## 一种通用的数据可视化模型设计与实现

## 宋美娜 崔丹阳 鄂海红 欧中洪

(北京邮电大学计算机学院 PCN&CAD 中心 北京 100876)

摘要 随着互联网、物联网、云计算等信息技术的迅猛发展,大量非结构化信息不断涌现,对数据可视化与可视化分析的需求不断增加,同时可视化开发成本不断提高。数据可视化模型的研究对于可视化开发必不可少。基于此背景,根据数据信息的分类研究提出一种通用、模块化的数据可视化模型,将可视化流程分为可视化结构映射、可视化空间布局、可视化视图映射三个主要模块,并基于数据可视化模型利用开源的 Echarts 图表工具包和D3 函数库形成可视化工具 DVL. js(Data Visualization Layout)。该工具可应用于 Web 平台,包含丰富的图形展示方式如柱状图、折线图、时序图、饼图、雷达图、层次树、力导向图、标签云等,这些展示方式都提供了完备的数据与交互接口,能应对不同场合的可视化需求。

关键词 数据可视化模型 可视化布局 图形插件

中图分类号 TP3 文献标识码 A DOI: 10.3969/j. issn. 1000-386x. 2017.09.008

# DESIGN AND IMPLEMENTATION OF A GENERAL INFORMATION VISUALIZATION MODEL

Song Meina Cui Danyang E Haihong Ou Zhonghong

( PCN&CAD Center, School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract With the rapid development of the Internet, IoT, cloud computing and other information technology, large volumes of unstructured data are emerging. However, as the increasing demand for data visualization and visual analysis, the cost of visual development is increasing. Hence, researches on data visualization models are essential for visualization development. In this context, based on the classification of data information, this paper presents a generic, modular data visualization model which is divided into three modules, including visualization structure mapping, visualization spatial layout, and visualization view mapping. And based on the data visualization model, the DVL js tool (Data Visualization Layout) is created by using the open source Echarts graph toolkits and the D3 function libraries. The DVL js toolkits can be applied to Web platforms, contain rich graphical displays such as bar charts, line charts, time sequence diagrams, pie charts, radar charts, and hierarchical tree, force directed graph and tag clouds. All these display methods provide complete data and interactive interfaces, which can meet the visualization requirements of different occasions.

Keywords Data visualization model Visualization layout Graphical widgets

## 0 引 言

数据可视化被定义为采用图形语言,利用人类对视觉可视模式的快速识别能力对异质性数据进行视觉呈现,主要研究人、计算机表示的信息以及它们的相互

影响的研究领域,是近年来的研究热点。

Card 等 <sup>11</sup> 首次提出可视化参考模型描述从数据源到可视化数据集的可视化映射过程。可视化映射表示从基于数学关系的数据表映射为被人视觉感知的图形属性结构,是可视化参考模型的核心。可视化映射中的数据结构需要是真实并保持数据的原貌,形成的

收稿日期: 2016-07-04。国家科技支撑计划项目(2014BAK15B01)。宋美娜,教授,主研领域: 云计算与大数据。崔丹阳,硕士生。鄂海红,副教授。欧中洪,副教授。

可视化表征或隐喻要易于用户快速感知,能够充分表达数据的趋势性、相似性、差别性等特征。

Shneiderman <sup>②</sup> 根据数据信息的特征将数据可视化分为一维、二维、三维、多维、层次、网络、时序等可视化类型。研究者围绕上述数据类型提出了众多的数据可视化新方法和新技术,随着大数据时代与互联网、社交网络等领域的交叉融合,逐渐催生了几类特征鲜明的数据可视化类型,例如文本标签图可视化、地图可视化。

周宁等<sup>3</sup> 提出了数据可视化系统的 RDV 模型,将 数据可视化过程分解为数据源、可视化数据、可视化对 象三种数据空间,以及映射和数据析取两个过程,并基 于代数结构原理介绍数据可视化整体结构。在知识可 视化和信息可视化的比较 4 中研究分析了可视化的基 本思想。刘晓平等 

以协同设计中静态约束信息的映 射研究中提出了可视化映射过程: 可视化结构映射、可 视化空间映射、可视化元素映射和可视化属性映射,整 个映射过程围绕静态约束数据信息展开。李晶等<sup>8</sup> 将 多维数据可视化分为数据信息的图形可视化和结构可 视化两部分。夏治坤 <sup>17</sup> 从研究多维数据可视化的直观 性、交互性、智能性等角度提出了符合用户需求的多维 用 HighCharts 工具设计了一种以统计图为主的可视化 查询系统。赖晓文等 <sup>®</sup> 以 SVG 技术为核心, 研究实现 一种具有标准面向对象接口的应用于电力系统方向的 可视化方法库集成平台,提高了电网数据可视化分析 效率。

本文提出了一种通用的数据可视化模型,该模型描述了原始数据经过处理转换为包含信息类别、区间、数值属性及可视化目标的可视化结构,并在适合的空间中选择匹配的布局方法形成可视化对象,通过可视化视图映射形成可视化图形的映射过程;在可视化的整个流程中用户都可以介入来调节模型以完成特定的任务需求。与文献 [3-9] 相比,该数据可视化模型具有如下主要特点:

- (1)将数据的基本类型与当前互联网的鲜明特征融合,可视化模型涵盖的数据更广泛。
- (2) 定义通用数据可视化结构,实现了对数据服务的支持。
- (3)引入可视化布局过程,提出可视化结构单一 布局映射和可视化结构多局映射,丰富了数据可视化 的表征形式。
- (4) 定义良好的可视化结构匹配机制,针对可视 化图形布局转换进行验证。
  - (5) 基于可视化模型实现了图形插件,可应用于

Web 系统开发,具有高度的灵活性和可移植性。

## 1 数据可视化模型定义

数据可视化模型定义为将可视化数据具有的属性特征映射为几何空间中可视化关键要素,完成可视化数据到可视化视图映射的过程。图 1 描述了数据可视化模型映射过程,以数据为中心的可视化映射包括结构映射、空间布局映射、视图映射,生成的数据实例包含可视化结构、可视化对象、可视化视图。以用户为中心的可视化控制包括视图交互、可视化空间布局转换、可视化结构转换、更新数据等操作。可视化数据编码的可视化映射流程概括如下:

- (1) 可视化结构映射: 将可视化数据根据数据特点、用户可视化目标转换为可视化结构的过程。
- (2) 可视化空间布局: 将可视化结构的类型与用户选择的可视化布局结合生成可视化对象的过程。
- (3) 可视化视图映射: 将可视化对象转换为可视 化元素和可视化属性以图形的方式呈现在用户界面的 过程。

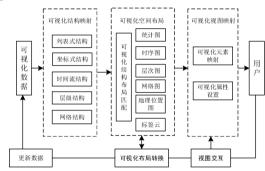


图 1 可视化映射模型

针对可视化映射模型中的过程映射函数流程如图 2 所示,可以分为三类:

- (1) 布局转换函数 FI: 对应可视化结构的布局和空间映射,可视化结构与可视化对象的对应关系可以是一对一、一对多的关系;
- (2) 视图映射函数 Fv: 对应可视化对象的视图 绘制;
- (3) 视图交互函数 Fi: 对应可视化视图的界面控制操作。

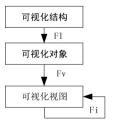


图 2 数据可视化映射流程

用户可以针对不同的任务,选择控制迭代可视化模型中的每一个环节,将新生成的可视化对象进行视图重绘。

## 2 数据可视化模型设计

### 2.1 可视化模型中的数据定义

数据可视化映射模型中数据以通用的数据结构作 用于可视化模型的映射过程。以可视化数据为输入, 根据数据具有的特点和用户可视化目标经过结构映射 生成可视化结构,以可视化结构和用户布局选择为输 入经过空间布局生成可视化对象,以可视化对象为视 图映射的输入经过图形绘制最后生成可视化视图。根 据数据在可视化过程中具有的类别属性、区间属性和 数值属性 [10] 将可视化数据格式定义为  $M = \{ title, de$ scribe, objects, property, relations, relationtype}。 title 为 可视化数据的名称描述; describe 为可视化数据的背景 描述; objects 表示可视化数据中由类别属性组成的集 合,由数据类别名称作为唯一标识; property 为数据的 区间属性组成的集合; relations 表示类别、区间与数值 属性关联关系组成的对象集合; relationtype 表示数据 类别与区间之间的关系类型。基于坐标式的结构中每 组关系可表示为 relation = { object [propertyvalue1, propertyvalue2,…]},由数据类别与信息区间共同确定 数据值属性;对于层次结构的数据结构表示为数据类 别与区间属性嵌套组成的对象集合,每一组关系可表 示为 relation = { name, children}, 其中 children 为当前 数据类别的区间属性组成的集合,当前的数据类别为 父级的区间属性集合; 对于网络结构的关系结构可以 表示为 relation = { source, target}, source 为源数据节 点, target 表示目标节点组成的数据集合。

#### 2.2 数据可视化模型

基于数据可视化流程图,数据可视化模型主要分为三部分:可视化结构映射、可视化空间布局、可视化视图映射。

#### (1) 可视化结构映射

可视化结构映射是将不同来源,不同格式的可视 化数据根据数据的类型、用户的可视化目标转换为可 视化结构的过程。可视化数据与可视化结构为一对一 的映射关系。

将 Shneiderman 提出的数据类型与大数据时代交叉融合,我们所设计的数据可视化结构包含:表征分组及分类关系的列表式结构,表征信息的程度、数值等属性的坐标式结构,其中包括直角坐标系,极坐标系、地

理坐标系等类型,表征数据层次、继承、因果等关系的层次结构,表征数据一对一、一对多、多对多等复杂关系的网状结构,表征数据随时间变换状态的结构时间流结构,表征文本属性值的标签字结构等类型。不同结构通过可视化结构中的 relationtype 表示,如表 1 所示。

表 1 可视化结构类型

relationtype 值	可视化结构类型	定义
1	一维结构	表示单一属性的统计数
		据集
2	二维结构	表示具有二维属性的统计
		数据集
3	多维结构	表示具有三维及以上属性
		的统计数据集
4	时间流结构	表征随时间变化的数据集
5	层次结构	表示具有层次、继承、因果
		等关系的数据集
6	网络结构	表示一对一、一对多,多对
		多等复杂关系的数据集
7	地理位置结构	与地理位置相关的数据集
8	标签字结构	表示文本统计结果的数
		据集

可视化结构映射是可视化模型的基础,只有正确、 合理、有效地表征可视化结构,才会使整个可视化过程 有意义。

#### (2) 可视化空间布局

可视化空间布局是指根据可视化结构的类型与用户的可视化布局选择生成可视化对象的过程。根据可视化结构的数据特点,可视化空间布局可以分为统计图布局、时序图布局、层次图布局、网络图布局、地理位置图布局、标签云布局。其中统计图布局根据数据的维度进一步表示为饼状图、柱形图、折线图、散点图、平行坐标图及雷达图等。

在可视化模型中,由可视化数据唯一确定的可视化结构并不能适应所有的可视化布局,根据可视化结构中的字段 relationtype 生成可视化结构布局匹配模块来验证可视化结构与选择的可视化布局是否匹配。可视化布局映射主要分为两种方式可视化结构单一布局映射和可视化结构多局映射。

定义 1 可视化结构单一布局映射是指可视化结构与可视化对象的一对一映射场景。

定义 2 可视化结构多布局映射是指可视化结构 与可视化布局的一对多的映射场景。

可视化布局与可视化结构对应关系如表 2 所示。 对于坐标式结构的数据根据数据的维度选择适当的空 间,保证可视化结构在空间内部完整、充分、有效地表示出来,属于可视化结构多局映射;层次结构和网络结构的数据可视化空间布局映射为一对一的方式。

表 2 可视化布局与可视化结构关系

可视化结构类型
一维、二维、多维结构、时间流
结构、地理位置结构、标签字
时间流结构
层次数据结构
网络数据结构
标签字结构
地理位置结构

根据数据类型特点和可视化布局的不同,将可视化对象定义为四种不同形式:第一种为统计图和地图位置图布局相关的可视化对象{ title, property, describe, objects, series: { name, data} ] };第二种为层次图布局的可视化对象{ name, children};第三种为网络图布局的可视化对象{ nodes: { id, name, value} ], edges: { source, target, weight} ] };第四种为标签云布局的可视化对象 { name, value} ]。

#### (3) 可视化视图映射

在确定可视化的结构与空间布局后,就会生成唯一的可视化对象进行视图绘制。可视化视图绘制是指将可视化对象转换为可视化元素以图形化的方式呈现在用户界面的过程。可视化视图映射主要包括可视化对象到可视化图形元素的绘制和可视化图形属性设置两个部分,通常这两个部分在可视化视图渲染过程中同步进行的。

可视化图形元素绘制中可视化元素一般要选用二维或者三维的基本几何图形来表示;可视化元素的关联一般选用直线或者弧线连接;针对地理位置、标签字等特殊的可视化元素选择相应地图和文字表现形式。可视化属性选择对于可视化结果表现的准确性、美观性、立体性至关重要。常见的可视化属性设置包括:使用颜色的色调、饱和度、亮度等标识不同类别对象,强化对象之间的区别;利用几何的长度、宽度、深度、大小、角度等特性与信息的数值属性建立对应关系表示信息的定量关系;利用动态的显示、运动、高亮、闪烁等特效表示可视化交互过程,强调信息,丰富用户的视觉感知。用户可以根据特定的任务驱动设置适合的可视化属性调整可视化视图。

可视化视图的映射一般采用基础的可视化工具实现可视化图形绘制和可视化属性的设置。在本文的视图映射部分针对统计数据相关的统计图布局,地理位置图布局使用 Echarts 工具完成,针对层次图、网络图

和标签云使用 D3 函数库完成。并在视图绘制的同时通过参数的形式设置可视化元素的属性。

## 3 信息可视化模型实现

本文以数据可视化模型为核心设计并实现了基于Web 的数据可视化平台。该平台利用 SpringMVC 框架实现可视化数据处理生成可视化结构; 根据 JSON 格式在网络传输与数据交换的优势, 使用 JSON 作为可视化结构实例贯穿可视化映射的始终; 利用 HTML5 和jQuery、Seajs、Echarts 图表工具和 D3 [1] 函数库实现浏览器端对可视化视图的空间布局与视图映射的支持。

基于数据可视化模型,利用 Echarts 和 D3 函数库实现了一种具有集成性、可移植性的可视化图形库插件 DVL. js。 DVL. js 中主要包含布局映射 layer 模块、数据转换 model 模块、可视化匹配 match 模块、图形绘制 draw 模块。 DVL. js 中布局映射模块描述了布局选择过程,根据 match 模块返回的是否匹配结果调用model 模块进行数据转换, model 模块返回相应的可视化对象,最后进行视图的绘制模块。布局映射是视图可视化的关键环节。图 3 描述了数据转换模块中可视化结构转换的过程,根据布局中选择的视图类型进行数据转换;图 4 描述了视图绘制模块的调用过程,该模块中目前实现的图形包括基于统计的柱形图、折线图、饼状图、雷达图、平行坐标图,基于时间关系的静态时序图和动态时序图,基于地图的可视化描点、热力图,动态和静态的标签云,层次可视图,网络力导向图。

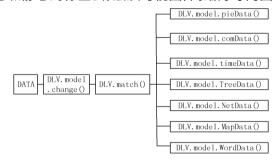


图 3 DLV. js 中 model 模块

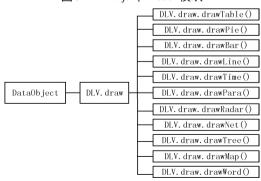


图 4 DLV. js 中 draw 模块

## 4 数据可视化模型应用

数据可视化平台支持多种不同的数据类型,数据来源主要有两种方式:从本系统数据库读取的数据和从外部导入的数据,对于外部使用接口形式导入的数据分为静态和动态两种方式。本文以使用接口导入静态数据为例介绍数据可视化模型映射过程。

使用外部接口 API 形式导入数据及其配置选项后形成的可视化结构如图 5 所示,根据可视化结构类型选择布局方式如图 6 右侧布局栏所示。进行匹配成功后调用 DVL. layer. setLayer()进行布局转换, DVL. layer. setLayer() 函数内部调用 DVL. model. change()生成可视化对象,将可视化对象作为 DVL. draw 模块的输入,调用相应的图形绘制函数生成可视化图形,图 6 描述了使用地理位置图呈现地区收入分布;可视化视图交互中用户可以选择不同的布局方式进行空间布局转换生成新的可视化图形,图 7 描述了使用柱状堆叠图后呈现的地区收入统计情况。

图 5 可视化数据接口



图 6 可视化图形实例 1

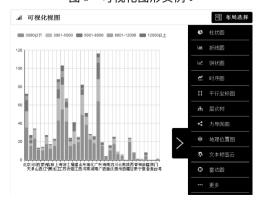


图 7 可视化图形实例 2

## 5 结 语

本文首先给出了一种通用的数据可视化数据结构定义,介绍了数据可视化模型中可视化结构映射、可视化空间布局、可视化视图映射三部分主要内容。并基于 SpringMVC 架构对可视化模型设计和借助对 Echarts 与 D3 工具包进行封装与二次开发的插件形成功能丰富的可视化图形方法,最终生成了通用的面向对象接口的可视化 Web 平台,为数据挖掘人员对挖掘结果的呈现提供了直观、生动的可视化方式,降低了开发成本,实现了数据可视化的快速生成与传播。基于 Echarts 和 D3 开发的可视化插件具有较强的可移植性和可集成性,可高效地应用于不同角色、背景的可视化分析平台。

虽然数据可视化模型提供了通用的可视化方案,涵盖了大部分的信息类型,支持十多种不同的可视化图形,但是如何进一步将不同的图形进行组合从而优化对信息特征和属性的表现,下一步工作会结合实际的应用场景进一步优化对可视化图形的设计。

## 参考文献

- Card S K, Mackinlay J D, Shneiderman B. Readings in information visualization: using vision to think [M]. Morgan Kaufmann, 1999.
- [2] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations [C] //Visual Languages, 1996. Proceedings. IEEE Symposium on. IEEE, 1996: 336-343.
- [3] 周宁,杨峰. 信息可视化系统的 RDV 模型研究 []. 情报学报,2004,23(5):619-624.
- [4] 周宁,陈勇跃,金大卫,等.知识可视化与信息可视化比较研究 ①.情报理论与实践,2007,30(2):178-181.
- [5] 刘晓平, 石慧, 毛峥强. 基于信息可视化的协同感知模型 []. 通信学报, 2006, 27(11): 24-30.
- [6] 李晶, 薛澄岐, 史铭豪, 等. 基于信息多维属性的信息可视 化结构 ①. 东南大学学报(自然科学版), 2013, 42(6): 1094-1099.
- [7] 夏治坤. 多维信息可视化模型 []. 科技情报开发与经济, 2006, 16(10): 170-172.
- [8] 陈先灏, 雒江涛. 面向电商数据的可视化查询系统 []. 计算机系统应用, 2016(4): 123-127.
- [9] 赖晓文,陈启鑫,夏清,等.基于 SVG 技术的电力系统可 视化平台集成与方法库开发 [].电力系统自动化,2012,36(16):76-81.
- [10] Ware C. Information visualization: perception for design [M]. Elsevier, 2012.

(下转第96页)

性,同时使模型具有较好的泛化能力。特别是对于 C4.5 决策树构建的目标函数,基于 SVR 的趋势寻优 算法显著提升了模型性能。

本方法非常适合中小企业在收集数据样本不足的情况下使用,即在有限的样本数量条件下获得较为科学和精准的分类结果。销售人员可以根据分类情况得到有价值的潜在客户预测,从而采用合适的个性化营销方案提升企业的商机转化效率。

## 参考文献

- [1] Ait-Mlouk A, Gharnati F, Agouti T. Multi-agent-based modeling for extracting relevant association rules using a multi-criteria analysis approach [J]. Vietnam Journal of Computer Science, 2016, 3(4):235-245.
- [2] Qiu J, Lin Z, Li Y. Predicting customer purchase behavior in the e-commerce context []]. Electronic Commerce Research, 2015, 15(4): 427-452.
- [3] Khosravifar B, Bentahar J, Gomrokchi M, et al. CRM: An efficient trust and reputation model for agent computing [J]. Knowledge-Based Systems, 2012, 30(2):1-16.
- [4] Tsui P T, Li F C, Pang A H, et al. Using innovative customer relationship management technologies to explore the business opportunities of an ageing population and provide better service [J]. SpringerPlus, 2015, 4(2):1-2.
- [5] Beheshti S, Benatallah B, Motahari-Nezhad H R. Scalable graph-based OLAP analytics over process execution data []. Distributed and Parallel Databases, 2016, 34(3): 379-423.
- [6] Tian X, Liu L. Does big data mean big knowledge? Integration of big data analysis and conceptual model for social commerce research []]. Electronic Commerce Research, 2017, 17 (1):169-183.
- [7] Loukis E, Kyriakou N, Pazalos K, et al. Inter-organizational innovation and cloud computing [J]. Electronic Commerce Research, 2016: 1-23.
- [8] Chan C C H, Hwang Y, Wu H. Marketing segmentation using the particle swarm optimization algorithm: a case study [J]. Journal of Ambient Intelligence and Humanized Computing, 2016,7(6):855-863.
- [9] Liu Q, Huang S, Zhang L. The influence of information cascades on online purchase behaviors of search and experience products [J]. Electronic Commerce Research, 2016, 16(4): 553-580.
- [10] Murthy S K. Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey [J]. Data Mining and Knowledge Discovery, 1998, 2(4): 345-389.
- [1] Bede B. Fuzzy Clustering [M]. Mathematics of Fuzzy Sets and Fuzzy Logic, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013: 213-219.
- [12] Akaike H. A New Look at the Statistical Model Identification

- [M] . Selected Papers of Hirotugu Akaike, Parzen E, Tanabe K, Kitagawa G, New York, NY: Springer New York, 1998: 215-222.
- [13] Caserta M, Lessmann S, Voβ S. A Novel Approach to Construct Discrete Support Vector Machine Classifiers [C] //Advances in Data Analysis, Data Handling and Business Intelligence-Proceedings of the, Conference of the Gesellschaft Für Klassifikation E. v. Joint Conference with the British Classification Society. DBLP, 2009: 115-125.

#### (上接第42页)

- [1] Bostock M, Ogievetsky V, Heer J. D3 data-driven documents
   [1] . IEEE transactions on visualization and computer graphics, 2011, 17 (12): 2301-2309.
- [12] 王恬, 李书琴, 王志伟. 农业信息搜索可视化平台研究 [1]. 计算机应用与软件, 2016, 33(3): 271-274.
- [13] 任磊, 杜一, 马帅, 等. 大数据可视分析综述 [1]. 软件学报, 2014, 25(9): 1909-1936.
- [14] 杜一,郭旦怀,陈昕,等.一种模型驱动的可视化生成系统 [1].软件学报,2016(5):1199-1211.
- [15] 孙扬. 多变元网络数据可视化方法研究 [D]. 国防科学技术大学,2010.
- [16] 杨峰,李蔚.信息可视化技术的分类研究 [J].图书与情报,2007(3):81-85.
- [17] 张建军,刘虎,倪芳英. 基于 SSH 与 Highcharts 整合架构的 Web 应用研究 [1]. 计算机技术与发展,2013,23(9): 245-247.
- [18] 杨莹. 基于简化光线投射 GPU 大规模地形实时可视化 [1]. 计算机应用与软件,2014,31(6):123-126.
- [19] 龚炳江,杨海梅. 敏捷 BPM 的可视化流程模型研究与实现 [1]. 计算机应用与软件,2013,30(10):199-201,254.
- [20] 彭国华,王波兴,黄运保. 辐射流体力学计算结果的可视 化处理 [1]. 计算机应用与软件,2015,32(4):56-58.
- [21] 葛彬,尚建嘎,余芳文,等. 一种融合 WebGIS 和 Web3D 技术的实时定位系统可视化框架 [1]. 计算机应用与软件, 2015,32(1):68-70,117.
- [22] Nazemi K, Burkhardt D, Hoppe D, et al. Web-based Evaluation of Information Visualization [J]. Procedia Manufacturing, 2015, 3: 5527-5534.
- [23] Satyanarayan A, Heer J. Lyra: An interactive visualization design environment [J]. Computer Graphics Forum, 2014, 33 (3):351-360.
- [24] Turkay C, Slingsby A, Hauser H, et al. Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data [J]. IEEE transactions on visualization and computer graphics, 2014, 20(12): 2033-2042.
- [25] Chen S, Yuan X, Wang Z, et al. Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data []]. IEEE transactions on visualization and computer graphics, 2016, 22(1): 270-279.