

# 一种面向层次和时序结构的多维可视化技术<sup>1</sup>

陆 伟 贺建根 王晓光 吴佳鑫

(武汉大学信息资源研究中心, 武汉 430072)

**摘要** 多维数据、层次数据和时序数据在许多应用和研究领域中都广泛存在, 并且实际中的数据经常同时具备这三种数据类型的结构特征。信息可视化是对数据进行分析的有力工具, 然而已有的可视化技术无法较好地分析复合结构特征的数据。针对具有层次和时序结构的多维数据可视化问题, 本文提出一种可视化技术, 该技术能够快速显示复合结构数据集的整体视图, 同时方便查看细节信息, 并提供各种交互手段进行可视分析。基于这种可视化技术, 开发 WHU-HTMVIS 可视化工具, 可以方便地构造层次时序多维数据可视化应用。对以上技术和工具所构造的应用实例的分析验证了该技术的有效性和易用性。

**关键词** 多维可视化 层次可视化 时序可视化 WHU-HTMVIS

## A Visualization Technique for Multidimensional Data with Hierarchical and Temporal Structure

Lu Wei, He Jianguo, Wang Xiaoguang, Wu Jiaxin

(Center for Studies of Information Resources of Wuhan University, Wuhan 430072)

**Abstract** Multidimensional data, hierarchical data and temporal data are widely used in many fields, and the actual data often simultaneously possesses structural properties of these three types of data. One of the prominent data analysis techniques is information visualization, but users tend to fail to analyze compound structure data with existing approaches. This paper presents a new visualization technique, trying to resolve the problems. This visualization technique provides good overviews for large-scale information and details for these focus objects. Meanwhile, it provides various ways of human-computer interaction to process visual analytics. It also presents a tool named WHU-HTMVIS for developing multidimensional data with hierarchical and temporal structure visualization application. A case study shows that the technique is efficient and easy to use.

**Keywords** multidimensional visualization, hierarchical visualization, temporal visualization, WHU-HTMVIS

## 1 引言

情报学的主要任务之一是要解决用户如何获取相关信息, 最终实现对信息认知的问题, 并且总是趋向简捷、方便、易用、省力地去解决这一问题。信息可视化是利用计算机支撑的、交互性的、对抽象数据的可视表示, 以增强人们对抽象信息认知的技术。信息可视化以各种更加符合人类认知模式的形式表达信息, 减少了信息与人类认知模式间的阻隔和认知负担, 使得信息的获取与理解变得简捷和方便。因而, 信息可视化技术在情报学界应用广泛, 并已渗透到情报学的各个领域, 如信息组织、信息检索、信息分析和信息服务等。

---

<sup>1</sup> 本文系教育部人文社会科学基地重大项目(项目编号:10JJD630014)和国家自然科学基金项目(项目编号:71173164 和 71003078)的研究成果之一。

由于信息的爆炸式增长和日益复杂化,信息可视化的任务也变得更加多样化。信息可视化的数据类型主要有一维数据、二维数据、三维数据、时序数据、多维数据、层次数据和网络数据七类<sup>[1]</sup>。在各类数据类型上发展出了各种可视化方法,通过这些可视化方法,增强了人们对海量数据的认知能力。但在实际运用中,数据的类型往往出现复杂化的情况,数据不再是单一的结构,而具有多重结构特征。例如,一个国家的进出口数据,既包括进出口货物类目这一多维数据类型,还包括进出口时间这一时序信息,同时货物类目是层次数据类型。面对这种复合类型的信息可视化问题,需要在已有可视化技术基础上开发新的技术。

## 2 相关研究

多维数据、层次数据和时序数据是三种常见而重要的数据类型,在信息可视化研究中存在大量关于这三种数据类型的可视化技术。多维可视化技术(Multidimensional Visualization)主要解决的问题是要反映多维信息及其各属性之间的关系,并在低维可视空间中展现多维抽象信息的多属性数据特征<sup>[2]</sup>,主要包括平行坐标系(Parallel Coordinates)<sup>[3]</sup>、Radviz (Radial Coordinate visualization)<sup>[4]</sup>、Andrews 曲线法<sup>[5]</sup>、Chernoff 面法(Chernoff Faces)<sup>[6]</sup>、星型坐标系(Star Coordinates)<sup>[7]</sup>等方法。层次可视化技术(Hierarchical Visualization)不仅能改善用户对层次结构数据及数据项之间关系的理解,同时也能够辅助信息的操纵<sup>[8]</sup>,它主要包括节点连接和空间填充两大类技术<sup>[8-10]</sup>。节点连接的代表有空间树(Space Tree)<sup>[11]</sup>、双曲树(Hyperbolic Tree)<sup>[12]</sup>、圆锥树(Cone Tree)<sup>[13]</sup>、径向树(Radial Tree)<sup>[14]</sup>、轴环柱型树(Collapsible Cylindrical Trees, CCT)<sup>[15]</sup>等;空间填充的代表有树图(Tree Maps)<sup>[16]</sup>、信息片(Information Slices)<sup>[17]</sup>、阶梯树(Step Tree)<sup>[18]</sup>、信息立方体(Information Cube)<sup>[19]</sup>等。时序数据是一种复杂类型的数据,它由随时间变化的序列值或事件组成。在可视化中必须专门考虑时间的作用,建立起与时间轴直接的、可视的关联<sup>[20]</sup>。时序可视化技术(Temporal Visualization)主要关注四个方面:时间点、时间间隔、时间关系和逻辑表达<sup>[21]</sup>。时序可视化的主要代表技术有主题河(ThemeRiver)<sup>[22]</sup>、LifeLines<sup>[20]</sup>等。

以上可视化技术增强了人们对多维数据、层次数据和时序数据的认知能力,辅助人类视觉感知系统摆脱三维空间定势的约束,帮助知识工作者理解和分析海量高维数据集,准确快速地发现数据集中隐藏的特征信息、关系信息、模式信息、趋势信息及聚类信息等。但这些技术多是针对单一的数据类型,无法对具有多重数据类型结构特征的数据进行可视化。本文试图在多维可视化技术的基础上,提出一种面向层次结构和时序结构的多维可视化技术。这种可视化技术能够克服传统多维可视化技术在层次和时序数据展示上的不足;而且可以适应在层次和时序结构特征任意缺失情况下的数据展示,因而该技术可以应用于多种数据结构类型。基于该技术,我们开发了一种称为 WHU-HTMVIS 的可视化工具,它可以方便地实现层次时序多维数据的可视化。

## 3 可视化技术

本文提出的层次时序多维可视化技术在改进 VIBE (Visualization By Example) 可视化技术<sup>[23, 24]</sup>的基础上,结合层次可视化和时序可视化技术,能有效的支持交互,提供多视图协同操作,能帮助用户从整体和细节层面增强对层次时序多维信息的认知。

### 3.1 图结构

层次时序多维可视化图  $G = (C, R, H, T, M)$  用以展示具备时序结构和层次结构特征的多维结构数据,其中  $C$  表示维度参考点,  $R$  表示数据记录节点,  $H$  表示层次结构关系,  $T$  表示时序关系,  $M$  表示多维关系。图 1 是层次结构图  $G_H = (C, H)$ 、多维信息可视化图  $G_M = (C \cup R, M)$  和时间序列可视化图  $G_T = (R, T)$  的结合。

$C$  和  $R$  是相互之间没有交集的有限点集。 $C$  点集的布局位置和  $M$  决定了  $R$  在图  $G_M$  中的布局。本文的层次时序多维可视化图结构如图 1 所示。

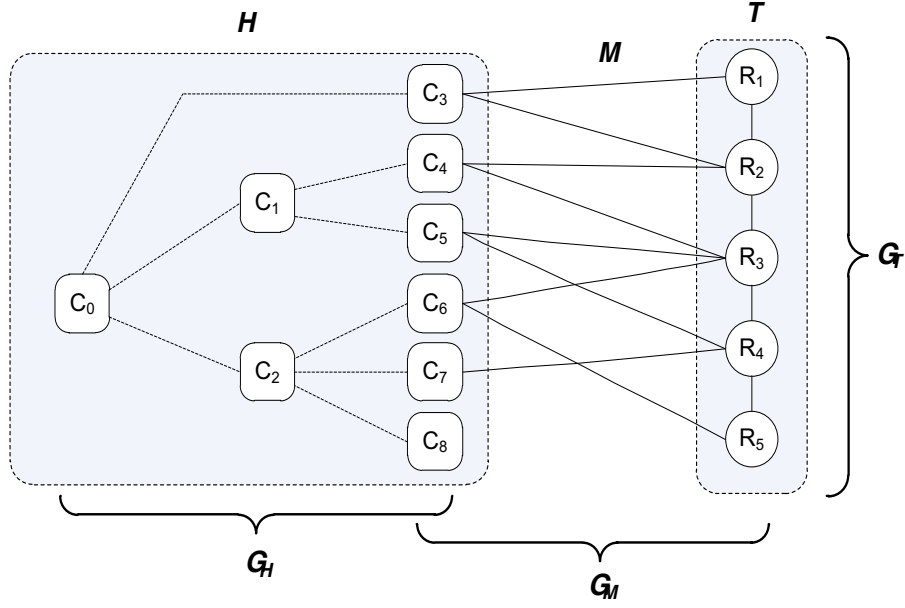


图 1 图结构

### 3.2 基于 VIBE 的多维可视化

将维度参考点描述为一组特殊的多维数据  $C_j(c_1, c_2, \dots, c_i, \dots, c_p)(j = 1, \dots, p)$ ，其中  $c_j = 1$ ， $c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_p = 0$ ，其中  $c_i$  表示第  $i$  个维度的信息， $p$  表示维度的个数，参考点  $C_j$  表示的实际含义为只存在第  $j$  维信息的情况，因而在可视化图中可以通过与各维度参考点的位置关系来感知数据记录  $R_s$  的多维信息。数据记录  $R_s$  描述为一组多维数据  $R_s(h_1, h_2, \dots, h_i, \dots, h_p)$ ，其中  $h_i$  表示数据记录  $R_s$  第  $i$  种维度的信息。本文将  $p$  个参考点  $C_1, \dots, C_p$  在视觉空间中映射为一个正  $p$  多边形，并以此为基础计算其他数据记录  $R_s$  在视觉空间中的坐标位置，参考点  $C_i$  的位置用  $V_i(x_i, y_i)$  表示，参考点的位置计算参见公式(1)：

$$P(C_i) = V_i(\cos(\frac{2\pi i}{p}) \cdot radius, \sin(\frac{2\pi i}{p}) \cdot radius) \quad (1)$$

数据记录  $R_s$  的位置由它和各维度参考点的相关性决定。数据记录与维度参考点的相关程度表示为该数据记录参考点向量 (Data Record Reference Point Vector)  $DDRRV(R_s) = (r_1, r_2, \dots, r_i, \dots, r_p)$ ，其中  $r_i$  是数据记录  $R_s$  与维度参考点  $C_i$  的相关性值。本文采用余弦相似模型计算相关性值，由于维度参考点  $C_i$  多维数据的特殊性，其计算过程也可有所简化，具体参见公式(2)：

$$r_i = \frac{\sum_{i=1}^p c_i \times h_i}{\left( \sum_{i=1}^p c_i^2 \times \sum_{i=1}^p h_i^2 \right)^{1/2}} = \frac{h_i}{\left( \sum_{i=1}^p h_i^2 \right)^{1/2}} \quad (2)$$

根据维度参考点的坐标位置和数据记录的  $DDRRV$ ，即可计算数据记录的节点坐标位置，维度参考点  $C_i$  坐标位置  $P(C_i) = V_i(x_i, y_i)$ ，则数据记录  $R_s$  节点坐标位置计算方法如公式(3)所示：

$$P(R_s) = V_s \left( \frac{\sum_{i=1}^p r_i x_i}{\sum_{i=1}^p r_i}, \frac{\sum_{i=1}^p r_i y_i}{\sum_{i=1}^p r_i} \right) \quad (3)$$

### 3.3 基于多线和动画的时序可视化

在多维可视化图中，单独的一个点不能表达时间序列信息。依据数据节点的时间序列关系和其坐标位置绘制曲线，即将数据节点按时间序列连接起来。本文采用三次样条插值法<sup>[25]</sup>来构造曲线，用一条平滑曲线来对各数据记录节点进行拟合，通过构造一组不同阶多项式，来形成一条把所有数据记录节点连接起来的平滑曲线，以清晰表达数据记录间的时序关系。为了区分各时序记录（时序记录包含多个数据记录节点），曲线颜色由计算机随机分配。数据节点的颜色代表时间信息，由蓝到黄再到红表示时间的先后关系，效果如图 2 所示。

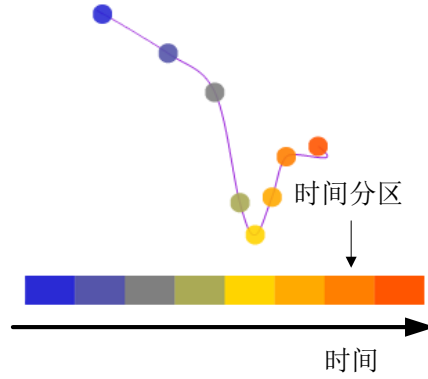


图 2 数据节点时序曲线

可视化图中的节点颜色，除了可以表达节点的时序信息的模式外（如图 2 所示），还有可以表达维度信息的颜色模式。如图 3 左上部分所示，每个数据记录节点（圆形节点）的颜色根据数据记录的维度信息和维度参考点（方形节点）的颜色确定，这样节点的颜色代表了一定的维度信息，节点颜色的计算公式如下：

$$color(R_k) = \frac{\sum_{i=1}^p r_i \times color(C_i)}{\sum_{j=1}^p r_j} \quad (4)$$

由于随数据量的增长，节点会有重叠的现象，不利于用户对信息的认知。鉴于此，本文利用基于多线的时序数据可视化技术（Polyline-based Time-varying Data Visualization）<sup>[26]</sup>增加了一个整体信息视图，有利于用户对信息的整体认知，发现信息趋势和异常。该视图将每条完整的时序记录集合的每个节点的维度信息颜色按时间等分在一条直线上，中间用渐变色过渡。每一条直线代表一条完整时间序列记录的多维信息，从左到右表示时间的先后关系，效果如图 3 左下部分所示。将所有代表时间序列数据记录的直线纵向排列到一起，组成了可视化图的时序整体视图，如图 3 右所示。

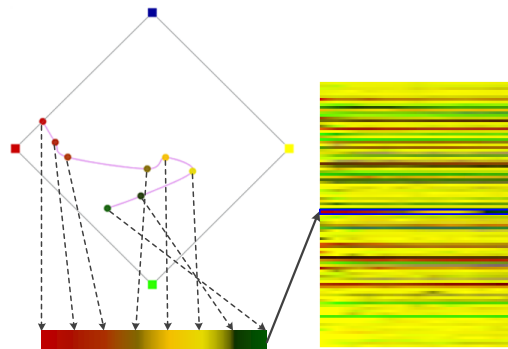


图 3 时序整体视图生成过程

动画是时序数据分析和时序数据趋势呈现的有力工具。在 WHU-HTMVIS 工具中，支持对时序多维数据的动画交互，并支持时序范围选择、暂停等交互操作。图 4 为 100 条时序记录动画的八个时间截面，节点颜色为时序信息模式。

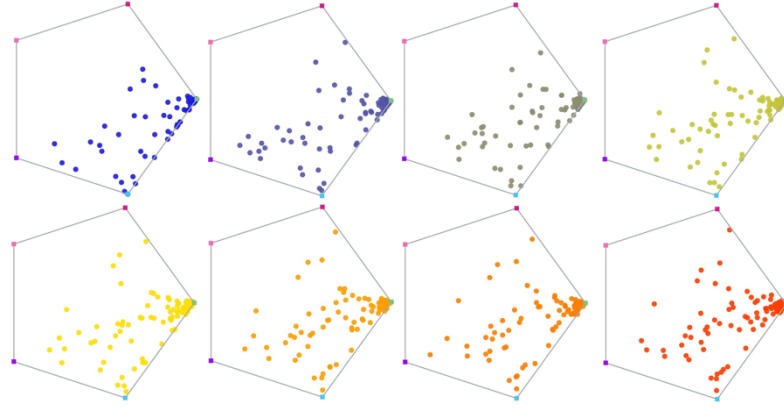


图 4 动画时序截图

### 3.4 基于交互和信息片的层次可视化

在可视化图中的维度参考点添加交互，将数据中的层次信息加以表达。当单击的参考点有下级参考点的情况下，重新调整整个可视化图中的参考点布局，并根据公式(3)重新计算所有数据记录的位置。参考点  $C_j(j=1, \dots, p)$  下有  $p'$  个下级参考点  $C_i'(j=1, \dots, p')$ ，参考点  $C_i'(j=1, \dots, p')$  的位置计算公式为：

$$P(C_i') = V_j \left( \cos\left(\frac{2\pi j - \pi}{p} + \frac{2\pi i}{p(p'+1)}\right) \cdot radius, \sin\left(\frac{2\pi j - \pi}{p} + \frac{2\pi i}{p(p'+1)}\right) \cdot radius \right) \quad (5)$$

例如，有  $\{A, B, C\}$  三个维度的多维空间，其中维度  $A$  包括  $\{A_1, A_2, A_3\}$  二级维度，时序记录  $R$  包含时间序列  $\{T_1, T_2, T_3\}$  内在  $\{A, B, C\}$  维度空间中的信息。在未展开  $A$  维度下的二级维度信息前，从图 5 中可以看到时序记录  $R$  在  $\{A, B, C\}$  维度空间中的时序演化过程。时序记录  $R$  在  $T_1$  时刻处于  $B, C$  两维度参考点之间，到  $T_2$  和  $T_3$  时刻演化到  $A$  维度参考点附近。 $T_2$  和  $T_3$  时刻时序记录  $R$  的位置说明，在  $T_2$  和  $T_3$  时刻  $R$  在  $\{A, B, C\}$  维度空间中维度信息差异不大。点击维度节点  $A$  之后，图中显示了时序记录  $R$  从  $T_2$  到  $T_3$  演化的更多细节，在新的维度空间中  $R$  在  $T_2$  时刻与  $T_3$  时刻的维度信息差异较大。加入这种交互方式后，将层级结构的多维信息加以可视化表达，方便可视分析和挖掘隐藏信息。

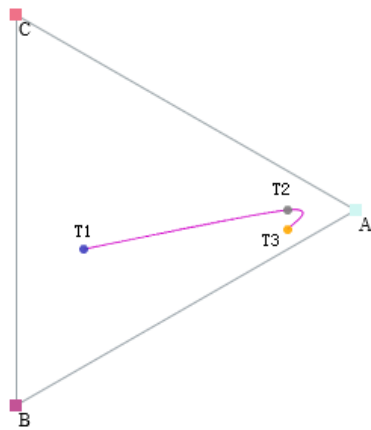


图 5 点击维度节点  $A$  之前

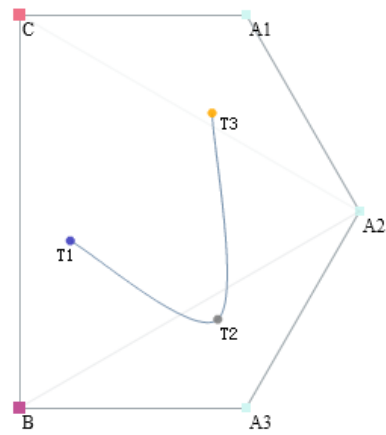


图 6 点击维度节点  $A$  之后

除了通过交互方式在多维视图中表达可视化层次信息外，对于细节层次信息可视化本文采用信息片技术实现，将具有层次结构的细节信息以信息片的方式显示。当用户点击一条时序记录集合后，显示用





为检索式，以 1994 年-2010 年为时间范围，采取“精确匹配”模式。每年每个关键词都检索一次文献数据库（万方数据知识服务平台），所以共检索数据库约 1700 次，每次检索都将其文献检索结果在个学科分类的发文量信息提取记录下来。采集结果共有 35267 条记录。

### 5.2 数据结构

数据采用 XML 格式存储（参见图 10），再经可视化程序读取解析 XML 生成可视化图。

```
<Keyword keyword="图书馆">
  <year year="1997" number="0"/>
  <year year="1998" number="787">
    <record discipline="文化、科学、教育、体育" number="707">
      <record subDiscipline="信息与知识传播" number="677"/>
      <record subDiscipline="科学、科学研究" number="12"/>
    </record>
  </year>
</Keyword>
```

图 10 数据存储格式

### 5.3 实验结果

图 11 为是 WHU-HTMVIS 工具对 1994 年-2010 年我国图书馆学情报学领域前 100 个核心关键词随时间在五个学科分类间的演化过程的可视化。从可视化图中可以看到图情领域的核心关键词在学科空间中的演化发展过程以及演变趋势。该图中选取了前五个学科分类维度，包括“文化科学教育、工业技术、经济、医学卫生、政治法律”五个学科分类。图情领域研究主题主要集中在“文化科学教育”学科分类，同时“文化科学教育”分别与“工业技术”、“经济”、“医学卫生”之间的学科空间中存在 A、B、C 三个节点聚集的条状区域，这说明我国图情领域的研究内容在“文化科学教育”与“工业技术”，“文化科学教育”与“经济”，“文化科学教育”与“医学卫生”几大学科分类间存在广泛的跨学科研究。

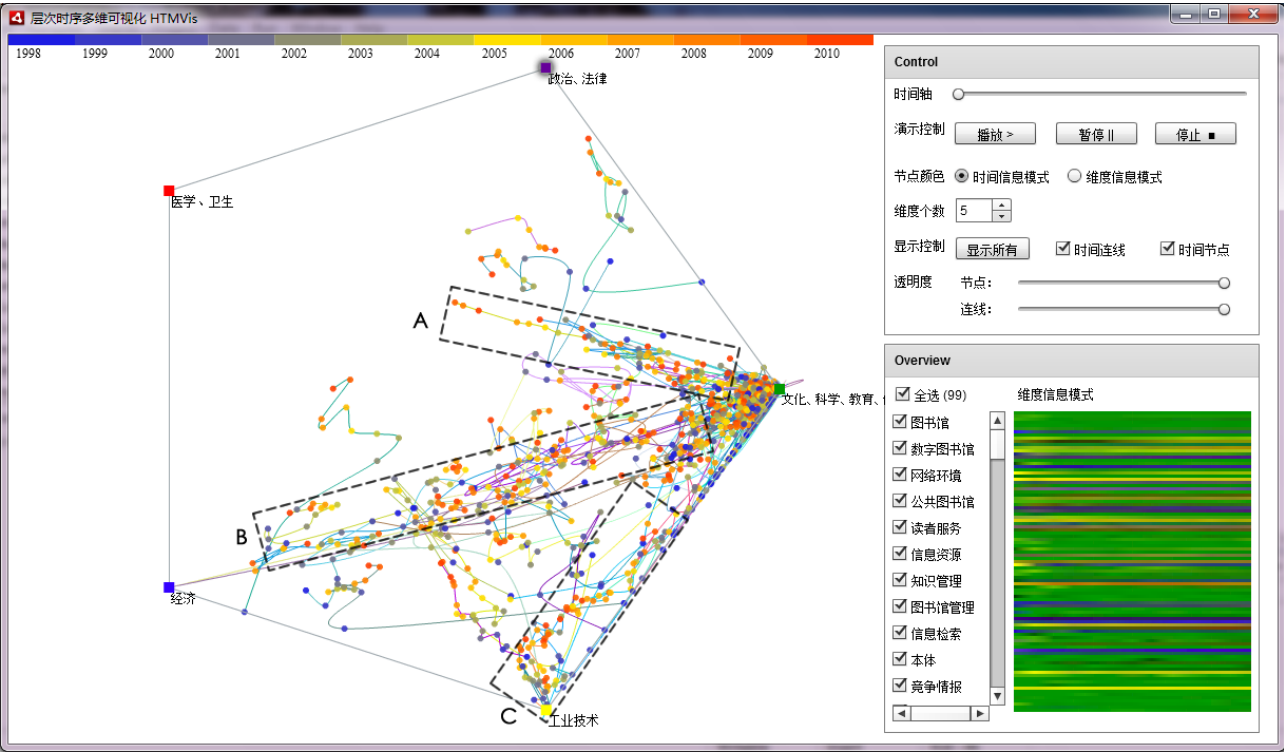


图 11 WHU-HTMVIS 软件界面

图 12 为通过交互方式选取了其中几个关键词的数据，每条曲线代表了相应关键词在学科空间中的发展过程。在可视化图中有几种不同的发展模式，稳定型关键词在学科空间中发展比较稳定，如公共图书馆和竞争情报；不稳定型关键词在学科空间中发展变化一直较大，如知识产权和著作权；后期稳定型关键词在经过一定发展后呈现较稳定的态势，如本体、信息化和 Web 2.0 等；稳定发展型关键词一直处于发展变化中，但发展较为稳定且有规律，如信息系统、文献和 WTO 等。

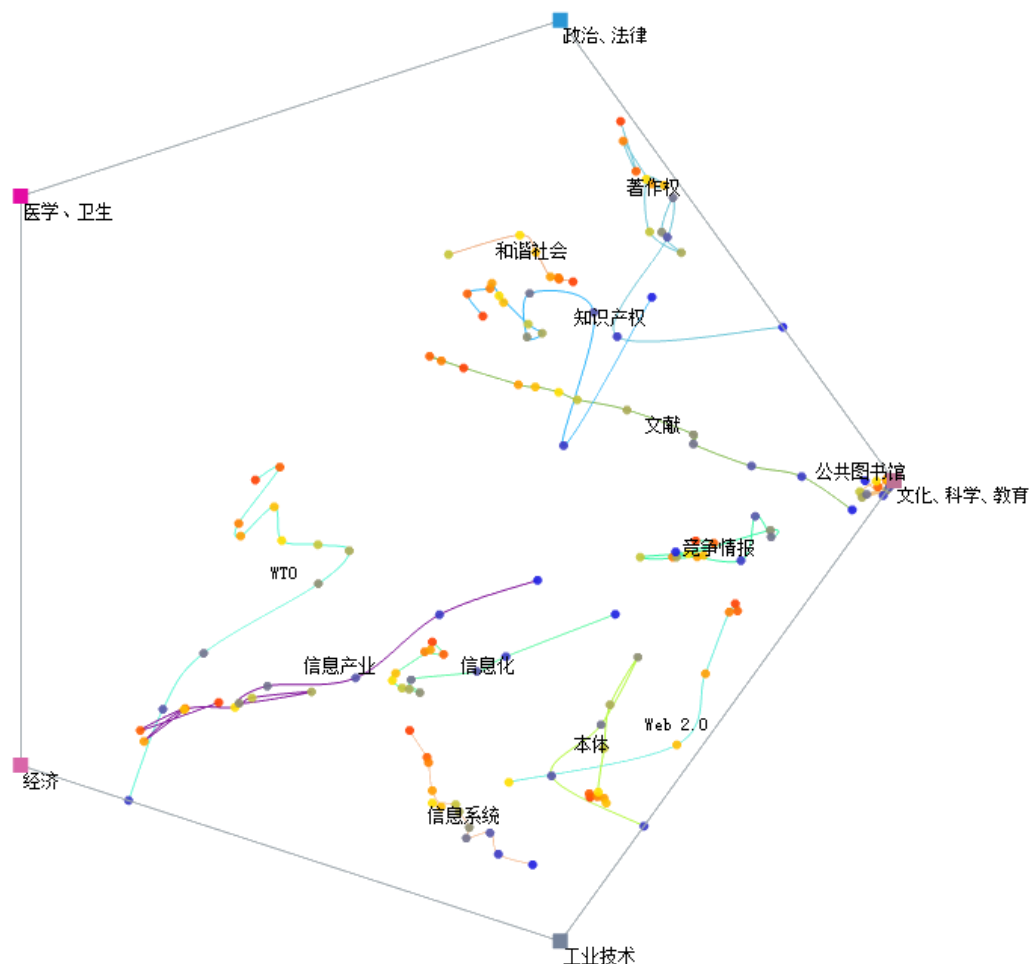


图 12 部分关键词演化曲线

通过 WHU-HTMVIS 可视化工具的交互操作，本文选取了活动在“文化科学教育”学科分类附近的“数字图书馆”、“网络环境”等 59 个关键词，如图 13 所示。这 59 个关键词大多为我国图情领域的传统研究主题和比较核心的研究主题，这些研究主题在图情领域的发文量比例大，在学科空间中呈现比较稳定的演变状态。该可视化图也说明我国图情领域的传统研究主题在大学科层面上跨学科性不强，研究主题发展的学科来源较为单一。图 13 上的节点重合度较高，各关键词在时间序列上学科空间中的演变程度小，但在大学科空间中的重合和稳定并不意味着在二级学科空间中有同样的情况。在图 14 中在保留大学科空间的背景下，引入二级学科空间，展示的各关键词在“文化科学教育”之下的二级学科空间中的演化过程。在图 14 中“科学、科学研究”学科分类附近存在大量纯蓝色的节点（初始年份节点），这说明我国图情领域有些研究主题是从“科学、科学研究”领域起步，后逐渐转移到“信息与知识传播”领域，这种变化在一定程度上揭示了图情学科的定位与研究范式的演化。



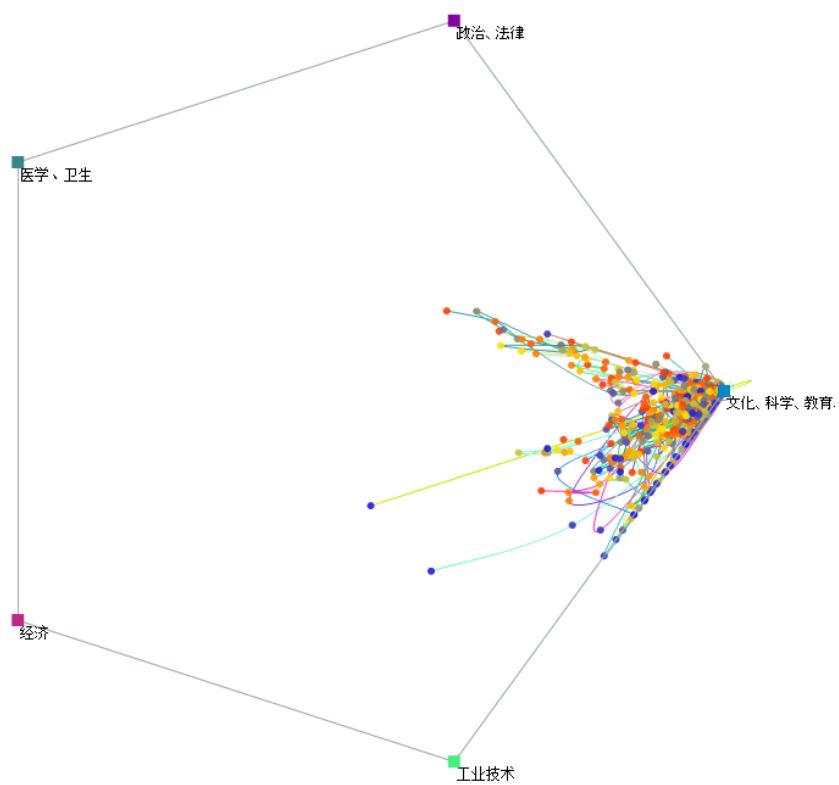


图 13 “文化科学教育”附近关键词

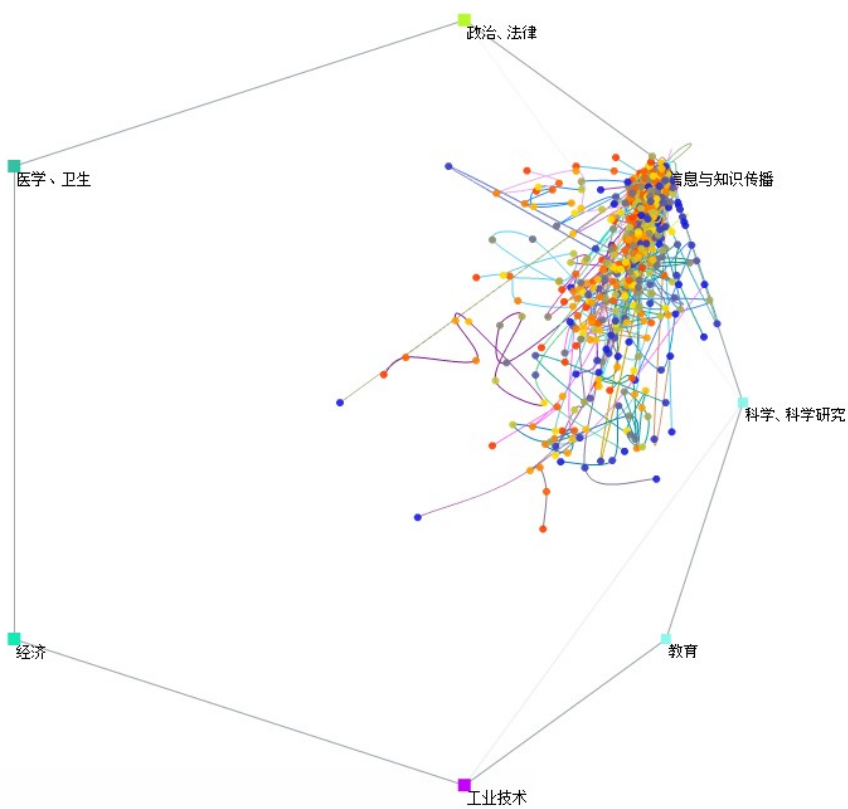


图 14 二级学科空间

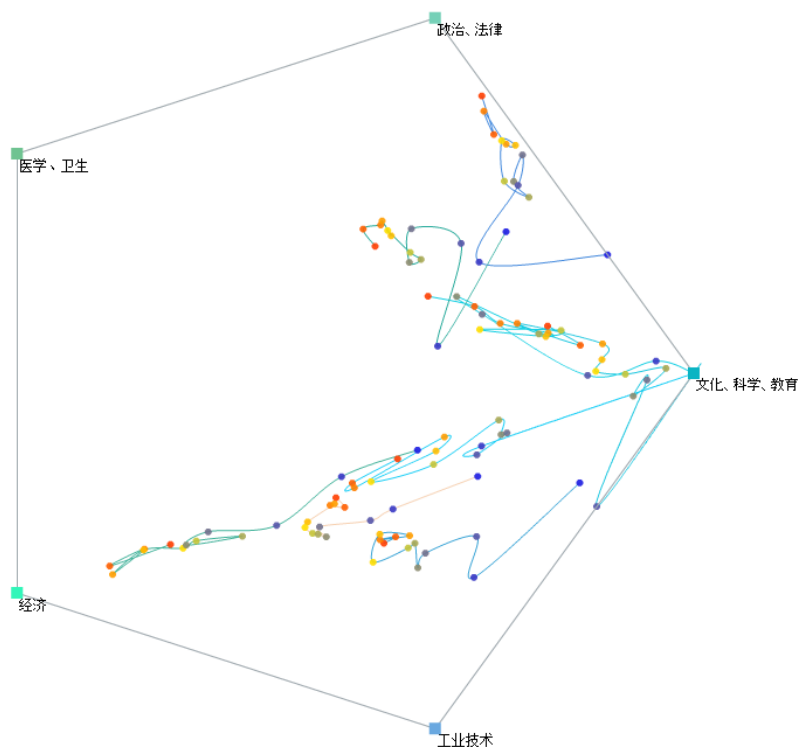


图 15 “文化科学教育” 输出关键词

此外，我们还选取了“知识产权”、“信息化”、“著作权”、“信息管理”、“文献计量学”、“医学图书馆”、“信息产业”、“知识创新”等 8 个关键词来分析图情学科研究主题随时间向其它学科分类扩散的情况，如图 15 所示。这 8 个关键词几乎都存在从“文化科学教育”学科分类逐渐向“经济”、“医学卫生”和“政治法律”三个学科分类转移的过程。这种趋势说明图情领域的部分研究主题已经走出了自身的小圈子，在其它研究领域得到了传播或应用，但更深层次的解释还有赖于后续的研究。

## 6 结论

本文针对现有可视化技术在复杂结构数据展示上的不足，提出了一种新的多维可视化技术。根据数据的多维属性，使用 VIBE 方法将数据映射至二维空间；利用基于多线的时序信息可视化技术和计算机动画技术实现了时序信息的可视化；利用人机交互和信息片技术实现对层次信息的可视化；多个可视化视图通过人机交互协同操作。基于该技术，我们开发了一款可视化软件 WHU-HTMVIS，通过对图情领域的关键词数据的实际应用，验证了该技术的有效性和易用性。目前，该技术还存在一些不足，在节点、曲线等可视化元素优化布局上，尤其是显示元素重叠上还存在一定问题。下一步，除了改进布局算法外，我们还将进一步完善人机交互界面，提高 WHU-HTMVIS 软件的易用性。

## 参 考 文 献

- [1] Ben S. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. 1996.
- [2] 孙扬, 封孝生, 唐九阳, 等. 多维可视化技术综述. 计算机科学. 2008(11): 1-7.
- [3] Inselberg A. The plane with parallel coordinates. The Visual Computer. 1985, 1(2): 69.
- [4] Hoffman P, Grinstein G, Marx K, et al. DNA visual and analytic data mining. Phoenix, Arizona, United States: IEEE Computer Society Press, 1997.

- [5] Andrews D F. Plots of High-Dimensional Data. *Biometrics*. 1972, 28(1): 125-136.
- [6] Chernoff H. The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association*. 1973, 68(342): 361-368.
- [7] Kandogan E. Star Coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. 2000.
- [8] 肖卫东, 孙扬, 赵翔, 等. 层次信息可视化技术研究综述. *小型微型计算机系统*. 2011(1): 137-146.
- [9] 王威信, 明春英, 王宏安, 等. 基于Venn图的层次信息可视化. *计算机学报*. 2007(9): 1632-1637.
- [10] 景民昌, 孙洁丽. 大型层次信息可视化技术研究. *情报科学*. 2008(4): 541-545.
- [11] Plaisant C, Grosjean J, Bederson B B. SpaceTree: Supporting Exploration in Large Node Link Tree, Design Evolution and Empirical Evaluation. *IEEE Computer Society*, 2002.
- [12] Lamping J, Rao R, Pirolli P. A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. Denver, Colorado, United States: ACM Press/Addison-Wesley Publishing Co., 1995.
- [13] Robertson G G, Mackinlay J D, Card S K. Cone Trees: animated 3D visualizations of hierarchical information. New Orleans, Louisiana, United States: ACM, 1991.
- [14] Yee K, Fisher D, Dhamija R, et al. Animated Exploration of Dynamic Graphs with Radial Layout. *IEEE Computer Society*, 2001.
- [15] Dachsel R, J, Ebert R. Collapsible Cylindrical Trees: A Fast Hierarchical Navigation Technique. *IEEE Computer Society*, 2001.
- [16] Shneiderman B. Tree visualization with Tree-maps: A 2-d space-filling approach. *ACM Transactions on Graphics*. 1991, 11: 92-99.
- [17] Andrews K, Heidegger H. Information Slices: Visualising and Exploring Large Hierarchies using Cascading, Semi-Circular Discs.
- [18] Bladh T, Carr D, Scholl J. Extending Tree-Maps to Three Dimensions: A Comparative Study. 见: *Computer Human Interaction*. Springer Berlin / Heidelberg, 2004: 3101, 50.
- [19] Rekimoto J, Green M. The Information Cube: Using Transparency in 3D Information Visualization. 1993.
- [20] Plaisant C, Milash B, Rose A, et al. LifeLines: visualizing personal histories. Vancouver, British Columbia, Canada: ACM, 1996.
- [21] Chittaro L, Combi C. Visual Representation of Temporal Intervals and Relations: Information Visualization Aspects and their Evaluation. *Society Press*, 2001.
- [22] Havre S, Hetzler B, Nowell L. ThemeRiver: Visualizing Theme Changes over Time. Washington, DC, USA: *IEEE Computer Society*, 2000.
- [23] Olsen K A, Korfhage R R, Sochats K M, et al. Visualization of a document collection: The vibe system. *Information Processing & Management*. 1993, 29(1): 69-81.
- [24] Zhang J. Visualization for Information Retrieval. 见: *Encyclopedia of Database Systems*. Springer US, 2009, 3410.
- [25] 陈文略, 王子羊. 三次样条插值在工程拟合中的应用. *华中师范大学学报(自然科学版)*. 2004(4): 418-422.
- [26] Imoto M, Itoh T. A 3D Visualization Technique for Large Scale Time-Varying Data. 2010.
- [27] Cockburn A, Karlson A, Bederson B B. A review of overview+detail, zooming, and focus+context interfaces. *ACM COMPUT. SURV*. 2008: 1-31.