

# 信息可视化和可视分析: 挑战与机遇

## ——北戴河信息可视化战略研讨会总结报告

戴国忠<sup>①</sup>, 陈为<sup>②</sup>, 洪文学<sup>③</sup>, 刘世霞<sup>④</sup>, 屈华民<sup>⑤</sup>, 袁晓如<sup>⑥</sup>, 张加万<sup>⑦</sup>, 张康<sup>⑦\*</sup>

① 中国科学院软件研究所, 北京 100190

② 浙江大学计算机辅助设计与图形学国家重点实验室, 杭州 310058

③ 燕山大学电气工程学院, 秦皇岛 066004

④ 微软亚洲研究院, 北京 100080

⑤ 香港科技大学计算机科学与工程系, 香港

⑥ 北京大学信息科学技术学院, 北京 100871

⑦ 天津大学软件学院, 天津 300072

\* 通信作者. E-mail: kzhang@tju.edu.cn

收稿日期: 2012-07-24; 接受日期: 2012-12-23

**摘要** 由天津大学软件学院发起并召集, 东北大学秦皇岛分校和燕山大学承办的信息可视化战略研讨会于2012年6月9~10日在北戴河成功举行. 本文总结了此次会议的讨论内容和达成的共识, 对当前信息可视化和可视分析领域的挑战和国际上这方面的研究方向做了大致的分析, 并对今后信息可视化和可视分析的大方向提出一些建议, 以及对具体行动提出一系列倡议.

**关键词** 信息可视化 可视分析 数据可视化 人机交互 大数据

### 1 引言

微软研究院科学家、图灵奖得主 Jim Gray 在 2007 年提出了“以数据为基础的科学研究第四范式”的概念, 研究方法已经从“我应该设计个什么样的实验来验证这个假设?” 逐渐发展为“从这些已知的数据中我能够看到什么相关性?” 这是因为, 在过去几年里, 数据量以年增长率 30% 的速度增长, 而数据的种类也在不断增多, 其中包括文本数据、图像数据、传感器数据和音频及视频数据. 数据的特征变得更加多样化, 包括混合性、空间性、时序性和二义乃至冲突性.

信息可视化 (Information Visualization, 简称 InfoVis) 在国内外都仍处于初级阶段, 在国内研究人群属极少数, 其成功应用更是屈指可数. 正是在这种大形式下, 我们认为有必要举办一次战略性讨论会. 由天津大学软件学院发起并召集, 东北大学秦皇岛分校和燕山大学承办的信息可视化战略研讨会于 2012 年 6 月 9~10 日在北戴河成功举行. 本研讨会旨在增进国内信息可视化同行间的相互交流与合作, 整合资源, 实现共赢. 对于一些信息可视化领域内的重大挑战和应用前景达成一定程度的共识. 参会人多以邀请为主, 召集和主办人邀请了国内在信息可视化、可视分析和人机交互领域里著名专家

**引用格式:** 戴国忠, 陈为, 洪文学, 等. 信息可视化和可视分析: 挑战与机遇——北戴河信息可视化战略研讨会总结报告. 中国科学: 信息科学, 2013, 43: 178-184, doi: 10.1360/112012-488

和最活跃的科研人员. 邀请的准则之一是他们在相关领域的国际一流期刊上发表过具有影响的文章. 组委会预先拟定了 4 个讨论主题:

- 1) 信息可视化的最新前沿和挑战;
- 2) 信息可视化在相关领域中的定位;
- 3) 适用于国内的重大攻关项目及应用方向;
- 4) 建立固定的交流平台, 如会议、刊物、网上论坛等.

会议的形式为围绕上述 4 个主题作 10~15 分钟自由发言和讨论, 以战略和大方向为主线. 与普通会议的最大区别是不讨论技术细节而是以战略眼光看待信息可视化领域的主要问题和应用前景. 会议的大部分时间用于围绕 4 个主题的热议讨论, 以及分析信息可视化及其相关领域里的热点. 总结了可视化与用户之间的关系、历史演变、作用与趋势.

## 2 何为信息可视化? 其最新前沿和挑战

本节首先简单介绍什么是“信息可视化”、它的定义和历史. 这里需要强调的是, 信息可视化指的是“大规模非数值型”与“抽象”数据. Jock Mackinlay 曾于 2000 年定义: “Information visualization (信息可视化) focuses on abstract, nonphysical data such as text, hierarchies, and statistical data.”<sup>[1]</sup> 信息可视化是一个跨学科领域, 旨在研究大规模非数值型信息资源的视觉呈现. 与科学可视化 (scientific visualization) 相比, 信息可视化侧重于抽象数据集, 如非结构化文本或高维空间中的点 (这些点可能并不具有固有的 2 维或 3 维几何结构). 信息可视化的演化过程是从树和图到文本, 再到多媒体, 以便最大限度地利用人们的多通道和分布式认知功能以及形象思维功能, 达到意会 (sense-making). 信息可视化为人们提供了从阅读局部信息到纵观全局信息、从表面到本质和从内容到结构的有力工具. 我们认为现今信息可视化的 3 大发展方向分别是跨媒体、大数据时代与社交网络.

目前在信息界所面临的巨大挑战是信息爆炸, 即所谓的“大数据”(Big Data). 如前所述, 信息可视化是建立在数据基础上的. 要想从海量数据中获取有用的知识, 信息可视化还必须借助于近 10 年取得长足进步的机器学习和数据挖掘 (包括文本挖掘和图挖掘) 的方法, 以及自然语言处理技术.

同时, 用户界面对于可视化技术的成功应用起着关键作用. 更具体地说, 在设计信息可视化系统时, 我们应该考虑人机界面所要提供的如下功能:

- 概观: 对完整的数据集给出一个全貌,
- 放大: 把相关兴趣点放大,
- 过滤: 过滤掉不需要的信息,
- 需求驱动: 只对用户选择想看的内容展示细节,
- 相关: 展示数据项之间的相关性,
- 历史: 保留交互动作的历史记录, 以便支持复原、重放、和渐进细化,
- 提取: 给定查询参数, 提取部分所需信息,
- 以上提到的相关领域之间的关系如图 1 所示.

## 3 信息可视化在相关领域中的定位

目前应该首先分析清楚信息可视化是技术还是学科? 是否可以提出一个可视化模型? 信息可视化

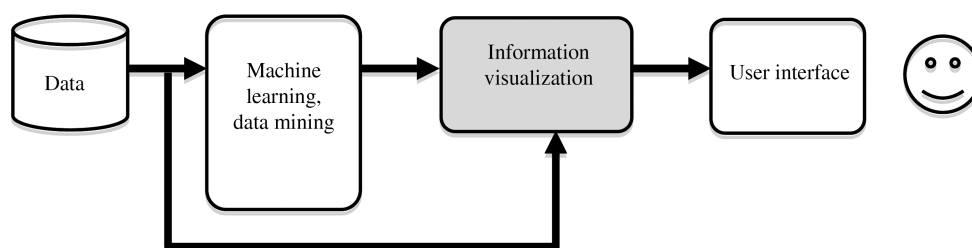


图 1 信息可视化的简单概念图

Figure 1 Conceptual illustration of information visualization

的根在哪里? 在世界主流人机交互会议 CHI 上, 2011 年收录了 14 篇信息可视化论文, 2012 年收录了 8 篇. 这说明信息可视化与人机交互是密切相关的. 这些收录的论文大致概括了如下内容:

- 交互/界面: 用新颖的交互技术来进行可视化数据的操作,
- 用户行为: 通过用户实验来探索可视分析中用户的行为, 从而指导可视化系统的实现,
- 行业可视化: 将信息可视化技术应用在特定领域, 如汽车、帐单、文献、医院、病例、土地凭证、大屏幕等,
- 社交网络: 把信息可视化技术和机器学习有机地结合起来, 将网络上的大量数据以用户可理解的方式呈现出来.

我们还分析了世界主流信息可视化会议的投稿情况, 其中包括: IEEE Visualization (Vis), IEEE Information Visualization (InfoVis) 和 IEEE Visual Analytics Science and Technology (VAST). 2012 年, Vis 收到 152 篇投稿, InfoVis 178 篇, VAST 104 篇. 从中我们可以看到, 信息可视化在可视化这个大领域内已成为最重要的发展方向. 根据过去 4 年信息可视化会议 (即 InfoVis'2008~2011) 的分会 (session) 情况, 我们可以分析哪些信息可视化方向正受人们关注. 总结这 4 年 InfoVis 会议为各分会分配的个数, 我们可以看到下面这些分会经常出现:

- Graph,
- Text,
- Multi-dimensional,
- Evaluation,
- Geo, Map, Space & Time,
- System,
- Applications.

这说明, 图 (Graph) 的可视化在在信息可视化中占有极为重要的地位, 而文本 (Text)、多维 (Multi-dimensional)、评估 (Evaluation) 和地理、地图、时空方面的可视化也是信息可视化的重要研究方向. 与信息可视化密切相关的是近年来发展起来的可视分析 (Visual Analytics), VAST 会议于 2007 年首次出现并迅速得到关注. 过去 4 年可视分析会议 (VAST'2008~2011) 的分会中出现一些新的词语:

- Collaboration,
- Analytic process,
- Computational,
- Sensemaking.

由此可见, 协同 (Collaboration) 和分析过程 (Analytics) 是可视分析中的重要方向; 而计算 (Computational) 和意会 (Sense-making) 也成为不可忽略的可视研究方向.

可视分析概念近几年才被提出, 有很多工作可以开展, 此领域的研究现状还需要梳理. 多数专家认为, 可视分析是信息可视化的自然延伸, 而可视分析更接近应用和实用情形. 一般认为可视分析是数据挖掘、计算机图形学与人机交互 3 个方面的综合. 但信息可视化和可视分析到底应该采取那个框架或模型, 还需要进一步去探索和理解.

会上有与会者提出对信息可视化的必需性和重要性如何理解? 信息可视化不一定是必需的, 但是可以极大地提高效率, 有些领域的信息可视化是必不可少的, 比如股票和地理信息系统 (GIS) 都已被广泛应用并成为不可缺少的工具. 这些成功应用的可视化方法有一个共同点, 即都非常简单易懂. 这就给我们提出了一个需要思考的问题: 目前什么样的领域需要更为复杂的可视化方法来解释和分析, 而这些方法又可以被大众所接受? 这样的领域是否存在? 如存在则仍需要进一步挖掘.

会上还讨论了信息可视化与知识工程之间的关系. 这里的关键是, 必须深刻理解信息转化成知识的必要条件. 有与会者提出一个尖锐的问题: 是否存在不能被可视化的知识? 知识表示本身就很难, 但可视化在知识表达上一定能发挥重要作用. 找到既直观又可以准确表达知识的方法 (可能与信息可视化不同) 是知识可视化的主要挑战. 知识可视化应该是信息可视化的一种升华, 其前景被普遍看好.

#### 4 国内相关研究进展

近几年来, 我国高校和研究院所在可视化的研究方面取得了令人瞩目的进步. 包括浙江大学、北京大学、清华大学、中科院软件所、中科院计算所、中科院超算中心、北京应用物理与计算数学研究所、天津大学、山东大学等单位都开展了卓有成效的可视化研究. 北京大学提出了一系列高维数据可视化方法, 在面向地震、交通时空的可视分析方法与系统方面做了较有影响的工作. 浙江大学的研究人员在高维时变数据的语义理解可视化、医学影像数据可视化等方面做出了创新工作. 中科院超级计算中心、北京应用物理与计算数学研究所的工作集中在天文、生物、气候、地学和国防安全等领域的大规模科学计算数据的可视化与分析, 在 GPU 机群上实现了宇宙结构形成的 TB 级数据的可视化.

我国在可视化方向的研究工作在国际学术界的显程度日益提升. 例如, 在 2011 年 IEEE 可视化年会, 中国内地发表了 2 篇科学可视化论文, 香港科技大学和微软亚洲研究院联合发表了一篇信息可视化论文. 北京大学提出了一种体可视化系统, 它支持“所见即所得体可视化 (WYSIWYG: What you see is what you get, volume visualization)”的新颖可视化交互手段<sup>[2]</sup>. 类似于 2D 图像的绘画操作, 这个系统开发了一套完整工具. 通过在要绘制图像的上部刷几笔线条, 允许对直接绘制的颜色、透明度、对比度、亮度和其他可见属性操作. 浙江大学的“面向色盲的有效体渲染 (an efficient volume rendering approach for dischromats)”<sup>[3]</sup>, 在可视化中针对色盲用户提出相应的颜色方案. 其核心思路是根据色盲用户颜色感知特征建立颜色传输函数的优化问题并求解, 然后引入色盲用户友好的颜色混合方法, 使得色盲用户在通过交互的直接体绘制系统浏览体数据时仍然能够获得较清晰的分类信息. 香港科技大学和微软亚洲研究院提出了 TextFlow<sup>[4]</sup>, 一个可视化和文本挖掘技术的无缝集成. 通过交互、主题挖掘模型和可视化提取 3 层特征, 如主题演化趋势、关键事件和关键词的关系, 帮助用户精炼分析结果和不断地洞察数据.

在华跨国公司研究机构和国内的民族企业都加速在可视化方向研究的布局. 微软亚洲研究院、IBM 中国研究院等跨国公司在华研究机构都对可视化, 特别是信息可视化和可视分析方向给予了高度重视, 在文本可视化等方向开展了具有国际影响的工作. 淘宝网成立了数据可视化部门, 并与浙江大学 CAD&CG 国家重点实验室合作开发了面向在线商业数据的可视化组件库 (<http://datavlab.org/>).

总体而言,我国在可视化方向的研究工作在各方面开展较为均衡呈现发展加速的趋势.在新兴的可视分析方向,各地也开展了富有成效的研究,取得了令人鼓舞的成果.更多的机构和单位也在开展相关研究,一方面尖端研究已具有国际影响,另一方面研究基础也日益扩大.

## 5 如何推动我国的信息可视化与可视分析研究

我们认为,信息可视化应该服务于大众,研究人员需要进一步分析重大国家需求及其与个体需求之间的矛盾和统一,以便做到信息可视化研究既能为国家和大众需求做贡献,又可以提高自身的学术水准.一方面,政治经济数据的可视化和可视分析面向政府重大需求,可为政府决策提供便利的工具.而另一方面,信息可视化的运用也可以促使政府和社会的运行更加透明化<sup>1)</sup>.

信息可视化在文化教育资源中的应用有很多成功案例,如儿童读物中所用的各种图画和图形表达已被充分证明对儿童的学习和理解起着不可低估的作用.所以可视化在儿童教育软件中的应用应该是我们关注的重点之一.推而广之,信息可视化在文化教育中的应用前景很好,但是评价体系不好,影响力不是论文可以评价的.文章阅读信息可视化是一个可以考虑的研究方向.

管理科学中的诸多领域,如市场营销、品牌管理、金融、人力资源、物流等,对于信息可视化的需求尤为突出.这些领域的许多概念极为适合用简单直观的图形来表示.直观且易用的可视化和可视分析系统将会为企业高管提供便捷高效的管理工具.

与会者还分析了信息可视化在智能交通、医学以及移动互联网等方面的应用现状.敦煌壁画保护的信息可视化也是一个有实际意义的研究,应对了国家对物质与非物质文化遗产保护的需求.可视化技术要在这些领域内成功应用,不同专业背景的研究人员的通力合作将是必不可少的.同时信息可视化的研究人员应该加强对于一些高层次且具有一定普遍意义的具体指导原则、领域模型和框架进行系统的研究.纵观国内现状,目前有很多领域用到信息可视化且涉及国家的重大需求,迫切需要信息可视化领域的研究人员参与.我们建议对这些应用领域进行统一规划,研制尽可能的通用平台为各行各业服务.

## 6 未来方向,倡议,并建立固定的交流平台

目前中国的信息可视化落后于国际先进水平,这与中国软件技术的普遍落后是密切相关的,还有大量工作要做.同时信息可视化更强调交叉学科.当数据量无限上升时可视化必然存在作用.需要逐步探索信息可视化发展方向,最好是从具体向普适的发展方向.

我们认为,在未来数年间与信息可视化发展精密相关的3大趋势为:新媒体技术(包括跨媒体)、大数据时代(Big Data)和社交网络.在此趋势下,下列研究方向将会大有发展前景:

- 类似于软件工程的设计和导原则,即信息可视化系统的 Life Cycle;
- 诸如 Windows、Android 和 iPhone 的集成开发环境;
- 以可视分析为主导作用的重大发现;
- 满足政府需求的各种应用,如维稳、灾害防治;
- 在商务智能(business intelligence)中的应用.

1) T. Munzner, <http://www.cs.ubc.ca/~tmm/>

与会者一致同意, 信息可视化今后需要在大数据框架下发展, 而且应该逐步探索出可视分析的理论框架体系. 我们需要推动并逐步构建信息可视化领域的学科文化, 通过展示信息可视化的重要性及其应用的广泛性, 并得到计算机科学领域内以及社会的认可. 我们对国内从事信息可视化的研究与应用人员提出如下倡议:

- 从文章/算法到建立可以真实、广泛使用的系统;
- 关注国家重大需求, 认识国家需求和个体需求的矛盾和统一性;
- 关注中国社会, 让信息可视化为社会做出应有的贡献;
- 推广信息可视化课程设置与教材建设、开展数据和信息可视化竞赛;
- 将信息可视化与互联网文化 (社区、微博、博客) 紧密结合, 以普及其知识和重要性.

我们还要进一步增强战略性部署, 如关键产品研发、以及整合资源. 最后, 我们与读者共享现阶段国内外有关信息可视化和可视分析的交流平台和出版物, 以便更多的科研人员得以了解并参与.

期刊: IEEE Transactions on Visualization and Computer Graphics

会议:

- IEEE Information Visualization (InfoVis);
- IEEE Visual Analytics Science and Technology (VAST);
- Pacific Visualization (PacificVis);
- European Visualization (EuroVis);
- Visual Information Communication and Interaction (VINCI)

国内的近期活动:

- Special Focus on Engineering Visual Analytics 将在 Science China Information Sciences 2013 年第 5 期出版;
- 大数据时代的虚拟现实与图形学研讨会, 浙江大学, 2012 年 5 月;
- 图形学、数据分析与可视化暑假班, 浙江大学, 2012 年 7 月 2~6 日;
- PKU Summer School, 北京大学, 2012 年 8 月 13~24 日;
- 中国计算机学会, 可视分析前沿讲习班, 北京, 2012 年 8 月 17~19 日;
- 数据可视化与可视分析研讨会 (<http://www.cad.zju.edu.cn/home/vagblog>), 浙江大学, 2012 年 9 月 25 日;
- VINCI'2012 (<http://www.cad.zju.edu.cn/home/chenwei/VINCI2012>), 浙江大学, 2012 年 9 月 27~28 日;
- VINCI'2013 (<http://se.tju.edu.cn/vinci2013/vinci/Welcome.html>), 天津大学, 2013 年 8 月 17~18 日.

**致谢** 我们对天津大学软件学院为此次研讨会提供的全额资助、燕山大学和东北大学秦皇岛分校的会务工作表示由衷的感谢, 感谢所有与会者的积极参与和发言, 感谢天津大学软件学院的李春和李杰的会议笔录.

## 参考文献

- 1 Mackinlay J D. Opportunities for information visualization. IEEE Comput Graph Appl, 2000, 20: 22-23

- 2 Guo H Q, Mao N Y, Yuan X R. WYSIWYG (What You See is What You Get) Volume Visualization. *IEEE Trans Visual Comput Graph*, 2011, 17: 2106–2114
- 3 Chen W F, Chen W, Bao H J. An efficient direct volume rendering algorithm for dichromats. In: *Proceedings Visualization/Information Visualization*, 2011, Vol. 17
- 4 Cui W W, Liu S X, Tan L, et al. TextFlow: towards better understanding of evolving topics in text. In: *IEEE Transactions on Visualization and Computer Graphics*, 2011

## Information visualization and visual analytics: challenges and opportunities

DAI GuoZhong<sup>1</sup>, CHEN Wei<sup>2</sup>, HONG WenXue<sup>3</sup>, LIU ShiXia<sup>4</sup>, QU HuaMin<sup>5</sup>, YUAN XiaoRu<sup>6</sup>, ZHANG JiaWan<sup>7</sup> & ZHANG Kang<sup>7\*</sup>

1 *Institute of Software, Chinese Academy of Sciences, Beijing 100190, China;*

2 *State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310058, China;*

3 *Institute of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China;*

4 *Microsoft Research Asia, Beijing 100080, China;*

5 *Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hongkong, China;*

6 *School of Electronics Engineering, Peking University, Beijing 100871, China;*

7 *School of Software Engineering, Tianjin University, Tianjin 300072, China*

\*E-mail: k Zhang@tju.edu.cn

**Abstract** Sponsored and organized by School of Software Engineering, Tianjin University, and co-organized by Yanshan University and Northeast University at Qinhuangdao, 2012 Workshop on Information Visualization was successfully held in Beidaihe, China, 9–10 June 2012. This article reports and summarizes the discussion details, including the challenges and current major research directions in information visualization and visual analytics. The article also sets the future directions and proposes various action items for promoting and advancing the cutting-edge research in information visualization and visual analytics.

**Keywords** information visualization, visual analytics, data visualization, human-computer interaction, big data