

交互式电力数据可视化与分析

方思桐,汪隶鋈,高 敏,钱荣涛,陈雪羿,申丽铭,张力丹,刘日晨,王 琼

(南京师范大学计算机科学与技术学院,江苏 南京 210023)

[摘要] 并排法是传统的比较可视化中较为经典的方法,即将多个集合成员或不同的结果在一个显示空间中并排显示. 本文所使用的是具有时序性的异构、高维数据,包括全国 GDP 数据及中国西南某省 43 个企业或商场的用电数据,以及天气数据. 考虑到数据量庞大,并排可视化方法不具有空间协调性与视觉直观性,因此,本文使用重叠法与显式编码的比较可视化技术,使用户可自由选择日期,将多天的采样点数据重叠在一个显示空间中,避免了视觉混乱,使数据分析更加简洁高效.

[关键词] 电力数据可视化,可视分析,交互式探索,比较可视化

[中图分类号] TP311 [文献标志码] A [文章编号] 1001-4616(2019)03-0096-11

Interactive Power Data Visualization and Analytics

Fang Sitong, Wang Lijun, Gao Min, Qian Rongtao, Chen Xueyi,
Shen Liming, Zhang Lidan, Liu Richeng, Wang Qiong

(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

Abstract: Power data have the characteristics of time-sequence, heterogeneity and high-dimension, while the parallel visualization method universally used in the analysis of power data lacks of both spatial coordination and visual intuition. In view of these properties and limitations of traditional methods, an interactive visualization method is designed for power data comparison. This method makes use of the technology of superposition and explicit encoding comparison visualization to compare and analyze multi-day sample data in a carefully designed circular layout, avoiding visual confusion, improving the effect of multi-view comparison, making the process of data analysis more concise and efficient. In the experimental part, taking the data of GDP, electricity consumption and weather data of 43 enterprises or shopping malls in a province of China as examples, this paper validates the method we presented and compares it with the traditional method. The result indicates that the proposed method is more intuitive and flexible in data analysis.

Key words: power data visualization, visual analytics, interactive exploration, comparative visualization

信息可视化(InfoVis)技术,是一种以直观方式传递抽象信息的方法^[1],是理解复杂数据的不可或缺的工具^[2],其研究的数据主要包括数字和非数字数据,如文本和地理信息. 本文采用信息可视化技术,针对中国西南某省 43 个企业或商城的用电数据实现可视化,以增强用户认知,发现纷繁复杂的数据信息中的本质特征和规律.

本文数据存在显著的不确定性、数据的复杂性和可变性,如:所有企业的用电数据分布在 2010 年 6 月 1 日至 2012 年 3 月 9 日之间,并且企业数据天数分布非连续且不同;一天中每个企业每隔 15 min 一次的采样点集合;该省 17 个地点两年的天气记录(每天的天气状况,气温高温值和低温值). 因此,更需要对数据集合进行视觉调查,这就需要有效的比较可视化方法.

一般将常用的视觉比较可视化分为 3 类:并列(juxtaposition),即将多个集合成员或不同的结果在一个显示空间中并排显示;重叠(superposition),即将多个集合成员或不同的结果在一个显示空间中重叠显示;显式编码(explicit encoding),即计算多个集合成员或不同结果之间的差异,使用颜色、不透明度、大小、图表等编码差异^[3]. 本文所使用的数据具有明显的时序特征,通常使用折线图这类传统的并排法、比较法

收稿日期:2019-07-05.

基金项目:国家自然科学基金(61702271、41471371、61702270).

通讯联系人:王琼,副教授,研究方向:计算机图形学. E-mail:329789995@qq.com

进行刻画. 这种方法容易实现,但需要用户记忆图形特征进行比较,且并不适用于展示本文中由于粒度较小而产生的海量数据. 同样,此情形下显式编码所描述的差异表征会失去其直观性的优势. 因此本文使用重叠比较可视化方法,将多天的 96 个采样点数据以环形折线形式同时重叠在一个显示空间中,以距离圆心的远近作为采样点数据大小的参考,使用户视感更加直观. 此外,本文还采用了显式编码技术,主要采用颜色进行编码,形成良好的对比区分效果.

1 相关工作

为了便于科学研究中的数据分析,可视化技术应运而生. 折线图作为可视化方法中的经典工具,是一种特殊的频率图,可用于表示离散随机变量的频率分布^[4]. 该工具简单易于实现,已经广泛应用于各种领域. 早在 1999 年,出于将非语言信息转换为语言信息的目的,Kobayashi 等人展示了一个用自然语言解释汇率变化行为的折线图^[5]. 但由于折线图功能的单一性,Demir 等人开发了多图表,将条形图和折线图叠加在一起,以呈现集合成员的统计特性^[6]. 随着大数据时代到来,包括折线图在内的并排可视化方法展现出其视觉混乱(visual clutter)的缺陷,如急诊室或手术室监测心率等信息可视化检查时,用户只有很少的时间“一目了然”地查看可视化视图,由于这种缺陷,往往没有足够的时间检查折线图^[7]. 因此,在并排可视化的基础上,重叠可视化与显式编码可视化显得尤为重要. 为了方便研究两个时间序列之间的时变相关性,Wilson 等人采用了窗口互相关(wcc)技术,该技术使用了显式编码可视化方法. 该技术产生了两个时间序列之间的二维相关矩阵(列代表时间窗口的位置,行代表不同的滞后),其中不同的相关值映射不同的颜色^[8]. Biswas 等人为了将气候系统的不确定性量化,使用了并排、重叠、显式编码可视化方法,在其设计的多元尺度图(MDS)中,点的大小可根据所选参数的灵敏度值进行缩放,点的颜色则反映等级(较亮的颜色代表较高的等级). 通过交互选择,用户可以在“多点”模式下操作,通过在点周围绘制一条曲线来选择任意天数. 另一窗口的折线图视图随之更新,以显示用户所选日期的所有 5 个参数的敏感度值^[9].

2 实例与实验分析

2.1 设计

针对本文的海量用电数据,结合用户提供数据时的描述,本文的设计需要满足以下几个需求:

需求 R1:构建图表显示用电量的变化趋势,并允许自由切换地区;

需求 R2:允许切换时间,对不同日期的用电量进行对比分析;

需求 R3:分析天气状况、GDP 等与用电量之间的可能关系.

本文针对以上需求,进行了 3 个视图的设计.

2.1.1 信息可视化工具

本文主要使用了可视化库 d3.js. D3(Data-Driven Documents),即数据驱动文档,是一个 JavaScript 库,利用了广泛实施的 SVG、HTML5 和 CSS 标准,实现在 WEB 浏览器中生成动态、交互式的数据可视化. 使用 D3 构建的海量网站有着炫目的样式、流畅的动画与灵敏的可交互性.

D3 允许将输入数据与所选择的文档对象模型(DOM)绑定,并应用数据驱动的转换来生成、修改内容. D3 强调组件生态系统的建设,由技术复用降低了工具整体的复杂性;降低了对控制流的封装,提高了对调试的支持;专注于转换计算而非呈现过程,在实现动画和交互的同时提高了性能. 这些性质使得 D3 的灵活性相较于传统可视化工具大大提高^[10].

D3 的使用中,由数据绑定建立数据与图形元素之间的映射. 为了对元素生命周期进行精确控制,D3 引入了重要的概念,即对元素的输入、更新和退出(enter,update and exit):数据图形需要进入界面,并在输入时设置其生命周期内保持不变的属性;动态属性则在每次更新时重新计算^[10],实现图形的更新;当图形表示的数据不再有效时,则使其从页面退出.

2.1.2 映射方法与布局设计

可视化映射是整个可视化流程的核心^[11],即将数据表映射为可视化结构,由空间基、标记,以及标记的图形属性等可视化表征组成^[12]. 作为对 Protovis 的替换,D3 的 SVG 模块提供了适合绘制可视化映射的各种形状^[11],通过将各种图形函数绑定到 SVG 路径元素的“d”属性,可以实现自定义的组件. 在此基础上

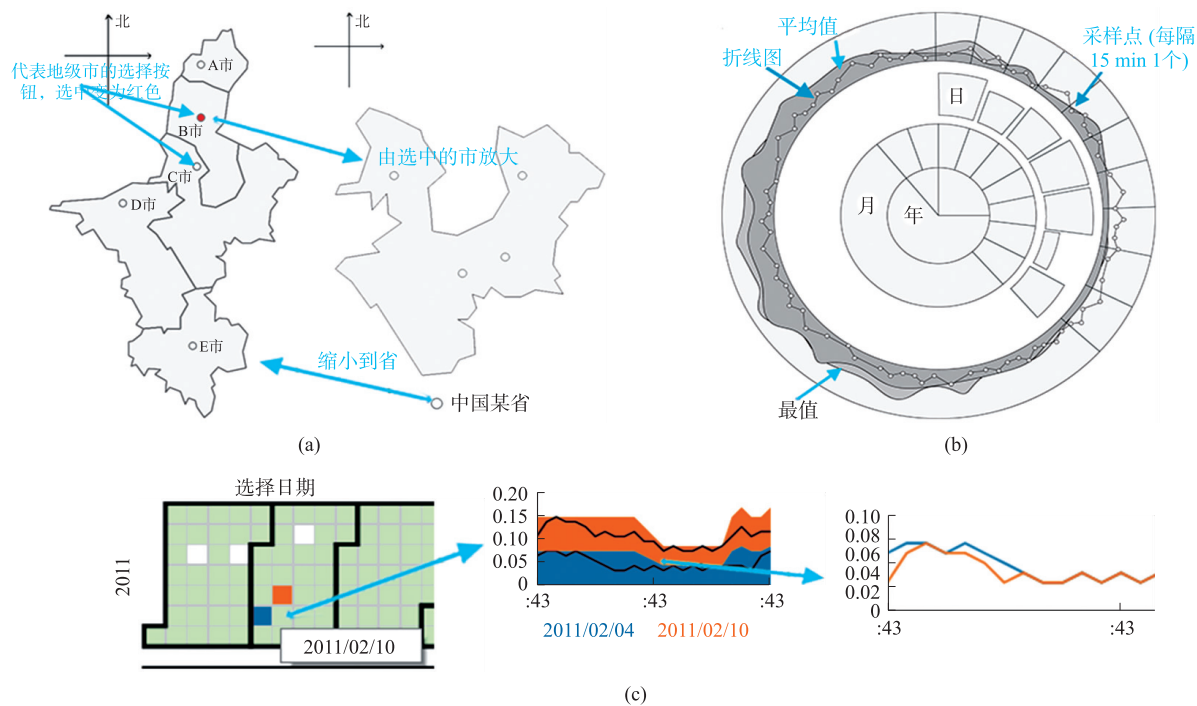
上,如何建立逻辑清晰的模型应当被仔细考虑.

可视化映射模型需满足两个基本条件:一是需要真实地表示并保持原数据的样貌;二是可视化映射形成的可视化表征需要易于理解、同时又具有丰富的表达能力,即能够充分展现数据的特征^[13]. 所要展现的用电数据具有 3 个属性:地理位置、采样时间以及天气状况^[14],因此本文将从这几个角度进行映射.

(1)地理位置属性映射

本文总共监测 43 个企业的用电信息,并且拥有相关坐标数据. 由于企业分布于市、市又隶属于省的层次关系,本文提出的设计首先绘制省市地图,再将企业坐标投影到地图上. 如图 1(a)所示,通过缩放,允许用户对默认的省级地图进行进一步地探索,将所关注的范围从全省逐步转移到所感兴趣的市或具体企业上,并对相同页面下的其他组件进行同步更新.

同样,对于全国 GDP 数据,在本文中也由全国地图信息绘制中国地图,并采用颜色插值的显式编码比较可视化技术展示数据的总体情况. D3 支持从文件读取数据并生成 SVG 图像,并具有地理坐标的转换系统,因此能准确地呈现所需要绘制的地理信息.



(a) 左边为默认情况下的全省地图,单击所想要观察的市则可进行局部放大,并允许选择该市内具体企业进行观察,此过程中同页面组件将同步更新. (b) 环形图,从中心向外的圆环分别代表年、月、日,最外圈为 24 h 的刻度,通过单击可选择想要观察的具体日期,并同步更新外圈的折线图与环形面积区. (c) 日历图,在左边的二维日历表格中可选择多个日期并自动标注为不同颜色,同时将采用相同颜色编码更新堆叠图与折线图.

图 1 部分组件设计
Fig. 1 Part of the designs of components

(2)采样时间属性映射

本文使用的数据由电力公司每隔 15 min 采集一次用户的用电信息组成,跨度约为 3 年. 这样密集地交互构成了智能电网中的用户侧大数据^[15],具有鲜明的时序特征.

时序数据是指按照时间顺序取得的一系列观测值^[16],具有自然的时间顺序,且通常是单向的,因此通常使用折线图绘制以展示其规律. 然而,考虑到本文所使用的用电数据具有较细的粒度,且需要更为灵活的交互手段以允许用户转移兴趣点,显然不适合使用经典的折线图进行刻画. 因此,本文着重考虑了相关组件的设计,力求在保持原始数据特征的前提下提供易于理解的视图.

环形图(circos). Ben Shneiderman 曾提出经典的信息搜寻理论,即概要优先、按需探索细节^[16]. 遵从这样的理念,环形图组件被提出. 环形图最初被设计用于研究基因组分析,目前被广泛应用于各个领域^[18]. 本文数据采样的最小单位为分钟,最大单位为年,由此可以构造一个层层嵌套的环形图组件;其最

内层圆环为年单位,向外分别为月、日、分钟,允许用户指定自己感兴趣的日期进行分析。

本文数据共涉及 43 家企业,每家企业一天之内就会产生多达 96 条记录。当时间分布跨度为数年时,如此庞大的数据将难以由折线图这样的简单图表所组织,并且传统的并排可视化比较法将不再适用。而本文所设计的环形图采用嵌套的样式,类似于钟表,让用户从内层向外指定所关注的年、月、日,在此限定条件下绘制一条或多条环形折线图,采用重叠比较法展示一天或多天的采样点分布,形成清晰的对比关系。

由于环形图可叠加多层信息的灵活性与可扩展性,环形图内可以嵌套月用电量最高值及月均值等信息,采用显式编码的比较可视化法绘制相应环形面积区,因此可以证明其足以承载多方面的数据信息。其具体设计如图 1(b) 所示。

日历图(calender)。同样为用户提供关注时间范围的手段,日历图将日期映射到二维平面内,在其他辅助视图内编码时序信息。如图 1(c) 所示,用户首先点击二维的日历图表上的小方格指定感兴趣的具体日期,此时同步更新承载用电量数据的堆叠图(stacked area graph)与折线图。当用户选择多个日期时,堆叠图上也对应产生多组数值,一方面通过折线刻画每组数值(即用电量)随时间(即一天中的 24 h)的变化规律,另一方面通过不同颜色的面积对比不同日期的用电量大小。

(3) 天气属性映射

散点图(scatter-plot)。散点图是一种经典的图表,独立地测量产生变量对 (X_i, Y_i) ,并使用笛卡尔坐标显示两个变量值,其中 X 和 Y 之间不需要具有函数关系。事实上,在所有形式的统计图形中,散点图几乎被认为是整个统计图形历史上最通用、最具多态性和最有用的发明^[19]。散点图可以用于推导变量之间的各种相关性,探索经验数据的规律性,在本文中用于承载用电量数据 (Y_i) 与天气数据 (X_i) 的变量对。其中, X_i 为某一天的天气属性, Y_i 为某一天的 96 个用电量采样数据之和。在数据量充足的情况下,通过观察点的分布趋势,用户可以初步窥见这两个变量间的可能关系。作为扩展,可以对点进行彩色编码,实现不同时间段数据的区分对比。

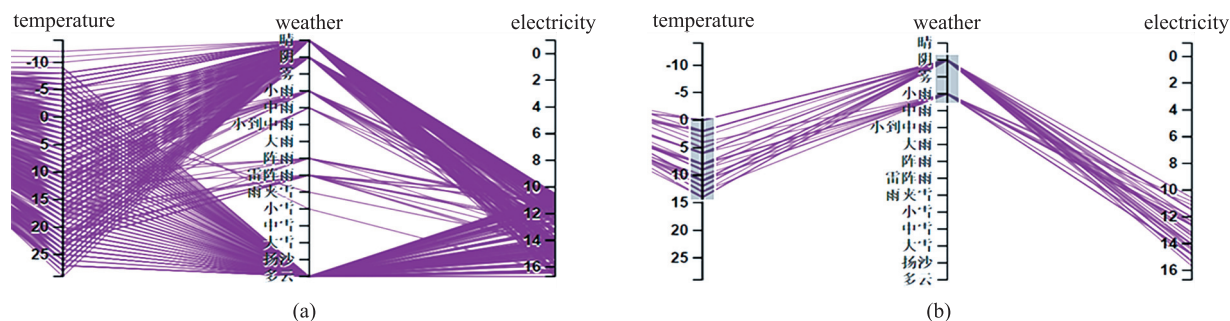
平行坐标(parallel-coordinates)。为了探寻天气、温度与用电量等多维数据之间的关系,本文采用经典的平行坐标视图进行设计。

平行坐标通常被用于高维(多元)数据的分析。绘制一个由 N 条平行、垂直且间距相等的轴,每条轴代表一个变量。一条多段线跨越所有的轴连接为一条折线,代表一个数据组。这样,在多条折线组成的空间中,可以通过线的密度推测各个维度变量之间的关系,如图 2(a) 所示。

由于相邻变量之间的关系比非相邻变量的关系更容易被观测到,轴的顺序对于推测数据特征至关重要。本文中轴的排序设置为年-月-日-温度-天气-(日)用电量。

平行坐标具有数据密集时难以辨认的缺陷。不过,采用一种经典的焦点+上下文(focus+context)技术可以针对此问题进行优化,即刷选(brush)交互技术。这种技术允许用户标记他们感兴趣的子集^[20],从而过滤冗余数据。在我们的组件中,用户可以在默认的上下文中进一步缩小自己感兴趣的日期范围并限定天气状况,筛选出所要重点观察的数据,效果如图 2(b) 所示。

综合以上的考虑,当用户将所要观测的时间范围缩小到具体日期级别时,主要采用折线图的形式绘制 96 个采样数据;当用户选择宏观考察时,主要通过环形图面积来反映用电总量或平均值,并由平行坐标图



对图(a)采用刷选技术可以过滤冗余数据,达到类似于图(b)的效果。

图2 平行坐标视图及刷选技术示例

Fig. 2 Parallel-coordinates plot and the examples of brush technology

及散点图来探寻天气、温度、用电量之间的关系.通过组合以上的自定义组件,即可反映用电量随时间的变化趋势以及与其他多个变量的关系,并且支持比较可视化.

2.2 实现

2.2.1 数据描述

本文的数据主要包括以下几个方面的数据:地理位置信息数据,用电量数据,天气状况数据,以及GDP数据.地理位置信息数据主要包括各省市的地图数据,用于转换绘制为SVG地图,以及43个企业的坐标数据,被投影于所绘制的地图上;43家企业的用电量数据跨度为2010年6月-2012年3月,每天每15 min采样一次,以天为单位存储为一条记录;同时间段内的天气状况数据包括天气描述(如,多云)及具体温度,同样以天为单位存储为一条记录;GDP数据为2010-2011年的平均GDP,每个省的数据形成一条记录.

数据存储采用的是逗号分隔值文件格式,即CSV(comma-separated values),能够为D3中的CSV解析器所处理.在使用Python搭建的微型服务器之上,D3可以从本地加载数据、完成页面元素之间的转化.该库使用格式不可知处理^[21],降低了数据转换可能造成的资源消耗.

2.2.2 实现过程

整个开发过程围绕D3的“输入、更新和退出”模式,将数据与图形之间形成映射关系,藉由D3的基础图形绘制库函数实现关键组件的设计,并最终实现完整视图的构建.这样的机制为页面元素提供了便捷的控制,并实现了流畅的动态效果.值得一提的是,本文工作实现部分只是基于这些现有的可视化库,在具体实现中,进行了大量的编码与扩展,因为这些可视化库提供的功能都是针对最基本的可视化组件.

此外,通过支持事件监听器,即接收针对特定元素的用户输入事件的回调函数,系统能够很好地支持用户交互.事件监听器可以通过命名空间共存于元素上.当每个元素完成动画制作时,由已注册的监听器完成连续的转换和后期动画清理(即删除要从页面上拿走的元素),并使用统一的计时器保证每个元素的排他性,平滑地改变对象属性或样式^[10].

具体实现过程将以重点构建的环形图与平行坐标为例进行说明^[14].

环形图绘制.环形图的绘制,首先将日期切分为年、月、日数组,用`d3.svg.arc()`创建与之对应的`arc1`、`arc2`、`arc3`圆弧,并用`d3.svg.area()`创构建每个月用电量的最大值区域`area1`和平均值区域`area2`.新建`year_circle`(年圆环)、`month_circle`(月圆环)、`day_circle`(日圆环)、`elecarea`(月用电量最高值环形面积)、`elecavg`(月用电量均值环形面积)图形元素,并将其`svg.path`元素的`d`属性分别赋值为`arc1`、`arc2`、`arc3`、`area1`和`area2`.

环形图最外圈的`elec`圆弧用于表示一天24 h的圆周时钟刻度.其上的`elec_circle`图形为一条圆周折线图,刻画一天中的96个采样数据点.`Circos`元素绑定了各种交互事件,例如触发`clickDay`时将同步更新`elec_circle`;触发`clickMonth()`时将根据所选择的月份刷新`area1`和`area2`,并默认选择日期为此月1号、触发`elec_circle`的更新.

平行坐标图绘制.用`d3.scale.ordinal()`有序比例尺创建的X轴由6个维度组成,分别为`year`(年)、`month`(月)、`day`(日)、`temperature`(温度)、`weather`(天气)以及`electricity`(用电量).用`d3.scale.linear()`线性比例尺创建的Y轴与这6个维度对应.轴交互图形元素与数据绑定,并为轴交互图形绑定刷选交互方式,便于用户过滤、筛选数据.

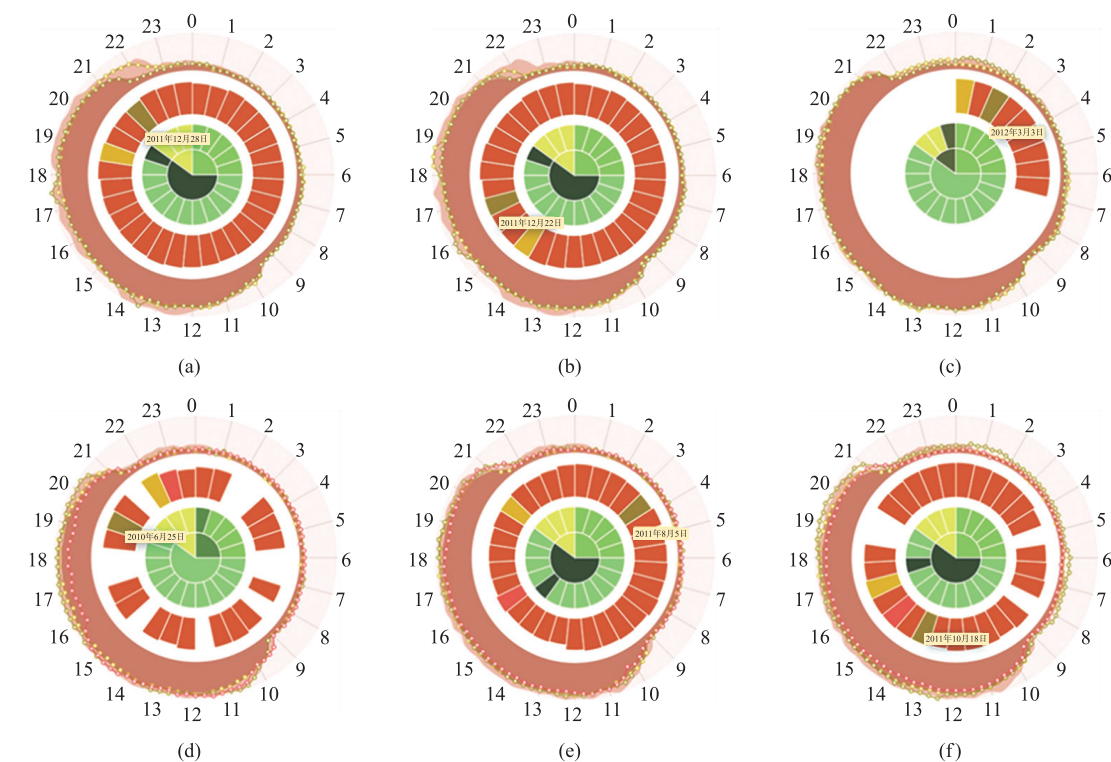
3 结果与评估

3.1 可视化结果的探索与分析

3.1.1 比较可视化结果与分析

如图3所示,环形图将3~4个环状图进行叠加,并基于用电数据的时序性特征,使用户可从内环至外环,依次选择年月日,查看某企业用电数据的情况.当用户由内而外依次选择年月日时,当前月用电数据的平均值与最值则映射为圆环之间的环形面积区大小.最外层折线上的圆圈代表了该天的96个采样点的数据.鼠标悬停在日期圆环上超过1秒时,可显示出年月日信息.

为了增加对比功能,本视图采用了比较可视化技术.如图3所示,该视图采用了重叠法,当用户选择两次或三次日期后,将所选择的若干天的用电数据的平均值与最值映射为环形面积区大小,并在最外层画有2~3条分别代表所选日期的96个采样点的折线,即将选择日期的采样点数据在视图中重叠显示.如



图(a)(b)(c)为选择 2 d 的环形图;图(d)(e)(f)为选择 3 d 的环形图.

图 3 环形图
Fig. 3 Circos plot

图 3(a)(b)(c),是用户选择 2 d 的情况,用户选择 1 d 后,该天的高亮颜色与折线颜色对应同为暗黄色,选择另一天后,该天的高亮颜色为墨绿色,与第一次选择的高亮颜色形成对比,方便用户查看. 用户根据折线上的采样点距圆心距离即可分析出,不同日期同一时刻的采样点差距. 图 3(d)(e)(f)则反映了选择 3 d 查看采样数据的情况,步骤与选择两天相同,但第 3 d 的高亮颜色为玫红色.

例如,在图 3(a)中,用户先选择 2011 年 12 月 25 日,后选择 2011 年 12 月 28 日,则 2011 年 12 月 25 日所在区域高亮为暗黄色,2011 年 12 月 28 日所在区域高亮为墨绿色,这两天对应的 96 个采样点折线也变为暗黄色与墨绿色,并由折线可轻易分析出 2011 年 12 月 25 日与 2011 年 12 月 28 日在 22 时的用电量差距较大.

3.1.2 整体到局部式数据探索

本文所有视图都采用了整体到局部(overview-to-details)技术,即视点控制技术,通过放大可视化结构或改变视点,使得细节可见. 本文同时使用整体与局部两个窗口,整体可提供局部视图的信息,并可以控制改变局部视图^[22].

如图 4 所示,该视图分为 3 个部分,自左向右为:环形图,企业分布图,散点图.

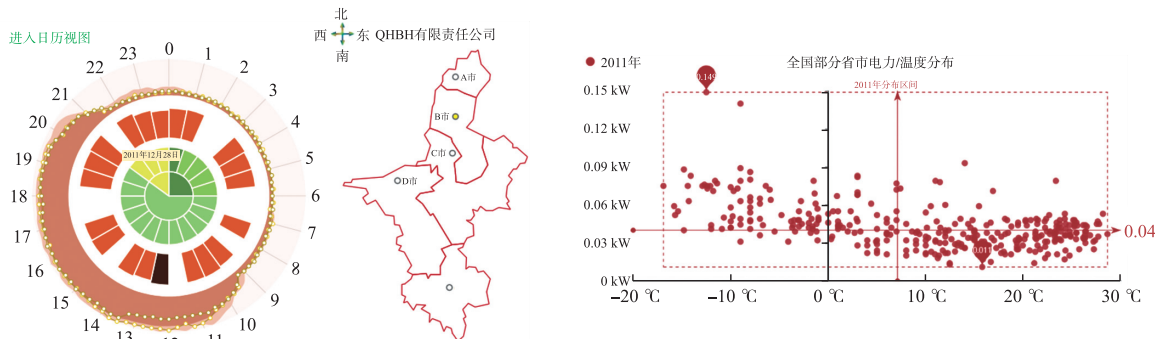


图 4 视图 1 效果图
Fig. 4 The rendering of view 1

环形图与企业分布图. 环形图主要可用于比较可视化, 已在前文详细阐述. 该视图将用电量沿环形时间轴可视化, 并与企业分布图相关联. 由于使用了整体到局部技术, 默认情况下, 企业分布图为省份概要图, 而当点击某市标记圆圈时, 可使标记变黄, 并使该市的地图放大显示.

散点图与企业分布图. 散点图展现了某省市温度与该地企业用电量的关系, 同样与企业分布图关联, 当点击某市标记时, 散点图切换至该市数据. 当鼠标移至散点图时会展现出此点的温度与用电量信息, 并计算出并显示出温度与用电量的平均水平. 考虑到数据过于繁杂的可能性, 本文使用了焦点+上下文技术, 如图 5 所示, 可选择某一区域放大观察. 例如, 观察图 5 可以得出结论: 该省市气温低时用电量较气温高时更多, 数据多集中于 20 ℃ ~ 30 ℃, 此时用户可放大该区域观察.

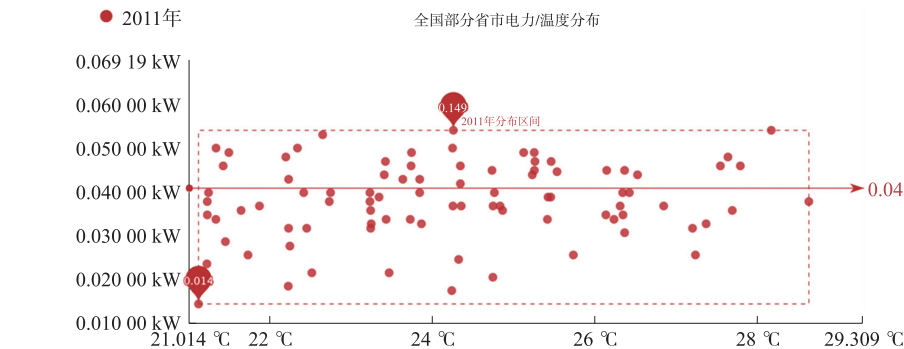


图 5 散点图: 焦点+上下文示例

Fig. 5 Scatter plot: an example of focus+context

继视图 1 以环形图、企业分布图与散点图展示用电量与日期、温度、地理位置之间的关系后, 为了使用户可选择对比 3 d 以上的用电数据. 本文在视图 2 中使用日历视图、堆叠视图与折线图展示不同日期、不同地区的用电量, 并形成有效对比.

如图 6 所示, 该视图由左方的日历图、堆叠图, 右方的企业分布图组成. 视图同样应用整体到局部技术, 用户可通过企业分布图中选择位于某市的企业来更新日历视图、堆叠视图.

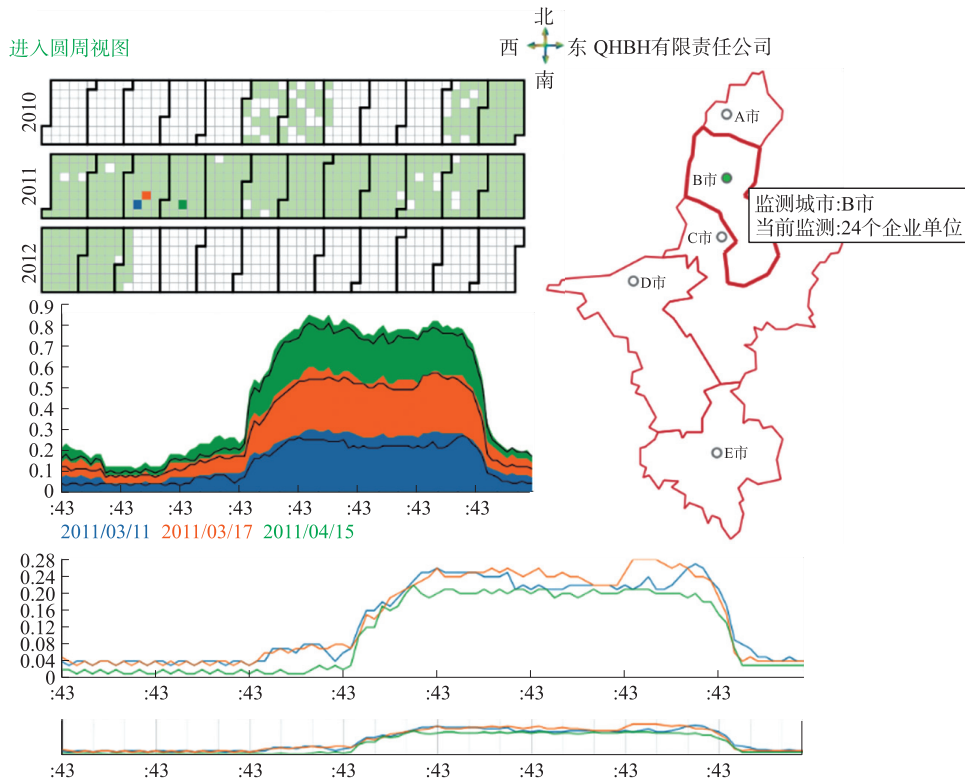


图 6 视图 2 效果图

Fig. 6 The rendering of view 2

日历视图与企业分布图. 日历视图以日历形式显示时间与用电量的关系,并与企业分布图相关联,当用户在企业分布图中点击代表某市的标志,则标志变绿,日历视图也随之切换至该企业数据. 如图 6 所示,该日历视图可展示 2010-2012 年之间的数据,透明为数据缺失部分,用户可通过点击非透明方格来切换日期. 结合显式编码技术,本视图可选择多个日期进行比较,点击两次即可取消选中. 如图 6 所示,选中的日期分别为 2010/06/01、2011/7/14 与 2011/8/30,且方格分别变为蓝、橙、绿色,以便于在堆叠视图中查看.

堆叠视图. 用户在日历视图指定日期后,堆叠图随之切换为所选中企业的用电量数据. 该视图采用了比较可视化中的重叠法,可根据日历视图所选日期进行数据的堆叠. 当用户在日历视图中取消选中时,便取消该天的堆叠,此过程中,坐标轴始终是自适应的. 如图 6 所示,该视图显示了 2010/06/01、2011/7/14 与 2011/8/30 的用电量数据,且将 3 d 数据分别用蓝、橙、绿显示,与日历视图一致,便于用户查看. 考虑到用户选择天数过多时,易造成视觉混乱,本文在下方的折线图添加了焦点+上下文技术,可通过刷选操作缩小范围,使数据分析更加高效.

视图 1 与视图 2 可使用户从不同角度分析出用电量与温度、日期、地理位置的关系. 此外,本文设计了视图 3,用以展示全国各省的 GDP 与用电量之间的关系.

如图 7 所示,视图 3 由左上方中国地图、右上方的动态转换图(donut transitions)与下方的平行坐标图组成.

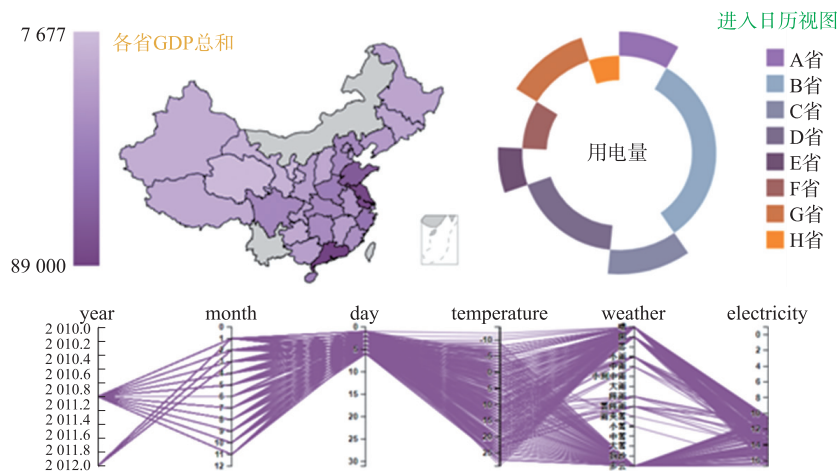


图 7 视图 3 效果图

Fig. 7 The rendering of view 3

中国地图与平行坐标图. 中国地图使用了颜色插值技术,将各省的 GDP 水平映射为不同深浅的颜色,以显式编码技术十分直观地对比了各省的 GDP 水平差异. 同时,中国地图与平行坐标图也采用了整体到局部技术,当选择地图上的某省时,平行坐标图则会切换为该省的年、月、日、温度、天气与用电量数据,每一条折线都代表某一时间点的用电量与温度、天气关系.

如图 7 所示,尽管数据量大时,会出现折线重叠的情况. 但用户可根据需要将数据密集且不易分析部分过滤. 或者当用户只希望关注某个特定区间的数据时,同样可以采用此交互工具,其结果如图 8 所示.

动态转换图. 该视图为动画视图,使用了复杂的链式转换技术,用于更新圆环图中的值. 该动态图具

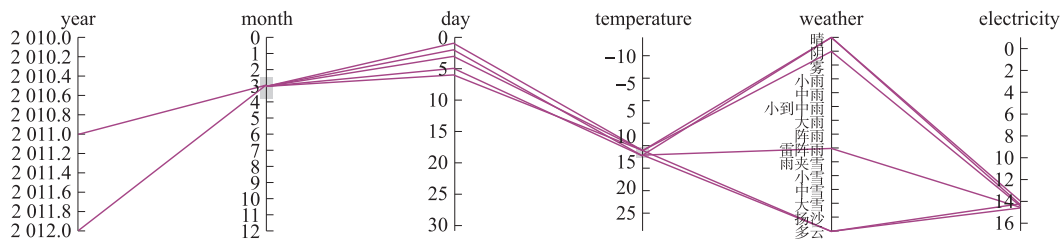
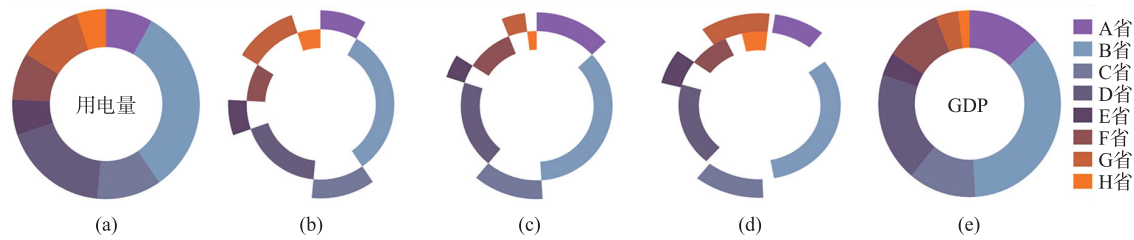


图 8 平行坐标折线图:过滤效果

Fig. 8 Parallel-coordinates plot: the effect of filtering

有两个状态,如图 9(a)(e)所示,并在两状态之间进行过渡变换. 两个状态分别为各省市的 GDP 比例与用电量比例,图 9(b)(c)(d)均为过渡阶段的部分截图. 若用户需要针对某省观察其用电量与 GDP 关系,可根据过渡状态时,代表该省的圆弧是否增大减小,直观分析两者关系. 如:虽然 G 省用电量占据相当比例,但在从用电量到 GDP 的过渡状态中,代表 G 省的圆弧明显缩小,即 GDP 所占比例明显减小,可得出 GDP 与用电量水平并不成正比.



图(a)为显示用电量状态;图(b)(c)(e)为过渡状态;图(e)为显示 GDP 状态.

图 9 动态转化图. 该图为动画

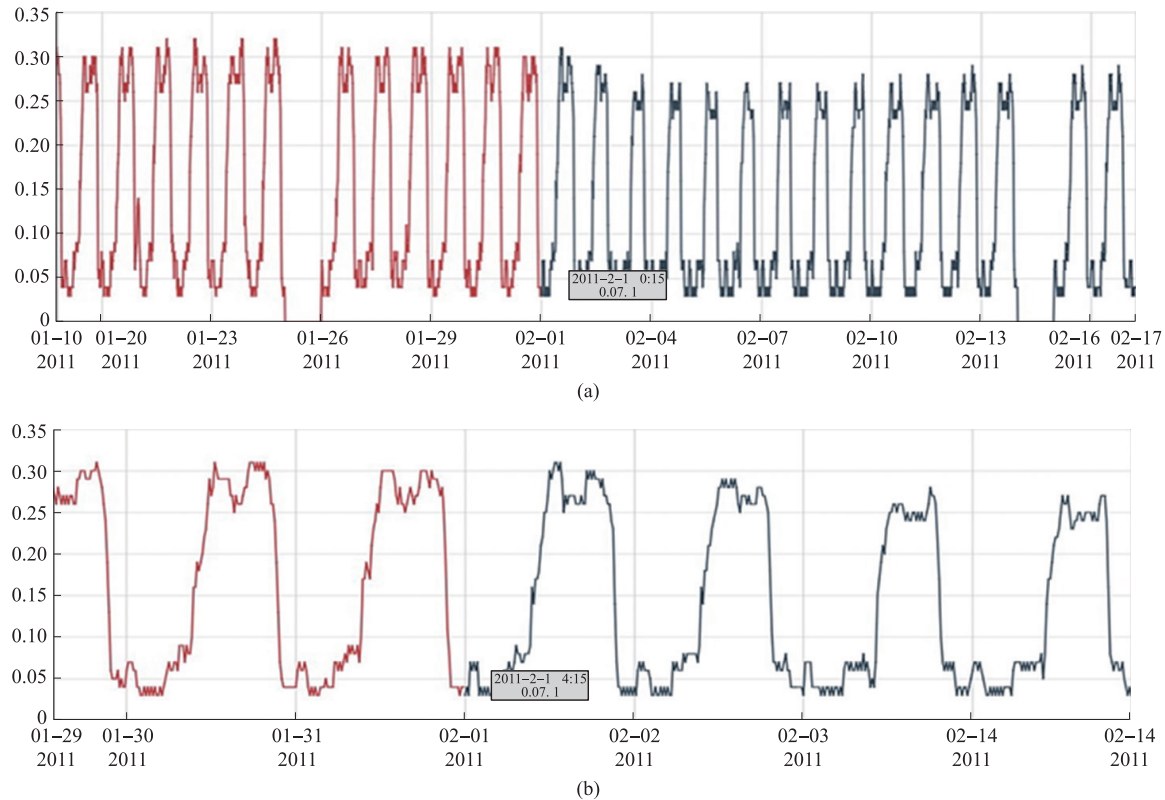
Fig. 9 Dynamic transformation diagram, which is animated

3.2 方法评估与讨论

折线图是一种图表类型,为数据可视化的传统工具,它将信息显示为一系列由直线段连接的数据点,称为“标记”. 折线图通常用于显示数据随时间间隔(时间序列)变化的趋势,因此通常按时间顺序绘制折线图. 图 10(a)(b)当中的折线图只导入了两个月范围的用电量数据,并且采用了焦点+上下文技术. 对比实验则基于这种折线图与本文所设计的环形视图进行.

3.2.1 基于环形图的时间选择评估

如 10 图所示,使用折线图对比多天数据很不直观. 例如用户需比较 1 月 30 日与 2 月 4 日的数据,只能通过纵轴的刻度来进行对比,过分依赖于用户的观察能力与大脑的分析能力. 而若日期相隔较远,用户



图(a)为上下文状态下的视图,对其采用刷选技术后进入如图(b)所示的焦点状态

图 10 传统折线图,该图导入了两个月的用电数据

Fig. 10 Traditional line chart, which imports two months' electricity consumption data

使用时将会大概率出现观察误差,为用户的使用带来很大不便。

本文中的环形图则有效解决了这个问题。该视图可任意选择某年、月、日进行对比,无论所选日期相隔多久,都能够根据最外圈的重叠折线进行数据比较分析,大大提高了用户分析数据的便捷性。

3.2.2 视觉直观性

由图 10 的对比,易得出当数据在焦点视图下,观察数据点较为直观,但一旦由焦点视图转移到上下文视图中,便造成了视觉混乱(visual clutter),使用户观感不佳,分析效率低下。

本文中的环形布局视图将用户不需要的数据隐藏,只针对用户的兴趣点进行可视化,很大程度上减少了视觉干扰,使用户能够更加高效快捷地工作,进行对数据的观察判断。

3.2.3 空间布局协调性的探讨

针对折线图令人眼花缭乱的观感与过度占用界面空间的缺点,本文所设计的视图使用焦点上下文、整体到局部、按需探索细节(details-on-demand)、刷选等技术进行探索。焦点上下文技术是一种放大显示画面中某块小的局部区域的透镜技术,放大区域的周围放到背景显示。焦点为某市或某地区的一些特定商场,上下文为不同省份乃至全国,允许隐藏用户不需要的信息,放大用户需要的信息,使界面整洁高效。按需探索细节技术,可在总览整体信息的情况下,交互式地选择一部分信息展示其细节^[22]。刷选技术则是用于高亮用户所感兴趣的子集。本文使用了这些技术后,可使在更少的空间占用情况下,更加简洁高效地展示给用户更多的信息,形成更加有效的对比效果。

4 结束语

本文针对具有时序特征的海量用电数据,设计了一种新颖的重叠比较可视化方法,将多天的 96 个采样点数据以圆周折线的形式重叠在同一个显示空间中,以距离圆心的远近作为采样点数据大小的映射,避免了传统的并排方法所可能造成的认知阻碍。系统具有丰富的交互手段以允许用户转移兴趣点,筛选所需要观察的数据信息,在保持原始数据特征的前提下提供了易于理解的视图。实验评估证明了本文提出的方法相较于传统的并排比较法(折线图)更高效。

然而,当用户选择天数过多时,依然可能造成圆周折线的重叠问题。这类密集数据问题,部分已在前期工作中解决^[23-29],剩余的问题将在以后的工作中进行进一步的探讨。

5 致谢

本文工作得到了国家自然科学基金(基金号:61702271、41471371、61702270)的支持。同时对北京大学机器感知与智能教育部重点实验室 PKUVIS 研究组成员张江、艾宏伟等人对该工作的前期开发提供的帮助与建议表示感谢,对相关电网公司提供数据表示感谢。

[参考文献]

- [1] KEIM D A. Information visualization and visual data mining[J]. IEEE transactions on visualization and computer graphics, 2002,8(1):1-8.
- [2] 舒清雅,刘日晨,洪帆,等. 集合模拟可视化进展[J]. 软件学报,2018,29(2):506-523.
- [3] GLEICHER M,ALBERS D,WALKER R,et al. Visual comparison for information visualization[J]. Information visualization, 2011,10(4):289-309.
- [4] DODGE Y. The concise encyclopedia of statistics[M]. USA:Springer Science & Business Media,2008.
- [5] KOBAYASHI I. Toward text based information processing:with an example of natural language modeling of a line chart[C]// IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028). Tokyo,Japan:IEEE,1999.
- [6] DEMIR I,DICK C,WESTERMANN R. Multi-charts for comparative 3d ensemble visualization[J]. IEEE transactions on visualization and computer graphics,2014,20(12):2694-2703.
- [7] RYAN G,MOSCA A,CHANG R,et al. At a Glance:pixel approximate entropy as a measure of line chart complexity[J]. IEEE transactions on visualization and computer graphics,2019,25(1):872-881.

- [8] KÖTHUR P, WITT C, SIPS M, et al. Visual analytics for correlation-based comparison of time series ensembles[J]. Computer graphics forum, 2015, 34(3): 411–420.
- [9] BISWAS A, LIN G, LIU X, et al. Visualization of time-varying weather ensembles across multiple resolutions[J]. IEEE transactions on visualization and computer graphics, 2017, 23(1): 841–850.
- [10] BOSTOCK M, OGIEVETSKY V, HEER J. D³ data-driven documents[J]. IEEE transactions on visualization and computer graphics, 2011, 17(12): 2301–2309.
- [11] 陈为, 张嵩, 鲁爱东. 数据可视化的基本原理与方法[M]. 北京: 科学出版社, 2013.
- [12] CARD S K, MACKINLAY J D, SHNEIDERMAN B. Readings in information visualization: using vision to think[M]//Readings in information visualization. San Francisco: Morgan Kaufmann, 1999.
- [13] 任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. 软件学报, 2014(9): 1909–1936.
- [14] 张力丹. 用电数据可视分析的应用研究[D]. 南京: 南京师范大学, 2018.
- [15] WANG P, RAO L, LIU X, et al. D-pro: dynamic data center operations with demand-responsive electricity prices in smart grid[J]. IEEE transactions on smart grid, 2012, 3(4): 1743–1754.
- [16] 郑斌祥, 席裕庚, 杜秀华. 基于离群指数的时序数据离群挖掘[J]. 自动化学报, 2004, 30(1): 70–77.
- [17] SHNEIDERMAN B. The eyes have it: a task by data type taxonomy for information visualizations[C]//Proc IEEE Symposium on Visual Languages. Boulder, Colorado: IEEE Computer Society, 1996.
- [18] KRZYWINSKI M, SCHEIN J, BIROLI I, et al. Circos: An information aesthetic for comparative genomics[J]. Genome research, 2009, 19(9): 1639–1645.
- [19] FRIENDLY M, DENIS D. The early origins and development of the scatterplot[J]. Journal of the history of the behavioral sciences, 2005, 41(2): 103–130.
- [20] DOLEISCH H, GASSER M, HAUSER H. Interactive feature specification for focus+context visualization of complex simulation data[C]//Nordic Conference on Human-computer Interaction: Extending Boundaries. Reykjavik, Iceland: ACM, 2010.
- [21] HEER J, BOSTOCK M. Declarative language design for interactive visualization[J]. IEEE transactions on visualization and computer graphics, 2010, 16(6): 1149–1156.
- [22] 张卓, 宣蕾, 郝树勇. 可视化技术研究 with 比较[J]. 现代电子技术, 2010, 33(17): 133–138.
- [23] LIU R, CHEN S, JI G, et al. Interactive stratigraphic structure visualization for seismic data[J]. Journal of visual languages and computing, 2018, 48(1): 81–90.
- [24] LIU R, GUO H, YUAN X. Seismic structure extraction based on multi-scale sensitivity analysis[J]. Journal of visualization, 2014, 17(3): 157–166.
- [25] LIU R, GUO H, YUAN X. A bottom-up scheme for user-defined feature comparison in ensemble data[C]//ACM SIGGRAPH Asia Symposium on Visualization in High Performance Computing. Kobe, Japan: ACM, 2015.
- [26] LIU R, GUO H, ZHANG J, et al. Comparative visualization of vector field ensembles based on longest common subsequence[C]//IEEE Pacific Visualization Symposium. Taipei, China: IEEE, 2016.
- [27] LIU R, GUO H, YUAN X. User-defined feature comparison for vector field ensembles[J]. Journal of visualization, 2017, 20(2): 217–229.
- [28] GAO M, WANG L, JIA J, et al. Interactive geological visualization based on quadratic-surface distance query[J]. Journal of electronic imaging, 2019, 28(2): 1–10.
- [29] GAO M, XIANG Y, WANG L, et al. Histogram-based nonlinear transfer function edit and fusion[C]//The 10th International Conference on Image and Graphics. Beijing, China: CSIG, 2019.

[责任编辑: 黄 敏]