

平行坐标综述

赖楚凡 袁晓如

摘 要:

关键词:

Parallel Coordinates

Chufan Lai Xiaoru Yuan

Abstract:

Keywords:

1 引言

在现实世界中，同一事物往往有多个不同的属性，要完整地描述该事物，需要记录其各个属性的信息。而这类具有多属性（维度）信息的数据，就称为多变量数据（Multivariate Data），或称高维数据（High-dimensional Data）。在传统的数据分析中，高维数据一般利用统计方法加以处理，并通过表格的方式呈现。但传统方法过于依赖自动算法，既无法结合用户的知识与判断，也不利于对抽象的高维信息的认知和理解。而信息可视化技术（Information Visualization）通过对高维数据的视觉呈现，辅以交互式的数据处理手段，能使用户充分参与到数据分析过程中，既加深了用户对数据的理解，也提高了分析的效率和可靠性。

高维数据一直是信息可视化领域的研究重点之一，其独有的两个特性：即维度数高、抽象性强，为可视化设计带来了巨大挑战。一方面，维度数的增长使得数据的信息量急剧增大，有限的显示空间只能表达一部分数据特征，而自动算法也会因为所谓的“维数灾难”（过高的复杂度、过低的采样率等 [5]）而难以应用。另一方面，由于人无法感知高于三维的拓扑空间，高维的数据特征（如聚类、流型等）往往抽象而难以理解。

针对这些挑战，领域内提出了各类高维可视化方法[24]，包括降维投影[20]、散点图矩阵[10]、雷达图[32]、星型坐标[41]等等。而由Inselberg等人提出的平行坐标 [34]（如图1）也因其简单直观、可拓展性强等优点而被广泛应用于高维数据的可视分析中。

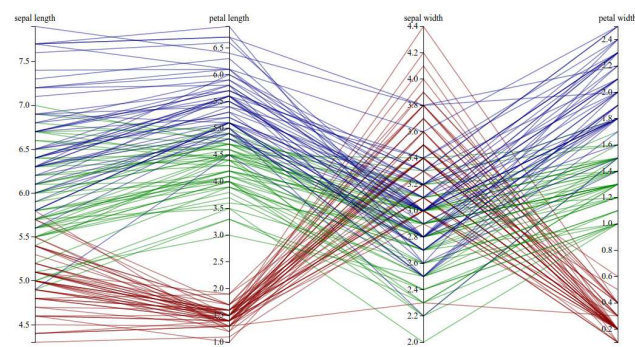


图 1: 平行坐标示例

然而，相比于其他高维可视化形式，平行坐标的设计也存在种种缺点，譬如展现的信息量较少、容易出现视觉混杂（Visual Clutter）、数据特征不直观等等（详见第3节）。针对这些问题，领域内的学者们做了大量的研究工作，并提出了各种方法对平行坐标进行改进和延展。

本文基于一些已有的综述论文 [24] [28] [54] [60]，并结合近年来的最新进展，将对平行坐标方面的研究工作进行比较、总结和展望。在文章

的余下部分里：第2节介绍了平行坐标的基本原理，并阐述其设计上存在的主要问题；第3节就平行坐标的不同部分，回顾了领域学者们为解决前述问题所作出的研究努力；第??节则介绍了平行坐标在不同领域内的应用现状；第4节基于已有的研究和应用，探讨了仍然存在的挑战，以展望未来的研究动向和应用前景；最后，第5节对全文内容进行总结。

2 基本原理及主要问题

在这一节里，我们将简要介绍平行坐标的设计，并逐步阐述其存在的几个主要问题。作进一步的探讨前，这里先简要介绍Jarke提出的的可视分析模型 [57]（如图2）。在一个可视分析流程中，数据（Data）驱动了可视化展示（Visualization），用户对视图进行感知（Perception）并从中获取知识（Knowledge）。基于已有知识，用户可以通过交互探索（Exploration）来改变视图的参数设置（Specification），以产生新的可视化来发掘更多知识。由此可见，可视分析流程主要由四部分组成，分别是数据、视图、视觉感知和交互。本节余下内容以及第3节的内容都将按这四个部分展开。

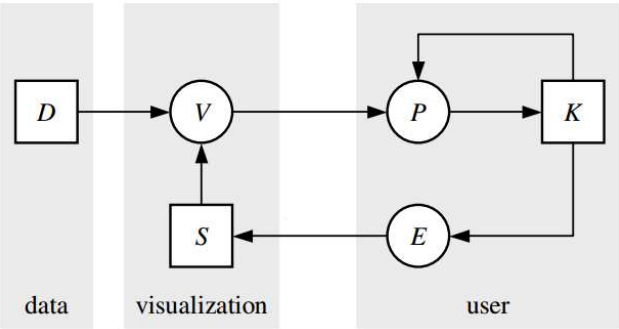


图 2: 可视分析模型。本图修改自Jarke 等人的论文 [57]。

2.1 数据

假设一个数据集有 m 个样本，每个样本有 n 个维度，则它可以记成一个 $m \times n$ 的矩阵，其中每一行是一个样本，每一列代表一个维度。这样的矩阵又称为数据矩阵（data matrix），是平行坐

标所处理的典型高维数据。数据的维度大致分为两种，即数值型维度（Continuous Dimension，具有无数种取值，如身高、体重）和类别型维度（Categorical Dimension，只有有限的取值，如国籍）。而一般的坐标轴只适用于数值型维度，类别型数据需要转换成数值型来表示，但往往会造成一定的误导（详见 2.2 节），这使得平行坐标不适合表现类别型数据。其次，高维数据的产生往往伴随着其他类型的信息，譬如高维时空数据、高维网络数据、高维流场数据等等。如何拓展平行坐标以表现这些混合高维数据，也是研究的重点之一。除了不同的数据类型，数据的隐私性、不确定性，对大数据的处理等等，都是在数据方面的研究会涉及的课题。

2.2 视图

正如第1节所提到，人无法感知高于三维的拓扑关系。而传统的笛卡尔坐标系限定了各坐标轴必须相互垂直，这使得高维的坐标空间极为抽象和难以理解。平行坐标解决这一难题的方法，顾名思义，就是将各个维度的坐标轴平行放置，如图3所示。一个高维数据在各个坐标轴上均有取值，连接各取值点所形成的一条折线，即表达了这个数据。事实上，基于线和点的转化关系，平行坐标和笛卡尔坐标之间有很好的几何对称性，有兴趣的读者可参阅相关文献 [34] [35]。

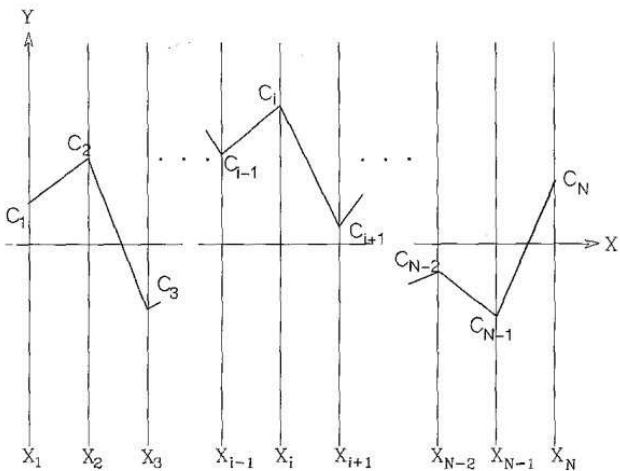


图 3: 平行坐标的构造原理。本图来源于Inselberg 等人的论文 [34]。

和其他高维可视化方法相比，平行坐标有几个突出的优点。首先，由于坐标轴的数量只受限

于屏幕的横向分辨率，它相比于散点图矩阵等可以容易容纳更多维度，具有很强的可扩展性。其次，它完整地表达了数据在每个维度上的取值，相比于降维投影、雷达图等具有较低的不确定性和信息损失。最后，平行坐标构造简单、无需自动算法辅助，既方便使用也容易理解。

但相对地，平行坐标也存在一些明显的局限和缺点：

其一，当一个维度上有多个相同的取值时，多条折线汇合到一个点，直接观察将无法区分不同的数据走向。而在类别型的数据中，类似问题出现于每一个轴，导致大量线段完全重叠，用户甚至无法辨别其中的数据量（见图4）。这正是平行坐标不适合表现类别型数据的原因。

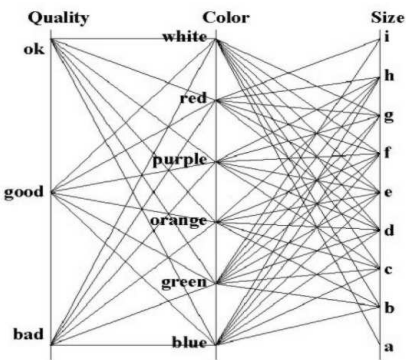


图 4: 类别型数据中的线段重合问题。本图来源于Rosario 等人的论文 [51]。

其二，每个数据用一条折线来表示，这会比散点形式占用更多的显示空间。当数据量增大时，视图很容易发生图元堆叠、导致视觉混杂（如图 5），用户将无法从中得到有用的信息。由此可见，平行坐标对数据规模的扩展性较差。

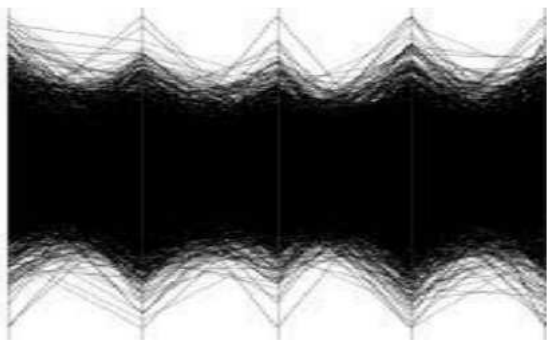


图 5: 视觉混杂问题。本图来源于Artero 等人的论文 [2]。

其三，平行坐标能够表现相邻维度间的关系，但信息量偏低。假设数据有 n 个维度，它们

两两组合可得到 $\frac{n(n-1)}{2}$ 个二维子空间，散点图矩阵能容纳其中的所有信息。但在平行坐标里，用户无法观察两个不相邻维度间的关系,因此只能看到 $(n-1)$ 个二维子空间。换言之，平行坐标只表现了 $\frac{2}{n}$ 的维度关系，不恰当的坐标轴排序很容易掩埋重要的信息（如图6）。

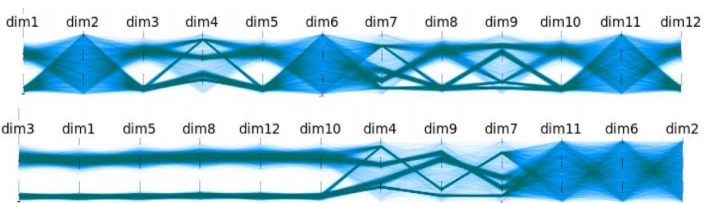


图 6: 不同坐标轴顺序对平行坐标的影响。其中上、下图分别为重排序前、后的视图，可见排序后各个维度的相关性和差异性更为明显。本图修改自Ferdosi 等人的论文 [19]。

最后，除了排序上的变化，坐标轴的放置还存在许多可能性。平行坐标的出现启发了大量对非垂直坐标系、甚至是非直线坐标轴的研究，这些都将在第 3 节中加以讨论。

2.3 感知

针对平行坐标的感知研究主要有两种，一是视觉感知的有效性评估，二是数据特征的刻画和提取。

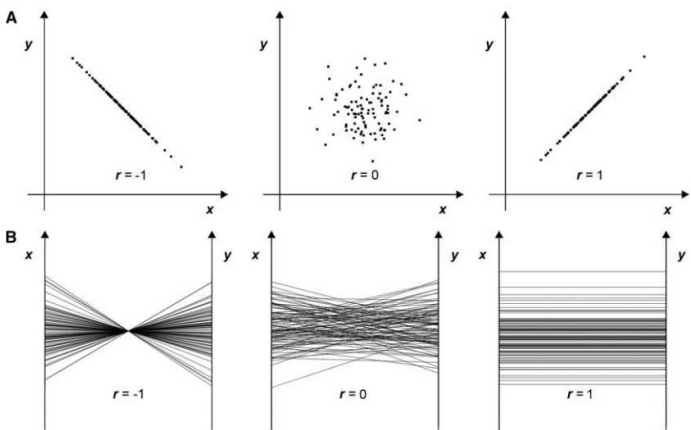


图 7: 维度相关性的解读。A、B两行分别显示了不同的相关关系下数据在散点图和平行坐标中的分布。本图来源于Li 等人的论文 [44]。

在平行坐标里，能观察到的数据特征主要有两种，即聚类信息和相关关系。一方面，相似的折线形成线簇，代表着不同的聚类，反映了数据

趋势、异常点等行为。但要观察到聚类，必须先处理好前面提到的视觉混杂问题，使隐含的聚类能被凸显出来。另一方面，相邻坐标轴之间的线段走向能反映这两个维度的相关关系（如图7）。但用户对相关性的感知是否准确，和其他方法相比是否更有效等等，都需要通过实验来评估。除此之外，还有很多有意义的数据特征无法通过简单的聚类或相关来描述。而已有的特征也需要相应的自动提取算法，来遍寻和评估不同的参数设置（如不同的维度排序），以找到特征明显、信息量大的视图。

2.4 交互

基于已有的观察，用户通过交互来改变视图，以发掘更多信息。平行坐标提供两种基本的交互功能，即数据刷选和坐标轴排序。对于刷选，用户需要给每个坐标轴限定取值范围[46]。这相当于在高维空间中建立一个超立方体（Hypercube）包围盒，同时满足各维度条件的数据才会被选中。这样的方式虽然逻辑简单，但操作繁琐、缺乏灵活性，用户既无法选中任意形状的数据区域，也无法根据其他数据特征来选择数据。除了刷选，用户还可以手动调整坐标轴的顺序。但这在数据维度较高时并不现实，用户很难通过交互来发现合适的排序。

3 平行坐标的相关研究

针对前一节中提出的各种问题，领域学者们做了许多研究工作来对平行坐标进行改进和延展。这一节将详细地回顾这些工作，并对它们进行比较和总结。为了保持行文结构、方便读者对照理解，这一节依然按数据、视图、感知和交互四个部分来展开。

3.1 数据

对数据的研究主要关注三个方面：不同的数据类型、不确定性与数据隐私、以及绘制效率。

3.1.1 不同的数据类型

正如 2.2 节所述，由于折线重叠问题，平行坐标并不适合表现类别型数据。在这种数据里，大量样本具有相同的类别，它们相互之间无法被区分。但事实上，这些数据的分析重点并不在于单个样本，而是同一类样本的集合，因此需要在平行坐标里加入表达集合关系的元素。一方面，可以对类别型维度进行量化[51]，以表达不同类别之间的关系（如图8），但用户依然无法判断每个类别内的数据量。

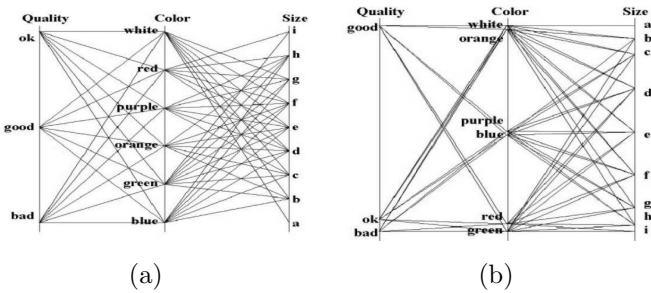
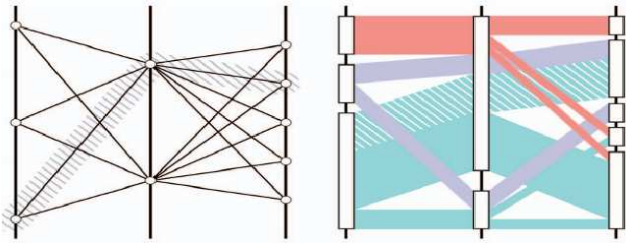
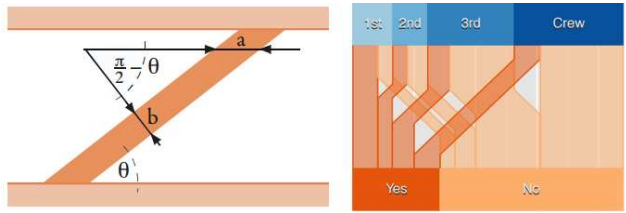


图 8: 类别型维度在量化前后的对比。本图源自Rosario等人的论文[51]。



(a) 平行集合



(b) 统一角度图

图 9: 平行集合及其改进版本，用以在平行坐标框架下表现类别型数据。图(a)、(b)分别来自论文[42]和[33]。

另一方面，样本数量可以用折线宽度来表示，即平行集合（Parallel Set）[42][6]的形式，各个集合之间的交错、包含等逻辑关系也更为明显（如图9(a)）。而Hofmann等人[33]指出，斜线段的宽度并不对应它所代表的数据量（比较图9(b)中的a、b两段）。他们提出用统一的角度来绘制折

线，即统一角度图（Common Angle Plot）的形式，使用户能通过直线部分的宽度来准确感知集合的大小。

除了类别型数据，时变数据也是一类重要的拓展，具体方法可分为静态和动态两种。其中静态的展示需要引入时间轴、并同时绘制所有时刻的数据，只是对时间轴的处理各有不同（见图10）。

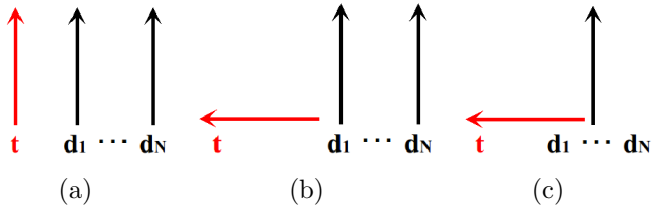


图 10: 时变平行坐标的三种静态展示方式: (a).时间轴平行于属性轴; (b).时间轴在坐标平面内垂直于属性轴; (c).时间轴垂直于坐标平面。

最简单的方式莫过于将时间轴看作一个平行轴，但各个时刻的数据被同等对待，不利于凸显数据的变化信息。Johansson等人 [39]提出利用不透明度和色调来区分数据变动的时间范围（如图 11）。用户可以直观地看出变化的频率和趋势，但不同时刻的数据依然会相互遮挡。

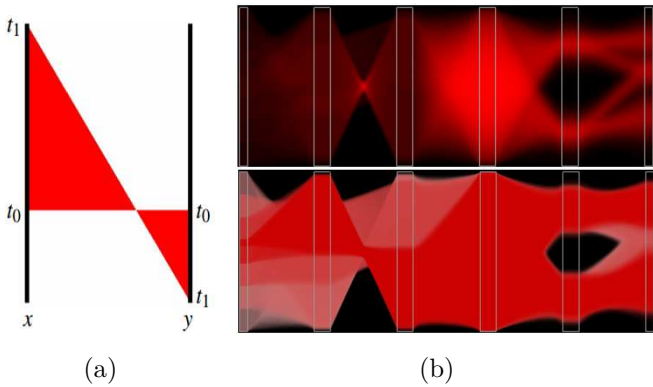


图 11: 用不透明度和色调刻画数据变动。其中折线变化用多边形表示（如图(a)），不透明度刻画了变化的频率，不同的色调则代表不同的时刻（如图(b)）。本图来源于Johansson 等人的论文 [39]。

为防止遮挡，也可以结合折线图（Line Chart），用多个轴表示一个维度的时变信息 [14] [13]（如图12），但同一维度的重复出现会降低视图的可扩展性。

进一步地，Wegenkittl等人 [59]令时间轴垂直于坐标平面，提出了三维时变平行坐标的形式

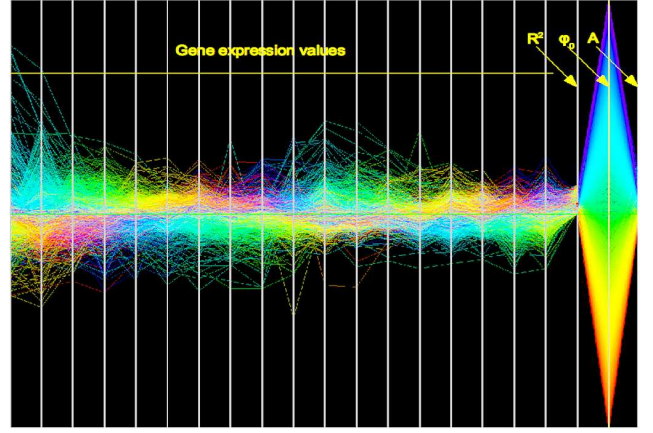


图 12: 在平行坐标中嵌入折线图。本图来源于Dietzsch 等人的论文 [13]。

（如图 13）。其中每条折线沿时间轴演化会形成一个曲面，该曲面即表达了时变高维的所有信息。但多个不同的曲面很容易交错重叠而难以观察，使其不适用于数据量大的情形。

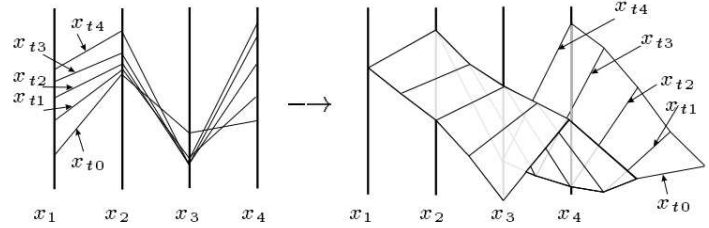


图 13: 三维时变平行坐标。图中 x_1 至 x_4 表示不同维度， t_1 至 t_4 为不同的时刻。本图来源于Wegenkittl 等人的论文 [59]。

除了静态展示，还可以用动画来表现数据变化 [4]。但视觉的瞬变使用户难以同时跟踪多个维度的变动，在维数较高、变动剧烈的情况下尤为如此。为每个维度提供额外的时变直方图 [7]可以缓解这一问题，但这又割裂了时变信息与维度关系。总的来说，以上方法都各有优劣，但还没有一个较为全面的方式来展现时变平行坐标。

还有一类拓展在于连续型数据。在科学数据（Scientific Data）里，所有样本可能都来自同一个数据域（Data Domain，如流场、物体表面等）。平行坐标能够逐个描绘这些样本，但无法展现通过插值重构出来的连续数据域，即所谓的“连续型数据”（Continuous Data）。绘制连续型数据的概念最早由Bachthaler等人 [3]提出并应用在散点图中。他们基于离散的样本，建模估计、并以颜色表现了视图各处的密度。基于这一研究，Heinrich等人利用散点图和平行坐标的几何

对称性，发展出连续型平行坐标 [27] 的形式（如图14）。后来的研究 [43] 则进一步揭示了这两种连续视图之间的密切联系。

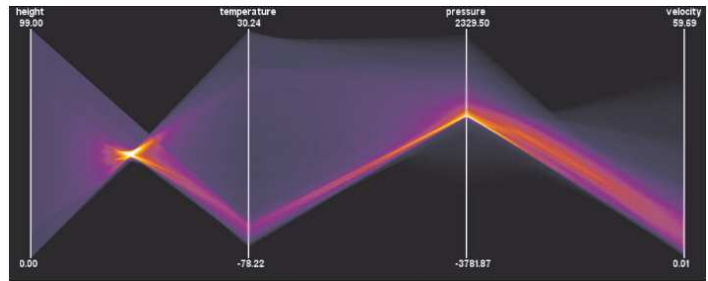


图 14: 连续型平行坐标。本图来源于Heinrich 等人的论文 [27]。

相比于传统形式，连续型平行坐标可以在样本较少的情况下重构出精度较高的原数据域，而且能避免视图混杂问题（见 3.1 节），也具有较好的可扩展性和渲染效率。但它无法给出单个样本的信息，也不支持传统的交互和分析方法（如刷选、聚类），在可用性上有一定的局限。

3.1.2 不确定性与数据隐私

无论什么类型的数据，都会在其采集和处理的过程中，因为种种原因（测量误差、样本随机性等）产生不同程度的不确定性。从结果推测的原数据因而会呈现概率分布，而这些分布则需要通过样本重构来得到。前面关于连续型数据的研究 [3] [27] 都假定数据域是被充分均匀采样的，并采取线性插值的方法来进行重构。但这些假定都基于格点型的数据（Grid Data），对于不充分、非均匀采样的情形并不适用。

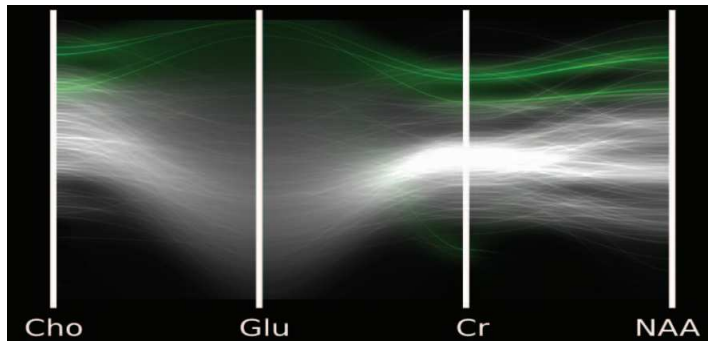


图 15: 平行坐标表现数据不确定性。本图来源于Feng 等人的论文 [18]。

对于一般的数据，Feng [18] 等人提出用高斯核密度估计（Kernel Density Estimation）来进行重

构，并利用连续型平行坐标的技术来绘制数据分布（如图15）。视图中密度越高的区域，其数据分布越集中，不确定性也越低。除此之外，他们还对该技术作了曲线形式的拓展，并加入了基于密度的刷选、高亮等交互方式。

不确定性往往是分析人员极力避免的因素。但反过来，我们也可以制造“额外的”不确定性，通过隐藏样本数据来保护其中的隐私信息。上述基于密度的方法给出了原数据域的分布，其信息量甚至大于样本数据，因此不适合隐私保护。Dasgupta 等人 [12] 提出通过聚类来封装底层数据（如图 16），并重新设计了相应的交互手段，以抵御不同形式的隐私攻击。但他们设置的信息保护机制过于复杂，导致视图难以理解，大大降低了其可用性。相比于隐藏单个数据，另一种方案在于只显示部分样本，或是对重构的数据域进行再采样 [29] 来获得虚构的样本集。

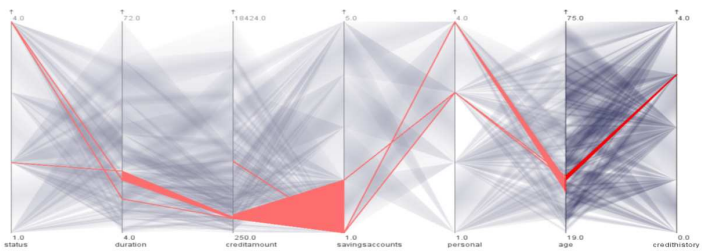


图 16: 通过聚类保护数据隐私，其中高亮的部分由于存在不确定性而难以理解。本图来源于Dasgupta 等人的论文 [12]。

3.1.3 绘制效率

当数据量较大时，直接绘制的效率较低，容易导致交互过程的延迟。Rosenbaum 等人 [52] 提出建立数据的精度层级（Data Hierarchy），在绘制完成前通过低精度小规模的数据来提供概要信息，以平衡信息量与绘制效率（如图 17）。

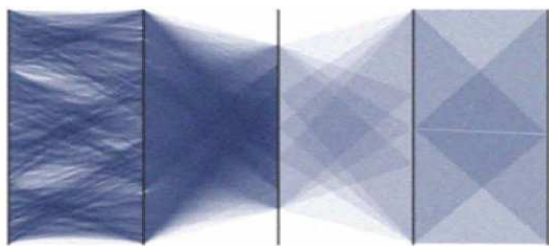


图 17: 递进式平行坐标，用户可以根据需要指定数据的精度。本图源自Rosenbaum 等人的论文 [52]。

但在视图精度不断提高的过程中，每个层级的数据都需要重新获取与绘制，使得自顶向下绘制的总体效率比直接绘制更低。Heinrich等人[29]提出自底向上地渲染连续型平行坐标的方式。他们对数据域进行再采样，通过不断增加的样本来重现、而非直接计算密度分布，从而用户在采样的早期就可看出数据分布的特点。

3.2 视图

如2.2节所述，对视图设计的研究主要关注三个问题，分别是数据重合、视觉混杂以及坐标轴的排序。

3.2.1 数据重合

数据重合的主要问题是无法在视觉上区分多个数据。Graham等人[23]提出用曲线代替折线，从而在重合点处，用户可以通过曲率的连续性来跟踪数据（图18(a)）；将重合点展开则可以进一步减轻这一问题（图18(b)）。

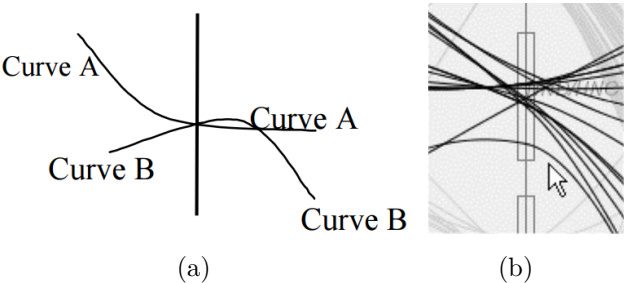


图 18: 通过曲率连续性区分不同数据。本图源自Graham等人的论文[23]。

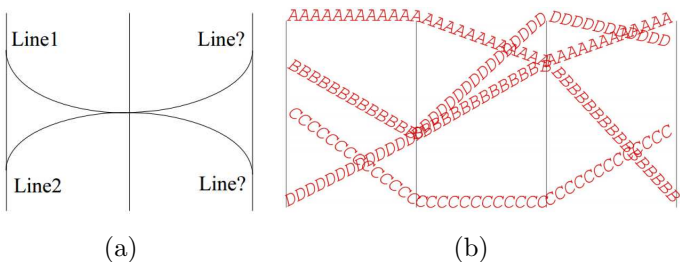


图 19: 平行坐标中的数据标签。本图源自Zhou等人的论文[68]。

而Zhou等人[68]指出，在重合点处相切的曲线依然无法区分。他们提出用数据标签代替线段（如图19），从而不论是直线或曲线，用户在放大

视图时，总能通过标签的不同来区分数据。但标签相比于线段更容易相互遮挡、影响可读性，使得该方法不适合数据量大的情形。

Dang等人则利用三维形式，将线元沿纵坐标堆叠起来（如图20），这与线宽表示数量的方式[42]有异曲同工之处。但三维视图容易产生遮挡、交互难度大，其实际效用还有较大争议。

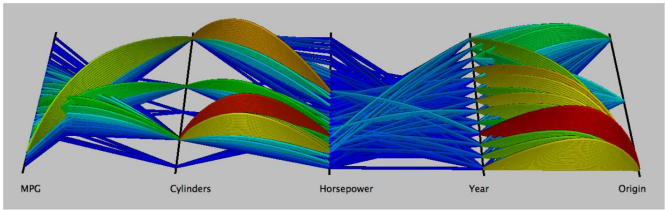


图 20: 堆叠线元以解决数据重合问题。本图源自Dang等人的论文[42]。

3.2.2 视觉混杂

视觉混杂问题是平行坐标研究的重点之一。要在有限的空间里表现大量的图元，主要有四种方法：刻画密度、视觉概要、聚类表达、以及数据压缩。

密度分布可以是图元或数据的密度，并借由不透明度或色调来表达。表现图元密度的最简单方法，是调节折线的不透明度，从而折线越密集的区域可见性越高（如图11(b)）。也可以基于屏幕像素来统计折线堆叠的情况，再利用不透明度[2][39]或色调[47]来呈现（如图21）。

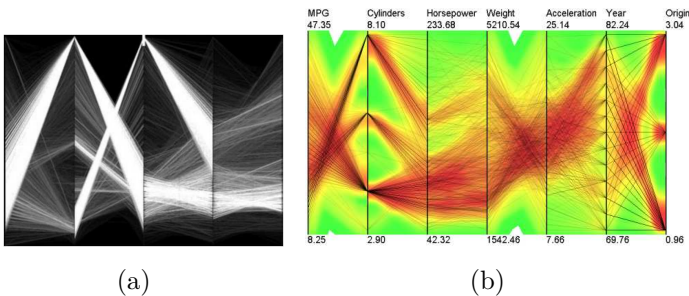


图 21: 利用不透明度或色调表现折线密度

然而在平行坐标里，对视图作密度估计并不反映数据域本身的密度。Zhou等人[67]提出用抛雪球算法来检测数据聚类，并逐步显现其中的密度分布（如图22）。

进一步地，连续型平行坐标的相关研究[27][43][18][29]直接对数据域作密度估计，并揭示了

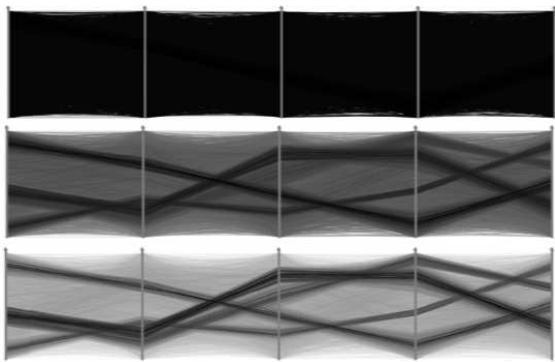
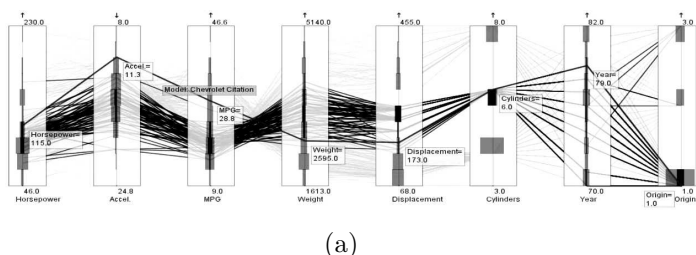


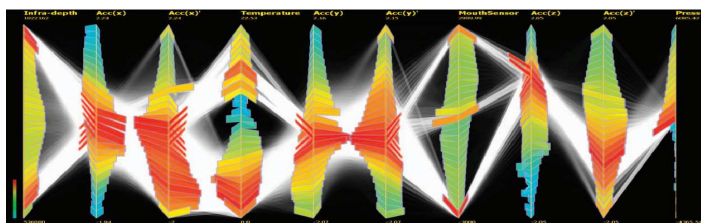
图 22: 利用抛雪球算法逐步凸显数据密度。本图修改自Zhou等人的论文 [67]。

相应的绘制方式（见图 14、图 15）。如此一来，既无需费时绘制所有数据，避免了混杂问题，又能较好地保留数据分布的信息。然而在上述方法中，低密度区域的信息会因为较低的可见性而被忽略。Johansson等人 [38]利用传递函数（Transfer Function）来调节数据密度与可见性的关系，从而用户可以根据分析需求来凸显不同密度的区域。

相比于表现密度，视觉概要的方法通过少量图元来概括数据分布。较简单的做法是分别描述各个维度的数据密度 [26] [8]（如图 23(a)），但这无法揭示维度之间的联系。要对折线进行概括，除了密度，还须考虑其方向。Geng等人 [22]提出了带角度的直方图（Angular Histogram），在表达数据密度的同时，通过柱形的角度来指示折线走向（如图 23(b)）。Muigg等人 [48]则利用噪声纹理（Noise Texture）传达方向信息（如图 24），并将这一技术推广到其他基于折线的视图中。



(a)



(b)

图 23: 利用直方图表现各维度上数据的密度和走向。(a)、(b)两图分别来自论文 [26]、[22]。

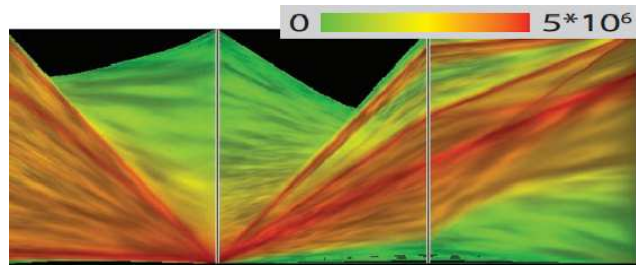


图 24: 利用噪声纹理表达折线的方向。本图修改自Muigg等人的论文 [48]。

然而很多时候，即使是经过概括的视图，也会因为较高的数据异质性而显得杂乱。可以先进行聚类、从根源上消除混杂，即为聚类表达型方法。这类方法可分为视觉聚类和数据聚类两种。

视觉聚类方法将相似的图元汇聚起来，在平行坐标里即是将折线“捆”在一起。Guo等人 [25]提出一种具有引力的局部算子，用户可以在视图中放置算子以对部分折线进行捆绑（如图 25）。

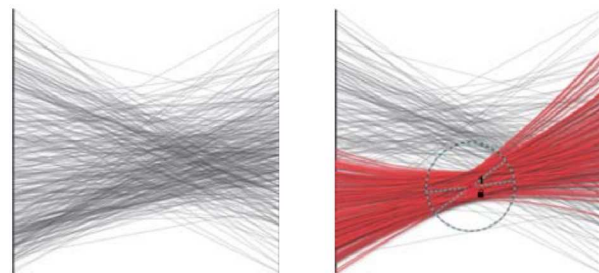


图 25: 局部捆绑折线的引力算子。本图修改自Guo等人的论文 [25]。

但交互聚类的方法往往单调繁复却不一定准确有效，可以引入自动算法。类似于抛雪球方法 [67]，Zhou等人 [66]为每条折线赋予引力，使得方向和位置相近的折线相互吸引而自动汇聚起来（如图 26）。

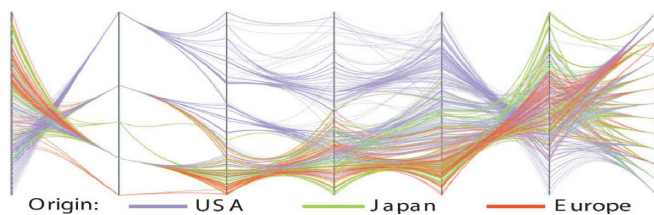


图 26: 折线的自动聚类。本图来自Zhou等人的论文 [66]。

另一方面，数据聚类方法先组织数据再进行表达，较为经典的研究是Fua等人 [21]提出的层次聚类（如图 27(a)）。但由于折线的位置方向各

异、容易混杂，视图中往往只标识各个聚类的中心及其范围 [53] [52]，无法给出单个数据的细节（如图 17）。为弥补这一缺点，可以检测出异常数据，并将聚类 and 异常点分开表达 [49]（如图 27(b)）。

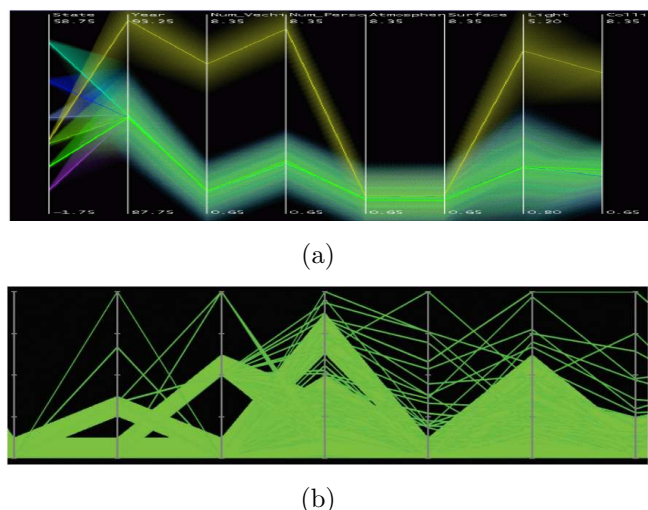


图 27: 聚类及异常点的表达。(a)、(b)两图分别来自论文 [21]、[49]。

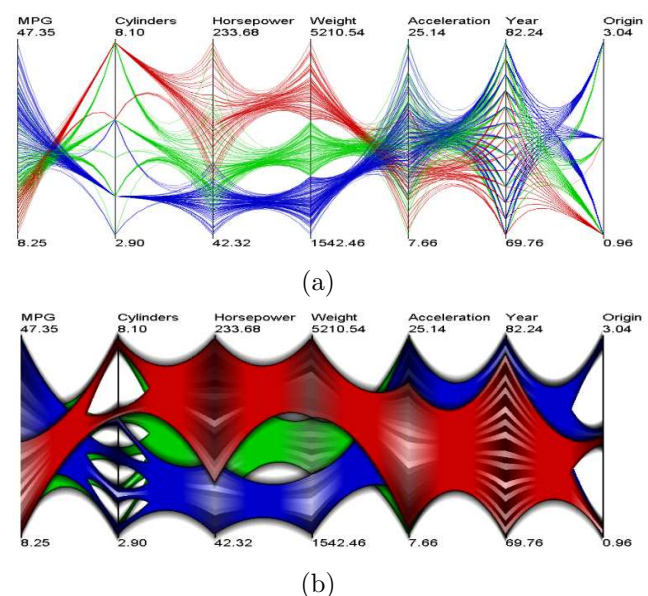


图 28: 阐释型平行坐标。本图来自McDonnell等人的论文 [47]。

而在曲线形式 [56] [23]出现以后，能够兼容单个数据和聚类的边缘捆绑（Edge Bundling）技术，就被广泛应用在聚类表达里。类似 [66]，McDonnell等人 [47]将每条折线捆绑到相应聚类的中心（如图 28(a)）。在此基础上，他们利用视觉概要的方法表达数据走势与密度，产生了适合用作数据展示的阐释型（Illustrative）平行坐标（如图 28(b)）。

而Heinrich等人 [30]指出，斜率在坐标轴处不连续，会违反使用曲线解决数据重合问题的初衷（见 3.2.1节）。他们改用Bézier曲线，并分离各聚类中心以避免相近的聚类相互遮挡（如图 29(a)）。借鉴平行集合的形式 [42]，Palmas等人 [50]进一步缩短曲线汇聚的长度，并利用条带来表现各维度聚类之间的关系。

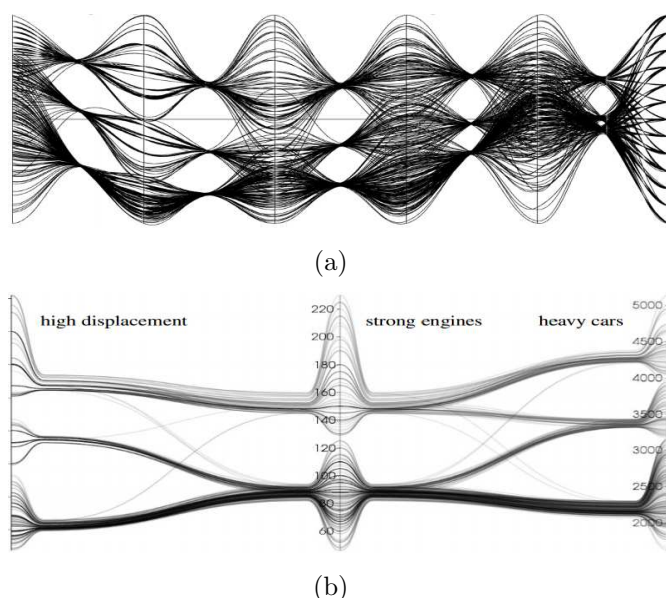


图 29: 基于Bézier曲线的边缘捆绑。(a)、(b)两图分别来自论文 [30] [50]。

鉴于散点图擅长表现数据聚类，Yuan等人 [64]在坐标轴之间嵌入散点图，并借助点的聚类来引导折线捆绑（如图 30）。该形式既表现了散点分布，也减轻了视觉混杂问题。但其效果取决于数据本身的聚类性能，且不适用于维度较高的情形。

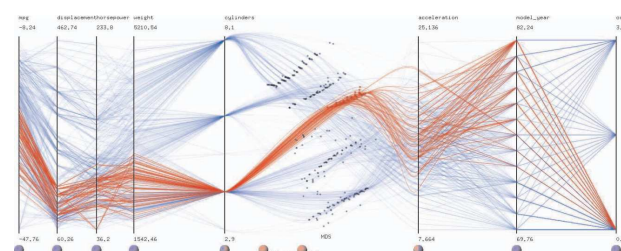


图 30: 折线的自动聚类。本图来自Zhou等人的论文 [66]。

数据压缩类的方法，即是通过少数具有代表性的样本来呈现数据分布。其中最简单的手段是数据采样，Ellis等人 [16] [15]提出采样透镜（Sampling Lens）的方式，用户可以交互地指定需要采样的视图区域（如图 31(a)）。但圆形的区

域并不适合选择折线，且随机采样得到的样本不一定具有代表性。Wong等人[61]使用小波变换来压缩数据（如图31(b)），并允许用户通过折线刷选来指定压缩的精度和范围，但构造的样本并不具有现实意义。Rosenbaum等人[52]进一步引入了递归区域划分技术（Recursive Interval Subdivision），使用中位数作为代表样本，并将其与小波压缩、层次聚类相比较。至于哪种压缩更能保留数据特征，Johansson等人[36]提出了一种基于距离变换（Distance Transformation）的衡量方法，用以定量地指导压缩方式的选择（如图32）。

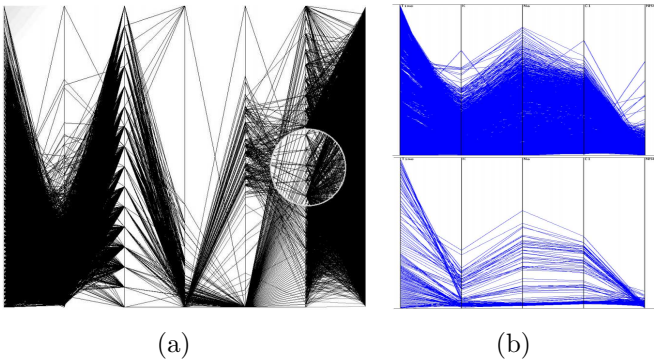


图 31: 数据压缩类方法。图(a)、(b)分别来自论文[16]和[50]。

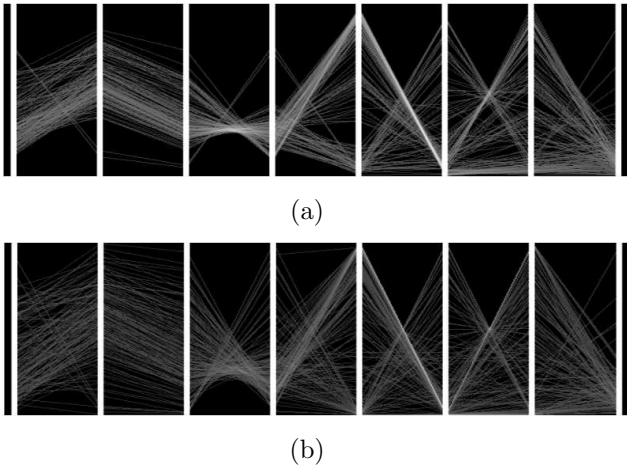


图 32: 不同的压缩方法在相同压缩率下的比较。其中图(a)丢失了低密度区域的信息，而图(b)则保留了更多的数据特征。本图修改自Johansson等人的论文[36]。

3.2.3 坐标轴排序

正如2.2节所述，坐标轴的顺序决定了哪些二维子空间会被显示，它对于整个视图的表达效果至关重要。

较为直接的做法是将相似的维度放在相邻的位置，使轴间的子空间具有较高的相关性。Ankerst等人[1]最早探讨了基于相关性的排序，并证明了无论在何种准则下，最优排序都是一个NP完全问题。尽管可以用各种自动算法（如蚁群算法[1]、贪心算法[40]、分支定界[11]等等）求得近似解，用户往往无法了解和参与自动排序的过程。为此Zhang等人[65]提出了基于相关性图的交互式排序方法。通过相关性图（图33右图），用户能够观察各维度之间关系的强弱。据此可以规划一条遍历图上各点的路径，而遍历的顺序即为坐标轴顺序。

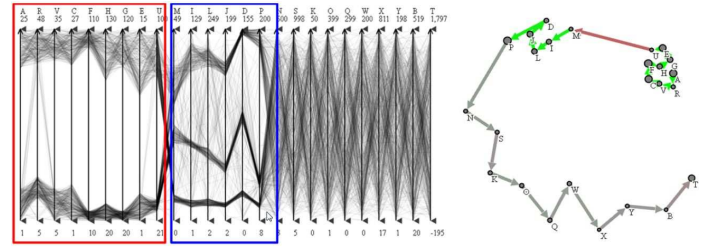


图 33: 交互式坐标轴排序。本图来自Zhang等人的论文[65]。

除了相关性，还可以进一步考察子空间的其它特征，譬如聚类性能。一方面，可以检测出具有重要聚类的高维子空间。在排序时保持其各个维度相邻，可以有效凸显其中的聚类信息[19]。另一方面，对于多种不同的排序，可以比较二维子空间的聚类特性[55]。轴间线段更趋一致的排序，更有可能反映潜在的数据规律（如图34）。

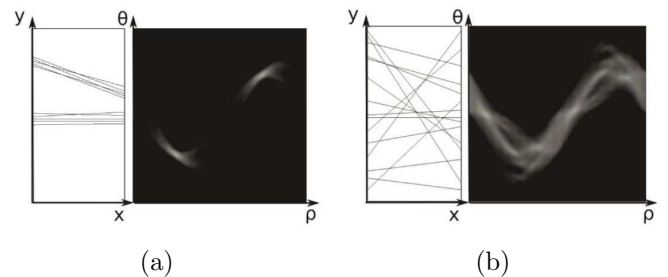


图 34: 通过特征空间比较轴间线段的聚类性，聚类明显的排序（图(a)）更能反映潜在的数据规律。本图源自Tatu等人的论文[55]。

当单个标准无法满足需求时，用户还可以组合多种准则来优化排序[40]。Dasgupta等人[11]结合线段交错性、汇聚性等七种准则，提出了较为完备的平行坐标评价标准。但如何定制准则、使其既直观易懂又容易计算，还有待进一步探究。

无论是何种准则，随着问题规模的增大，启发式算法（Heuristic Algorithm）的性能会不断下降。此外，过高的维度会压缩轴间空间使其不可见，排序也就失去了意义。因此在排序的同时，有必要降低数据的维度数。Yang等人 [62] [63] 基于相似性对各维度进行层次聚类，利用少数具有代表性的维度来展示数据。Zhang等人的交互式排序 [65] 则通过缩放相关性图来筛选维度。而Johansson等人 [40] 建立了不同的评判准则，让用户可以根据分析需求、有针对性地进行维度筛选。

除了排序和降维，坐标轴的布置方式也是一个有趣的研究课题。非平行、非平面的布局往往能表达更加多样的数据关系。其中一类拓展利用了三维空间，如 3.1.1 节提到的三维时变平行坐标（见图 13）。Johansson [37] 等人设计了环形的布局，以表达多对一的维度关系（如图 35(a)）。Fanea 等人 [17] 则结合平行坐标与星型图，并通过三维交互来提供两者之间的过渡（如图 35(b)）。但三维视图容易产生遮挡严重、交互困难等问题，因此更多的研究只关注平面布局的变化。

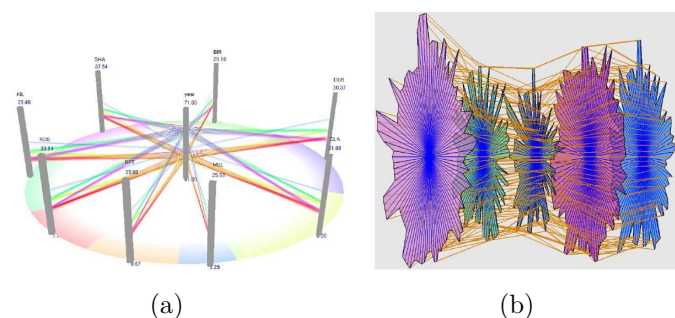


图 35: 平行坐标的三维拓展。图(a)、(b)分别来自论文 [37] 和 [17]。

为表现更多的子空间，Lind 等人 [45] 尝试了多对多平行坐标的形式（如图 36(a)）。但受限于视图复杂度，他们仅研究了七个变量的情形，并认为这种形式不具备良好的可扩展性。Claessen 等人 [9] 延展了这一概念，提出让用户自由放置和联结坐标轴，以产生各种基于轴的可视化形式（如图 36(a)）。但这既增加了用户的交互负担，也不可避免地造成了空间浪费，其合理性有待进一步探讨。

考虑到散点图矩阵能够表达所有子空间，Viau 等人 [58] 通过动画过渡融合散点图与平行坐标，尝试将后者嵌入到矩阵中。Heinrich 等人

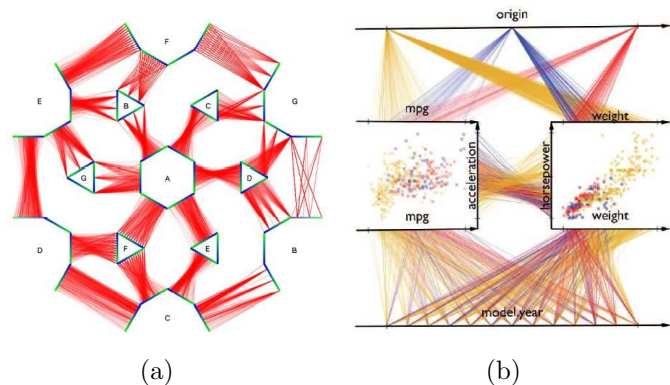


图 36: 坐标轴的自由布局。图(a)、(b)分别来自论文 [45] 和 [9]。

[31] 则进一步提出了平行坐标矩阵的形式（如图 36(b)），通过矩阵各行表现不同的排序，以提高视图的表达效率。相比于完全自由的布局，矩阵形式不存在空间浪费、视图旋转等问题，也更符合用户的使用习惯。但无论哪种形式，重复显示同一维度都将大大降低视图的可扩展性，也为交互设计带来了挑战。

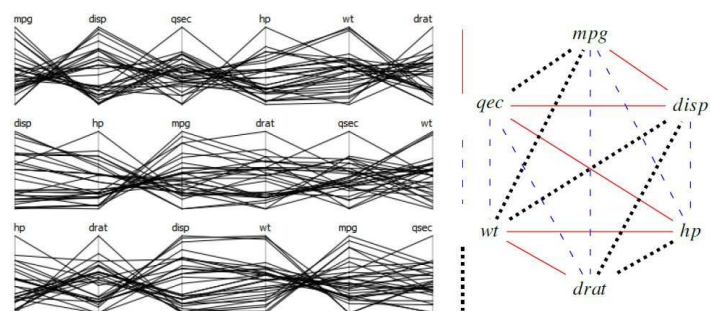


图 37: 平行坐标矩阵，表达了数据中所有的二维子空间。本图来自Heinrich等人的论文 [31]。

3.3 感知

3.4 交互

4 探讨与展望

5 结论

参考文献

- [1] M. Ankerst, S. Berchtold, and D. A. Keim. Similarity clustering of dimensions for an en-

- hanced visualization of multidimensional data. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*, pages 52–60. IEEE, 1998.
- [2] A. O. Artero, M. C. F. de Oliveira, and H. Levkowitz. Uncovering clusters in crowded parallel coordinates visualizations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 81–88. IEEE, 2004.
- [3] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1428–1435, 2008.
- [4] N. Barlow and L. J. Stuart. Animator: A tool for the animation of parallel coordinates. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 725–730. IEEE, 2004.
- [5] R. E. Bellman and S. E. Dreyfus. Applied dynamic programming. 1962.
- [6] F. Bendix, R. Kosara, and H. Hauser. Parallel sets: Visual analysis of categorical data. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 133–140. IEEE, 2005.
- [7] J. Blaas, C. P. Botha, and F. H. Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1436–1451, 2008.
- [8] M. t. Caat, N. M. Maurits, and J. B. Roerdink. Design and evaluation of tiled parallel coordinate visualization of multichannel eeg data. *Visualization and Computer Graphics, IEEE Transactions on*, 13(1):70–79, 2007.
- [9] J. H. Claessen and J. J. Van Wijk. Flexible linked axes for multivariate data visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2310–2316, 2011.
- [10] W. C. Cleveland and M. E. McGill. *Dynamic graphics for statistics*. CRC Press, Inc., 1988.
- [11] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1017–1026, 2010.
- [12] A. Dasgupta and R. Kosara. Adaptive privacy-preserving visualization using parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2241–2248, 2011.
- [13] J. Dietzsch, J. Heinrich, K. Nieselt, and D. Bartz. Spray: A visual analytics approach for gene expression data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 179–186. IEEE, 2009.
- [14] R. M. Edsall. The parallel coordinate plot in action: design and use for geographic visualization. *Computational Statistics & Data Analysis*, 43(4):605–619, 2003.
- [15] G. Ellis and A. Dix. Enabling automatic clutter reduction in parallel coordinate plots. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):717–724, 2006.
- [16] G. Ellis, E. Bertini, and A. Dix. The sampling lens: making sense of saturated visualisations. In *CHI’05 Extended Abstracts on Human Factors in Computing Systems*, pages 1351–1354. ACM, 2005.
- [17] E. Fanea, S. Carpendale, and T. Isenberg. An interactive 3d integration of parallel coordinates and star glyphs. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 149–156. IEEE, 2005.
- [18] D. Feng, L. Kwock, Y. Lee, and R. M. Taylor. Matching visual saliency to confidence in plots of uncertain data. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):980–989, 2010.

- [19] B. J. Ferdosi and J. B. Roerdink. Visualizing high-dimensional structures by dimension ordering and filtering using subspace analysis. In *Computer Graphics Forum*, volume 30, pages 1121–1130. Wiley Online Library, 2011.
- [20] I. K. Fodor. A survey of dimension reduction techniques. 2002.
- [21] Y.-H. Fua, M. O. Ward, and E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the conference on Visualization'99: celebrating ten years*, pages 43–50. IEEE Computer Society Press, 1999.
- [22] Z. Geng, Z. Peng, R. S. Laramée, J. C. Roberts, and R. Walker. Angular histograms: Frequency-based visualizations for large, high dimensional data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2572–2580, 2011.
- [23] M. Graham and J. Kennedy. Using curves to enhance parallel coordinate visualisations. In *Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on*, pages 10–16. IEEE, 2003.
- [24] G. Grinstein, M. Trutschl, and U. Cvek. High-dimensional visualizations. In *Data mining conference KDD workshop 2001*, pages 7–19, 2001.
- [25] P. Guo, H. Xiao, Z. Wang, and X. Yuan. Interactive local clustering operations for high dimensional data in parallel coordinates. In *Pacific Visualization Symposium (PacificVis), 2010 IEEE*, pages 97–104. IEEE, 2010.
- [26] H. Hauser, F. Ledermann, and H. Doleisch. Angular brushing of extended parallel coordinates. In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 127–130. IEEE, 2002.
- [27] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1531–1538, 2009.
- [28] J. Heinrich and D. Weiskopf. State of the art of parallel coordinates. *STAR Proceedings of Eurographics*, 2013:95–116, 2013.
- [29] J. Heinrich, S. Bachthaler, and D. Weiskopf. Progressive splatting of continuous scatterplots and parallel coordinates. In *Computer Graphics Forum*, volume 30, pages 653–662. Wiley Online Library, 2011.
- [30] J. Heinrich, Y. Luo, A. E. Kirkpatrick, H. Zhang, and D. Weiskopf. Evaluation of a bundling technique for parallel coordinates. *arXiv preprint arXiv:1109.6073*, 2011.
- [31] J. Heinrich, J. Stasko, and D. Weiskopf. The parallel coordinates matrix. *EuroVis-Short Papers*, pages 37–41, 2012.
- [32] P. Hoffman, G. Grinstein, and D. Pinkney. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM international conference on Information and knowledge management*, pages 9–16. ACM, 1999.
- [33] H. Hofmann and M. Vendettuoli. Common angle plots as perception-true visualizations of categorical associations. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2297–2305, 2013.
- [34] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.
- [35] A. Inselberg. *Parallel coordinates*. Springer, 2009.
- [36] J. Johansson and M. Cooper. A screen space quality method for data abstraction. In *Computer Graphics Forum*, volume 27, pages 1039–1046. Wiley Online Library, 2008.

- [37] J. Johansson, M. Cooper, and M. Jern. 3-dimensional display for clustered multi-relational parallel coordinates. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pages 188–193. IEEE, 2005.
- [38] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 125–132. IEEE, 2005.
- [39] J. Johansson, P. Ljung, and M. D. Cooper. Depth cues and density in temporal parallel coordinates. In *EuroVis*, volume 7, pages 35–42, 2007.
- [40] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):993–1000, 2009.
- [41] E. Kandogan. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In *Proceedings of the IEEE Information Visualization Symposium*, volume 650, page 22, 2000.
- [42] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *Visualization and Computer Graphics, IEEE Transactions on*, 12(4):558–568, 2006.
- [43] D. J. Lehmann and H. Theisel. Features in continuous parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1912–1921, 2011.
- [44] J. Li, J.-B. Martens, and J. J. Van Wijk. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*, 9(1):13–30, 2010.
- [45] M. Lind, J. Johansson, and M. Cooper. Many-to-many relational parallel coordinates displays. In *Information Visualisation, 2009 13th International Conference*, pages 25–31. IEEE, 2009.
- [46] A. R. Martin and M. O. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Proceedings of the 6th Conference on Visualization’95*, page 271. IEEE Computer Society, 1995.
- [47] K. T. McDonnell and K. Mueller. Illustrative parallel coordinates. In *Computer Graphics Forum*, volume 27, pages 1031–1038. Wiley Online Library, 2008.
- [48] P. Muigg, M. Hadwiger, H. Doleisch, and E. Gröller. Visual coherence for large-scale line-plot visualizations. In *Computer Graphics Forum*, volume 30, pages 643–652. Wiley Online Library, 2011.
- [49] M. Novotny and H. Hauser. Outlier-preserving focus+ context visualization in parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):893–900, 2006.
- [50] G. Palmas, M. Bachynskyi, A. Oulasvirta, H. P. Seidel, and T. Weinkauff. An edge-bundling layout for interactive parallel coordinates. In *Pacific Visualization Symposium (PacificVis), 2014 IEEE*, pages 57–64. IEEE, 2014.
- [51] G. E. Rosario, E. A. Rundensteiner, D. C. Brown, M. O. Ward, and S. Huang. Mapping nominal values to numbers for effective visualization. *Information Visualization*, 3(2):80–95, 2004.
- [52] R. Rosenbaum, J. Zhi, and B. Hamann. Progressive parallel coordinates. In *Pacific Visualization Symposium (PacificVis), 2012 IEEE*, pages 25–32. IEEE, 2012.
- [53] H. Siirtola. Direct manipulation of parallel coordinates. In *Information Visualization, 2000. Proceedings. IEEE International Conference on*, pages 373–378. IEEE, 2000.
- [54] H. Siirtola and K.-J. Räihä. Interacting with parallel coordinates. *Interacting with Computers*, 18(6):1278–1309, 2006.

- [55] A. Tatu, G. Albuquerque, M. Eisemann, P. Bak, H. Theisel, M. Magnor, and D. Keim. Automated analytical methods to support visual exploration of high-dimensional data. *Visualization and Computer Graphics, IEEE Transactions on*, 17(5):584–597, 2011.
- [56] H. Theisel. Higher order parallel coordinates. In *VMV*, pages 415–420, 2000.
- [57] J. J. Van Wijk. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 79–86. IEEE, 2005.
- [58] C. Viau, M. J. McGuffin, Y. Chiricota, and I. Jurisica. The flowvizmenu and parallel scatterplot matrix: Hybrid multidimensional visualizations for network exploration. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1100–1108, 2010.
- [59] R. Wegenkittl, H. Löffelmann, and E. Groller. Visualizing the behaviour of higher dimensional dynamical systems. In *Visualization’97., Proceedings*, pages 119–125. IEEE, 1997.
- [60] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization.
- [61] P. C. Wong and R. D. Bergeron. Multiresolution multidimensional wavelet brushing. In *Visualization’96. Proceedings.*, pages 141–148. IEEE, 1996.
- [62] J. Yang, M. O. Ward, and E. A. Rundensteiner. Visual hierarchical dimension reduction for exploration of high dimensional datasets. 2002.
- [63] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 105–112. IEEE, 2003.
- [64] X. Yuan, P. Guo, H. Xiao, H. Zhou, and H. Qu. Scattering points in parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1001–1008, 2009.
- [65] Z. Zhang, K. T. McDonnell, and K. Mueller. A network-based interface for the exploration of high-dimensional data spaces. In *Pacific Visualization Symposium (PacificVis), 2012 IEEE*, pages 17–24. IEEE, 2012.
- [66] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. In *Computer Graphics Forum*, volume 27, pages 1047–1054. Wiley Online Library, 2008.
- [67] H. Zhou, W. Cui, H. Qu, Y. Wu, X. Yuan, and W. Zhuo. Splatting the lines in parallel coordinates. In *Computer Graphics Forum*, volume 28, pages 759–766. Wiley Online Library, 2009.
- [68] H. Zhou, P. Xu, Z. Ming, and H. Qu. Parallel coordinates with data labels. In *Proceedings of the 7th International Symposium on Visual Information Communication and Interaction*, page 49. ACM, 2014.