



医学数据学导论

第 15 章：健康大数据可视化技术

编著：薛付忠 & 井明

学院：医学大数据学院

时间：Dec 1, 2022

版本：0.1



迟序之数，非出神怪，有形可检，有数可推。——南北朝祖冲之《述异记》

目录

第 15 章 健康大数据可视化技术	1
15.1 数据可视化原理	1
15.1.1 数据可视化基本概念	1
15.1.2 数据可视化原理	2
15.1.3 常用可视化工具	2
15.2 时间维度数据可视化技术	3
15.2.1 队列数据可视化	3
15.2.2 纵向数据可视化	3
15.2.3 函数型数据可视化	4
15.2.4 流数据可视化	4
15.3 时空维度数据可视化技术	4
15.3.1 流行病地图	4
15.3.2 地理信息系统	4
15.3.3 GIS 中常用的可视化技术	4
15.4 多维度数据可视化技术	4
15.4.1 电子病历多维数据可视化	4
15.4.2 人群数据多维可视化	4
15.4.3 生物组学多维数据可视化	4
15.5 网络数据可视化技术	4
15.5.1 流行病学网络可视化	4
15.5.2 生物组学网络可视化	4
15.5.3 生物通路可视化	4
15.5.4 网络动态可视化	4
15.6 数据可视化应用	4
15.6.1 展示数据的复杂结构	4
15.6.2 展示数据的复杂分布	4
15.6.3 展示数据的复杂关系	4
15.6.4 展示数据的复杂动态	4
15.6.5 展示数据的血缘关系	4

第 15 章 健康大数据可视化技术

15.1 数据可视化原理

15.1.1 数据可视化基本概念

在计算机学科的分类中，利用人眼的感知能力对数据进行交互的可视表达以增强认知的技术，称为可视化。可视化将不可见或难以直接显示的数据转化为可感知的图形、符号、颜色、纹理等，增强数据识别效率，传递有效信息。

基于计算机的可视化系统提供数据集的可视化表示，旨在帮助人们更有效地执行任务。当需要增强人类能力而不是用计算决策方法代替人类时，可视化是合适的。可能的视觉习语的设计空间是巨大的，并且包括如何创建以及如何与视觉表示进行交互的考虑。Vis 设计充满了权衡取舍，设计空间中的大多数可能性对于特定任务都是无效的，因此验证设计的有效性既必要又困难。可视化设计者必须考虑三种截然不同的资源限制：计算机、人类和显示器的资源限制。可以根据用户需要它的原因、显示的数据以及习语的设计方式来分析 Vis 的使用情况。

Vis 允许人们在不确切知道需要提前问什么问题时分析数据。

现代时代的特点是通过访问比以往任何时候都多的数据来做出更好的决策。当人们对数据有明确定义的问题时，他们可以使用来自统计学和机器学习等领域的纯计算技术。★一些曾经由人类完成的工作现在可以通过基于计算机的解决方案完全自动化。如果一个全自动的解决方案被认为是可以接受的，那么就不需要人工判断，因此你不需要设计一个 vis 工具。例如，考虑股票市场交易领域。目前，已经部署了许多用于高频交易的系统，它们可以在特定市场条件下做出买卖股票的决定，例如，当达到特定价格时，根本不需要人工进行耗时的检查在循环。你不会想要设计一个 vis 工具来帮助人们更快地进行检查，因为即使是增强型人类也无法每秒推理数百万只股票。

然而，许多分析问题没有明确说明：人们不知道如何解决问题。有很多可能的问题要问——从几十到几千甚至更多——人们事先不知道这些问题中哪一个是正确的。在这种情况下，最好的前进路径是一个人类参与的分析过程，您可以在设计中利用人类视觉系统的强大模式检测特性。当您的目标是增强人类能力而不是完全取代循环中的人类时，Vis 系统适用。

您可以为多种用途设计 vis 工具。您可以通过帮助设计人员设计纯计算的未来解决方案，制作一个用于过渡使用的工具，其目标是“让自己摆脱工作”。你也可以制作一个长期使用的工具，在短期内不打算取代人类的情况下。

例如，您可以创建一个 vis 工具，这是在开发正式的数学或计算模型之前更清楚地了解分析需求的垫脚石。这种工具将在过渡过程的早期以高度探索性的方式使用，甚至在开始开发任何类型的自动解决方案之前。除了工具本身之外，针对特定现实世界领域问题设计 vis 工具的结果通常是对用户任务的更清晰的理解。

在过渡的中间阶段，您可以构建一个针对纯计算解决方案设计人员的 vis 工具，以帮助他们改进、调试或扩展该系统的算法，或了解算法如何受到参数变化的影响。在这种情况下，您的工具针对的是与最终系统的最终用户截然不同的受众；如果最终用户需要可视化，它可能会使用非常不同的界面。回到股票市场的例子，一个更高层次的系统，它决定在不同情况下使用多种交易算法中的哪一种，可能需要仔细调整。帮助算法开发人员分析其性能的 vis 工具可能对这些开发人员有用，但对最终购买该软件的人没有用。

您还可以结合其他计算决策为最终用户设计 vis 工具，以阐明自动系统是否根据人类判断在做正确的事情。该工具可能用于在过渡后期做出部署决策时的临时使用，例如，在委托机器学习系统花费数百万美元交易股票之前查看机器学习系统的结果是否值得信赖。在某些情况下 vis 工具在做出决定后就被放弃了；在其他情况下，vis 工具继续长期用于监控系统，以便人们在发现不合理行为时可以采取行动。

与这些过渡性用途相比，您还可以设计长期使用的 vis 工具，在这种情况下，人们将无限期地停留在循环中。一个常见的案例是科学发现的探索性分析，其目标是加速和提高用户生成和检查假设的能力。图 1.1 显示了一个

vis 工具，旨在帮助生物学家通过分析 DNA 序列变异来研究疾病的遗传基础。尽管这些科学家大量使用计算作为他们更大工作流程的一部分，但短期内完全自动化癌症研究过程的希望并不大。

您还可以设计用于演示的 vis 工具。在这种情况下，你支持那些想向别人解释他们已经知道的事情的人，而不是探索和分析未知的东西。例如，《纽约时报》结合新闻报道部署了复杂的交互式可视化。

15.1.2 数据可视化原理

数据可视化大致可分为信息可视化、科学可视化和可视化分析三大类。基础的数据可视化可以仅仅是输入几组数据，生成简单的条形图或直线图等等，但是，随着数据量的增大，可视化目标会发生改变，可视化系统的复杂度也会相应增加。数据可视化从基础到复杂，都可以用下面的可视化模型来概括。

1990 年 Robert B. Haber 和 David A. McNabb 提出的数据可视化流程已经非常先进，整个流程是线性的。它把数据分成五大阶段，分别要经历四个流程，每个过程的输入是上一个过程的输出。从图上看非常直观，很好理解。为了更清晰的表达可视化的原理，以上可视化流程整体也可以总结为三步：分析 - 处理 - 生成。

(1) 分析进行一个可视化任务时，首先要作的是分析，分析又分为三个方面：任务、数据、领域。

首先我们要分析我们这次可视化的出发点和目标是什么。我们遇到了什么问题、要展示什么信息、最后想得出什么结论、验证什么假说等等。数据承载的信息多种多样，不同的展示方式会使侧重点有天壤之别。只有清楚表达以上问题，才能确定可视化任务要过滤什么数据、用什么算法处理数据、用什么视觉通道编码等等。

其次我们要分析我们的数据，这是至关重要的一步。因为每次可视化任务拿到的数据都是不同的，数据类型、数据结构均有变化，数据的维度也可能成倍增加。抽象的数据类型如何对应现实中的概念，不同的数据类型如何进行视觉编码，这些我们在下一篇数据模型中进行介绍。

最后我们针对不同的领域，也要进行相应的分析。毕竟术业有专攻，可视化的侧重点要跟着领域做出相应的变化。

(2) 处理处理可以分为两部分：对数据的处理和对视觉编码的处理。

1. 数据处理在可视化之前我们要对数据进行数据清洗、数据规范、数据分析。

数据清洗和规范是必不可少的步骤。首先把脏数据、敏感数据过滤掉，其次再剔除和我们目标无关的冗余数据，最后调整数据结构到我们系统能接受的方式。

数据分析中最简单的方法当然是一些基本的统计方法，如求和、中值、方差、期望等等；复杂的方法有数据挖掘的各种算法，这是又一个领域了，在此不赘述。

最后的可视化结果中我们肯定不可能把所有的数据统统展示出来，于是又涉及到包括标准化（归一化）、采样、离散化、降维、聚类数据处理的方法，这些概念之后可以单独写篇文章来介绍。

2. 设计视觉编码视觉编码的设计是指如何使用位置、尺寸、灰度值、纹理、色彩、方向、形状等视觉通道，以映射我们要展示的每个数据维度。

(3) 生成这个阶段基本上就是把之前的分析和设计付诸实践，在制作或写代码过程中，再不断调整需求、不断地迭代（有可能要重复前两步），最后产出我们想要的结果。

15.1.3 常用可视化工具

一个完整流程的可视化工作，包括数据采集、数据清洗、数据分析和可视分析四个步骤，其中还需要考虑基础平台搭建和架构设计，以及可视化软件和系统评估选型等工作，根据数据蕴含的意义和可视任务的类型，选择合适的可视化软件或工具可以保障可视化任务的高效进行。根据可视化对象的类型和领域的不同，可以将可视化软件分为医学可视化、科学可视化、信息可视化和可视分析四类，其中，医学可视化软件最早被广泛应用于临床医学影像数据的展示和分析任务，本节将主要介绍医学可视化软件的常用工具，3D Slicer、Osirix 和 VolView 是三个最具代表性的软件系统。

(1) 3D Slicer 简称 Slicer，是一个免费的、开源的、跨平台的医学图像分析与可视化软件，广泛应用于科学研究与医学教育领域。Slicer 可以运行于 Windows、Linux 和 MacOS 平台之上，具备以下功能：

- 支持 DICOM 图像，并支持其他格式图像的读写。

- 支持三维体数据、几何网络数据的交互式可视化。
- 支持手动编辑、数据配准与融合和自动图像分割。
- 支持弥散张量成像和功能磁共振成像的分析和可视化，提供图像引导放射治疗分析和图像引导手术的功能。

(2) Osirix 软件是 MacOS 平台上最成功的开源医学图像软件，同时支持移动平台 iOS，可以运行在苹果 iPhone 手机和 iPad 平板上。Osirix 集 PACS 工作站和图像处理软件于一体，提供了高效的二维和三维图像的处理功能，为放射成像、功能影像、三维成像和分子影像等研究提供支持，可用于显示、浏览、解析和后处理由 MRI、CT、PET、PET-CT 等医疗设备产生的 DICOM 数据。它完全兼容 DICOM 标准，与现有医学图像浏览软件形成互补。Osirix 也支持很多其他图像和视频格式，如 TIFF、JPEG、PDF、AVI、MPEG 和 MOV。除此之外，Osirix 针对多模态、多维图像的浏览和可视化进行了优化设计，支持四维（三维图像序列加上时间序列组成，如心脏跳动周期的 CT 数据）和五维（三维图像数据加上功能影像组成，如心脏的 PET-CT 数据）数据的展示。Osirix 在功能、操作和性能上非常符合临床医生的要求。

(3) VolView 是美国 Kitware 公司开发的交互式三维数据场可视化商业软件，支持 Windows 和 Linux 操作系统。在生物医学领域，VolView 提供了医学影像处理功能，可完成各类医学影像数据的三维可视化，辅助医生进行手术规划和对病变部位定位等深入认识。在工业工程领域，可提供零件模型反求工程，探测零件的内部探伤等细微错误，并进行精确定位。VolView 的主要功能如下：

- 二维切片图像处理，包括浏览、放大、缩小和旋转等，同时提供图像格式之间的转换。
- 大规模体数据的格式转换和滤波处理等。
- 快速三维重建，包括三维轮廓重建、等值面重建和光线投射体绘制等。
- 对重建的三维对象交互操作，如旋转、放大、缩小、分割、局部编辑和测量等。

15.2 时间维度数据可视化技术

15.2.1 队列数据可视化

队列人群是指在特定时间段内具有共同特征的一群人。以学生群体为例，2022 年入学的学生群体可以看作一个队列，队列中的个体都有一项一致的学籍信息，即入学时间为 2022 年。在医学大数据分析领域，队列人群主要指符合相似病程经历的患者群体，例如，被诊断为患有肺癌的患者信息可以组成一个队列。

队列分析是一种行为分析，它在分析之前将数据集中的数据分解为相关组，这些群体或群组通常在规定的时间跨度内具有共同的特征或经历。以电子商务的客户分析场景为例，队列分析常用于识别群体在使用某个电商服务期间的保留、交互以及随时间的演变趋势。

15.2.2 纵向数据可视化

纵向建模分析在生物医学学科中已经得到广泛应用，其常用技术包括可视化和变更点分析 (CPA) 算法，纵向建模结果的解释通常涉及计算从业者和领域专家之间的协作。

15.2.3 函数型数据可视化

15.2.4 流数据可视化

15.3 时空维度数据可视化技术

15.3.1 流行病地图

15.3.2 地理信息系统

15.3.3 GIS 中常用的可视化技术

15.4 多维度数据可视化技术

15.4.1 电子病历多维数据可视化

15.4.2 人群数据多维可视化

15.4.3 生物组学多维数据可视化

15.5 网络数据可视化技术

15.5.1 流行病学网络可视化

15.5.2 生物组学网络可视化

15.5.3 生物通路可视化

15.5.4 网络动态可视化

15.6 数据可视化应用

15.6.1 展示数据的复杂结构

15.6.2 展示数据的复杂分布

15.6.3 展示数据的复杂关系

15.6.4 展示数据的复杂动态

15.6.5 展示数据的血缘关系