

# 基于 VaR 技术的多维数据交互可视化研究

陈式座, 樊明辉, 关玉梅

(福州大学 福建省空间信息工程研究中心 数据挖掘与信息共享  
教育部重点实验室, 福建 福州 350002)

**摘要:**传统多维数据可视化技术比如平行坐标法等,在表达多维数据时,对各个不同数据维之间的数据表达式会产生一定的混乱,同时对各维的相关性与交互表达比较困难。研究并设计了基于 VaR 技术的多维数据交互可视化小工具,其充分利用 VaR 技术所实现的丰富导航与选择、突出等工具来形象突出显示与表达多维数据中某一特定维的数据,同时应用 MDS 算法,在二维屏幕上各点的相对位置关系来表示多维数据中各维的相关性。

**关键词:** VaR 技术; MDS 算法; 多维数据; 交互可视化

**中图分类号:** TP311.13

**文献标识码:** A

**文章编号:** 1673-629X(2007)08-0005-03

## Research on Multidimensional Data Interaction Visualization Based on VaR Display Technique

CHEN Shi-zuo, FAN Ming-hui, GUAN Yu-mei

(Ministry of Edu. Key Lab. for Data Mining & Info. Sharing, Spatial Information  
Research Center, Fuzhou University, Fuzhou 350002, China)

**Abstract:** The traditional multi-dimensional data visualization technologies such as parallel coordinates in the expression of multidimensional data, the data among the various data-dimensional expression would produce a certain confusion and the relevance of the peacekeeping expressed difficulty with the rest of the world. In this paper, design a VaR technology multidimensional data world visualization tool, make full use of VaR technology to achieve the rich navigation and selection, highlighting the tools to image highlights and expression of multi-dimensional data of a particular dimension of data and apply multi-dimensional scaling algorithm in two-dimensional screen at various points in the relative position to show multidimensional data peacekeeping relevance.

**Key words:** value and relation technique; multi-dimensional scaling arithmetic; multidimensional data; interaction visualization

### 0 引言

对多维数据的可视化可以让人易于从复杂的数据中发现规律,从而去发现挖掘多维数据之间的关联性、隐含信息和发展趋势,利用数据信息的直观表达,进而做出及时和正确的判断和决策。

### 1 面向像素可视化技术

面向像素数据可视化技术(Pixel-oriented Visualization Technique)是很重要的一种数据可视化方法,该方法是由德国慕尼黑大学的 Keim 于 1996 年提出,同

时他以此方法为例开发了 VisDB 可视化系统<sup>[1]</sup>。面向像素技术的基本思想是将每一个数据项的数据值对应于一个带颜色的屏幕像素,对于不同的数据属性以不同的窗口分别表示。面向像素技术的特点在于能在屏幕中尽可能多地显示出相关的数据项,对于高分辨率的显示器来说,可显示多达  $10^6$  数量级的数据<sup>[1]</sup>。

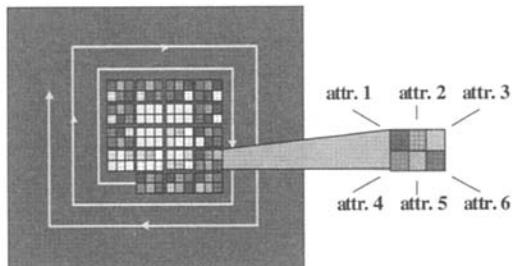


图1 面向像素可视化示意图<sup>[2]</sup>

图1就是一个典型的利用面向像素技术实现多维数据进行交互可视化表达的方法,该方法是将每一

收稿日期: 2006-10-12

基金项目: 福建省教育厅重点项目(K04001); 福建省青年科技创新项目(2004J005)

作者简介: 陈式座(1981-),男,福建永春人,硕士研究生,研究方向为空间数据可视化与空间决策支持系统;樊明辉,博士,研究方向为空间数据挖掘、地理信息网络共享等。

个数据值映射成一个有色的像素并将数据值按照大小集中显示在分离的窗口中,数据值的范围也根据一个固定的颜色表映射成了像素。面向像素每个维度里面的信息采用不同颜色的像元符号表示,代数据集值的像元从屏幕的中央螺旋地向四周展开。

## 2 VaR 可视化技术

### 2.1 VaR 可视化技术概述

文中所介绍 VaR(Value and Relation)技术是面向像素可视化技术中典型的一种<sup>[3]</sup>。VaR 技术将多维空间中的每一个维用 2D 空间中的一个像元符号(Glyph)表示。VaR 技术将多维数据中的数据以图像(主要是像素)的形式表示,在屏幕中尽可能多地表现出多维数据所包含的信息(包括各个维度之间相互的关系,每个维度的数据项大小等等),并通过分析和开发工具发现其中隐含或未知信息(各个维度之间的关联性、相似性等等)<sup>[4]</sup>。

VaR 多维数据可视化工具可以实现以下功能:

(1) 多维数据各个维度相关性(关联性等等)的可视化表达(在二维坐标上表示出来);

(2) showing name: 移动鼠标到符号上的时候,相应的维的名称将显示在状态栏上;

(3) Layer reordering: 鼠标点击符号,该像素将突出显示出来;

(4) Manual Locating: 用户可以手工拖动符号到屏幕上任意位置,将相互重叠(Overlap)的像元符号分开;

(5) Extent Scaling: 缩放功能,可以突出某个维度的信息。

VaR 可视化技术在对多维数据的交互可视化的可视化表达具有良好的效果。通过 VaR 技术可以将多维数据库中所有大数据量的数据经过预处理以后,采用 MDS 算法将多维数据在平面上可视化各个维度自己的相关性,挖掘多维数据中所包含的隐含信息进而将每个维度中大数据量进行可视化。

### 2.2 MDS 算法

VaR 技术在实现对多维数据的交互可视化表达中运用了经典的 Multi-dimensional Scaling(以下简称 MDS)算法,现对于 MDS 算法进行简要分析。MDS(多尺度扩展分析)技术是“探索”和“观看”多维数据的强有力的方法,最早 Torgerson 在 20 世纪 50 年代中期提出,后由 R. N Shephard 加以发扬,再经 C. H. Combs, T. Bennett 和 J. B Green 及其合作者则将此技术用于市场研究中(R. J. KUO, 2002)。

MDS 算法通过低维空间(通常是二维空间)展示

多个研究对象之间的相互关联性,利用其在二维平面之间的距离与相互关系来反映研究对象之间的相似程度或者相关程度<sup>[3]</sup>。

MDS(多尺度扩展分析)是市场研究中非常重要的一种方法,它通过低维空间(通常是二维空间)展示多个研究对象——商品之间的相互关联性,利用平面距离来反映研究对象之间的相似程度,并产生一张能够看出这些商品相关性的直观图(主要是在二维平面上显示)<sup>[5]</sup>。

MDS 算法可以分解为如下计算步骤:

a) 给定  $N$  个对象维及其它们之间的相对距离( $N \times N$  维矩阵):

$$d_{xy} = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (1)$$

b) 根据计算出来的两两维度之间的欧氏距离形成一个对称的矩阵:

$$\begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{12} & 0 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{2n} & \cdots & \cdots & a_{(n-1)(n-1)} \\ a_{1n} & \cdots & \cdots & 0 \end{pmatrix} \quad (2)$$

c) 根据对称的距离矩阵计算出该对称矩阵的对应的特征值与特征向量,将最大的前两个特征值  $\lambda_1, \lambda_2$ , 及两个特征值对应的特征向量  $\alpha_1, \alpha_2$  分别求出;

d) 将所得到的最大的前两个特征值绝对值的开方所组成的矩阵与对应的特征向量组成的矩阵相乘,得到一个  $n$  行 2 列的矩阵,这个矩阵就是反映各维之间相对关系的二维坐标值,在二维坐标系中得出  $k$  - dimension 空间中的  $N$  个点;

$$\begin{matrix} & X \text{ 坐标} & Y \text{ 坐标} \\ \begin{pmatrix} \sqrt{|\lambda_1|} & 0 \\ 0 & \sqrt{|\lambda_2|} \end{pmatrix} * \{\alpha_1, \alpha_2\} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \vdots & \vdots \\ \beta_{1n} & \beta_{2n} \end{pmatrix} \end{matrix} \quad (3)$$

这样,就可以根据计算结果得出二维坐标值,从而在二维平面上显示多维数据中各个维度之间的相关性。图 2 中反映出了各点的相对位置关系就代表多维数据中各维之间的相关性,该图应用 MDS 算法实现了从数据到散点图位置关系的转化的一一对应。

## 3 基于 VaR 技术的多维数据交互可视化

基于 VaR 技术的多维数据交互可视化方法,利用 MDS 算法计算最大的两个特征值对应的特征向量的做法,提供当前时间粒度上属性值的平面投影,生成表

示各个属性维信息的平面散点位置图。在散点的基础上,以欧氏距离为度量依据,提供基于距离的聚类技术以发现存在于空间属性数据集中的主要类别,并以此为基础,寻找空间属性数据集中的离群点。此外 VaR 数据可视化技术同时也支持三维或二维散点图表示并根据需要灵活地进行维度切换;支持散点图矩阵,根据需要即时进行平面剪切形成系列二维经典散点图矩阵(ScatterPlot Matrix),在二维平面上展示信息子空间的变化。VaR 可视化技术支持动态交互工具:提供投影、映射、开窗、缩放以及链接 & 弹刷(Linking & Brushing)等动态交互工具,实现与多维散点图紧密集成。将经典二维散点图进行三维扩展,以适应空间信息中多属性数据的三维可视表达<sup>[6]</sup>。

散点图根据需要即时进行平面剪切形成系列二维经典散点图矩阵(Scatter-Plot Matrix),实现相应的投影、映射、开窗、缩放以及链接 & 弹刷(Linking & Brushing)等动态交互工具,实现与多维散点图紧密集成。例如,有一组与长春这个地理位置相关属性数据—气象数据有温度、湿度、风力等。若屏幕上对应长春的像素为 a,那么这些数据就显示在像素 a 上。这个方法既直观又把数据与地理位置的相关性很好地表现出来<sup>[7]</sup>。

图 3 所表达的就是一个数据的可视化表达,各个维度中数据值的大小同时也表现出来,其大小与颜色成对应关系,其中每个小像元对应一个数值,像元从每个像素的屏幕中央螺旋地向四周展开,各个维度的信息集中于一个区域显示。而各维的相互关联性就各点在二维屏幕上的分布情况而得出。可以在该散点图进行放大等操作来突出表现重点维度的信息。图 4 表示的就是对于某一维的全部数据可视化效果。

4 结 论

文中深入讨论了面向像素可视化技术中典型的具有代表性的一种—VaR 可视化技术。该技术既能很好地在二维屏幕上表达各维数据之间的相关性,又能

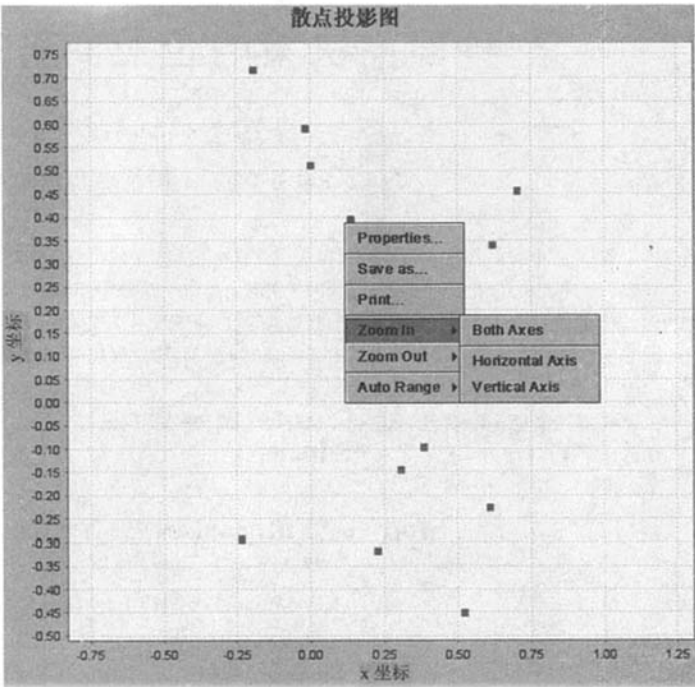


图 2 经 MDS 算法得出的散点图

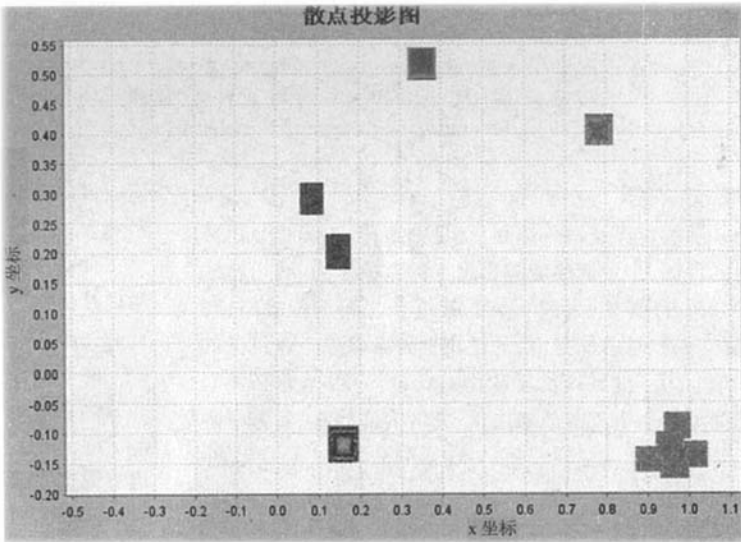


图 3 基于 VaR 技术多维数据交互可视化散点图

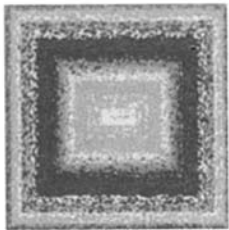


图 4 某一维度所有数据可视化效果

(下转第 164 页)

如果检索 GetElement(查找元素),与其相关联的前驱有 Data(数据)、DataElement(数据元素)、DataStructure(数据结构)和 LinearList(线性表),后继有 ListDelete(线性表删除)。检索 GetElement 知识点后得到的结果如图 3 所示。

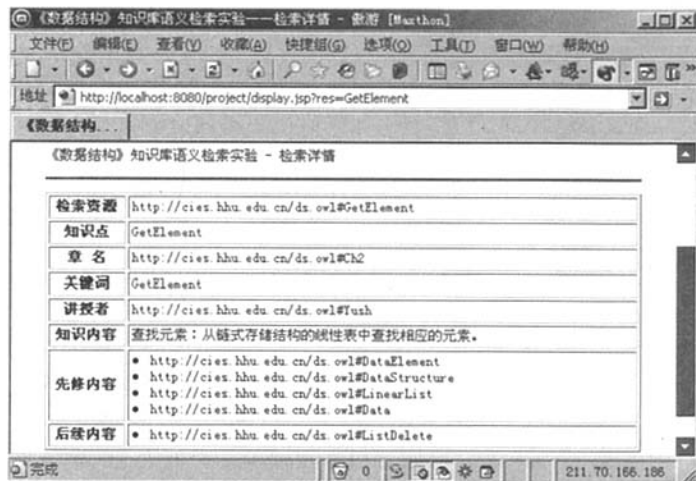


图 3 查询 GetElement 得到的检索详情

基于本体的语义检索与传统检索方式比较,语义检索是在知识(概念)层面上进行检索,刻画了概念之间的内在联系,可以发掘出那些隐含的、不明确的信息和概念,在查全查准方面可以更好地满足用户要求。如查询某章的所有内容时,与传统的基于关键字是不同的,它是通过概念之间的包含(类与子类、类与实例等)关系来进行查找的,可以有效地控制检索的范围,从而不仅可以提高查全查准率,还可以大大地提高系统的检索速度。

### 3 小 结

知识库是 ITS 的基础,知识库的知识表示问题是所有人工智能和智能化系统必须首先解决的问题之一。文中将本体与教学知识库建模结合起来,探讨知识共享和互操作问题,并研究了基于本体的语义检索技术。其中还有很多具体的工作要做,如本体的正确性检查,在大规模压力测试下,检索系统的性能评估等,将作为进一步工作的方向。

#### 参考文献:

- [1] Mizoguchi R, Jacqueline. Using Ontological Engineering to Overcome Common AI - ED Problems[J]. International Journal of Artificial Intelligence in Education, 2000, 11: 107 - 121.
- [2] 邓志鸿,唐世渭. Ontology 研究综述[J]. 北京大学学报:自然科学版, 2002, 38(5): 730 - 738.
- [3] Gruber T R. A translation approach to portable ontologies[J]. Knowledge Acquisition, 1993, 5(2): 199 - 220.
- [4] Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing[J]. International Journal of Human - computer Studies, 1995, 43: 907 - 928.
- [5] Borst W N. Construction of Engineering Ontologies for Knowledge Sharing and Reuse [D]. Enschede: University of Twente, 1997.
- [6] Fensel D. The semantic web and its languages[J]. IEEE Computer Society, 2000, 15(6): 67 - 73.

(上接第 7 页)

利用 VaR 技术丰富的交互表达与变换工具,将各维的具体数据值采用像元进行表达,挖掘多维数据中的隐含信息与异常点,表现多维数据中各维的相关性,具有深入研究意义与价值。

#### 参考文献:

- [1] Keim D A, Kriegel H - P. Visdb: Database exploration using multidimensional visualization[J]. IEEE Computer Graphics & Applications, 1994, 14(5): 40 - 49.
- [2] Keim D A. Designing pixel - oriented visualization techniques: Theory and applications[J]. IEEE Transactions on Visualization and Computer Graphics, 2000, 6(1): 1 - 20.
- [3] Feldt N, Pettersson H, Johansson J, et al. Tailor - ma de Exploratory Visualization for Statistics [R]. Sweden: NVIS -

Norrköping Visualization and Interaction Studio, Linköping University, 2000.

- [4] Yang Jing, Patro A, Huang Shiping, et al. Value and Relation Display for Interactive Exploration of High Dimensional Datasets[C]//Rundensteiner E A. Information Visualization, 2004. INFOVIS 2004. IEEE Symposium. [s. l.]: [s. n.], 2004: 73 - 80.
- [5] Wickelmaier F. Sound Quality Research Unit[C]//An Introduction to MDS, IEEE Symposium. Denmark: Aalborg University, 2003: 57 - 65.
- [6] 刘 勤,周晓峰,周洞汝. 数据可视化的研究与发展[J]. 计算机工程, 2002(23): 61 - 63.
- [7] 冯玉才,刘 嘉. 大量空间数据可视化的算法[J]. 计算机工程, 2003(29): 79 - 81.