第28卷 第8期 计 算 机 工 程 2002年8月
Vol.28 № 8 Computer Engineering August 2002

· 发展趋势／热点技术 · 文章编号：1000—3428(2002)08 —0001—02 文献标识码：A 中图分类号： TP317.4;TP391.41

数据可视化的研究与发展

刘 勘，周晓峥，周洞汝
（武汉大学(水电校区)计算机学院，武汉 430072）

摘 要：针对数据可视化是可视化技术在大型数据库的应用中提出的新的数据分析和处理技术。该文介绍了数据可视化的概念和发展状况，然后针对大型数据集介绍了几种数据可视化技术以及它们的代表方法，并对数据可视化和科学计算可视化进行了分析和比较，最后探讨了数据可视化技术的研究发展方向。

关键词：数据可视化；数据开发；数据开发分析

Data Visualization Research and Development

LIU Kan, ZHOU Xiaozheng, ZHOU Dongru
(Department of Computer Science, Wuhan University,Wuhan 430072)

【Abstract】Data visualization is the new technique of data analysis and process, which aims at presenting the data sets with image and graph with the assistance of such methods and techniques as interactions, multi-dimension analysis and data mining. This article firstly introduces the concept of data visualization and its development, and then introduces several types of data visualization techniques and their representative methods used in large data sets. It also covers an analysis and comparison between data visualization and scientific computing visualization. In the end, this article discusses on the future work of data visualization.

【Key words】Data visualization; Data exploration；Exploratory data analysis

1 数据可视化的基本概念

可视化(Visualization)技术是利用计算机图形学和图像处理技术，将数据转换成图形或图像在屏幕上显示出来，并进行交互处理的理论、方法和技术。它涉及到计算机图形学、图像处理、计算机视觉、计算机辅助设计等多个领域，成为研究数据表示、数据处理、决策分析等一系列问题的综合技术。可视化技术最早运用于计算科学中，并形成了可视化技术的一个重要分支——科学计算可视化(Visualization in Scientific Computing)。科学计算可视化能够把科学数据，包括测量获得的数值、图像或是计算中涉及、产生的数字信息变为直观的、以图形图像信息表示的、随时间和空间变化的物理现像或物理量呈现在研究者面前，使他们能够观察、模拟和计算。科学计算可视化自1987年提出以来，在各工程和计算领域得到了广泛的应用和发展。

近年来，随着数据仓库技术、网络技术、电子商务技术等的发展，可视化技术涵盖了更广泛的内容，并进一步提出了数据可视化的概念，所谓数据可视化是对大型数据库或数据仓库中的数据的可视化，它是可视化技术在非空间数据领域的应用，使人们不再局限于通过关系数据表来观察和分析数据信息，还能以更直观的方式看到数据及其结构关系。数据可视化技术的基本思想是将数据库中每一个数据项作为单个图元元素表示，大量的数据集构成数据图像，同时将数据的各个属性值以多维数据的形式表示，可以从不同的维度观察数据，从而对数据进行更深入的观察和分析。

数据可视化技术包含以下几个基本概念：

(1)数据空间：是由n维属性和m个元素组成的数据集所构成的多维信息空间；

(2)数据开发：是指利用一定的算法和工具对数据进行定量的推演和计算；

(3)数据分析：指对多维数据进行切片、块、旋转等动作剖析数据，从而能多角度多侧面观察数据；

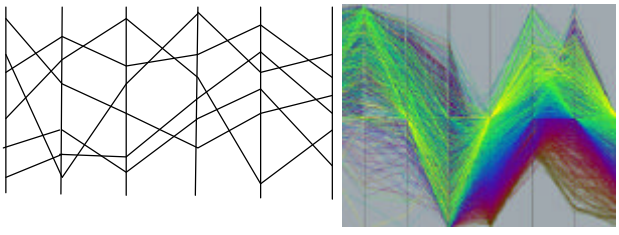
(4)数据可视化：是指将大型数据集中的数据以图形图像形式表示，并利用数据分析和开发工具发现其中未知信息的处理过程。

2 数据可视化的主要技术

目前数据可视化已经提出了许多方法，这些方法根据其可视化的原理不同可以划分为基于几何的技术、面向像素技术、基于图标的技术、基于层次的技术、基于图像的技术和分布式技术等等。下面介绍的是几种主要的技术。

2.1 基于几何的技术

基于几何的可视化技术包括Scatter plots、Landscapes、Projection Pursuit、Parallel Coordinates等等，是以几何画法或几何投影的方式来表示数据库中的数据。平行坐标法是最早提出的以二维形式表示n维数据的可视化技术之一[1]。它的基本思想是将n维数据属性空间通过n条等距离的平行轴映射到二维平面上，每一条轴线代表一个属性维，轴线上的取值范围从对应属性的最小值到最大值均匀分布。这样，每一个数据项都可以根据其属性值用一条折线段在n条平行轴上表示出来。



(a) 5个6维数据的平行坐标表示 (b) 平行坐标法开发的大型数据集

图1 平行坐标法示例

作者简介：刘 勘(1970～)，男，博士生，主要研究方向为信息可视化和数据挖掘技术；周晓峥，博士生；周洞汝，教授、博导

收稿日期：2001-08-02

开发人员和操作人员是透明的。它由加脱/密处理、语法分析、数据库接口等模块组成。

加脱/密处理模块是数据库加脱/密引擎的核心模块，包括数据库加脱/密引擎的初始化、内部专用命令的处理、加密字典信息的检索、加密字典缓冲区的管理、SQL命令的加密变换、查询结果的脱密处理、加脱密算法实现等功能子模块，另外还包括一些公用的辅助函数。

语法分析模块的功能是将SQL命令转换成易处理的树形式的语法结构。该模块由词法分析器、语法分析器、语法错误处理、语法树转换成SQL命令等子模块组成。在进行语法分析时，先对SQL命令进行词法分析，分割成各个词法单位，再输入语法分析器，得到一棵语法树。语法分析模块还包括语法树反向生成SQL命令的功能函数，用于将经过加密变换后的语法树转换成新的SQL命令。

数据库接口模块将所有访问数据库的操作封装在一起，屏蔽了各类数据库的特性，使得加脱密处理模块不必关心实际使用的是哪种数据库。该模块包含两部分接口，一是前端数据库客户访问数据库加/脱密引擎的接口函数，二是数据库加/脱密引擎访问后台数据库服务器的接口函数。

4 结论

本系统采用在DBMS外层实现数据库加密系统的方法，使得系统对数据库最终用户完全透明；而且，数据库加密系统完全独立于数据库应用系统，不需要改动数据库应用系统就能实现加密功能；同时，该系统采用分组加密法，二级密钥管理，实现了“一次一密”，具有很高的安全性；更加重要的是该系统在客户端进行数据加脱密运算，不会影响数据库服务器的系统效率，数据加/脱密运算基本无延迟。

进一步的工作是在数据库加密系统的基础上实现更强的用户身份认证功能(如IC卡、Skey等)、与操作系统的安全机制紧密衔接，从而保证数据库加密系统在一个更加安全的环境下运行。

参考文献

1 Bobrowski S.王 焱,王 磊译.Oracle 8体系结构.北京:机械工业出版社,2000

2 高品均,陈荣良加脱密引擎北京:计算机世界周报,2000

☆☆

(上接第2页)

学计算可视化主要是数据场的可视化，包括点数据场、标量场、矢量场和张量场的可视化等。数据可视化致力于在二维平面上显示数据的多维属性，分析并发现其中的关联和走势。科学计算可视化可以利用基本的参考模型，先将原始数据转换为几何数据，再将几何数据转换为图像数据，最后进行映射和绘制。数据可视化没有基本的模型可以遵循，可视化系统可给出带有多变量的图形化分析数据，帮助分析员进行信息发现，然后查看到那些无论系统计算能力有多强，机器算法都难以确定的模式和关系。下表给出了科学计算可视化和数据可视化的主要区别。

表1 数据可视化和科学计算可视化比较表

	科学计算可视化	数据可视化
数 据 源	计算和工程测量中的数据	大型数据集（库）中的数据
数据处理过程	数据预处理—>映射（构模）—>绘制和显示	数据提取—>数据多维显示—>数据分析和挖掘
主要应用方法	线状图、直方图、等值线（面）绘制、体绘制等	平行坐标法、面向像素法、树图、枝形图等
作 用	提供集成、方便的数据处理工具，对海量数据进行模拟和计算	直观地表达数据和数据之间的关系，获得对其内在信息的洞察
应 用 领 域	医学、地质、气象、流体力学等	商业、金融、企业管理等

数据可视化和科学计算可视化虽然应用的领域和处理方法都不尽相同，但因为两者的基础都是利用计算机图形图像处理技术来分析和显示数据源，所以在具体的技术方法中仍有值得互相借鉴的地方。如数据可视化技术中的Chernoff-face方法就是利用科学计算可视化中对多维点数据场的可视化技术。Issei Fujishiro等人开发的GADGET/IV系统也是其相应的科学计算可视化系统GADGET在信息可视化中的扩展。

4 数据可视化技术的发展方向

首先，可视化技术必须同数据挖掘有更紧密的联系。目前的数据可视化技术中的数据挖掘和分析功能难以运用数据挖掘的公式和算法，对可视化的数据反映出的结构和特点难以把握和证实。而数据挖掘和数据分析工具本身并不包含可视化技术，所以研究数据可视化技术和数据挖掘技术之间更加紧密的结合是提高数据可视化功能的一个重要方面。其次，可视化系统需要提高数据可视化技术的人机交互能力。数据可视化系统的人机交互不仅是要求用户提出有效的数据查询，而且需要在数据库管理系统的帮助下分析、评估用户的查询条件和查询结果，并指导用户提出更符合实际数据关系的查询。另外，可以先开发针对某一类特定数据的可视化系统。针对某一类型的数据，如银行信贷数据、股票数据、公司人事信息等，开发相对应的数据可视化系统，使相应的数据得到充分的显示和分析，然后再考虑推广到更广泛的数据中应用。最后，对前面介绍了几种主要的数据可视化技术和方法都有进一步扩展和完善的空间。如Cushion Treemaps方法解决了树图中对平衡树的问题。Harri Siirtola则对平行坐标法的交互功能进行了改进，并增加了平行坐标法的动态求和及聚类功能[5]。同时，很多研究技术还没有开发出相应的产品，这也是需要完善的地方。

参考文献

1 Inselberg A,Dimsdale B.Parallel Coordinates:A Tool for Visualizing Multi-dimensional Geometry.Visualization '90,San Francisco,CA, 1990:361

2 Keim D A.Pixel-oriented Visualization Techniques for Exploring Very Large Databases.Journal of Computational and Graphical Statistics, 1996,5(1):58

3 Pickett R M,Grinstein G G.Iconographic Displays for Visualizing Multidimensional Data,Proc.IEEE Conf. on Systems,Man and Cybernetics.IEEE Press,Beijing and Shenyang,1988-05:514

4 Shneiderman B.Tree Visualization with Treemaps:A 2D Space-Filling Approach.ACM Transactions on Graphics,1992,11(1):92

5 Siirtola H.Direct Manipulation of Parallel Coordinates.Proceedings of the IEEE International Conference on Information Visualization (IV- 2000), London,2000-07:373