

面向可视化系统设计与开发的嵌套增量模型

崔迪¹ 胡万祺² 郭小燕³ 陈为²

(宁波工程学院电子与信息工程学院 宁波 315211)¹

(浙江大学 CAD&CG 国家重点实验室 杭州 310058)² (甘肃农业大学信息科学技术学院 兰州 730070)³

摘要 在数据时代,可视化系统开发的需求越来越大,如何制定可视化系统开发的范式已经成为一个热门的研究问题。现从软件工程的角度出发,提出针对可视化系统开发的嵌套增量模型,帮助和指导用户进行可视化系统开发,并结合一个手机基站数据可视化系统作为实例,诠释嵌套增量模型的开发流程,阐述如何针对人群移动模式分析这一具体场景从而做好需求分析、数据处理、可视化设计和增量迭代等工作。

关键词 信息可视化,可视化系统开发,嵌套增量模型,手机基站

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2017.06.045

Nested Incremental Development Model for Design and Implementation of Visualization Systems

CUI Di¹ HU Wan-qi² GUO Xiao-yan³ CHEN Wei²

(School of Electronic and Information Engineering, Ningbo University of Technology, Ningbo 315211, China)¹

(Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, China)²

(School of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China)³

Abstract It remains a challenging problem to design and implement effective visualization systems with a well-studied fashion in the data era. This thesis proposed a nested incremental model (NIM) for visualization developers from the perspective of software engineering. The key idea and flowchart of our nested incremental model is demonstrated with a visualization prototype for visual exploration of mobile base station data. The experiments verifies the effectiveness and efficiency of the proposed NIM.

Keywords Information visualization, Visualization systems development, Nested incremental model, Mobile base station

近年来,数据的爆炸性增长导致可视化系统开发的需求越来越大。这对可视化工作者而言既是机遇也是挑战,如何规范化地进行可视化系统的设计和开发,提高开发效率,是可视化学者所关心的问题。在信息可视化领域,许多学者致力于可视化标准设计流程研究,提出相关可视化设计框架。Haber 和 MacNabb 于 1990 年提出了一个科学可视化的流水线模型^[1],用一条串行路线描述了从数据空间到可视空间映射的 4 个阶段:数据分析、数据过滤、可视映射和可视化绘制。Card, Mackinlay 和 Shneiderman 于 1999 年对上述可视化流水线模型做了改进,将用户的反馈加入到可视化流程的任何阶段,并将流水线改成了回路,即著名的最优信息可视化流程模型^[2]。Tamara Munzner 提出了一个四层嵌套模型^[3-4],分别是领域问题层(定义问题和任务)、数据/任务抽象层(确定任务需要的数据以及对数据进行的操作)、可视化编码/交互层以及算法实现层。C. Stoite 等人提出了可视化循环模型^[5],欧洲学者 Daniel Keim 等人提出了可视分析学标准流程^[6]。但无论在任何一个可视分析框架中,人都是核心要

素^[7],通过设计回路模型将用户对可视设计的评判与要求向前反馈,从而指导与优化可视系统的设计,通过调参等方法将专家知识引入可视系统的设计中。本文采用需求分析、数据处理、可视映射与交互、可视渲染 4 个层次作为基本框架进行可视的研究与开发,在整个设计框架中,人的主观意识与专家经验参与可视化设计的过程。

可视系统设计过程的本质是一个软件开发过程。软件开发流程中常用的模型包括瀑布模型、快速原型模型、螺旋模型、增量模型等^[8]。瀑布模型由 Winston Royce 于 1970 年提出^[9],其核心思想是按设计工序将问题化简,各项活动按照顺序逐项解决,当前活动接受上一项活动的工作结果。快速原型模型需要迅速建立一个可反映用户主要需求的软件原型,用于理解和澄清问题,使开发人员与用户达成共识,最终在确定的客户需求基础上开发客户满意的软件产品,当用户认为这个软件原型符合需求时,开发人员再进入线性顺序的开发过程。螺旋模型是使用原型及其他方法来尽量降低软件开发项目的风险的。增量模型是将软件拆分成一系列的增量构件

到稿日期:2017-01-18 返修日期:2017-03-12 本文受国家自然科学基金(61422211)资助。

崔迪(1985-),女,博士生,讲师,CCF 会员,主要研究方向为大数据分析、智能信息处理等,E-mail:sukey1006@126.com;胡万祺(1992-),男,硕士生,主要研究方向为数据可视化;郭小燕(1976-),女,博士,副教授,主要研究方向为智能优化算法,E-mail:guoxy@gsau.edu.cn(通信作者);陈为(1976-),男,博士,教授,主要研究方向为可视化,E-mail:chenwei@cad.zju.edu.cn。

来完成分析、设计、编码和测试的,每个增量并行地开发,增量模型可使软件开发较好地适应用户需求的变化,从而降低开发风险,因此越来越受到开发者的青睐。为适应用户需求的动态变化,增强可视系统的并行性与可扩展性,本文提出了一种针对可视化系统开发的模型——嵌套增量模型,将完整的可视化系统开发流程分解成多个低耦合、高内聚的增量构件,每个构件均包括需求分析、数据处理、可视映射和交互、可视化渲染 4 个完整的可视化流程。

1 嵌套增量模型

1.1 可视化系统的特点

对于可视化系统优劣的评价,在很大程度上依赖于人的主观意识,因此会导致用户需求并不具备一般软件系统明确性,常常会因人而异,需求的频繁变化会给可视化系统造成沉重的设计负担与风险。在可视化系统开发中,其核心是可视化设计,它具有复杂的设计流程,每一个流程的设计环节都会影响最后可视化系统的效果。因此对可视化系统进行合理的解构可以提高系统的扩展性和灵活性,有效地引导开发者进行可视化设计,快速地实现可视化原型,降低由于需求不稳定而造成的不必要的设计成本与风险。

1.2 嵌套增量模型基本框架

嵌套增量模型的本质是将一个大问题分解为若干相对独立的小问题,每一个小问题都可看作可视系统的“增量”,每一个增量都包含一个完整的可视系统设计的线性序列。将系统的核心功能定义为第一个增量(核心增量),实现用户的主体需求,在用户的试用和评估后,将用户的建议和尚未实现的需求共同作为下一个增量,如此循环,直至最终满足用户需求,并产生完整的可视化系统。基于嵌套增量模型设计框架进行可视化设计的过程是在核心增量解决基本问题的基础上,增加新的增量对系统功能进行不断增强和补充的过程。增量模型的设计流程如图 1 所示,其基本框架如图 2 所示。

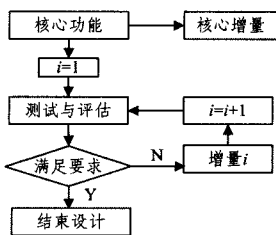


图 1 增量嵌套模型的设计流程

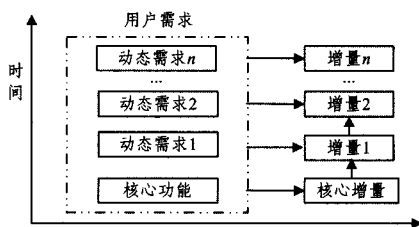


图 2 增量嵌套模型的基本框架

1.3 增量构件的组成

每个增量构件中都包含规范的可视化设计和实现的 4 层嵌套流程,从外到内分别为需求分析、数据处理、可视化映射与交互以及可视化渲染;用户的反馈和测试在嵌套结构之外,

但能作用于任何一层,并影响该层内的所有层,增量模型的构成如图 3 所示。

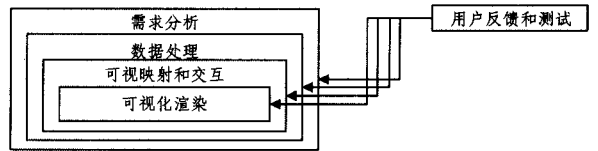


图 3 增量构件的组成

1.3.1 需求分析

增量构件最顶层是需求分析,需求理解的全面、正确与否将直接影响最后的可视化效果,需求分析主要从用户领域、认识原始数据、展示需求、归纳整理 4 个方面展开。

(1) 用户领域

可视化系统要做的不仅仅是数据结果的简单呈现,而是要辅助领域专业人员对领域问题进行理解与分析。了解目标用户的基本业务、兴趣点能更好地进行可视化系统功能的设计。

(2) 原始数据

原始数据与用户希望的可视化的数据往往存在一定的差距,因此全面了解原始数据的格式和每个数据对象具体的含义可更好地理解用户意图,全面体现用户需求。开发者对于原始数据应关注的细节有:数据是样本数据还是真实数据;实际数据量的大小,即开发的可视化系统所要支持处理的数据量级;数据的获取方式是在可视化系统本地存储,还是通过网络调用合作方的 api 获取。另外需充分认识原始数据的性质,如果其与网络、图等相关,则常常使用图论的算法进行数据处理;如果其与地理位置相关,则会偏向于基于 GIS 系统进行数据处理。

(3) 展示需求

可视化设计是一个相对主观的流程,目标用户可能会对最终的展示效果提出不同的需求。开发者须用专业的可视化知识去说服和引导目标用户,尽量去除不合理需求并修改不完善的需求,避免影响最终实现的系统功能和效果。这要求开发者对可视化系统展示业务非常熟悉,并具有较好的沟通能力,同时能正确理解合作方对于可视化展示的最终期望。

(4) 归纳整理

将用户想法与需求具体为可视化展示和交互功能,形成需求文档和系统原型,并将其在开发者和目标用户多次讨论与迭代修订后定稿,用于指导后续的可视化系统设计。

1.3.2 数据处理

原始数据需经过预处理得到清洁、简化、结构清晰的数据,建立可视化图表所需要的数据结构,并将其输入到可视化映射模块中的过程称为数据处理。数据处理一般分为数据清洗、数据入库和数据变换。

(1) 数据清洗

原始数据一般都含有噪声、不完整信息或错误信息。从应用角度来讲,未经清洗的数据可能会导致分析结果错误或使数据分析方法变得混乱。因此,在对数据进行可视化之前,要先做数据清洗,以提高数据质量,提升可视化效果。通常采用的方法有:用常量填充缺失值、删除或修改错误值、删除冗余值等。现阶段较好的数据清洗工具提供强大的数据诊断功

能和交互界面,让用户快速且有效地完成数据清洗的工作,如Wrangler^[10]和OpenRefine^[11]。

(2) 数据入库

完成数据清洗后进行数据存储,以便后续可视化系统开发过程中的查询和使用。合适的存储方式能大幅提高数据访问效率,可视化系统开发中涉及到的数据存储组织方式主要包括文件存储和数据库存储两种。文件方式的优点是灵活性强,开发者可以按照任意格式对所存储的数据进行格式组织,有利于开发者从存储底层开始对存储过程进行调整和优化。但这种方式会造成访问过程繁琐、数据约束添加难、数据存储高级功能设置难度大,为安全控制和管理带来不便。数据库存储方式相对于文件存储是一种更加高效和安全的存储管理方式,其提供了丰富的数据查询和分析功能。可视化常用的数据库有MySQL, MongoDB等。

(3) 数据变换

原始数据经过清洗、入库后,根据需求做进一步的处理和变换,以得到简化、结构清晰的数据,并输出到可视化映射模块的过程称为数据变换。常用的数据变换类型包括数据滤波、数据降维、数据采样和聚类等。原始数据中的噪声可能会遮盖数据本身的特征,对用户形成误导,通过数据滤波可对原始数据去噪,常用的滤波器包括低通滤波、高通滤波和频带滤波等。人类所能感知的空间仅限于三维,高于三维的数据超出了可视化可以显示的维度。因此在对高维数据进行可视化之前,一般要进行降维处理,常用的降维方法有线性降维和非线性降维。多维尺度分析(MultiDimensional Scaling, MDS)、主成分分析(PCA)和非负矩阵分解(NMF)属于线性降维,ISOMAP和局部线性嵌套(LLE)属于非线性降维。高维、大尺度和多变量数据导致可视化的信息超载,通常用采样和聚类两种方法解决信息超载的问题。采样通过抽取部分数据代替整体数据进行分析以减少查询数据时间,使数据分析者可以快速高效地获取数据报告,常用的采样方法有随机取样、技术取样、加权取样等。将类似的采样点放在同一类别并在可视化系统中显示的方法称为聚类,通过聚类可以简化数据,在可视化系统中隐藏具体的采样点,只展示数据中最重要的结构,从而去除视觉遮盖。通常在社交网络中需对不同人群进行聚类,在生物领域中需对基因序列进行聚类,在多媒体领域中需对文本、图像和视频的进行聚类。常用的聚类方法有K-means^[12]、层次聚类^[13]、基于密度的聚类^[14]等。

1.3.3 可视映射与交互

可视映射和交互是可视化设计的核心内容。可视设计的目标是让用户能够在最短的时间内获取数据的整体信息和部分细节信息,甚至能够预测用户在使用可视化系统时的行为和期望。可视化映射的直观性决定了可视化结果被用户接受的难易程度。

(1) 可视映射

在设计可视化映射时,必须精心选择图元和视觉通道,确保用户很容易地理解可视化所要展示的内容。图元通常是用一些点、线、面、体等来映射数据的几何图形。一维的图元是点;二维的图元有曲线和平面,如矩形、圆形等;三维的图元有三维的面和体,如立方体、球体等。图元的设计是可视设计的一个重要工作,既要考虑图元反映数据的能力,也要考虑用户

理解图元的能力和效率。图元设计过于简单则表达能力不足,过于复杂则用户理解起来困难,因此需掌握一种平衡。视觉通道用来控制图元的视觉特征,通常可用的视觉通道包括图元的位置、大小、形状、方向、填充颜色以及边框颜色等。不同用户的感知能力虽有差别,但也有规律可循。当使用不同的可视化元素或者视觉通道时,用户对数据感知的准确性不同,著名可视化学者 Mackinlay 总结了不同数据类型的可视化视觉通道选择的优先级^[15]。如图4所示,图元的位置是最能准确反映各种类型数据的视觉通道;颜色对数值型数据的表达不佳,但适合显示类别型数据;长度、角度等对数值型数据有很好的效果,但不适合编码有序型和类别型数据。

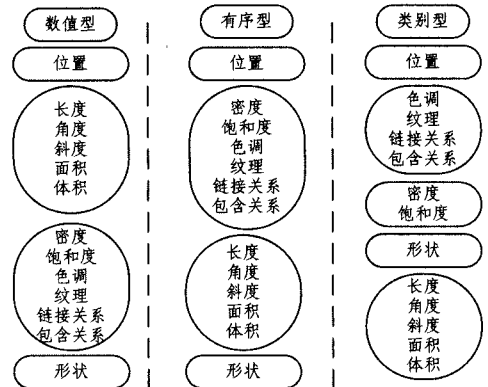


图4 基本数据类型使用的可视化编码方式,优先级自上而下

(2) 可视交互

静态的可视化图表能将简单数据中的信息传达给用户,但对于复杂的数据,需要用户通过交互才能获取更多的信息。可视化交互是指将用户探索数据的意图传达到可视化系统中,以改变可视化图表的显示。对于可视化系统,可视化交互的完善设计影响着系统的功能性和实用性。可视化交互方法多样,如用鼠标点击、键盘输入、触摸屏等实现选择、过滤、变换视图等可视化任务。面向不同的应用场景,开发者可使用不同的交互模型,以下3个交互模型在设计可视化交互时经常用到。

1) 概览+细节模型:由于人在同一时刻只能关注有限的数量,当数据量太大时较难绘制出太多细节,因此调整所绘制数据的细节粒度来满足大量数据可视化的要求。概览+细节模型先给用户展示数据集的概览,定位关注数据的入口,用户通过过滤或者视图变换将所关注的数据的细节进行可视化。目前概览+细节模型的典型应用有平面地图可视化,例如谷歌地图、百度地图、高德地图等。

2) 聚焦+上下文模型:针对在同一视图中无法显示所有数据这一问题提出的交互模型。一般用一个视图绘制当前聚焦数据的细节内容,另一个视图绘制上下文,适度展示当前聚焦数据以外的其他数据。

3) 多视图联动模型:该模型可对同一数据同时采用多种形式的可视化视图表示,并允许用户在所有视图内进行交互,任一视图中的交互结果都更新到其他视图。多视图给用户提供了探索数据的多个入口,帮助用户更好地定位感兴趣的数据信息。

这3种模型各有其应用场合,可视设计中常常结合使用

多种模型,本文进行手机基站数据可视系统时采用 3 种模型相结合的方式

1.3.4 可视化渲染

可视化渲染是可视化嵌套模型的最后一层,也是可视化的最终呈现。SVG 和 Canvas 是两种基于 Web 的重要绘制方法,它们各自有不同的应用场景。

SVG(可缩放矢量图形)[16]采用 xml 文本来描述矢量化的图形,浏览器可直接渲染。SVG 图形绘制后具有文档对象模型(Document Object Model, Dom)实体,因此可完全支持 Dom 的所有接口规范,包括事件驱动、css 样式指定等,具备较强的动态交互性。常用的 SVG 绘图库有 d3.js, Raphael.js 等。

Canvas[17]是 html5 新标准里面的新元素,最早由 Apple 引入 WebKit,用于 Mac OS X 的 Dashboard,后又在 Safari 和 Google Chrome 实现,主要功能是设计客户端矢量图形。Canvas 在浏览器中可渲染画布实体,把一个绘图 API 展现给客户端脚本,从而将绘制目标展现在画布上。Canvas 图形的事件响应需用 javascript 等脚本语言实现。现阶段很多 Canvas 库已实现事件响应,可将高层的 API 提供给用户直接使用,如 fabric.js 和 zRender 等。Canvas 通过 WebGL 和 GLSL 的辅助实现三维空间绘制,目前最常用的 WebGL 可视化库有 Three.js,deck.gl 等,Canvas 往往更适用于大规模的可视化绘制[18]。

2 手机基站数据可视化系统开发

本文选择手机基站数据作为可视化系统的嵌套增量开发模型的研究案例,从 2014 年 1 月和 2 月的 700 万部手机的 140 亿条记录和 25000 个基站塔有所关联的数据中抽取了 2014 年 1 月 21 日至 2014 年 1 月 27 日一周的数据,并将研究范围限制在温州市鹿城区和龙湾区之间。原始数据集以文本格式存储,每行记录一个用户在所有基站出现的数据,以“时间基站 ID”的形式保存,描述该用户在一周内的移动轨迹。设计可视化系统描述不同用户的移动轨迹,该系统将聚类转移轨迹设计为核心模块,通过与用户的沟通分别设计地图视图、特征值视图、查询视图 3 个增量模块,如图 5 所示。

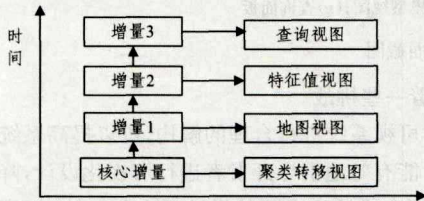


图 5 手机基站数据可视化嵌套增量模型

2.1 核心增量构件:聚类转移模型视图

该系统通过对手机基站数据的可视分析来研究人群移动轨迹,因此该系统的核心主视图应展示不同人群移动轨迹关系,对应可视系统的核心增量构件。

2.1.1 轨迹状态划分

基于手机基站数据进行人群轨迹分析时首先将人群轨迹划分为两种状态:静止状态与移动状态。其状态的界定借鉴了 Zheng 停留地点的研究[19],将在足够长的时间内活动范围

小于某个定值的一段轨迹划分为静止状态,反之划分为移动状态。对分割好的轨迹片段进行聚类,以得到不同移动模式的人群,本文采用 K-Means 算法进行聚类,设计聚类特征向量为 $P(x_1, x_2, \dots, x_8)$,其中 x_1, x_2, \dots, x_8 为特征值,分别代表比例熵、时间相关比例熵、矩心位置、半径、家庭位置、活动半径、平均速度、轨迹长度,其含义描述如表 1 所列。

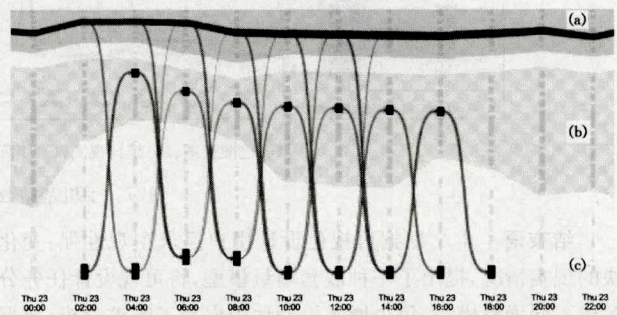
表 1 聚类特征向量

Table with 2 columns: 特征值 (Feature Value) and 描述 (Description). It lists 8 features: 比例熵, 时间相关比例熵, 矩心位置, 半径, 家庭位置, 活动半径, 平均速度, and 轨迹长度, each with a brief explanation of what it represents in the context of movement data clustering.

对这些特征值进行归一化后,采用 K-Means 方法对轨迹片段进行聚类。为了达到最好的可视化效果,不断进行调参,并得到最终的聚类为 20。

2.1.2 转移模型

对手机基站数据进行聚类之后,继续提取人群的移动模式随着时间的变化而转移和演化的数据。一个用户在不同时间段内的轨迹片段可能会被聚到不同的类中。因此,从聚类层面上看,不同时刻不同聚类间会发生人数转移。采用每隔 2 小时进行一次采样的方式,获取当前时刻的各个轨迹片段所属的聚类,计算出前一时刻和后一时刻的聚类间的转移概率。本文对于一周数据共获得 84 个采样时间点,可得到一个 20 * 84 的转移矩阵。聚类转移模型成为最终直接输入到可视化映射模块的数据对象,本文采用平行坐标轴的自定义视图来展现聚类转移模型,如图 6 所示。



注:(a),(b),(c)区域内的运动轨迹片段占比依次递减。

图 6 聚类转移模型可视化

2.1.3 可视图元

该视图选取矩形和有向线段作为主要图元,每个时刻的一个聚类采用一个矩形表示,矩形的长度表示该时刻该聚类中所含有的轨迹片段的个数;矩形的颜色编码该聚类中运动的轨迹所占的比例,颜色越深表示该时刻下该聚类中运动的片段越多。20 个聚类在同一时刻上堆叠放置,堆叠顺序按照运动的轨迹所占有的比例进行排列,根据运动片段所占比例将显示视图划分为 3 个部分,如图 6 所示,(a)区域代表运动

片段所占比例大于 0.9; (b) 区域代表运动片段所占比例大于 0.1 小于 0.9; (c) 区域代表运动片段所占比例小于 0.1。这 3 个区域按照颜色梯度编码由深至浅。节点和节点间用有向路径表示前后两个时刻聚类间的转移概率, 路径的粗细编码了转移概率的大小。

2.1.4 交互设计

由于每个时刻下聚类的排序不确定, 本文通过交互设计获取某个聚类在不同时刻下的演变信息。具体为: 用户双击任一时刻下的某聚类, 视图中会出现一条灰色的流图, 将每个时刻的同一聚类连接, 并高亮该聚类的所有节点以及所有相关转移链接; 通过时间刷在聚类转移图中选取一段时间, 选取成功后其他视图相应更新; 双滑块空间让用户选取转移概率区间, 对图中的转移边进行过滤。

2.2 其他增量构件

通过聚类转移可视化使用户可以观察某个聚类的演化, 比较相同时刻不同聚类的轨迹数量, 但由于它是一个高度抽象的模型, 因此会丢失很多信息。在可视系统设计的过程中, 为了使系统能够提供更多功能、体现更多内涵, 必须根据用户

的动态需求构建其他增量构件。

增量 1: 地图视图

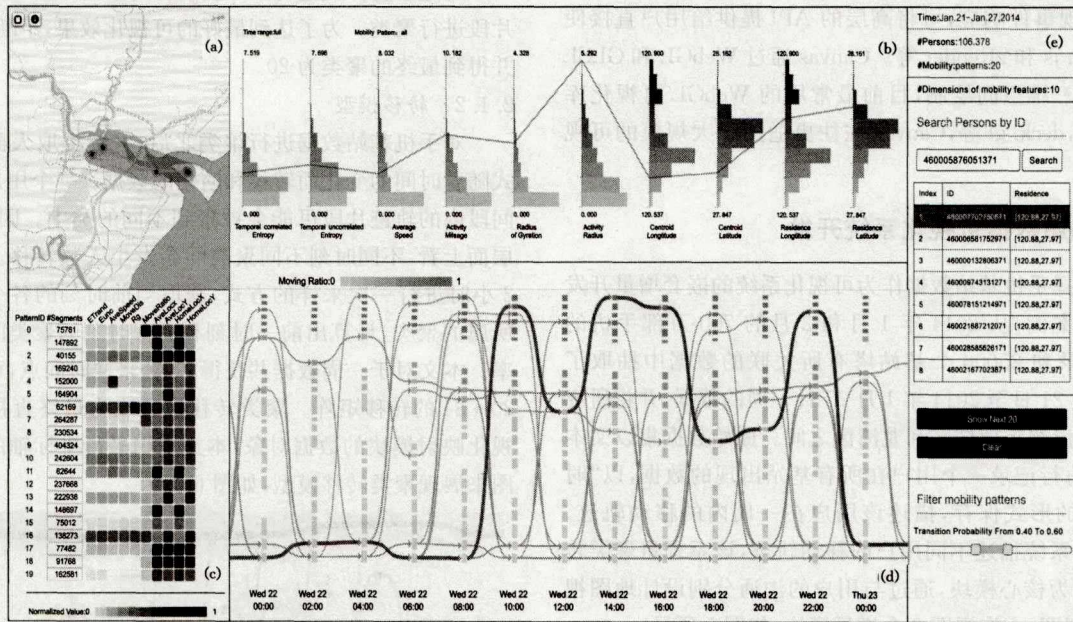
地图视图与聚类转移视图相对应, 用来将聚类中一些代表性的轨迹对应在实际地理空间中, 在实际地图中形成相应的移动路线, 从而提供更加直观具体的视觉信息, 使用户能更清晰地了解某个聚类的转移路线, 如图 7(a) 所示。地图视图主要使用开源项目 OpenStreetMap^[20] 来实现。

增量 2: 特征值视图

对于某些聚类, 用户除关心其移动轨迹外, 常常对聚类包含的轨迹数量以及聚类特征值感兴趣, 因此在统计视图中通过直方图对每个聚类中的特征值进行可视化, 如图 7(b) 所示。在矩阵视图中用一个小矩形代表某个聚类的各个特征值, 颜色越深代表其值越大, 如图 7(c) 所示。

增量 3: 查询视图

将单个人的轨迹数据在系统中进行可视化, 有助于使用目标用户作为特例来验证已发现的人群移动模式, 在查询面板中输入用户 ID 时, 可显示其在不同聚类中的移动轨迹, 如图 7(e) 所示。



(a) 地图视图; (b) 统计视图; (c) 矩阵视图; (d) 聚类转移模型视图; (e) 查询面板

图 7 手机基站数据可视化系统界面截图

结束语 本文根据可视化设计用户需求主观性强、变化快的现实情况, 提出了一种嵌套增量模型, 将可视设计任务分解成不同增量模块, 每个增量模型块中嵌套了需求分析、数据处理、映射和交互 4 个可视化步骤。用户的核心需求对应于可视系统的核心增量构件, 根据用户的动态需求不断增加新的增量构件, 从而逐渐完善系统功能, 满足用户需求, 避免传统设计模型由于用户需求的不稳定而造成不必要的开发成本。最后以手机基站数据可视化系统作为开发实例验证了嵌套增量模型的实用性。该系统将聚类转移轨迹设计为核心增量构件, 通过与用户的沟通分别设计地图视图、特征值视图、查询视图 3 个增量构件, 从而保证系统的完备性。与传统的设计方案相比, 使用嵌套增量模型开发的可视系统的可维护性有大幅提升, 但嵌套增量模型在可视系统的设计开发过程

中也面临着一些挑战。

(1) 对可视系统进行合理的解构, 可以提高系统的扩展性和灵活性, 能有效地引导开发者进行可视化设计, 降低由于需求不稳定而造成的不必要的设计成本与风险, 但模块的划分不当将很大程度地影响系统的稳定性与效率, 增量模块划分常用的方法有: 横向拆分功能模块、纵向拆分架构模块、通过界面划分组件模块、通过数据划分加工组件等, 通常根据具体的问题灵活采用一种或几种方法进行问题的解构, 指导增量模块的划分。

(2) 嵌套增量模型属于非整体开发模型, 以渐增方式开发系统, 如果增量构件之间的耦合情况尚未处理好, 则会使可视系统设计过程失去整体性。

因此在利用嵌套增量模型进行可视设计的过程中, 对于

一个复杂的可视化系统选用适当的方法进行增量模块的划分,同时进行增量模块之间无缝耦合从而提高系统整体的内聚性为下一步的研究方向。

参 考 文 献

- [1] HABER R B, MCNABB D A. Visualization idioms: A conceptual model for scientific visualization systems[J]. *Visualization in Scientific Computing*, 1990, 74: 93.
- [2] CARD S K, MACKINLAY J D, SHNEIDERMAN B. Readings in information visualization: using vision to think[M]. Morgan Kaufmann, 1999.
- [3] MUNZNER T. Visualization Analysis and Design[M]. CRC Press, 2014.
- [4] MUNZNER T. A nested model for visualization design and validation[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2009, 15(6): 921-928.
- [5] STOLTLE C, HANRAHAN P. Polaris: A system for query, analysis and visualization of multi-dimensional relational databases[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2002, 8(1): 52-56.
- [6] KEIM D, ANDRIENKO G, FEKETE J D, et al. Visual Analytics: Definition, Process, and Challenges[J]. *Lecture Notes in Computer Science*, 2008, 4950: 154-175.
- [7] 陈为, 沈则潜, 陶煜波. 数据可视化[M]. 北京: 电子工业出版社, 2014: 6.
- [8] 海藩. 软件工程导论[M]. 北京: 清华大学出版社, 2003.
- [9] ROYCE W W. Managing the development of large software systems[J]. *Proceedings of IEEE WESCON*, 1970, 26(8): 328-338.
- [10] KANDEL S, PAEPCKE A, HELLERSTEIN J, et al. Wrangler: Interactive visual specification of data transformation scripts[C]// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011: 3363-3372.
- [11] MORRIS T, GUIDRY T, MAGDINIE M. OpenRefine: A free, open source, powerful tool for working with messy data[R/OL]. Technical report, The OpenRefine Development Team, 2015. <http://openrefine.org>, 2015.
- [12] MACQUEEN J. Some methods for classification and analysis of multivariate observations[J]. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, 1(14): 281-297.
- [13] SIBSON R. SLINK: an optimally efficient algorithm for the single-link cluster method[J]. *The Computer Journal*, 1973, 16(1): 30-34.
- [14] ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[J]. *Kdd*, 1996, 96(34): 226-231.
- [15] MACKINLAY J. Automating the design of graphical presentations of relational information[J]. *ACM Transactions on Graphics (Tog)*, 1986, 5(2): 110-141.
- [16] FERRAILOLO J, JUN F, JACKSON D. Scalable vector graphics (SVG) 1.0 specification[M]. Iuniverse, 2000.
- [17] FULTON S, FULTON J. HTML5 canvas[M]. O'Reilly Media, Inc. 2013.
- [18] JOHNSON D W, JANKUN-KELLY T J. A scalability study of web-native information visualization[C]// *Proceedings of Graphics Interface 2008*. Canadian Information Processing Society, 2008: 163-168.
- [19] ZHENG Y, CAPRA L, WOLFSON O, et al. Urban computing: concepts, methodologies, and applications[J]. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2014, 5(3): 38.
- [20] HAKLAY M, WEBER P. Openstreetmap: User-generated street maps[J]. *IEEE Pervasive Computing*, 2008, 7(4): 12-18.
- [21] GEEM Z W, TSENG C L. New Methodology, Harmony Search and Its Robustness[C]// *Late-Breaking Papers of Genetic and Evolutionary Computation Conference (GECCO-2002)*. New York, USA, 2002: 174-178.
- [22] KANG S L, GEEM Z W. A new structural optimization Method based on harmony search algorithm[J]. *Computes and Structures*, 2004, 82(9/10): 781-798.
- [23] LEE K S, GEEM Z W. A new meta-heuristic algorithm For continuous engineering optimization: harmony Search theory and practice[J]. *Computer Methods in Applied Mechanics and Engineering*, 2005, 194(36-38): 3902-3933.
- [24] GEEM Z W, LEE K S, PARK Y. Application of harmony search to vehicle routing[J]. *American Journal of Applied Sciences*, 2005, 2(12): 1552-1557.
- [25] KOUNTOURIOTIS P A, OBRADOVIC D, MANDIC D P, et al. Multi-step forecasting using echo state networks[C]// *International Conference on Computer As a Tool*. 2005: 1574-1577.
- [26] BUEHNER M, YOUNG P. A Tighter Bound for the Echo State Property[J]. *IEEE Transactions on Neural Networks*, 2006, 17(3): 820-824.
- [27] SKOWRONSKI M D, HARRI J G. Automatic speech recognition using a predictive echo state network classifier[J]. *Neural Networks*, 2007, 20(3): 414-423.
- [28] LIN X, YANG Z, SONG Y. Short-term stock price prediction based on echo state networks[J]. *Expert Systems with Applications*, 2009, 36(3): 7313-7317.
- [29] HOLZMANN G, HAUSER H. Echo State networks with filter neurons and a delay & sun readout[J]. *Neural Networks*, 2010, 23(2): 244-256.
- [30] OZTURK M C, XU D M, Principle J C. Analysis and design of echo state networks[J]. *Neural Computation*, 2007, 19(1): 111-138.
- [31] TIAN Z D, GAO X W, LI S J, et al. Prediction Method for Network Traffic Based on Genetic Algorithm Optimized Echo State Network[J]. *Journal of Computer Research and Development*, 2015, 52(5): 1137-1145. (in Chinese)
- [32] 田中大, 高宪文, 李树江, 等. 遗传算法优化回声状态网络的网络流量预测[J]. *计算机研究与发展*, 2015, 52(5): 1137-1145.

(上接第 231 页)