

●周 宁 陈勇跃 金大卫

大型层次化信息的可视化方法研究*

摘 要 对层次信息的特性进行研究并选择合适的可视化方法不仅对复杂信息可视化应用设计有重要作用,而且能满足多种可视化用户的需求。研究大型层次化信息在有限屏幕上有序显示的方法和步骤,有助于解决信息量增加带来的信息查询困难问题。根据层次信息特有的结构特征,可视化描述可以通过节点连线图和矩形嵌套图来表示,并可通过相应的信息布局设计和视图导航的两个步骤来实现。图5。参考文献13。

关键词 可视化 层次化信息 树图 视图导航

分类号 G354

ABSTRACT The analysis of characteristics of hierarchical information and the proper selection of visualization methods will not only have important effects on the visualization applications of complex information, but also meet user's diversified needs for visualization. To study the methods and steps for orderly display of large-scale hierarchical information on limited screens will help solving the problems of information search difficulties caused by the increase of information. According the structural characteristics of hierarchical information, visualization description can be represented with node/link diagrams and tree-maps, and can be realized in the two steps of information layout design and viewport navigation. 5 figs. 13 refs.

KEY WORDS Visualization. Hierarchical information. Tree map. Viewport navigation.

CLASS NUMBER G354

根据抽象信息间关系不同,信息一般分为线性信息类型、层次信息类型、多维信息类型、网络信息类型等^[1]。其中层次信息是最普遍、最重要的一种,在客观世界中广泛存在,如磁盘目录结构、文档管理、图书分类、人类社会的族谱、社会组织机构、地址簿等。层次关系几乎无处不在,而且很多数据信息都可以通过一定的抽象转化为层次信息,对层次信息的特性进行研究并选择合适的信息可视化方法不仅对于复杂信息可视化应用的设计有举足轻重的作用,而且能够满足多种可视化用户的需求。

传统描述层次信息的方法是将它们组织成一个类似于树的节点连线图。这种树状节点连线图结构简单、直观且便于普通用户掌握。但是一般情况下,树结构中层次节点增长的速度很快,通常层次节点的数目随着层次的增长呈指数增长。在普通的二维空间中,树形结构的分支很快就会拥挤交织在一起,变得混乱不堪^[2]。这是层次信息可视化过程中的主要问题。本文在分析层次信息表示方法的基础上,着重研究了大型层次化信息怎样在有限屏幕上有序显示的方法和步骤,解决由于信息量增加带来的信息查询困难。

1 层次信息的可视化表示

1.1 节点连线图

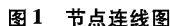
在层次信息可视化中,数据包含的潜在信息结构可以用不同的可视化结构描述。一般把这种层次结构描述成节点连线图(Node/Link Diagrams),即树图。节点用来表示数据,连线表示数据间的关系。计算机领域中这样规范树结构只有一个节点没有父节点,除此之外,每个节点至少有一个父节点。通常我们把树描述成 $T=(V,E)$,其中 v 是节点的集合,如图1所示。

1.2 矩形嵌套图

典型的矩形嵌套图(Tree-map)是Shneiderman^[3]等人提出的一种表示层次信息的可视化模型,此模型的构想比较简单。在这种可视化结构中,层次结构的每个节点被表示为一个矩形,矩形面积表示相应节点的权重,图2是图1描述的层次结构的矩形嵌套图。按照次序嵌套矩形来表示层次关系,即表示一个父节点的所有子节点的矩形被该父节点的矩形包围着。

矩形嵌套图的优点是可以有效利用计算机屏幕空间,且能够很容易实现。但是,它丧失了层次结构的直观性。并且对处于同一层次上的不同父节点的子节点的关系也丧失了。

* 本文为国家自然科学基金项目(70473068)的研究成果。



2 大型层次化信息的可视化步骤

大型层次信息的可视化分为两个重要步骤。第一步就是把相关数据绘制到一个几何平面中,我们称它为绘制图表和信息布局问题。第二步是可视化导航,就是交互式的逐步改变视图,找到用户需要的目标信息^[4]。

2.1 在可视化中的布局设计问题

层次结构的布局设计在信息的组织结构中不仅有重要作用,而且还是一种用简单的形式对复杂的结构表述的一个重要方法。

层次可视化的研究粗略的分为两个主要的分支：连线方法 (the connection approach) 和圈地方法 (the enclosure approach)。它们对层次信息的可视化是两种有效的方法，每种方法的使用主要依靠在特殊应用领域的数据的属性。图 1 显示了一个分类层次结构的例子，它使用的是典型的连线方法。图 2 显示的是圈地方法。

连线方法使用节点链接图表来明确显示信息的关系,它能使用户一眼就能看出信息之间的关系,但是它对空间的利用效率不是很高,大部分的空间作为背景而被浪费。

圈地方法对优化显示空间是一个很好的解决方法。它通过一种新的填充空间的概念,把层次化信息绘制成一个二维的矩形空间,保证百分之百地使用显示空间。当前使用最广的 enclosure 技术有 Tree-Maps, Cushion Tree-Maps^[5], Squarified Tree-Maps^[6], Sun-

burst^[7], InterRing^[8] 和 Venn diagram^[9]。矩形嵌套图和它的变异方法在有限的空间里能显示大量的信息,它能用于具有大量信息的层次结构的可视化,在信息可视化领域有很大的作用,如:在股票市场数据、文件系统可视化、图像的浏览器等等中都获得了成功应用。虽然这些方法能可视化显示大量数据的层次化结构,但是,它们一般不能明确显示信息的相关关系,必须花费额外的认知努力来感知和理解它们的相关关系。

为了能继承这两种方法的优点,有一个折中的办法,就是把连线方法和圈地方法结合起来,这些技术包括 Space-Optimized Tree (SO-Tree) 和 FlexTree。SO-Tree 继承了 Johnson 和 Shneiderman 提出的填充空间 (space-filling concept) 的概念,在可视化大量的层次化信息时,最大化地利用了显示空间^[10-11]。从技术上来说,这种技术使用了多变性区域划分的递归方法来布置节点和子树。SO-Tree 还使用了节点链接图表,来加强对层次结构的理解。但是多边形对感知者来说不是一个好的形状。Song 等人提出了多倍焦距技术的 FlexTree^[12],它把填充空间的概念也引入节点链接图表中,加强了可量测性。这种技术通过减少在同等水平下节点的高度和空白空间来压缩树的宽度,增加了显示空间,但是这种压缩的布局通常不清晰也不美观。

2.2 视图导航问题

层次结构的导航是通过结构中存在的层次链接,从一个节点到达另一个节点。当层次结构非常大时,在有限的屏幕上很难全面地显示整个信息,但是通过导航和兴趣节点的详细视图,能使用户很容易访问上下文信息并能定位到自己所需要的信息。在庞大信息导航过程中,它能使用户在查看信息时感知自己所处的位置以及移动的方位。当通过信息空间进行导航时,它能辅助用户进一步决定自己下一步怎么走。

为了在大型层次结构中较容易地访问上下文和定位信息,许多优化的导航方案被开发出来以适应不同的几何布局技术。目前大部分的导航技术可以粗略分为 3 种方法:焦点 + 上下文 (focus + context), 变焦 + 过滤 (Zooming + filtering), 和增量探索视图 (in-

cremental exploration)^[13]。

焦点+上下文是一种视图方法,它给用户显示一个小的焦点区域的详细信息和整个内容的整体视图。典型的焦点+上下文技术有:鱼眼视图(Fisheye Views),聚焦显示视图(Polyfocal Display),双焦点视图(Bifocal Lens),透视墙视图(Perspective wall),双曲线视图(Hyperbolic Browser)等。

变焦+过滤视图是在显示中通过减少上下文数量来进行操作。首先通过一维或者多维的数值范围从数据中选出一个子集,形成记录表,然后根据记录表过滤信息,从而达到减少上下文的目的。常用的变焦+过滤技术有星域显示视图(Starfield Display),树图(Tree-Maps),Pad++或最近的版本 Piccolo 等。

增量探索视图是随着用户对信息空间的探索增加,增量地显示整个层次的一小部分。该技术能处理巨大的数据集,但在同一时间不可能在屏幕上显示完整的层次。

比较这3种方法,在导航时焦点+上下文技术对解决访问上下文信息有更大优势。它使用了放大+嵌入(enlarge+embedded)或者放大+混合(enlarge+blended)的概念(如图3),也就是使用变形或语义放大技术去放大信息结构的某一部分(用户关注的那部分),形成局部信息的详细视图。人们也把详细视图称为焦点视图(focus view)。然后,把这个局部的焦点视图嵌入到整个上下文信息结构视图中去。这个方法非常实际有效,在许多的可视化工具中被广泛使用并被商业化。但是它有一个明显的局限就是区域的分化。整个几何区域,视图被分成两个部分:一个是局部结构的图像,另一个是整个图像剩余部分的图像,因而限制了在焦点视图中显示的信息数量。一些焦点+上下文方法还在局部和整体视图间提供了一个过渡区域,以适应人的在鉴别局部和整体视图中的认知过程(如图3)。

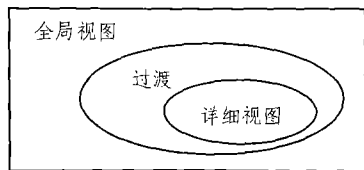


图3 导航视图

变焦+过滤视图和增量探索视图在导航过程中对访问上下文信息没有提供标准的策略。但是这些技术提供了特别的机制,如:路标节点和历史路径,在某种程度上去显示上下文信息。

3 大型层次化信息的可视化实例分析

现在我们以知识检索模型实例中双曲树可视化工具来分析一下大型层次化信息可视化设计的具体过程,如图4。

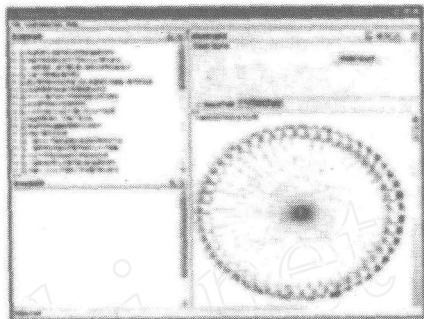


图4 双曲树可视化全局视图

要实现大型层次化数据的可视化须包括以下两点:根据用户的偏好和信息间的关系把信息结构自动排列在屏幕上;根据用户偏好平滑地实现人机交互,提供信息导航功能。在实例中我们围绕这两个目标实现层次信息结构的可视化,其关键技术包括布局、映射、提供动态视图。

3.1 布局

设计中,为了合理使用空间,双曲树节点布局间的连线应尽可能小,保证节点不重叠。排列算法从根节点开始按照从上到下顺序计算每个节点的位置。先把根节点的坐标设为双曲平面中心坐标,然后把根节点周围的扇形区域分配给根节点的子节点。这样每个第二层子节点(设根节点为第一层)都有自己的扇面,依次类推,有子节点的每个节点都把自己的子节点放到自己的扇面里,从而可以使用递归算法计算树中每个节点在双曲几何中的位置。

具体在双曲空间中节点的布局是通过 Klein 模型来实现的,如图5在QPR扇形中, r 表示树节点扇形区域的起始角度, d 表示扇形区域的大小。假设图中 P 有3个子节点(如果不是叶节点,这3个子节点是子树的父节点,排列算法关键在于确定 P 子节点(子树) P_1, P_2, P_3 的扇形区域。引出 PQ', PQ'' 为每个子节点分配扇形区域,节点 P_1, P_2, P_3 位于扇形区域中间,且距 P 相同的距离。为使子树间不重叠,过 P_1, P_2, P_3 画 PQ, PQ', PQ'' 的平行线,这些平行线确定了子树的子扇形区域的大小。因为在双曲空间中过直线外一点有多条直线和已知直线平行,所以可以找到一条合适的平行线使所得扇形角度尽可能地与父节点一样大。如果某个节点的扇形区域为 (r_p, d_p) ,并且有 n 个子节点,那么第 k 个子节点的扇形区域为:

$$(r, d) = (r_p + k \times d_p / n, d_p / n)$$

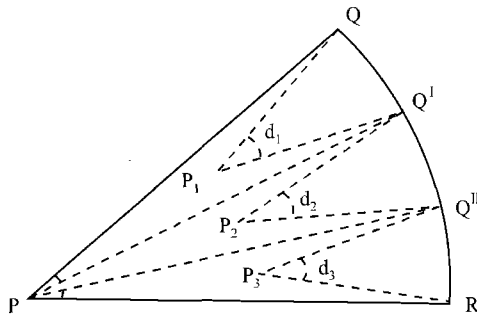


图5 双曲空间 Klein 模型布局排列算法

在双曲空间中,每个子节点得到的扇面都会尽可能与其父节点一样大,即通过排列算法计算,双曲空间中的节点不论层次大小,得到的双曲空间区域都可以一样大。所以双曲树的每一个节点都有足够的区域显示其详细信息。用户通过焦点变换可以获得任意节点的详细信息。

3.2 映射

双曲空间只是一个抽象的空间,人们直观上很难理解。一旦树结构在双曲平面上排列完毕,就必须按某种方式把它映射到二维平面中。更形象的说就是使用照相机从外部给整个双曲空间拍全景图。把整个无限大的双曲平面映射到二维平面上的投影方法可以很好地提供“焦点+上下文”视图。在双曲树中,通过采用焦点+上下文技术很好地实现了焦点变化时平滑过渡,能够帮助人们浏览大型层次信息且有效导航。

在双曲线可视化实现过程中,人们通过 Klein 投影技术,把双曲空间映射到二维平面中形成一个有限圆面。这个圆面就是用来显示层次信息、结构的区域。双曲空间焦点周围的区域在圆面上大面积显示,远离焦点的区域以远景图的形式慢慢“消失”在圆面边缘上。在圆面上,随着半径的增加,节点对象尺寸变得越来越小,而圆面边缘对应双曲空间上有无限的空间,可以存放大量节点对象,如图4。

在双曲空间中,通过布局为每个节点分配了相同大小的显示区域。任何一个节点被置为视图焦点时,此节点都会在圆面中有较大的显示区域,而远离新焦点的区域会消失在圆面的边缘上。

3.3 动态视图

在显示视图中,通过拖动鼠标浏览视图时,所有点的位置都会发生变化,那么节点位置移动的大小不仅取决于节点在层次信息结构中的位置,还取决于鼠标移动的起始点。在实例中,双曲方法在实现的时候,先把结构中所有节点在双曲空间中进行移动,然后映射到二维

平面上。因此首先要计算移动起始点在双曲空间中的位置,以便计算双曲空间中的节点的移动位置。

参考文献

- 1,2 周宁,张玉峰,张李义. 信息可视化与知识检索. 北京: 科学出版,2005
- 3 Johnson B, Shneiderman B. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. The 1991 IEEE Visualization 1991 (Piscataway, NJ, USA). IEEE Computer Society Press
- 4 Ware C. Information Visualization: Perception For Design. Morgan Kaufmann Series in Interactive Technologies, Morgan Kaufmann; San Francisco, CA, 2000, 438pp.
- 5 van Wijk JJ, van de Wetering H. Cushion treemaps: visualization of hierarchical information. IEEE Symposium on Information Visualization (InfoVis '99) 1999 (San Francisco, CA, USA). IEEE Computer Society Press; Silver Spring, MD, 1999:73-78.
- 6 Bruls M, Huizing K, van Wijk JJ. Squarified Treemaps. Joint Eurographics and IEEE TCVG Symposium on Visualization 2000 (Vienna, Austria). Springer; Berlin, 2000:33-42.
- 7 Stasko J, Zhang E. Focus + context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. IEEE Symposium on Information Visualization (InfoVis '00) 2000 (Salt Lake City, UT, USA). IEEE Press; New York, 2000:57-65.
- 8 Yang J, Ward MO, Rundensteiner EA. InterRing: an interactive tool for visually navigating and manipulating hierarchical structures. IEEE Symposium on Information Visualization (InfoVis '02) 2002 (Boston, MA, USA). IEEE Computer Society Press; Silver Spring, MD, 2002:77-84.
- 9 Herman I, Melancon G, Marshall MS. Graph visualization in information visualization: a survey. IEEE Transactions on Visualization and Computer Graphics 2000 (6):24-44.
- 10 Quang Vinh Nguyen, Mao Lin Huang. A Space-Optimized Tree Visualization. Proceedings of the IEEE Symposium on Information Visualization 2002 (InfoVis '02)
- 11 Chaomei Chen. Information Visualization: Beyond the Horizon. Springer, 2004
- 12 Song H, Curran EP, Sterritt R. Multiple foci visualization of large hierarchies with FlexTree. Information Visualization Journal 2004 (3):19-35.
- 13 Quang Vinh Nguyen, Mao Lin Huang. EncCon: an approach to constructing interactive visualization of large hierarchical data. Information Visualization 2005 (4):1-21

周宁 教授,武汉大学信息管理学院博士生导师。通讯地址:武汉。邮编 430072。

陈勇跃 武汉大学信息管理学院情报学博士研究生。通讯地址同上。

金大卫 武汉大学信息管理学院博士研究生,中南财经政法大学讲师。通讯地址同上。

(来稿时间:2006-12-12)