

数据可视化技术应用趋势与分类研究

张 浩, 郭 灿

(扬州大学 新闻与传媒学院, 江苏 扬州 225002)

摘 要:数据可视化技术应用需求的增长和用户的不断扩大推动了新的可视化技术应用工具和平台的发展。在追溯数据可视化的概念缘起与发展历程基础上,着重分析了数据可视化技术应用的发展趋势,并对目前出现的若干种数据可视化技术应用的新技术工具和服务平台加以功能分类和比较,归纳出主要的功能类型,分别是呈现多维内容信息的可视化新应用、揭示数据关联趋势的可视化新服务、有助于发散形象思维的可视化新工具和提供社区交流空间的可视化新平台。

关键词:数据可视化;可视化技术;可视化应用类型;可视化工具;可视化平台

中图分类号:TP393

文献标识码:A

文章编号:1672-7800(2012)005-0169-04

1 数据可视化的发展历程

数据可视化技术的起源可追溯到 20 世纪 50 年代计算机图形学的时代。当时人们利用计算机创建了首批图

形图表。1987 年,由布鲁斯·麦考梅克等人编写的美国国家科学基金会报告《科学计算之中的可视化》对于这一领域具有重大的奠基意义。此后科学可视化的概念出现,再之后,信息可视化也被提出。信息可视化旨在为许多应用领域之中对于抽象的异质性数据集的分析工作提供支

号找寻最优的阈值,在改进含参数的阈值函数中获取最优参数去噪等基于小波去噪方法具有广阔的研究前景。

参考文献:

- [1] 杨建国. 小波分析及其工程应用[M]. 北京:机械工业出版社, 2005.
- [2] DONOHO D L. De-noising by soft-thresholding[J]. IEEE Trans Inform, Theory, 1995(3).
- [3] 张文娟,周丹丹,王林. 小波阈值去噪的一种改进方案[J]. 电脑开发与应用, 2007(9).

- [4] 王知强. 一种基于新阈值函数的小波图像去噪算法[J]. 哈尔滨理工大学学报, 2011(4).
- [5] GRABISCH M, NICOLAS J M. Classification by fuzzy integral: performance and tests[J]. Fuzzy Sets and Systems, 1994(2/3).
- [6] DONOHO D L, JOHNSTONE I M. Ideal spatial adaptation via wavelet shrinkage[J]. Biometrika, 1994, 81.
- [7] 董长虹. Matlab 小波分析工具箱原理与应用[M]. 北京:国防工业出版社, 2004.
- [8] 林杰,付梦印,李道平. 自适应小波阈值去噪算法及在图像处理中的应用[J]. 兵工学报, 2011(7).

(责任编辑:杜能钢)

Image De-noising Based on Different Wavelet Thresholding Functions

Abstract: Threshold selection is the key to wavelet threshold de-noise. On the basis of analysis of the problem of the traditional threshold function, this paper focuses on the comparison of hard thresholding, soft thresholding, multilevel threshold function and a continuous soft threshold de-noising, and the experimental results analysis based on objective indicators of PSNR.

Key Words: Wavelet; Image De-noising; Multilevel Threshold; Continuous Soft Threshold

基金项目:全国教育科学“十一五”规划 2010 年度教育部青年专项课题(ECA100368)

作者简介:张浩(1977—),男,江苏扬州人,博士,扬州大学新闻与传媒学院讲师,研究方向为网络与新媒体技术应用;郭灿(1988—),女,江苏徐州人,扬州大学新闻与传媒学院硕士研究生,研究方向为教学设计、网络教育应用、教育信息化。

持。而数据可视化的概念的外延则被认为同时涵盖了科学可视化与信息可视化两大领域。数据可视化这一术语的出现统一了成熟的科学可视化领域与较年轻的信息可视化领域,与信息图形、信息可视化、科学可视化以及统计图形密切相关。当前,在研究、教学和开发领域,数据可视化是一个极为活跃而又关键的方面。

1.1 数据可视化的概念与标准

随着数据可视化技术应用的不断发展,目前对数据可视化的理解可以描述为是指通过特定软件工具以图表、地图、标签云、动画或任何使内容更容易理解的图形方式来呈现数据。当然数据可视化不仅能够帮助数据的最终呈现,对发现数据中新的信息也起到非常关键的作用。数据可视化的实质是借助图形化手段,清晰有效地传达与沟通信息,使通过数据表达的内容更容易被理解。

为了实现信息有效地传达,数据可视化需要追求兼顾美学形式与功能的需要,并通过直观地传达关键特征,实现对数据集的深入洞察。华而不实的数据可视化形式无法达到数据的有意义呈现。茱莉亚和诺哈认为可视化效果的优质体现需要关注 4 个要素,即形式新颖、内容充实、直观高效和符合审美。

随着数据可视化新工具的出现,使可视化技术可以实现用户与可视化数据之间的交互。新的可视化技术应用的标准还应该向以下 4 个方面努力:①直观化:直观、形象地呈现数据;②关联化:挖掘、突出呈现数据之间的关联;③艺术化:增强数据呈现的艺术效果,符合审美规则;④交互性:实现用户与数据的交互,增强用户对数据的控制。其中,直观化和关联化强调数据可视化的功能性,艺术化和交互性则强调数据可视化的美学设计和智能化。实现了直观化和关联化,就能够直观、形象地呈现数据,找出数据之间关键、独特的关联性,而智能、交互性强的动态数据可视化过程可实现信息的传达和双向沟通。

1.2 数据可视化技术的应用趋势

在以往的数据可视化中,想要充分利用数据的复杂性来提供有意义的呈现数据的方案,须对统计学、数据挖掘、图形化设计以及信息可视化多个学科领域有足够的了解。本弗雷提出必须把这些领域统一成一个简单的过程,他将可视化数据的流程归纳为以下 7 步骤:获取、分析、过滤、挖掘、表述、修饰和交互。研究者如果对统计学、数据挖掘、图形化设计等多个学科领域的知识了解不够,开展如此 7 个步骤就会困难重重。因此,鉴于数据可视化如此高端的技术要求,其用户群仅局限于数据可视化领域的专家和技术人员。然而,在这个信息时代,需要进行可视化数据应用的普通人群迅速扩大,促进了新的、更加智能的数据可视化工具的出现,为更大范围的人们提供了接触可视化思想的机会。

不同的用户需求必然导致可视化工具的发展,从以往单一发展技术专业的(依赖于特定领域的)、过程分散复杂的可视化工具逐步开始派生发展到跨专业的(不依赖于特定领域的)、功能集成的“傻瓜式”可视化工具。新的数据

可视化工具提供了一种快捷方式,从编制数据到阐明数据一条龙实现。最简单的可视化工具仿佛一个(无需知道内部工作原理的)黑箱。此类应用程序提供了一个用户界面,用户可以粘贴数据或者链接到数据源、电子表格、RSS,然后程序将数据转化成图形,该图形可能是地图动画、互动图表或者关键字云图。当然,操作技能更高一筹的用户可以使用多种技术工具来创建构想的可视化类型;然而重要的是,较新的技术工具允许非技术用户在不必学习如何使用复杂的模型和多媒体软件的情况下,能够对任一学科或多个学科进行数据的可视化呈现。近年来新出现的数据可视化服务和工具为大范围的人们提供了更多体验数据可视化的机会。

2 数据可视化技术应用分类

数据可视化从信息服务角度来理解,它所提供的实质是一种界面,无论数据可视化内部是什么样的结构,对于使用者而言,或者对于信息提供者而言,需要的是以简洁易懂、省时高效的方式呈现数据内容,理解数据含义。无论是哪种数据可视化技术工具,其最终落脚点都应是满足某种用户的需求,而对于数据可视化技术工具的提供者,他们的最终目标是如何高效地提供数据可视化服务并开发相应的可视化工具或平台。以下将从功能应用的维度对数字可视化技术应用中出现的新工具和新平台加以分类,并分别详细阐述其特点、应用案例和可视化过程。

2.1 呈现多维混搭信息的可视化应用

这类可视化工具主要是社交网络与地图工具的混搭应用,因此可称为混搭应用式的数据可视化工具。混搭应用式的可视化工具在使用社交网络的人群中很受欢迎。比如在 Facebook 中,用户可以创建“朋友地图(Friend maps)”,用数字化的球棒模型来显示朋友网络;Flickr 网站的地图功能让摄影师更加容易显示拍摄照片所在位置;世界各地的 Twitter 用户使用 Twitvision 发送的微博会根据其所在位置呈现在一张世界地图上,他们也可以使用如 Stweet 一样的工具,利用 Twitter 和谷歌街景(Google Street)的混搭功能,显示微博发言来自的那个街道的照片。在此类数据可视化应用中,用户获取的视觉信息不是单一维度而是多维的。例如 Twitvision 用户可以直观地看到某条微博的发布者名称及其地理位置等多个信息。除了与社交网络的应用混搭,谷歌地图和谷歌地球已经进入了教室以及其它将数据与地理信息关联的地方,如加拿大英属哥伦比亚大学的学生们创建了一个谷歌地球可视化应用,用于温哥华社区“食品安全”评估项目,以帮助人们找到当地健康食物源。

可视化工具下的数据可视化过程如图 1 所示。此类数据可视化应用只需将用户正常使用网页时发布、更新的多方面的数据,如照片、状态、地理信息等信息,进行后台相应的可视化处理,就可为用户提供一个可视化的显示,直观和关联地呈现多维的信息。在此过程中,用户可能甚

至都不知道数据可视化的存在,而扮演着更新数据的角色。用户无意识地参与了可视化数据的过程,享受可视化“服务方”提供的可视化结果,是数据可视化结果的见证者、数据可视化的体验者。

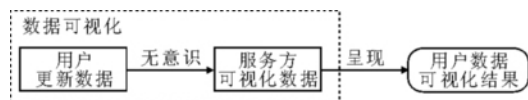


图1 呈现多维信息的数据可视化过程

2.2 揭示数据关联趋势的可视化新服务

这类数据可视化结果产生的可视化信息表达重心是挖掘数据之间的重要关联或者是呈现一个有理有据的数据发展趋势。这种情况通常是数据可视化服务方对某一数据集进行复杂、专业的可视化加工,最终为用户呈现一个动态的、图形化的、令人印象深刻的可视化结果。这类服务中最具代表性的是 Google 的一款实验室网络产品——谷歌趋势(Google Trends),通过它用户可以获知全世界对其所喜爱主题的关注度,查看某主题随着时间推移在 Google 上被搜索的频率。谷歌趋势通过分析 Google 网络搜索,计算用户输入的字词被搜索的次数,并将其与 Google 上随时间推移的搜索总量相比较,然后用图表向用户显示结果,即按线性比例绘制搜索量图表,并提供新闻引用量图表,显示用户的主题在 Google 新闻报道中的出现次数。如果发现某个特定字词的新闻报道引用量出现峰值,谷歌趋势会在图表中进行标记,然后自动选择编写时间接近该峰值的一篇 Google 新闻报道,并显示其标题。在搜索量和新闻量图表之外,谷歌趋势还显示了用户输入的首字词的最热门城市、地区和语言等信息。此外,值得一提的是 Google 用以描述日本 2011 大地震制作的可视化数据动画,成功跨越了各种语言交流上的障碍,形象、准确、高效地向全世界人民传达了地震的震区覆盖面、震幅大小等随时间变化的整体趋势信息。

国内的人人网不久前推出的尚处于开发阶段的好友档案服务,基于用户好友信息提供了在线分析服务,分析统计并图形化地呈现当前用户好友的男女比例、大学分布、地区分布地图、姓氏统计、对自己的关注度等。再如,中国天气网的天气预报,不仅将天气信息可视化地直接标识在地图上的相应地区,而且记录了地区最近整点的气温、相对湿度、降水量、风向风力等信息,并以动态图表的形式呈现给用户一个天气变化的走向。

此类数据可视化服务使得用户主动为所查询、检索的数据或信息选择一种或多种可视化的呈现方式成为可能。用户对其所检索数据存在有意或无意的期待偏向。有些用户乐于发现数据之间的关联、期待了解数据的发展趋势;而有些用户厌倦了阅读大篇幅的文字描述却难于发现所述重点的呈现方式,期待更加直观、高效的数据呈现方式。服务方针对用户的期待,对用户检索的数据、信息进行相应的可视化处理——挖掘数据之间的重要关联或者是呈现一个有理有据的数据发展趋势,并进行图形化处理,最终为用户呈现一个动态的、图形化的、高效的、令人

印象深刻的可视化结果。不同于混搭应用式的数据可视化服务,此类数据可视化服务提供了更多的图形化手段——图表、图片、动画等。此类数据可视化服务的数据可视化过程,如图2所示,用户不仅是数据可视化结果的体验者,而且可体会到数据可视化的存在和意义,并一定程度地参与到数据可视化的过程中。

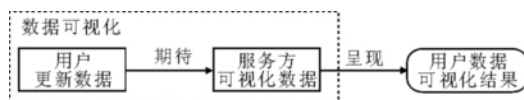


图2 呈现数据关联以及趋势的数据可视化过程

2.3 有助于发散形象思维的可视化新工具

概念图或思维导图(Mind Map)工具是这个类型中的代表。思维导图是一种图像式思维的工具,是利用图像式思考辅助工具来表达思维的工具,它帮助人们发散且不失关联地认识问题、挖掘信息。在西方学术界,由 Tufts 大学设计的概念图工具“视觉化理解环境”(VUE: Visual Understanding Environment)是一种比较流行的创造知识地图的工具,哥伦比亚大学的彼得教授鼓励他任教的跨学科课程“地球/人类系统”中的学生使用 VUE 来创造知识地图,将可持续发展的复杂问题与其各自所在的生物、物理和社会科学领域知识加以联系。其它比较流行的思维导图工具还有 FreeMind、Mindmanager、iMindMate 等,此类工具的简单易用使其用户群远不止于学术界。思维导图可以应用在家庭、教育和商业领域,包括记笔记、头脑风暴、总结、修正、理清想法。思维导图工具可以有效地组织大量资讯、与空间组织相结合、动态的层级组合,此外,还可以将提取的资讯节点互相结合。

使用此类数据可视化工具,用户可以积极、自主地为其所分析、组织的数据或信息设计一种可视化的呈现。与前述的两种数据可视化工具比较,此类工具的数据可视化过程如图3所示,用户对数据可视化有一定的了解且明确数据可视化的存在和意义,主动、积极地设计、参与可视化数据,在可视化数据的过程中加深自己对数据的理解深度、获取新知,并且及时将可视化过程中的反思反馈到原始数据的分析和组织过程中去。

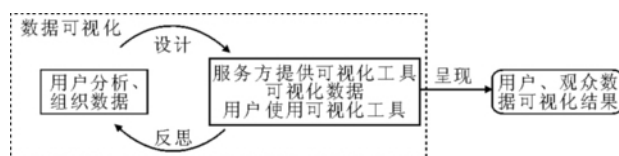


图3 辅助图像式思维的数据可视化过程

2.4 提供数据交流社区的可视化新平台

数据可视化平台通常提供了一套可视化工具集成,以供用户创造数据可视化实例,这套工具集提供了图像、图表、地图、动画等多种图形化手段以供选择,同时这样的平台还提供了让人们分享数据可视化应用的公共论坛。知名的数据可视化平台有 IBM 的 Many Eyes 网站以及纽约时报的数据可视化实验室(Data Visualization Lab)。Many Eyes 网站是数据可视化实例的集合,普通用户可以

浏览并参与讨论众多可视化实例和数据集,为网站现有的数据集创建可视化实例;注册用户还拥有更新数据集和可视化实例、上传用户的自定义数据、在论坛中心发帖或跟帖、选择观看其它可视化项目、跟踪记录自己的贡献和观看历史、记录自己在论坛中心的活动、查看他人写给你的评论的权限。《纽约时报》的数据可视化实验室是 IBM 的 Many Eyes 专门为纽约时报定制的数据可视化平台。纽约时报认为用户在诠释数据和处理信息的过程中能够带来深刻的见解和独特的洞察力,并且想法设法去挖掘新颖的呈现数据的方式。《纽约时报》的编辑们提供数据,读者为这些数据创建交互式表格、图表、地图以及其它形式的图形化呈现。时报网站的读者可以互相评论可视化实例、以图片的形式与他人共享这些可视化数据,还可以创建主题中心以便用户收集特定主题的可视化实例并且讨论特定题材。Google 也一直在努力创建 Google 用户的社区。遗憾的是,谷歌电子表格和谷歌动态图表在国内很难访问。

使用数据可视化平台时,用户需要循环、交替思考和解决如何对这些数据进行设计、统计分析和组织,以及如何选择和使用恰当的可视化工具及可视化数据。如图 4 所示,在使用数据可视化平台的过程中,了解数据可视化的人群不只限于数据可视化的用户,还有相当数量的关注数据可视化结果的社区观众。数据可视化平台提供了数据可视化应用的公共论坛,论坛中的部分用户既是数据可视化的设计者,同时又是可视化结果的观众。用户和观众在欣赏数据可视化结果的同时,将论坛中心的评价反馈到数据可视化的设计过程中去,以便进行优化。

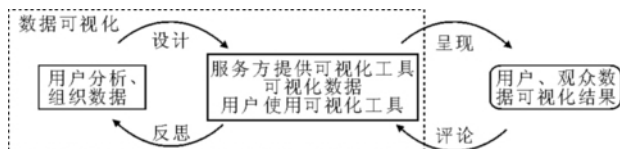


图 4 数据可视化平台下的数据可视化过程

2.5 四类数据可视化技术应用比较

以上分 4 类介绍了一些数据可视化的技术应用:呈现多维混搭信息的可视化应用、揭示数据关联趋势的可视化新服务、有助发散形象思维的可视化新工具以及提供数据可视化交流社区的可视化新平台,这些数据可视化的技术应用在用户参与程度、运用的图形化手段以及可视化过程上都有各自的特征。它们之间的比较见表 1 所示。

表 1 数据可视化新工具的比较

类型	维 度		
	用户参与程度	主要图形化手段	可视化过程
呈现多维混搭信息的可视化应用	体验	地图	直观化、关联化
揭示数据关联趋势的可视化新服务	参与、体验	图表、图像、地图、动画	直观化、关联化、艺术化、交互
有助发散形象思维的可视化新工具	创建、参与、体验	图表、图像	直观化、关联化、艺术化
提供数据可视化交流社区的可视化新平台	创建、参与、体验、评价	图表、图像、地图、标签云、动画	直观化、关联化、艺术化、交互

3 结束语

数据可视化技术已越来越普及,用户对单机版和基于网络的各种数据可视化技术应用工具的熟悉度在不断增长,对于 Web 数据获取的行为也愈加频繁。随着物联网的发展,全球用户对传感数据的访问量将进一步高速增长。关注数据挖掘和分析工具的发展具有十分重要的现实意义。随着用户对数据分析需求的增长,对数据可视化的要求也变得越来越强烈,他们会期望通过可视化映射来有效地呈现所想要获得的数据,更便捷地理解其中的含义进而做出决策,也希望体验和参与更加有趣的可视化呈现,这对于数据可视化服务方无疑也是个挑战。在如今不断增加的多媒体演示方法和日益增多的数据数量和类型相结合的环境下,普通用户也不断地被要求或主动参与到设计和创建可视化项目的过程中。对于当下数据可视化技术应用的梳理及分类将有助于更清晰地了解这个领域的现状及发展趋势并将其更合理地应用到教育、学习、生活等各个方面,笔者也将持续关注这一领域的发展并期待更多智能数据可视化平台的出现。

参考文献:

- [1] EDUCAUSE. 7 Things you should know about data visualization II[EB/OL]. <http://net.educause.edu/ir/library/pdf/ELI7052.pdf>.
- [2] FRITS H. POST, GREGORY M. Nielson and georges-pierre bonneau. date visualization; the state of the art[M]. Boston/Dordrecht/London: Kluwer Academic Publishers, 2003.
- [3] VITALY FRIEDMAN. Data visualization and infographics[J]. Graphics, Monday Inspiration, 2008(10).
- [4] JULIE STEELE, NOAH LLUUNSKY. Beautiful visualization[M]. Sebastopol: O'Reilly Media, Inc, 2010.
- [5] BEN FRY. Beautiful visualization[M]. Sebastopol: O'Reilly Media, Inc, 2007.
- [6] BC Food security map[EB/OL]. <http://blogs.landfood.ubc.ca/duncan/2008/08/01/bc-food-security-map/>.
- [7] 谷歌趋势[EB/OL]. <http://www.google.com/trends>.
- [8] 日本地震 Google 动画制作[EB/OL]. http://v.youku.com/v_show/id_XMjUxNjE1OTY4.html.
- [9] 中国天气网[EB/OL]. <http://www.weather.com.cn/static/html/weather.shtml>.
- [10] DANIEL BEEBY. Using visual understanding environment (VUE) as a concept mapping tool[EB/OL]. http://ccnmtl.columbia.edu/enhanced/solutions/using_visual_understanding_env.html.
- [11] Quick start; tour[EB/OL]. <http://www-958.ibm.com/software/data/cognos/manyeyes/page/Tour.html>.
- [12] The New York Times Data visualization Lab[EB/OL]. <http://open.blogs.nytimes.com/2008/10/27/the-new-york-times-data-visualization-lab/>.

(责任编辑:杜能钢)