， $H(C)$ 表示地理空间聚类的信息熵大小， $H(C_k)$ 表示第 k 维属性聚类的信息熵大小， $H(C, C_k)$ 表示第 k 维属性信息在地理空间聚类分析结果中的数据分布紊乱程度。基于互信息构建属性信息与地理空间位置之间的相关性度量标准，计算结果越小表明该聚类下属性信息的空间分布特征越明显。

为减少数据线的分布紊乱程度以增强用户对属性信息空间分布特征的感知，以相关性计算结果为依据自适应的确定平行坐标轴排列，将具有明显空间分布特征的属性信息抽象为平行坐标轴与地图相邻放置，通过数据线有效关联平行坐标系与地图视图。以图 4 中的空间聚类及属性分布为例，通过数据线构建地理空间聚类 c_1, c_2, c_3 与属性轴之间的关联关系，发现属性轴 A 上的属性信息与空间聚类的相关性相较于属性轴 B 上的属性信息更强，具有明显的空间分布特征，将属性轴 A 与地图相邻放置可以有效减少数据线的交叉重叠，帮助用户更好地观察和分析地理空间分布及多维属性的关联特征；反之则容易产生数据线分布紊乱，干扰视觉认知。

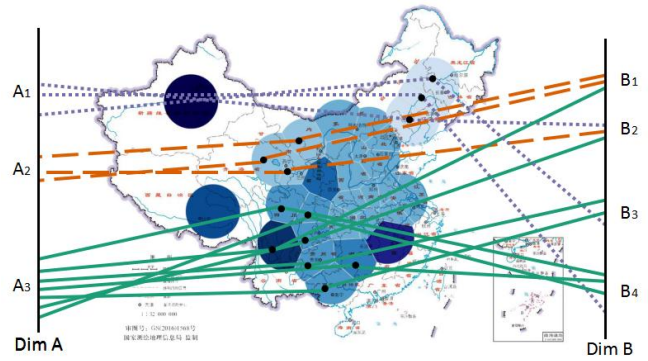


图 4 平行坐标轴关联地图示意图

Fig.4 Diagram with parallel coordinates and geographical map

本文利用相关矩阵图，通过颜色深浅程度直观的展示地理空间多维数据各属性间的相关性大小。相关矩阵图中颜色越深代表相关性越大，相反颜色越浅相关性越小。并对相关矩阵图的对角线矩阵块进行颜色映射，表示属性聚类结果以及地理空间聚类结果，类别越多颜色越深，可视化结果如下图 8 (a) 所示。并设计径向坐标图，通过连接地理空间多维数据中各属性信息的四分位数构建闭合多边形，直观的展示属性信息的数据分布情况，如图 8 (b) 所示。用三条闭合多边形曲线分别代表第一四分位数，第二四分位数和第三四分位数，当多边形间隔紧密时表明该属性信息数据分布较为集中，相反则说明该地理空间对象的属性信息分布较分散。

3.3 数据线布局优化设计

为避免大量数据线交叉重叠造成维度间信息视觉不连续，对数据线设置透明度，增强重要内容的显示跟踪属性信息在每个轴之间的趋势走向。构建连续分段的 Bézier 曲线代替原始的直线，以减少大量数据线交叉重叠带来的视觉紊乱。进一步对平行坐标系上的数据线进行堆叠处理，直观地展示属性信息聚类结果，如图 5 所示。然而堆叠效果在某些特定情况下容易造成信息位移，无法判断各数据间属性信息的相对大小。因此，用户可以根据不同的需求交互选择是否展示堆叠效果。

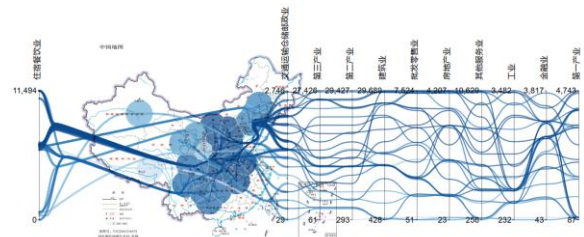


图 5 平行坐标堆叠可视化

Fig.5 Stacked visualization of parallel coordinates

地理空间多维属性数据中，不同属性信息的地理空间分布特征不尽相同，通过数据线直接关联地图视图和平行坐标系，会造成大量交叉重叠现象，对属性信息空间分布的理解和类别的探索都带来困扰。为降低数据线交叉造成的视觉紊乱，本文基于地理空间聚类结果，动态计算属性轴与地图视图之间数据线汇聚点位置，对数据线的布局进行优化处理。如图 6 (a) 所示，确定某一地理空间聚类中相距最远的两个地理空间对象，记作 a , b ，连接 a , b 构建长度为 d 的线段，并作线段垂直平分线。为确保数据线的汇聚位置与所有地理空间对象都不产生重合现象，在距离垂直平分线 d 处构建第一个汇聚点 c_1 。如图 6 (b) 所示，同样以地理空间聚类为依据，计算地图两侧属性轴上对应属性信息的平均值，作为汇聚点的纵坐标。该地理空间聚类对应属性信息的最大差值记为 w ，距离属性轴 w/λ 处记作该汇聚点的横坐标，以此构建第二个数据线汇聚点 c_2 。根据对应属性信息的分布密集程度，交互定义 λ 值，使汇聚点远离属性轴减少数据交叉。并编码连接线的多边形宽度来直观展示每条曲线中所包含的数据量大小。



(a) 汇聚点 c_1 (b) 汇聚点 c_2

图 6 数据线汇聚点确定示意图

Fig.6 Different bundling results

(a) bundling point c_1 (b) bundling point c_2

4 实验结果分析与讨论

本节主要介绍本文可视分析系统界面的交互设计，并通过实际的案例分析验证系统的可行性，并根据专家反馈信息进一步验证系统的有效性，并对本文系统的不足之处进行了讨论。

4.1 界面交互设计

图 7 出示了本文系统界面，主要由控制窗口、径向坐标视图、相关矩阵视图以及嵌入平行坐标的地图视图构成。在控制窗口用户可以交互选择所需研究的地理空间多维属性数据，以及空间位置属性区域划分和属性信息聚类阈值的调整，交互点击平行坐标显示键，显示地理空间多维属性信息的平行坐标。地图随控制窗口所选数据的地理空间尺度相应的放大缩小，使所需研究的地理空间铺满整个地图视图，并通过颜色映射帮助用户直观的区分地理空间聚类结果。系统还提供了相关矩阵视图和径向坐标视图，用户在嵌入平行坐标的地图视图中交互选择感兴趣的地理空间区域，相关矩阵图 and 径向坐标图相应的变化，多视图展示多维属性间的互信息大小以及各属性信息的数据分布密度，帮助用户更好地分析和探索地理空间多维属性数据存在的空间分布特征及关联模式。当用户根据实际案例需求交互调整空间位置距离阈值与属性信息距离阈值时，平行坐标轴排列顺序以及地图嵌入位置根据聚类结果自适应的发生变化以达到最佳可视化效果。

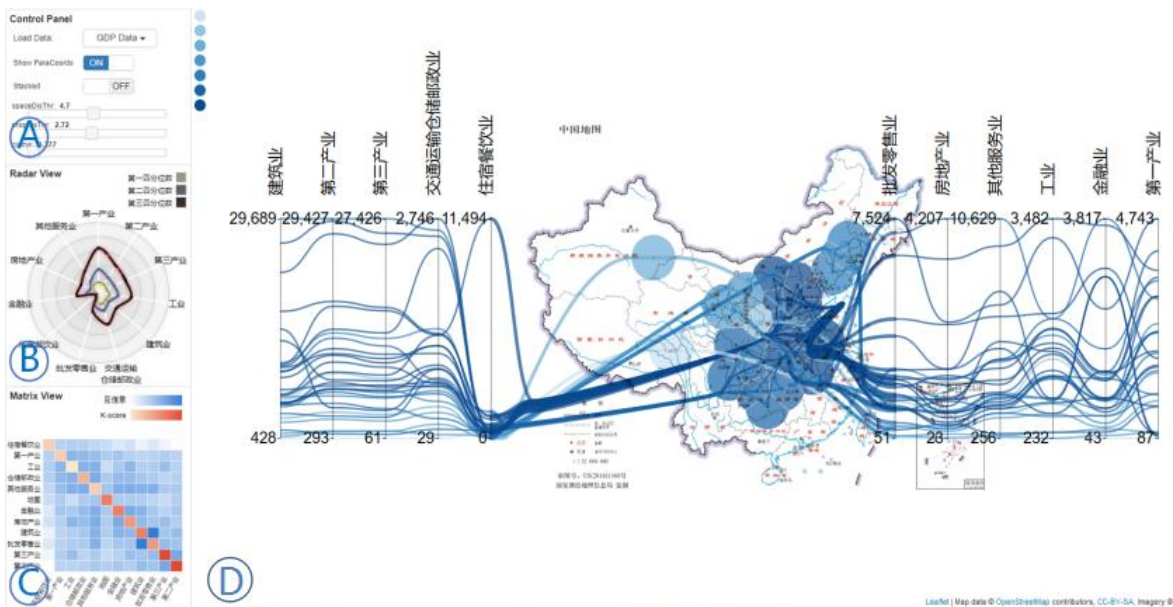


图 7 可视分析系统主界面

Fig.7 The visual interface of our visualization system

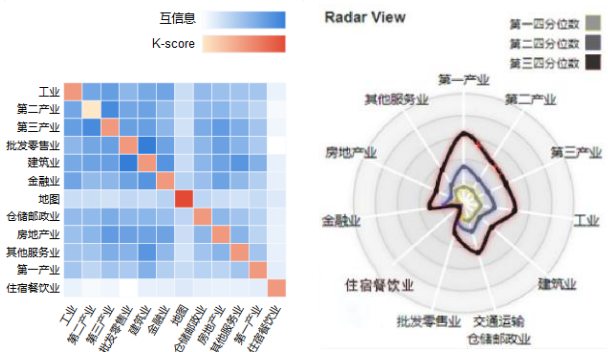
4.2 案例分析

本文引入 2 组具有明显地理空间多维特征的数据进行案例分析，包括中国生产总值统计数据 and 浙江省空气质量检测数据，通过实际的案例验证系统的有效性和实用性。

4.2.1 国内生产总值数据

国内生产总值数据集包含 11 个专题属性以及 32 个地理空间对象。属性包含第一产业，第二产业，第三产业，工业，建筑业，交通运输仓储邮政业，批发零售业，住宿餐饮业，金融业，房地产业以及其他服务业。

进行聚类分析，不同明暗程度的颜色代表不同的地理空间区域，其中辽宁省，吉林省和黑龙江省被划分为同一地理空间区域，印证了基于专题属性的地理空间聚类算法的有效性。相关矩阵图如图 8 (a) 所示，任意两个属性间的相关性越大对应矩阵块的颜色就越深，例如第一产业和住宿餐饮业对应的矩阵块颜色最深，表示对应属性间具有很强的相关性，相反，第二产业和第三产业对应的矩阵块颜色最浅，具有较弱的相关性。并根据相关性计算结果，以地图为中心对平行坐标进行轴排列。观察图 8 (a) 发现，相比于其余属性，交通运输仓储邮政业和住宿餐饮业与地理空间具有较强的相关性，因此对应的属性信息抽象为属性轴与地图相邻放置，再以交通运输仓储邮政业和住宿餐饮业为起始点循环比较与其具有最强相关性的属性信息以此构建有效关联平行坐标系的地图视图。如图 9 所示，交通运输仓储邮政业与地图相邻放置存在明显的空间分布特征。交通运输仓储邮政业与区域经济相互作用相互影响，良好的区域经济能带动交通运输行业的发展，而一个区域交通运输行业发展相对滞后，就会阻碍该区域的经济水平。改革开放后中国东部地区的经济发展水平领先于西部地区，沿海地区相较于内陆地区吸引更多的生产投资和人力资源，进一步促进了交通运输仓储邮政业的发展，因此该轴排列算法存在一定的合理性。生产食材以及其它一些生物材料的产业被称为第一产业，我国不同区域的主要农作物不同，南方以水稻为主，北方的主要农作物则为小麦，不同地形适合不同的植被生长，因此第一产业的区域分布特征不明显，对应属性轴离地图视图较远。



(a) 相关矩阵视图 (b) 径向坐标视图

图 8 属性可视化

Fig.8 Multiple attribute visualization

(a) Correlation matrix view (b) Radial chart

如图 9 所示，通过地图平行坐标关联视图界面展示地理空间多维属性数据集，并提供如图 8 所示的相关矩阵视图和径向坐标视图直观地展示多维属性间的相关性大小。由先验知识可知，辽宁省，吉林省和黑龙江省因相似的地理环境以及资源储备量形成了相似的结构被称之为东三省。由图 9 所示，基于专题属性的地理空间聚类算法对 32 个省份

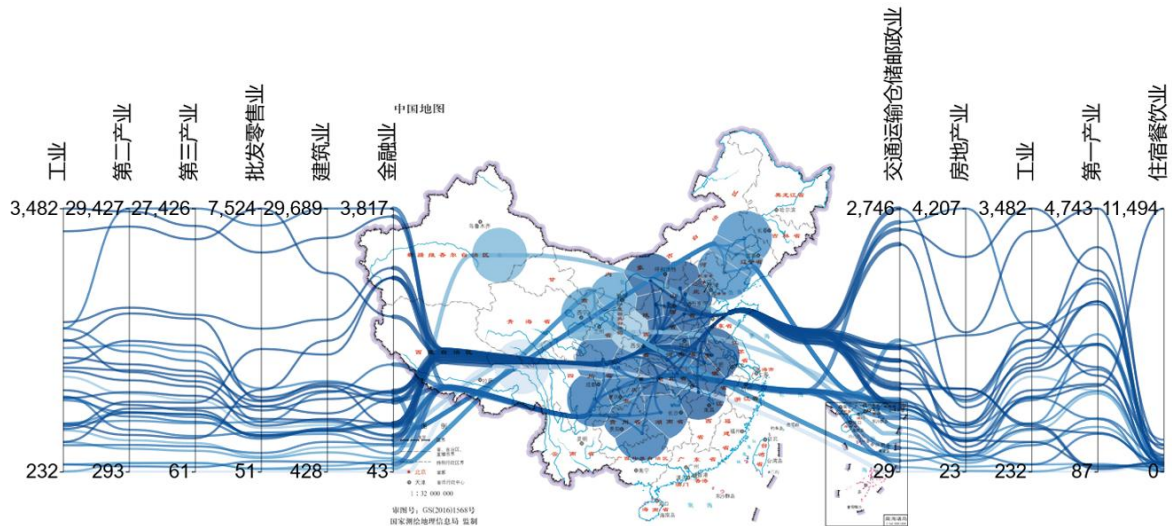
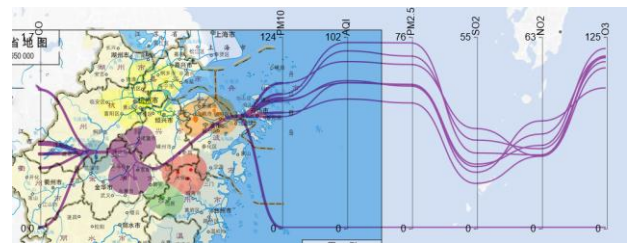


图 9 嵌入平行坐标的地图视图轴排列结果

Fig.9 Dimension rearrangement result of map embedded in the parallel coordinates view

4.2.2 浙江省空气质量监测数据

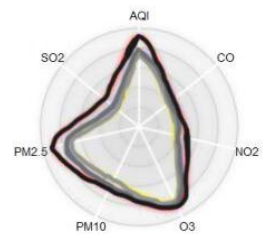
浙江省空气质量数据是典型的多维时空数据，记录了浙江省 53 个空气质量监测站点实时监控的 AQI, SO₂, NO₂, CO, PM_{2.5}, PM₁₀, O₃ 六个属性信息。如图 10 所示，在嵌入平行坐标的地图视图交互点击某一感兴趣地理空间区域，该区域对应的属性信息高亮显示，系统重新计算该高亮区域与各属性信息之间的相关性对平行坐标轴进行重排列。如图 10 (a) 所示，高亮显示的区域位于浙江省西南部，属于丽水市和金华市境内，属性信息越靠近地图数据分布越集中。观察图 10 (a) 可知该区域空气中 PM_{2.5} 和 O₃ 含量都较高，对应的 AQI 指数也较高，表明该区域的城市空间污染程度较强。且越靠近地图，属性信息的空间分布越集中，如 PM₁₀。PM₁₀ 主要在各工业过程中排放，因此该区域的工业发展可能较为相似，使得 PM₁₀ 的排放具有较为明显空间区域分布特征。对应的径向坐标图如图 10 (c) 所示，通过绘制该区域中各属性的四分位数展示属性信息的数据分布情况。为观察在不同时间段内同一区域的属性信息差异，高亮显示相同区域在 2014 年 6 月 5 号的空气质量检测数据，如图 10 (b) 所示。相较于 2014 年 6 月 4 日，此区域所包含的大量属性具有大幅度的下降，推测可能与 6 月 5 号当天是世界环境保护日有关。2014 年 6 月 5 日对应的径向坐标图如图 10 (d) 所示，图中三条四分位数多边形闭合曲线基本重合，表明对应区域各属性信息的数据分布较为集中。



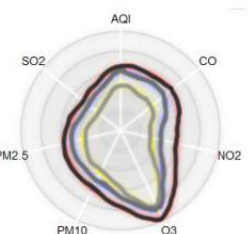
(a) 6月4日空气质量检测数据可视化



(b) 6月5日空气质量检测数据可视化



(c) 6月4日径向图



(d) 6月5日径向图

图 10 浙江省空气质量检测数据可视化结果

Fig.10 Visualization results of air quality data of Zhejiang Province (a) Map view of June 4th (b) Map view of June 5th

(c) Radial chart of June 4th (d) Radial chart of June 5th

4.3 专家反馈

为了进一步验证面向地理空间多维数据可视分析系统的有效性和实用性,本文邀请1位地理学专家和1位经济统计学专家对系统进行使用和评估,并与他们进行一对一访谈,收集他们的反馈意见总结如下:(1)交互式可视化设计:两位领域专家都认为该系统将经典的多维属性数据可视化方法平行坐标与地图进行了良好的结合,并针对两种视图的不同特点进行了大量的可视化设计方便用户观察分析。地理学专家评论说:“本文提出的地理空间多维数据可视分析方法是一种全新的可视化设计,将平行坐标与地图视图相结合提供一个全面的可视分析系统,可用于探索与地理空间位置相关的多维属性数据,并分析数据间存在的区域相关性,并通过多视图协同可视化展示,可以解决一些具有挑战性的问题。”经济统计学专家补充道:“该可视化系统结合经济统计模型构建交互式可视化系统,容易被不同背景的用户所理解。且通过多种可视化方式展示不同空间位置聚类情况下与多种属性间的互信息大小。相比于传统的统计分析软件具有更强的可扩展性,提供的系统交互功能可以帮助用户交互定义参数。”(2)适用性和改进:两位领域专家都认为该系统可以处理和分析具有地理空间分布特征的其他领域数据。地理学专家强调:“本文设计的可视分析方法适用于商业智能,可以根据地理空间聚类区分不同地区在不同时间段对不同服务的需求,帮助相关部门进行相应调度。”此外,经济统计学专家建议将我们的系统应用于涵盖更长时间的数据集,以便我们有足够的信息来获取人们的日常行为。此外,通过整合其他信息,可以针对各种应用进行更高级的分析任务。

4.4 不足之处

相比于传统地理空间多维数据可视分析方法,本文基于平行坐标轴动态排列提出了有效关联地图视图的地理空间多维数据可视分析方法,具有丰富的交互功能,可以直观的展示数据,并有效辅助用户探索地理空间多维属性数据的空间分布特征及分析区域关联模式,具有一定的实用性。但仍然存在以下不足:(1)本文借助地图和平行坐标分别展示地理空间数据的地理空间对象和多维属性信息,通过数据线构建属性数据间的关联关系,当采用平行坐标展示高维属性数据时,因面板限制等原因可能会造成视觉紊乱,地理空间分布特征难以识别等问题,因此,可以借助 Glyph 设计展示多维属性,综合描述地理空间多维属性的地理分布特征及

关联关系;(2)对于时序可变的地理空间多维属性数据的融合度不高,可以进一步构建时间序列分析模型,用于交互探索地理空间多维时序数据。因此,未来工作中将重点研究面向地理空间时序多维数据的可视分析方法,进一步实现地理多维属性数据的时序分布特征模式的分析与挖掘。

5 总结

本文研究一种基于平行坐标轴动态排列的地理空间多维数据可视分析方法,帮助用户探索地理空间多维属性数据的空间分布特征及其关联关系。针对地理空间多维属性提供基于专题属性的地理空间聚类分析方法,有效挖掘地理空间多维属性数据各要素之间的关联关系和分布模式;采用互信息度量地理空间聚类与属性聚类的相关性,有效确定地图嵌入位置及平行坐标轴排列顺序;自适应的计算视图连线的汇聚点位置,对数据线进行优化设计减少大量交叉重叠造成的视觉紊乱;提供大量的用户交互模式,各个可视化模块协同展示,为用户分析地理空间多维属性数据提供有效手段。且大量的可视化效果及专家反馈进一步验证了本文可视分析系统的有效性和实用性。

参考文献(References)

- [1] Wang M X, Zhang Z H, Yang W F. Space Geographic Data Visualization. In: Science Press, Beijing, 2012. 55-67. [王明孝, 张志华, 杨维芳. 地理空间数据可视化[M]. 科学出版社, 2012. 55-67]
- [2] Zhou Z G, Shi C, Shi L S. A Survey on the Visual Analytics of Geospatial Data[J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(5): 747-763[周志光, 石晨, 史林松,等. 地理空间数据可视分析综述[J]. 计算机辅助设计与图形学学报, 2018, 30(5):747-763.]
- [3] Becker, Richard A, Cleveland, et al. Brushing Scatterplots[J]. Technometrics, 1987, 29(2):127-142.
- [4] Tang J, Liu J, Zhang M, et al. Visualizing Large-scale and High-dimensional Data[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 37(7):287-297.
- [5] Heinrich J, Weiskopf D. State of the Art of Parallel Coordinates[J]. Eurographics, 2013, 34(1):17-25.
- [6] Matejka J, Anderson F, Fitzmaurice G. Dynamic Opacity Optimization for Scatter Plots[J]. 2015, 2707-2710.
- [7] Goodwin S, Dykes J, Slingsby A, et al. Visualizing Multiple Variables Across Scale and Geography[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(1):599.

- [8] Jolliffe I T. Principal Component Analysis[J]. *Journal of Marketing Research*, 2002, 87(100):513.
- [9] Kruskal J B, Wish M. Multidimensional Scaling[M]. *Book on demand*, 1978, 875-878.
- [10] Kohonen T. Self-Organized Formation of Topologically Correct Feature Maps[J]. *Biological Cybernetics*, 1982, 43(1):59-69.
- [11] Turkay C, Lundervold A, Lundervold A J, et al. Representative Factor Generation for the Interactive Visual Analysis of High-Dimensional Data[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2012, 18(12):2621.
- [12] Yuan X, Ren D, Wang Z, et al. Dimension Projection Matrix/Tree: Interactive Subspace Visual Exploration and Analysis of High Dimensional Data[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2013, 19(12):2625.
- [13] Sacha D, Kraus M, Bernard J, et al. SOMFlow: Guided Exploratory Cluster Analysis with Self-Organizing Maps and Analytic Provenance[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2017, 24(1):120-130.
- [14] Inselberg A. Parallel Coordinates: Visual Multidimensional Geometry and its Applications[M]. *Universitat Autònoma De Barcelona*, 2009, 98-101.
- [15] Wu W C, Xu J Y, Zheng Y X, et al. TelCoVis: Visual Exploration of Co-occurrence in Urban Human Mobility Based on Telco Data. *IEEE Transactions on Visualization & Computer Graphics*, 2015, 24(1):935-944.
- [16] Nguyen H, Rosen P. DSPCP: A Data Scalable Approach for Identifying Relationships in Parallel Coordinates[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2017, PP(99):1-1.
- [17] Wang J P, Liu X T, Shen H W, et al. Multi-Resolution Climate Ensemble Parameter Analysis with Nested Parallel Coordinates Plots[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2017, 19(1):81-90.
- [18] Qin H X, Wei X S. Research on edge binding technology in parallel coordinates. *Journal of Computer-Aided Design & Computer Graphics*, 2017, 7:1235-1244 [秦红星, 卫学仕. 平行坐标中的边捆绑算法. *计算机辅助设计与图形学学报*, 2017, 7:1235-1244]
- [19] Wu B, Cao W Q. A Force-Directed Skeleton-Based Bundling with Clustering in Parallel Coordinates. *Journal of Computer-Aided Design & Computer Graphics*, 2017, 10:1807-1815 [巫滨, 曹卫群. 平行坐标系聚类数据的力导向分段骨骼绑定绘制. *计算机辅助设计与图形学学报*, 2017, 10:1807-1815]
- [20] Wegman E J. Hyperdimensional Data Analysis Using Parallel Coordinates[J]. *Journal of the American Statistical Association*, 1990, 85(411):664-675.
- [21] Qu H, Chan W Y, Xu A, et al. Visual Analysis of the Air Pollution Problem in Hong Kong[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2007, 13(6):1408-1415.
- [22] Zhang Z Y, McDonnell K T, Mueller K. A Network-based Interface for the Exploration of High-Dimensional Data Spaces[J]. *Pacific Visualization Symposium*, 2012, PP(5):17-24.
- [23] Zhen L L, Li M Q, Cheng R, et al. Adjusting Parallel Coordinates for Investigating Multi-objective Search[J]. *Simulated Evolution and Learning*, 2017, 224-235.
- [24] Zhou Z G, Ye Z F, Yu J J, et al. Cluster-aware Arrangement of the Parallel Coordinate Plots[J]. *Journal of Visual Languages & Computing*, 2017.
- [25] Chen W, Zhu B, Zhang H X. BN-Mapping: Visual Analysis of Geoapatial Data with Bayesian Network. *Chinese Journal of Computers*, 2016, 7:1281-1293 [陈为, 朱标, 张宏鑫. BN-Mapping:基于贝叶斯网络的地理空间数据可视分析. *计算机学报*, 2016, 7:1281-1293]
- [26] Turkay C, Slingsby A, Hauser H, et al. Attribute Signatures: Dynamic Visual Summaries for Analyzing Multivariate Geographical Data[J]. *Visualization & Computer Graphics IEEE Transactions on*, 2014, 20(12):2033-2042.
- [27] Wang J P, Liu X T, Shen H W, et al. Multi-Resolution Climate Ensemble Parameter Analysis with Nested Parallel Coordinates Plots[J]. *IEEE Transactions on Visualization & Computer Graphics*, 2017, 23(1):81-90.
- [28] Lu M, Wang Z C, Liang J, et al. OD-Wheel: Visual Design to Explore OD Patterns of a Central Region[J]. *Visualization Symposium*, 2015, 87-91.
- [29] Chen Y, Lin X L, Zhao Y F, et al. SunMap: an Associated Hierarchical Data Visualization Method Based on Heatmap and Sunburst. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 7:1075-1083 [陈谊, 林晓蕾, 赵云芳, 等. SunMap:一种基于热图和放射环的关联层次数据可视化方法. *计算机辅助设计与图形学学报*, 2016, 7:1075-1083]
- [30] Yang Y L, Dwyer T, Goodwin S, et al. Many-to-Many Geographically-Embedded Flow Visualisation: An Evaluation[J]. *IEEE Trans Vis Comput Graph*, 2017, 23(1):411-420.
- [31] Li G Q, Deng M, Chen T, et al. A Dual Distance Based Spatial Clustering Method. *Acta Geodaetica et Cartographica Sinica*, 2008, 4:482-488 [李光强, 邓敏, 程涛, 等. 一种基于双重距离的空间聚类方法. *测绘学报*, 2008, 4:482-488]

第一作者简介:



周志光, 1983 年生, 男, 副教授, 主要研究方向为图形学与可视化。E-mail: zhgzhou1983@163.com。

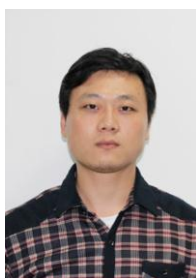


余佳珺, 女, 硕士研究生, 主要研究方向为可视化与可视分析。E-mail: yjj863352341@163.com。



郭智勇, 男, 在校学生, 主要研究方向为可视化与可视分析。E-mail: 15958130169@163.com。

通信作者:



刘玉华, 男, 讲师, 主要研究方向为可视化, 可视分析和社交网络。E-mail: liuyuhua@zufe.edu.cn。