

# 前沿 | 国际可视化盛会PacificVis2017的十个精彩案例

转载

[weixin\\_34391854](#) 于 2017-05-17 09:52:00 发布 278 收藏

文章标签: [大数据](#) [数据结构与算法](#) [设计模式](#)

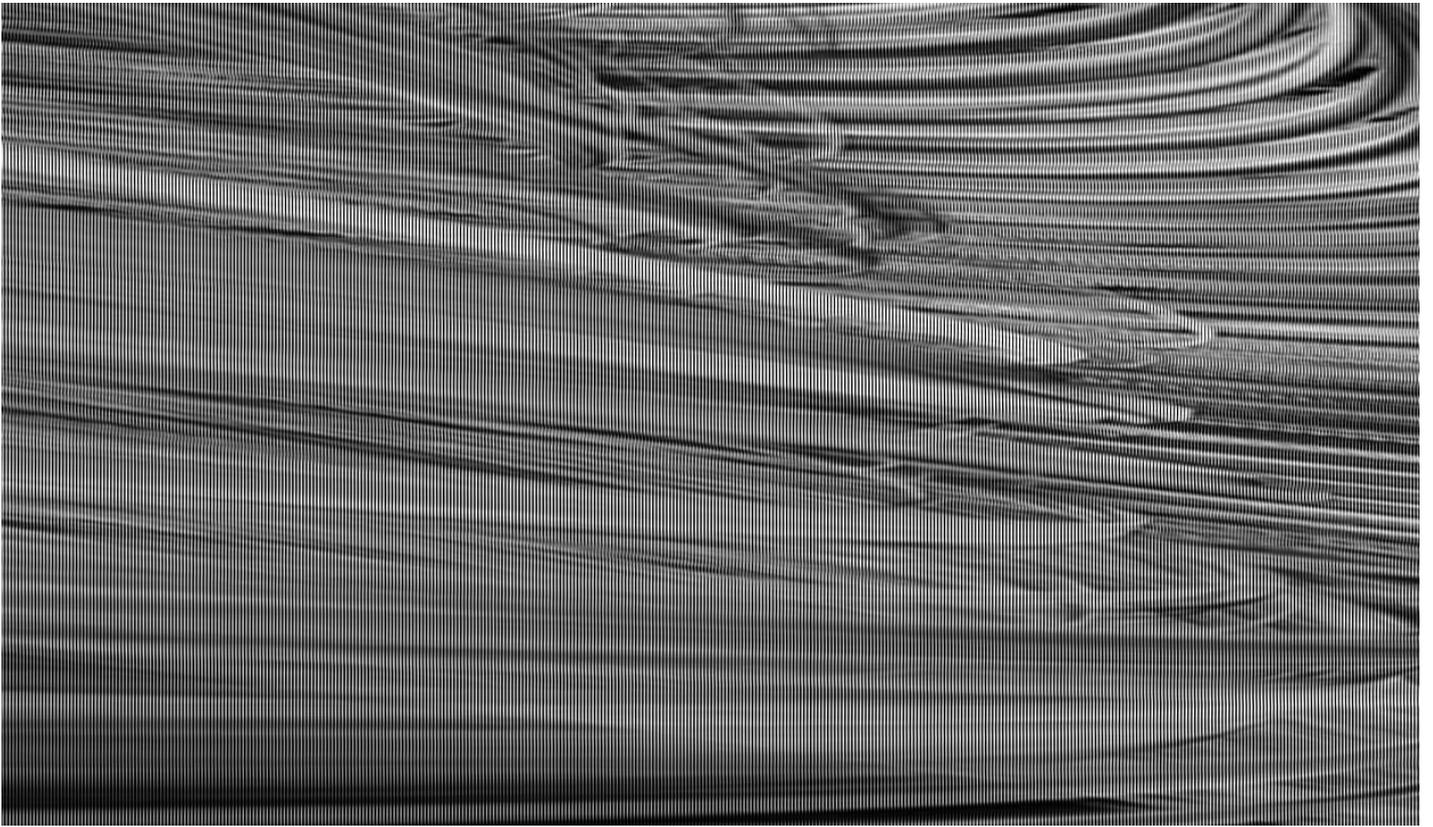
原文链接: <https://yq.aliyun.com/articles/82790>

版权



## 1. 可视分析中的新方向

马匡六教授, 美国加利福尼亚大学戴维斯分校



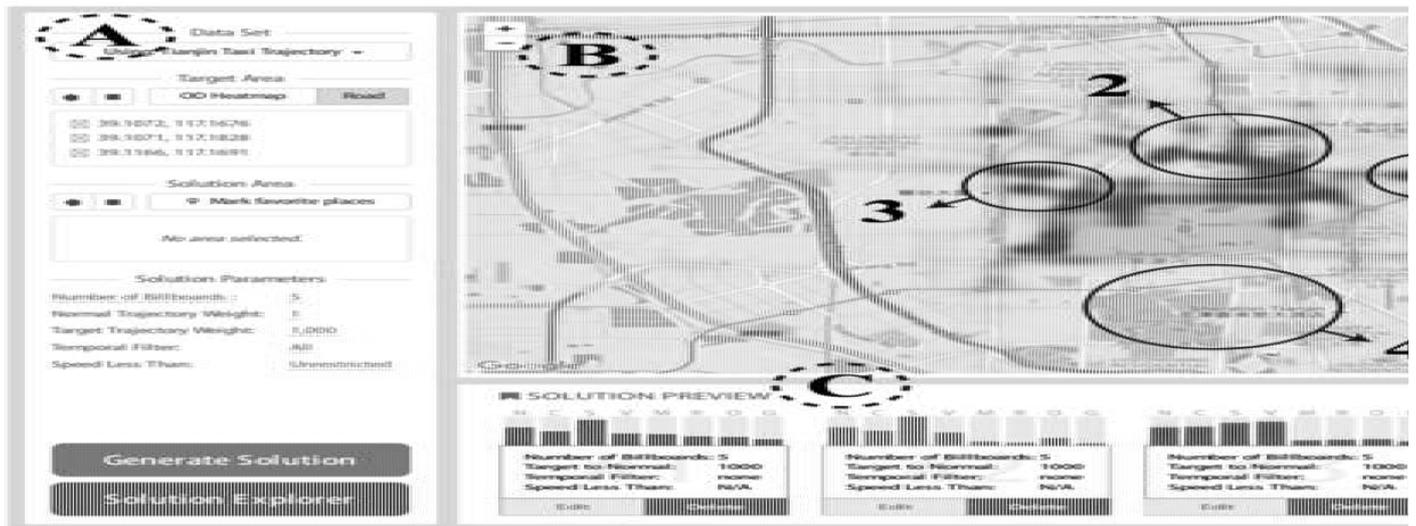
与PPPL科学家合作进行等离子体湍流的计算机模拟

马教授是国际可视化与可视分析研究领域的著名学术带头人之一，作为本次大会中Pacific VAST研讨会的第一位演讲嘉宾，主要从极大规模(Extreme Scale)数据的数据分析、事件序列数据分析这两个方面介绍了可视分析的研究新动向。第一个部分主要关注超大规模模拟所产生数据的可视分析，马教授指出使用原位可视化(In Situ Visualization)是大势所趋，应用场景包括科学可视化、在线数据或流数据可视化等等。其中，数据聚集操作是一类重要的方法。通过对粒子或者场数据进行概率分布函数(Probability Distribution Functions)计算并离散化为一维或二维直方图，模拟数据的信息能够得到很大程度的保留，并且能支持缩放、修改划分、修改直方图槽数量、以及利用直方图进行查询和选择等多种操作；把二维直方图按时间顺序堆叠起来还可生成3D体可视化进行时变属性分析。

第二部分主要介绍了事件序列数据可视化的几个案例，包括高性能计算中的性能、日志数据的可视化，如使用Behavioral Lines的云计算性能的可视分析，以及医院烧伤病人日志数据的可视分析等等。马教授还简短地提到一个新的研究方向是保留隐私的可视化，具体可参考论文Privacy Preserving Visualization for Social Network Data with Ontology Information。

## 2. 城市数据的可视分析

巫英才教授，浙江大学



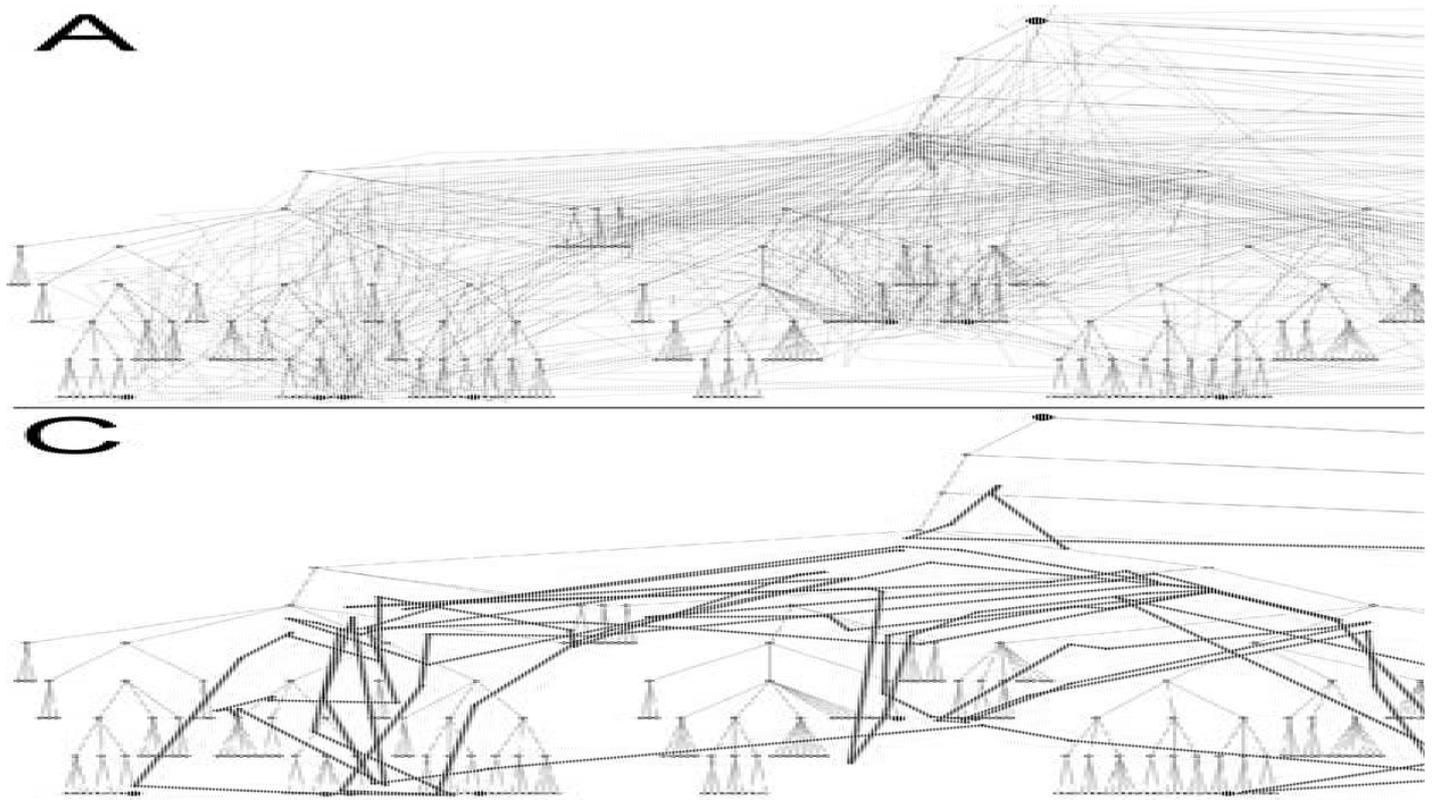
SmartAdP系统界面

巫英才教授首先介绍了城市数据分析面临的挑战，主要存在三个方面：（1）如何视觉提炼海量数据；（2）如何将专家知识融入分析中并获得见解；（3）如何将挖掘的知识有效的传递给用户。基于这些挑战，可视分析是非常重要的。相应地，巫教授设立了多层次可视化、富交互多角度探索，以及通过故事叙事的形式表现分析结果这三个可视分析目标。

随后他介绍了在城市可视分析方面的两个工作：SmartAdP 是一个利用出租车轨迹数据，综合车流量（车流量越大，越可能看到），车速（越慢越好），OD区域（潜在用户的途径路线），环境等多个因素，自动产生广告牌的摆放的解决方案，并提供可视分析界面支持用户进行探索和决策的工具，具体介绍可参考同名论文或北大的博文：用海量出租车轨迹数据选取广告牌放置位置；Route-Zooming 则是一种通过道路变形将道路信息嵌入到地图中的技术，实现尽可能少变形的前提下，对地图上的道路进行拓宽，然后将时空数据图表内嵌入道路中。

### 3. 眼动追踪与可视分析

Daniel Weiskopf, 德国斯图加特大学



几种不同的眼动轨迹图

近年来，硬件技术的进步和眼动追踪装置成本的降低使得这种分析技术可以被大量的研究人员使用。Weiskopf 教授一方面指出了眼动追踪技术在用户研究中的作用，即帮助研究用户是如何使用可视分析界面的。与传统的计算准确度和问卷调查等形式不同，该技术能有效捕捉用户注意力的分布情况。另一方面，则介绍了如何通过可视分析技术来辅助对（大量）眼动追踪数据的解读。其中，常用的两种分析方法包括Space-time cube轨迹显示法以及Map display轨迹显示法；而针对不同类型的眼动追踪数据，如关注区（AOI）、浏览轨迹（Scan Path）和聚焦点（Fixation）等，Weiskopf 教授也分别介绍了其分析方法。更多详情请参考其论文：Visual Analytics Methodology for Eye Movement Studies。

最后，Weiskopf 教授总结道，以上两种眼动追踪可视分析结合起来，可以称作是一种“Vis4Vis”方法——即“用于分析可视化作品的可视化（Visualization for visualization）”。这种可视化类别目前未得到明确定义和划分，因此他希望学术界能够更多地关注该方向的研究。

#### 4. 用可视化进行数据叙事

Bongshin Lee, 微软



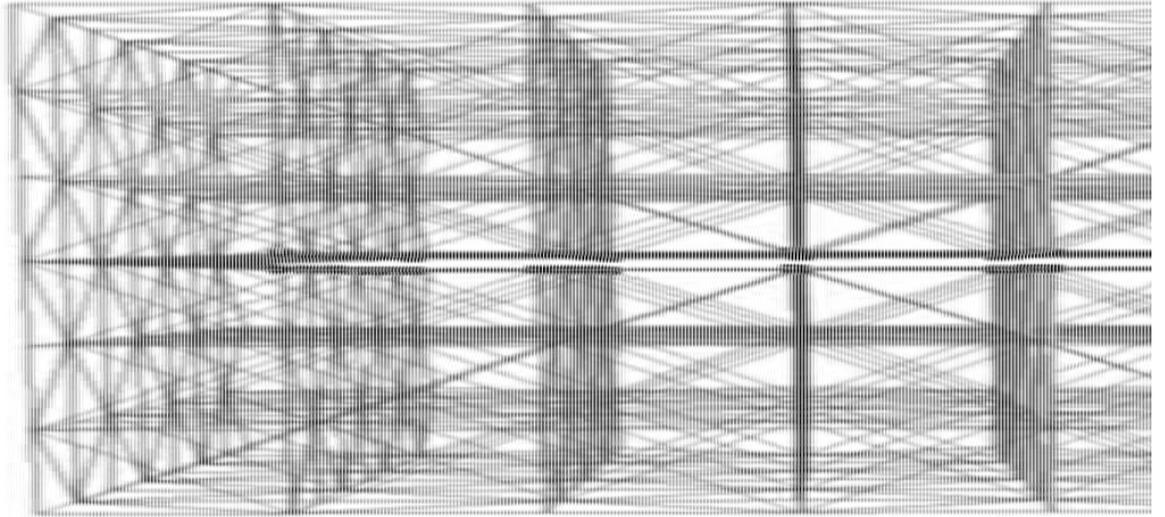
SketchStory工具：用自由手绘讲述更吸引人的数据故事

来自微软研究所的Bongshin Lee博士由汉斯·罗斯林教授经典的全球人口变化介绍等可视化案例引入，介绍了她在故事叙事（storytelling）方面自2012年以来的理论研究。在她的文章More than telling a story中，给出了对视觉数据故事（visual data story）的定义，概括出三大要素：数据驱动、视觉展示，以及故事片段之间的逻辑关系；而故事叙事的过程则由探索数据，制造故事，和叙述故事这三个阶段有机地组成。

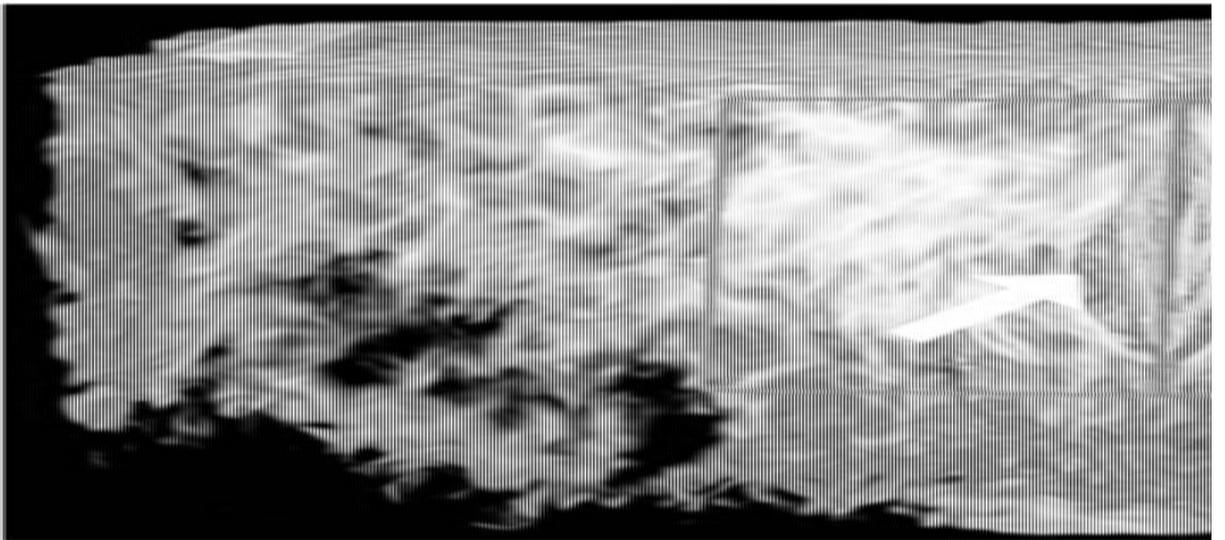
另一方面，Lee博士具体介绍了其团队开发的，几款帮助故事叙事的创作工具：SketchStory受白板动画（Whiteboard Animation）展示方式的启发，提出一种使用触控笔在电子板上描绘图表的故事叙述方式，帮助提升听众的参与感；DataClips是为非专业人士设计的数据视频创作工具，通过对纽约时报等74个数据新闻视频进行分析，建立数据新闻材料库，相比传统视频制作软件更易于上手，且能更灵活地修改数据；Timeline Storyteller是使用时间线为顺序的叙事工具，利用了对200余款同类可视化的总结及从中提取的设计模式，可以在线尝试。最后展望了未来的storytelling方向和挑战：包括如何融合故事的探索到展现的过程；如何评估数据驱动的故事和工具；数据故事的可读性；新颖的、浸入式的交互体验在storytelling的应用。

## 5. 虚拟牵引器（Virtual Retractor）

Cheng Li, 俄亥俄州立大学



A. The applied tetrahedra



C. Deform the vector of NEK dataset

Virtual Retractor: 使用基于物理变形的交互式数据探索系统

来自俄亥俄州立大学的可视化小组提出了Virtual Retractor这种创新的交互式数据探索系统，在保证数据特征的基础上对于可视化进行变形，从而解决科学可视化的3D空间中存在着遮挡的问题。该系统采用Retractor的隐喻，允许用户像切割和分割实际物体一样，直接操纵和探索数据。这个过程通过使带有空隙的四面体网格（tetrahedral mesh）变形来模拟由切割产生的切口来进行。分割操作扩大了空隙，允许用户观察原来被遮挡的内部结构。其中，网格（mesh）的构造考虑了局部数据属性，如局部数据密度或梯度。因此，变形将受到所选数据属性的影响。具有高数据属性值的区域更加坚固，难以变形，而属性值较低的区域将会变形。从而，变形可以保持数据中被感兴趣的特征。

## 6. 可视分析改变世界

David Ebert, 美国普渡大学



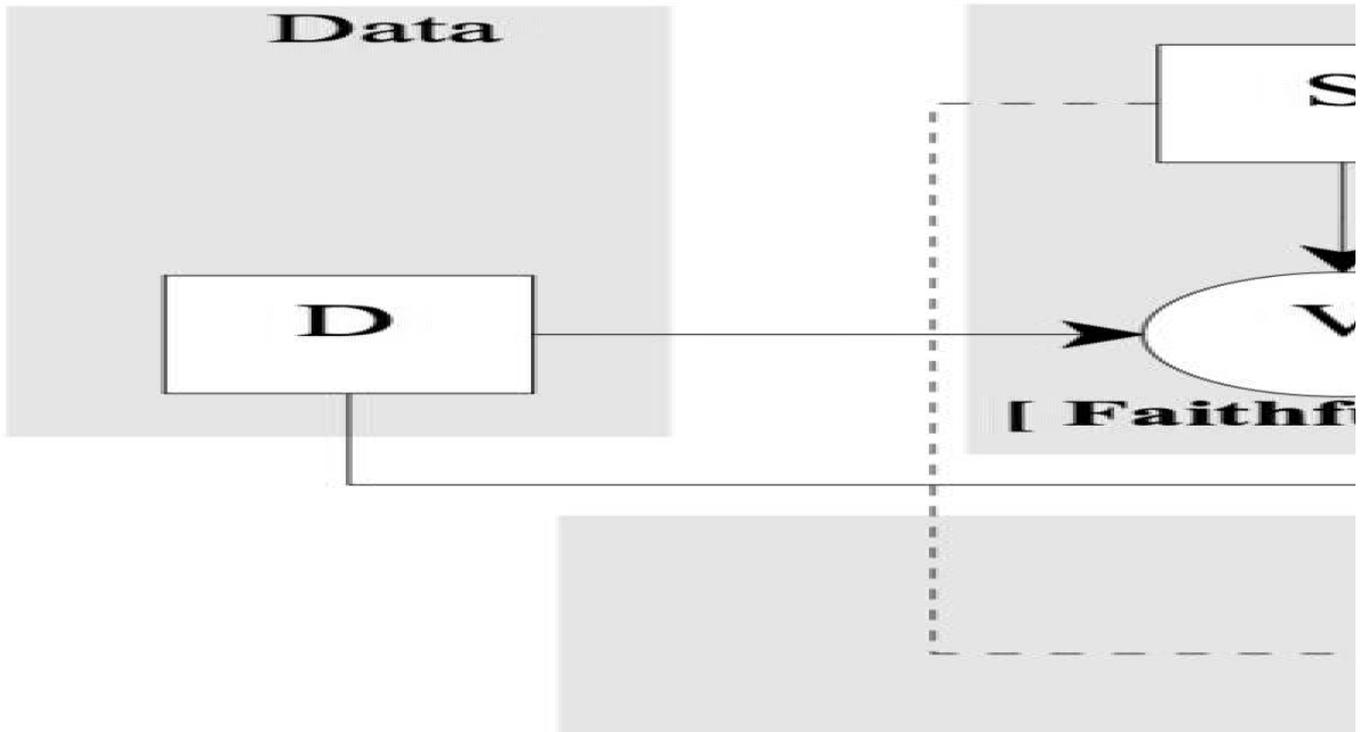
用于多个聚合尺度的微博数据分析的可视分析框架

David Ebert教授是可视分析领域的专家，在包括安全、国防、社交媒体等等领域都有相关的可视分析工作。本次演讲中，他首先总结了我們目前面临的解全球性难题，如能源、可持续发展、洁净水、人口资源等问题，并鼓励学者进行基于应用驱动的研究（Application-driven research），这样能有效推动科学发现和改善世界，而且也有助于获得研究资金和更多优质合作机遇。

在对可视分析进行释义之后，他介绍了人机合作设计Human Computer Collaborated Design (HCCD)，这种可以结合人类与计算机各自的优势、方便沟通，并有效地实现问题决策的方法；并举了一系列关于利用社交媒体、安全、资源分布等等方面的例子（可参见他早前的报告）。他强调，可视分析是辅助人、提升人的能力，而不是替代人的工作。最后，他还对最近兴起的深度学习、风险可视分析（Risk Visual Analytics）、因果可视分析（Causality Visual Analytics）以及VR/VA结合可视化等热点话题进行了讨论；并表示他对VR/VA在科学发现和决策分析领域的作用持观望态度，但基于人们对这类技术的热衷程度，他亦对此抱有期待。

## 7. 图可视化的新评估方法

Seok-Hee Hong, 悉尼大学



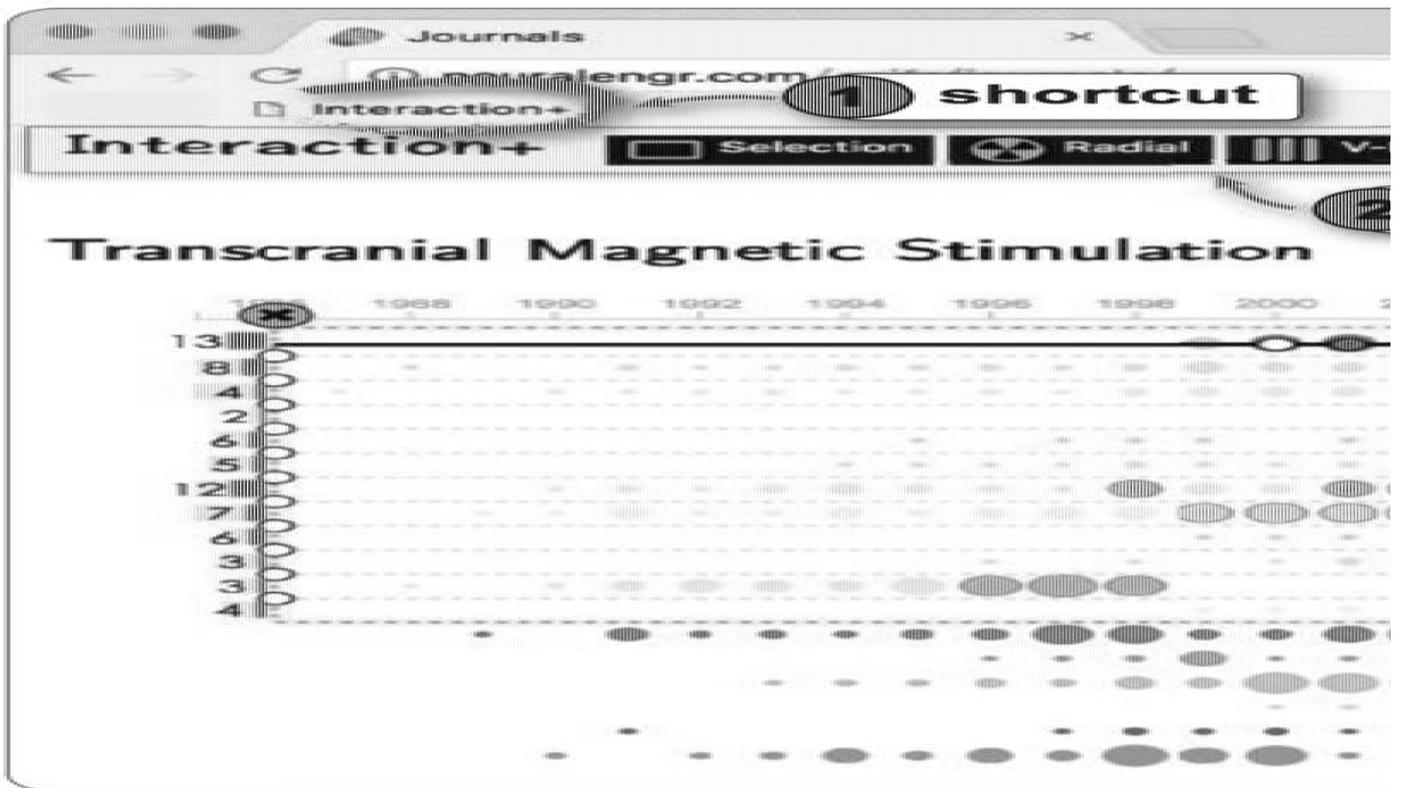
Hong教授及其合作者提出的图可视化模型

图可视化主题在今年所接受论文、海报张贴中都有着很大的比例，而Hong教授在演讲中介绍的忠实性（Faithfulness）指标在此次大会中多次被其他演讲者引用，可谓热中之热。现有的图布局评估方法，如基于边交叉的方法，主要关注图的可读性；但随着大数据时代的来临，原有的评估方法是不够的，因为包含大量数据的图通常过于复杂，导致用户难以直接通过感知边的交叉来评估图布局的优劣。

于是，他们针对大图提出了忠实性（Faithfulness）指标，它包括三大部分，Information faithfulness（图布局对信息表现的忠实性），Task faithfulness（图布局对任务的支持度）和Change faithfulness（图布局对数据微变化的敏感度）；同时提出了基于形状的度量方法（Shape-based Metric）来计算Task faithfulness，且通过实验证明了这种方法的有效性。有关对力导向图和边捆绑技术等现有可视化方法的忠实性（Faithfulness）评估，以及忠实性与可读性的比较，请参考其论文：Towards Faithful Graph Visualizations。

## 8. Interaction+: “让可视化动起来”

陆旻，北京大学



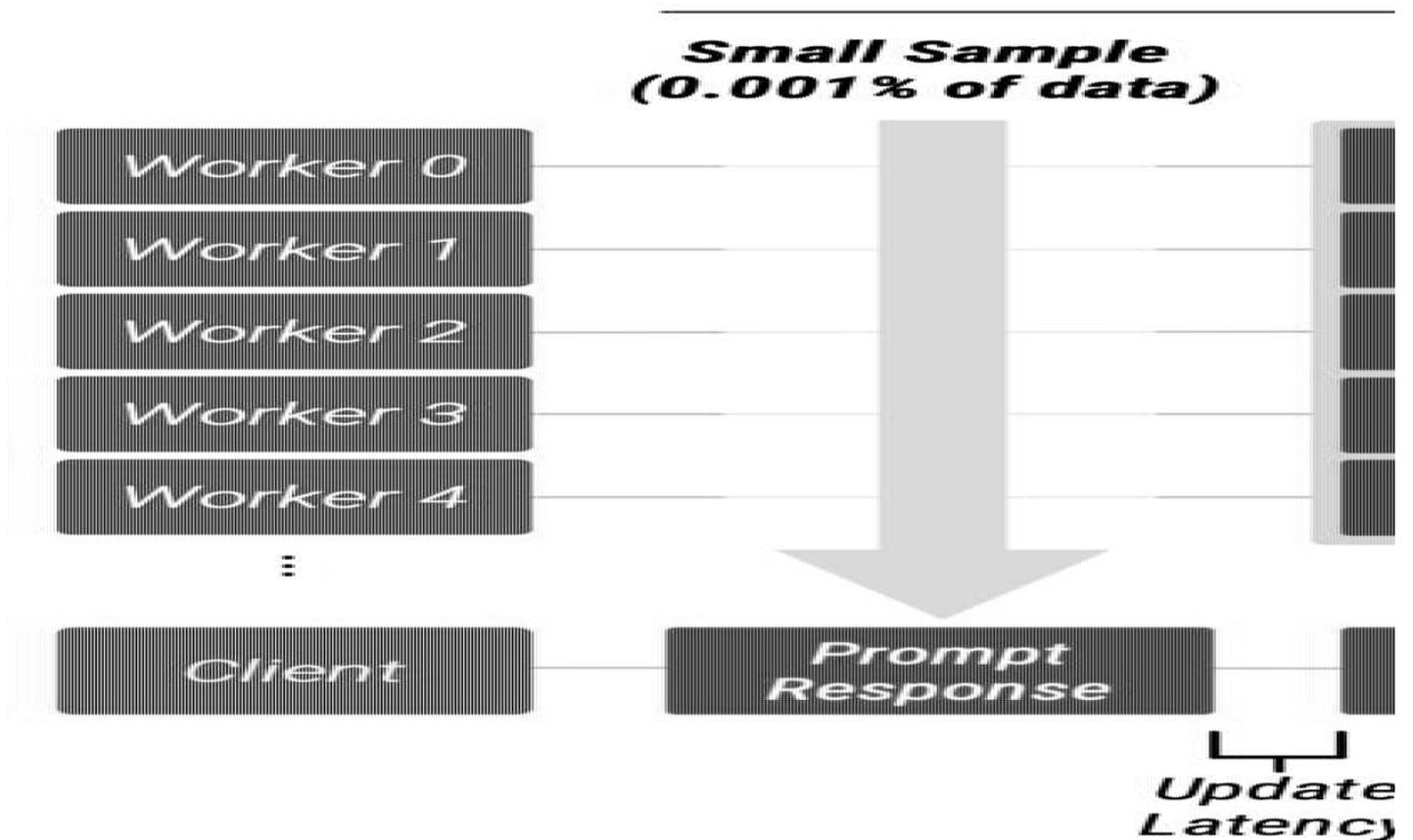
Interaction+ 基本交互界面

这个富有创意的工具早在2016年ChinaVis作为海报展出时就十分吸引人眼球，无需编程就能将网络上现有的可视化作品变成动态可交互的，这对可视化设计师甚至工程师都是提高工作效率的福音。如今，这个工具不仅能自动识别已有的网页可视化作品中的视觉图元及属性信息，然后对原可视化内容进行过滤、聚集和重排，还能让用户能在原可视化的基础上用辅助面板进行查询、比较、打标签等交互。

其实现原理是通过解析网页HTML文件物件模型 (DOM)，并对SVG元素进行提取和操作，因此该工具适用于任何网页上的SVG可视化作品，可以是程序员编写的程序也可以是艺术家制作的信息图。作者以纽约时报上的几个可视化作品为例，对工具进行了精彩的演示，视频详见北大公众号相关文章：“让可视化动起来”的网页交互工具Interaction+。

## 9. 基于Spark的大规模高维数据可视化

Jaemin Jo, 首尔国立大学



SwiftTuna中的大规模并行处理流程

来自首尔国立大学的Jaemin Jo首先分析了针对大规模数据的交互式探索的相关研究（见附录）。常见的方法是使用数据立方体（data cube）等预处理数据结构，来总结数据并提供低延迟响应。然而，由于查询中涉及很多维度，这种方案会占用大量的内存，且在查询前必须由数据构建特定数据结构。他所在的团队则利用亚马逊EC2集群，并且基于Spark的流式处理方式，开发了适用于大规模高维数据可视分析的系统SwiftTuna。为了支持响应查询，SwiftTuna利用增量处理方法，立即提供低保真响应（即快速响应）以及延迟的高保真响应（即增量响应）。该系统支持对高达含有40亿条记录和40个维度的真实数据集的实时探索，且增量响应的延时仅为几秒钟。具体实现详见论文：SwiftTuna: Responsive and Incremental Visual Exploration of Large-scale Multidimensional Data。

## 10. 增强现实（AR）可视化

Gerard Jounghyun Kim, 韩国高丽大学

# VIZ. ISSUES IN AR

Characteristics of AR	How it might affect the "utility" of visualization
Data insertion into 3D world	<ul style="list-style-type: none"><li>● Resulting augmented imagery might not be natural<ul style="list-style-type: none"><li>■ Does not look 3D,</li><li>■ Does not look in the right place</li><li>■ Does not have the right affordance or metaphor</li><li>■ Registration error is too large</li></ul></li><li>● Target data may be occluded</li><li>● Augmentation may be occluded</li></ul>
Dynamic background	<ul style="list-style-type: none"><li>● Augmentation may not be clearly visible</li><li>● Augmentation may disappear</li></ul>
Unchangeable target data (i.e. real objects fixed in place)	<ul style="list-style-type: none"><li>● Difficult to proper place the augmentation</li><li>● Difficult to represent structure of the collective target data</li><li>● Difficult to determine the proper size of augmentation</li></ul>
Target data with spatial contexts	<ul style="list-style-type: none"><li>● Augmentation may occlude the context (neighboring objects)</li></ul>

Kim教授总结的AR中的可视效果问题

随着AR技术的兴起，AR与可视化的结合也成为了可视化领域的新热点。Gerard Jounghyun Kim教授主要从AR领域研究者的角度，分享了他对AR可视化的看法。Kim教授认为AR可以被视为一种特殊的数据可视化，其数据不仅有现实数据，还包括了增强物数据；由于AR的特殊属性，在实现有效的可视化时会面临一些独特的问题，对此Kim教授总结和提出了AR可视化的三个主要要求，即（1）自然度 - Naturalness，（2）可见度 - Visibility，（3）持久性/稳定性 - persistance/stability。

自然度指的是增强物需要与用户对真实世界的认知模型保持一致，如正确的深度信息、相似的光照效果、合理的空间关系等等。可见度是指最大化信息的可见度，例如增强目标物体的显示、利用过滤法消除信息遮挡、平衡真实环境和增强物的关系等等。持久性和稳定性是指增强物应保持稳定，且易于查看与交互。Kim教授以钢琴乐谱增强现实可视化为例，指出用户在演奏的时候同时关注键盘与乐谱，需要经常转移视线，容易造成疲倦。

最后，Kim教授谈论了AR可视化中一些待解决的问题以及相应的发展方向，包括增强物的放置（如去除遮挡、决定尺寸位置）、三类要求衡量标准的制定，以及平衡增强物的自然性和可见性等等。

## 后记

以上内容是对笔者在参加会议期间个人比较感兴趣，或是比较新颖的一些主题整理的，限于篇幅还有很多精彩的演讲没有列举，如果感兴趣可以参考完整的会议项目。更多详情也可参考北京大学可视分析公众号的系列文章（详见附录：参考文献），本文中一些对专有名词及技术描述的翻译参考了该系列文章，特此声明和感谢。

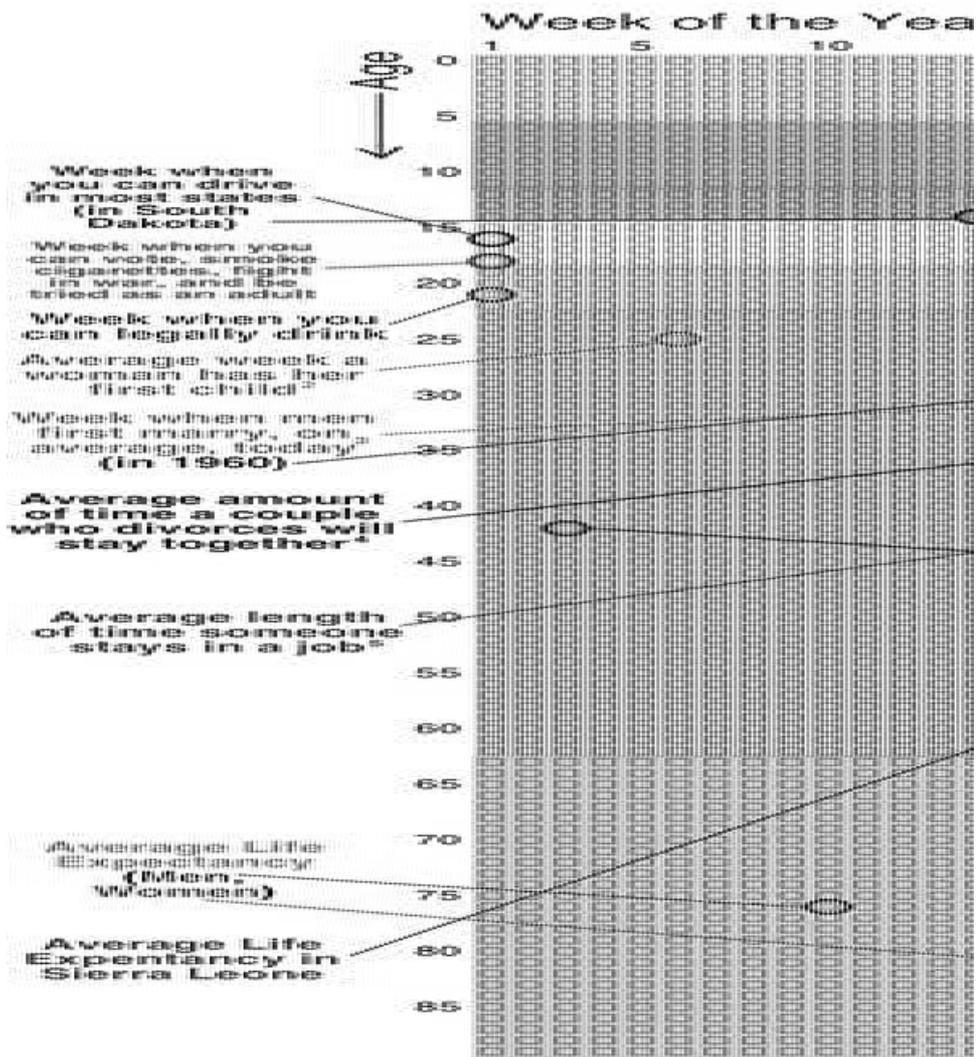
由于会议内容均以英文呈现，且有许多笔者不太熟悉的领域和名词，尽管笔者在写作过程中参考了大量资料，仍可能存在疏漏或不准确的地方，请多包涵，也欢迎批评指正。谢谢！

## 附赠实用链接

### 数据叙事实例及工具：

- 机器学习的视觉化介绍 | Kantar信息之美2015年获奖作品，以懂得区别旧金山和纽约市房屋的模型为例，在网页中将机器学习的基本过程一步步地用动态图展现，有（繁体）中文版。（<http://www.informationisbeautifulawards.com/showcase/743-a-visual-introduction-to-machine-learning>）
- 统计的乐趣 | 汉斯·罗斯林 (Hans Rosling) 最出名的一次演讲，在 4 分钟内，用 12 万个数据解释 200 个国家 200 年的发展，有中英字幕。（[http://v.youku.com/v\\_show/id\\_XNTM0ODg3ODg4.html?from=s1.8-1-1.2&spm;=a2h0k.8191407.0.0](http://v.youku.com/v_show/id_XNTM0ODg3ODg4.html?from=s1.8-1-1.2&spm;=a2h0k.8191407.0.0)）
- 新兴的故事叙事技术 | 颇值得一读的综述性英文论文，题目为 *Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories*.（<https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/MSR-TR-2016-14-Storytelling-Techniques.pdf>）
- SketchInsight 和 SketchStory | 这个视频展现了数据叙事中，两种白板动画型创作工具的结合使用：前者用于创作故事，后者用于展示故事。（<https://www.microsoft.com/en-us/research/video/sketchinsight-telling-stories-with-data-via-freeform-sketching/>）
- 以周计算的你的一生 (互动版), 阿拉伯之春时间线 | 两个时间线可视化的实例，分别采用了分段布局和曲线布局。（[http://www.huffingtonpost.com/wait-but-why/your-life-in-weeks\\_b\\_5366289.html](http://www.huffingtonpost.com/wait-but-why/your-life-in-weeks_b_5366289.html), <https://labs.coruscantconsulting.co.uk/life/weeks/>, <http://www.slow-journalism.com/sj-site/wp-content/uploads/2012/03/Arab-Spring-timeline.jpg>）

# The Life



以周计算的你的一生.png

