

数据可视化技术的研究与进展

任永功^{1,2} 于 戈¹

(东北大学计算机科学与工程学院 沈阳110004)¹ (辽宁师范大学计算机与信息技术学院 大连116029)²

摘 要 可视化技术对于分析和探究大规模的多维数据集变得越来越重要。本文首先介绍了数据可视化的概念及其发展历程,然后分类、对比分析了已存在的可视化技术,最后对数据可视化技术的未来发展方向进行了探讨。

关键词 数据可视化,可视化技术,可视化数据挖掘

Research and Development of the Data Visualization Techniques

REN Yong-Gong^{1,2} YU Ge²

(School of Computer Science & Engineering, Northeastern University, Shenyang 110004)¹

(School of Computer & Information Technology, Liaoning Normal University, Dalian 116029)²

Abstract Data visualization techniques are of increasing importance in exploring and analyzing large amounts of multidimensional information. This paper 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. In the end, the development and further directions of data visualization techniques are proposed.

Keywords Data visualization, Visualization large data sets, Visual data mining

1 数据可视化的定义

科学计算可视化^[1]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。科学计算可视化(Visualization in Scientific Computation)是对计算及数据进行探索,以获得对数据的理解与洞察。也就是说,科学计算可视化实现把计算中所涉及的和所产生的数字信息转变为直观的、以图像或图形信息表示的、随时间和空间变化的物理现象或物理量呈现在研究者面前,使他们能观察到模拟和计算,即看到传统意义上不可见的事物或现象;同时还提供与模拟和计算的视觉交互手段。

随着科学技术的发展,特别是计算机技术的迅猛发展,人类产生与获取数据的能力成数量级地增加。面对海量的数据,想通过人工地分析这些数据从而得以深刻地理解并进一步形成正确的概念和看法几乎是不可能的了。人们需要新的技术来帮助理解这巨大数量的数据。数据可视化技术正是在这样的背景下获得了人们越来越大的重视和迅速发展。所谓数据可视化是对大型数据库或数据仓库中的数据的数据的可视化,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示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。数据可视化技术凭借计算机的巨大处理能力及计算机图像和图形学基本算法以及可视化算法把海量的数据转换为静态或动态图像或图形呈现在人们的面前,并允许通过交互手段控制数据的抽取和画面的显示,使隐含于数据之中不可见的现象成为可见,为人们分析、理解数据、形成概念、找出规律提供了强有力的手段。

2 数据可视化技术特点及其分类

怎样来分析大量、复杂和多维的数据呢?答案是要提供像人眼一样的直觉的、交互的和反应灵敏的可视化环境。因此,数据可视化技术的主要特点是:

(1)交互性。用户可以方便地以交互的方式管理和开发数据。

(2)多维性。可以看到表示对象或事件的数据的多个属性或变量,而数据可以按其每一维的值,将其分类、排序、组合和显示。

(3)可视性。数据可以用图像、曲线、二维图形、三维体和动画来显示,并可对其模式和相互关系进行可视化分析。

对于那些本身就可以有两维或三维语义的数据进行可视化,在利用计算机进行可视化之前就已经使用多年了。自从计算机开始应用于可视化技术以后,人们发现了许多新颖的可视化技术,现有的技术也得到了改进,而且应用领域扩展到了大规模的数据集可视化及动态交互显示等方面。然而,对于大多数存储在数据库中的数据来讲,因为数据没用固定的二维或三维语义特性,所以至今也没有一种标准的方法把数据映射到笛卡尔坐标系中。通常,关系数据库被看作多维属性数据的集合,每维属性与坐标的维数相对应。数据可视化技术能用三维正交标准来描述^[2]:可视化技术、扭曲技术,交互技术,如图1所示。

图1所描述的正交性意味着:任何可视化的技术都与扭曲技术和交互技术协同工作。可视化技术分为基于几何技术、基于图标技术、面向像素技术、分层技术、基于图形技术、混合技

*)基金项目:辽宁省教育厅高等学校科研基金项目(A202122027)。任永功 讲师,博士生,研究方向为计算机视觉、数据可视化。于 戈 教授,博士生导师,主要研究领域为数据库系统理论与技术。

术等。众所周知的几何技术包括：散点矩阵^[3]、coplots^[3]、地形图^[4]、投影-截面组合视图技术^[5]、多维切片^[6]、平行坐标^[7]；基于图标技术包括枝形图^[8]、形状编码^[9]、彩色图标^[10]；面向像素技术包括：螺旋技术^[11]、递归模型技术^[12]、圆环分割技术^[13]；分层技术包括：多维堆积技术^[14]、树图^[15]、锥型树^[16]；基于图形技术包括簇形图、对称优化技术和分层图示^[17]。除可视化技术以外，对于数据的分析处理，交互技术和扭曲技术也是非常重要的。交互技术允许用户直接干预可视化过程，例如：映射^[18]、投影^[18]、过滤^[19]、缩放^[20]和连接 & 刷新^[21]。交互技术允许用户根据操作对象的特性动态改变可视化过程，而且也可以联合多种相对独立的可视化技术进行多样性操作，以便于提供更多的信息。扭曲技术的基本思想是显示所有数据的同时有选择地详细显示某局部数据，许多简单和复杂的扭曲技术已经产生并得以应用^[22]，例如：透视墙^[23]、双焦点透

镜^[24]、表格透镜^[25]、fisheye 视图^[26]、双曲线树^[27]和多维箱技术^[28]。为了更加详细而结构化地了解不同数据可视化技术的特性，对各种技术从可视化数据特性、可视化任务特性和可视化技术特性三方面做出了比较^[29]，如图2所示。

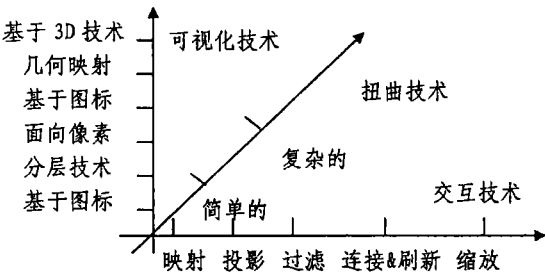


图1 可视化技术的分类图

测试项		聚	多元	变量	数据项	分类	可视化	学习
可视化技术分类		类	分析	个数	个数	数据	重叠	曲线
几何技术	散点矩阵	++	++	+	+	-	o	++
	山地图	+	+	-	o	o	+	+
	投影-截面组合视图	++	++	+	+	-	o	+
	多维切片	+	+	+	+	-	o	o
	平行坐标	o	++	++	-	o	--	o
基于图 标技术	枝形图	o	o	+	-	-	-	o
	形状编码	o	-	++	+	-	+	-
	彩色图标	o	-	++	+	-	+	-
面向像 素技术	独立于查询	+	+	++	++	-	++	+
	基于查询	+	+	++	++	-	++	-
分层技术	多维堆积图	+	+	o	o	++	o	o
	树图	+	o		++	++		o
	锥型树	+	+	o	+	o	+	
基于图 形技术	基本图形	o	o	-	+	o	o	+
	特殊图形	++	+	-	+	o	+	+

图2 可视化技术比较图

其中：++代表优，+代表良，o代表中，-代表较差，--代表差。很显然散点矩阵可视化技术，独立于查询可视化技术，基于查询可视化技术三种方法综合能力较强。

3 数据可视化的主要技术

我们早已熟悉一些显示技术，如：柱形图、条形图、折线图、散点图、面积图、圆环图以及曲面图、股价图等。目前国际上有许多种数据可视化技术，下面介绍的是几种主要的技术。

3.1 基于几何投影技术

基于几何投影技术的目的是在多维数据集中找到“有意

义”的投影，是以几何画法或几何投影的方式来表示数据库中的数据，主要技术包括 scatter plots、Landscapes、Projection Pursuit、Parallel Coordinates 等等。平行坐标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^[7]。其基本思想是通过使用相互平行而且等距的坐标轴将 K 维空间降维映射成二维来显示。这些坐标轴分别对应于不同的空间维，并且从对应维的最小值到最大值进行线性变化。每个数据项有一根折线表示，该折线与每个坐标轴都有一个交点，交点值就是该数据项在坐标轴对应的值，如图3所示。

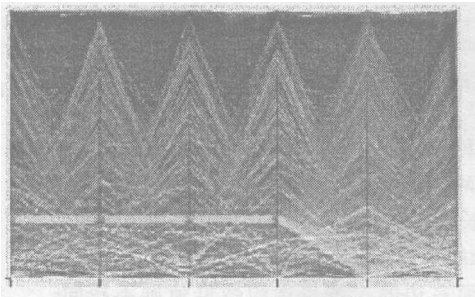
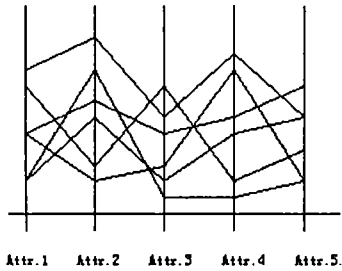


图3 平行坐标技术

利用平行坐标法开发的系统包括 Parallel Visual Explorer (IBM)、XMDX (Mam Ward)、AVS/Express (Van Wijk) 等等。这个方法的基本思想非常简单,对于较少的数据集能使用户在二维平面上看到每个数据的 n 维属性,对于大型数据集能反映出各维属性之间的关系和数据在各维属性之间的走向趋势。但是,因为折线可能重叠,所以在使用该技术时,同时显示在屏幕的数据项的个数是有一定限制的,一般大约1000个数据项。

3.2 面向像素技术

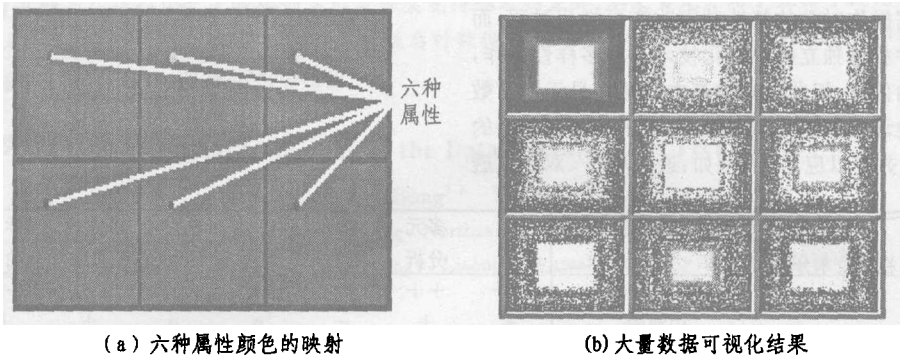


图4 面向像素技术图示

面向像素的可视化方法包含独立于查询的方法和基于查询的方法两种。独立于查询的方法将根据某些属性对数据排序,并使用一个填充屏幕的模式在显示器上排列数据值。对于根据某个属性进行自然排序的数据特别有效。基于查询的方法是对某个查询产生的数据相关性进行可视化,根据数据值同所查询的要求的符合程度来匹配颜色。针对每一个数据项的值 (a_1, a_2, \dots, a_n) 及查询要求 (q_1, q_2, \dots, q_n) 通过一个距离函数 $distances(d_1, d_2, \dots, d_n)$ 计算每个属性值与查询要求的匹配值,得到每一个数据的一个总的距离值 d_{n-1} 以反映数据项与查询要求之间的匹配程度,总的距离 d_{n-1} 越小越是用用户所希望看到的数据。查询数据结果按 d_{n-1} 的值由小到大从屏幕的中央螺旋地向四周展开。这样不仅能看到所查询的数据,而且对于数据从近似匹配到不匹配的走势也能直观地表现。用户可以发现属性之间的相关性、功能依赖以及其他一些有趣

的关系。

3.3 基于图标的技术

基于图标技术的基本思想是将每一个多维数据项映射成一个图标,即用简单图标的各个部分来表示 n 维数据的属性。基于图标的可视化技术包括 Chernoff-face、Shape Coding、Stick Figures 等,这种技术适用于某些维值在二维平面上具有良好展开属性的数据集。Chernoff-faces 是著名的图标显示技术,基本思想是数据项的两位被映射成两个用于显示的坐标维,而剩下的维则被映射成一张脸的各个器官——鼻子、嘴巴、眼睛的形状以及脸部本身的形状,如图5(a)。所示此种方法的优点是利用了人类类似于脸和脸部特征的敏感性,从而达到了良好的可视化效果。但是,能用 Chernoff-faces 技术进行可视化的数据集非常有限。

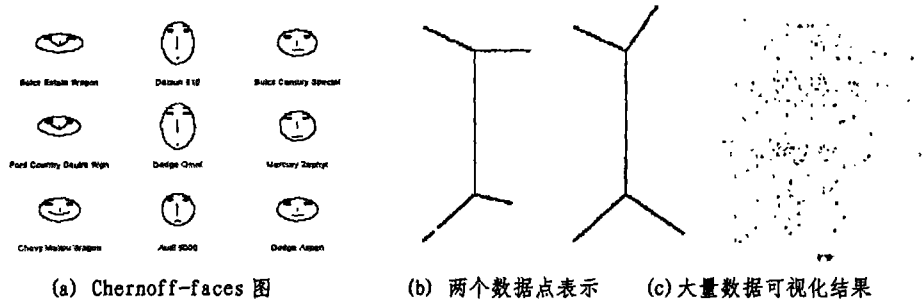


图5 图标技术可视化结果

枝形图也是一种有效的可视化方法^[8]。枝形图方法首先选取多维属性中的两种属性作为基本的 x - y 平面轴,在此平面上利用小树枝的长度或角度的不同表示出其他属性值的变化。例如图5(b)所示的两个数据点,它们对左边的二维属性含有相同的数据值,而右边的二维属性的数据值则不相同。ExVis 是根据枝形图方法开发的可视化系统,图5(c)显示的是以收入为横轴、年龄为纵轴的数据可视化结果。

3.4 基于层次和图形可视化技术

基于层次的可视化技术主要针对数据库系统中具有层次结构的数据信息,例如人事组织、文件目录、人口调查数据等。著名的技术有 n -Vision 技术、Dimensional Stacking^[14]、Treemap^[15]、Cone Trees^[16]等。它的基本思想是将 n 维数据空间划分为若干子空间,对这些子空间以层次结构的方式组织并以图形表示出来。例如,Dimensional Stacking 技术就是将 K 维空间细分成多个二维空间。树图是其中的一种代表技术。如图6所示,图中每一个结点都有一个名称和数值大小,父

结点是各子结点大小的总和。

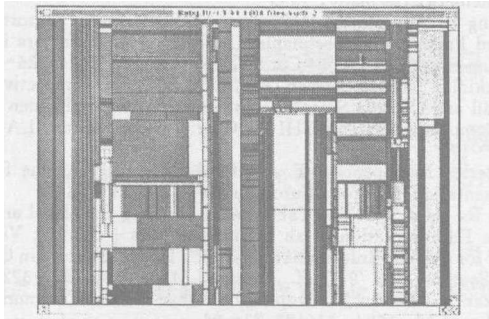


图6 树图可视化结果

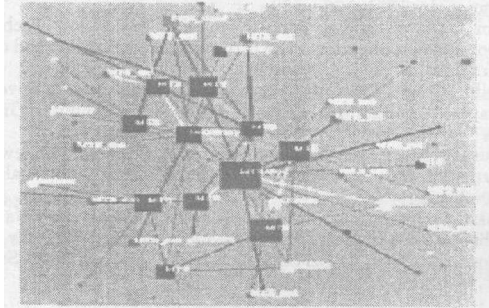


图7 基于图形技术效果图

树图根据数据的层次结构将屏幕空间划分成一个个矩形子空间,子空间大小由结点大小决定。树图层次则依据由根结点到叶结点的顺序,水平和垂直依次转换,开始将空间水平划分,下一层将得到的子空间垂直划分,再下一层又水平划分,依次类推。对于每一个划分的矩形可以进行相应的颜色匹配或必要的说明。

利用基于分层可视化技术开发的系统主要有:Hyperbolic Trees(Xerox),Info Cube(spn),Elastic Windows(HCIL-Maryland),TreeMap(HCIL-Maryland)等。

基于图形技术的基本思想是使用特殊的页面编排法、查询语言和抽象技术有效结合显示一个大的图形。主要例子有Hy+,Margritte,SeeNet等。可视化结果可参看图7。

3.5 其他的数据可视化技术

数据可视化还有很多的技术和方法。DVET(Data Visualization Environment Tool)系统利用了虚拟现实技术展示数据空间和空间上的点(数据)。HD-Eye^[28]算法结合了数据挖掘技术,先将数据分簇,再对感兴趣的簇可视化。Table Lens系统仍然以表的形式表现数据,但是以图示代替了表中的数字,并且给出观察的视点,易于用户选择和操纵数据表中的区域。Xgobi系统强调动态和交互技术,用户能同时以不同的可视化方法处理相同的数据。另外,基于3D技术的更多新方法和技术也正在被研究和开发。

4 数据可视化技术的发展方向

4.1 允许用户更方便地进行交互的动态可视化

目前商业软件中提供的可视化在本质上大多是静态的。数据被一次读取然后显示出来。用户很难和可视化图形进行交互。动态可视化允许用户对数据可视化进行旋转、缩放和筛选等操作。在三维可视化中,一些简单的交互方式如旋转、缩放增加信息展示的角度,从而可以提高用户获得有用认识的能力。更多更复杂的可视化可以让用户将多个可视化图形进

行合并,甚至可以把可视化作为一个查询的工具。刷图可以同步一个数据集上的两个或者更多可视化。例如:如果拥有同一个数据集的散点图和雷达图,那么当在某一个可视化图形上选择了一些数据的时候,系统会自动将这些数据在另一个可视化图形上的分布亮度中显示出来。这种强大的可视化联合方法让用户可以拥有多个单一数据集上的同步视图。如果一个可视化允许用户通过点击或者通过画一个矩形的形式选择特定的数据点,那利用这种功能将可以把这些数据点形成一个子集,这个选择的过程称为钻取。类似地,在散点图中,可以在感兴趣的点周围绘制一个多边形,从而形成一个可视化查询。因为可视化数据挖掘项目经常需要不断地重复之前所描述的一些步骤,所以从可视化图形中抽取和存储多个数据子集将是非常有好处的。动态可视化还可以延伸到对不断变化的数据进行可视化。

4.2 可视化数据规模增大和数据结构多样化

对于大数据的定义一直在不断地增长。现在很多的企业都拥有TB数量级的数据,虽然VDM(可视化数据挖掘)方法中有一个步骤是数据转换,包括删除一些不重要的字段和行,但是剩下的字段和行还是会非常多的。传统的可视化技术如柱图形在显示含有大量行和列的数据的时候会出现问题,除非这些数据预先经过了汇总。这是因为在这些可视化图形中,基本的可视单元是有一定大小的多边形(如柱和饼),从而限制了可以显示的信息数量。基于像素的技术,如环形细分,其可视化中的每个像素都包含了一定的信息,从而也就在屏幕上显示了更多的信息。大多数的可视化技术只能对含有固定数量字段的数据进行可视化。对非结构化的数据进行可视化,如文本或者含有不定数量字段的数据,使得数据挖掘项目在传统的领域得到进一步的应用,并且向新的领域拓展。文本文档的可视化是一个有趣的挑战,因为其内在的复杂性意味着利用可视化即可以对获得的认识进行验证,同时还可以浏览这些文档的集合。

4.3 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。在实际应用中,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的应用。

参考文献

- 1 李晓梅,黄朝晖,等编著.科学计算可视化导论.长沙:国防科技大学出版社,1996.6
- 2 Keim D A, et al. State-of-the-Art Report Visual DataMining: [State of the Art Report]. Eurographics Conference 2002, Saarbrücken, Sep. 2002
- 3 Cleveland W S. Visualizing Data. AT&T Bell Laboratories, Murray Hill, NJ, Hobart Press, Summit NJ, 1993
- 4 Wright W. Information Animation Applications in the Capital Markets. In: Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995. 19~25
- 5 Furnas G W, Buja A. Prosections Views: Dimensional Inference through Sections and Projections. Journal of Computational and Graphical Statistics, 1994, 3(4): 323~353
- 6 van Wijk J J, van Liere R D. Hyperslice. In: Proc. Visualization '93, San Jose, CA, 1993. 119~125
- 7 Inselberg A, Dimsdale B. Parallel Coordinates: A Tool for Visualizing Multi-Dimensional Geometry. In: Visualization '90, San Francisco, CA, 1990. 361~370
- 8 Pickett R M, Grinstein G G. Iconographic Displays for Visualizing Multidimensional Data. In: Proc. IEEE Conf. on Systems, Man and Cybernetics. IEEE Press, Piscataway, NJ, 1988. 514~519

- 9 Beddow J. Shape Coding of Multidimensional Data on a Microcomputer Display. In: Visualization '90, San Francisco, CA, 1990. 238~246
- 10 Levkowitz H. Color icons: Merging color and texture perception for integrated visualization of multiple parameters. In: Visualization '91, San Diego, CA, Oct. 1991
- 11 Keim D A, Kriegel H-P. VisDB: Database Exploration using Multidimensional Visualization. Computer Graphics & Applications, Sept. 1994. 40~49
- 12 Keim D A, Kriegel H-P, Ankerst M. Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data. In: Proc. Visualization '95, Atlanta, GA, 1995. 279~286
- 13 Ankerst M, Keim D A, Kriegel H P. Circle Segments: A Technique for Visually Exploring Large Multidimensional Data Set. In: Proc. Visualization '96, 1996
- 14 LeBlanc J, Ward M O, Wittels N. Exploring N-Dimensional Databases. In: Visualization '90, San Francisco, CA, 1990. 230~239
- 15 Johnson B. Visualizing Hierarchical and Categorical Data. [Ph. D. Thesis]. Department of Computer Science, University of Maryland, 1993
- 16 Robertson G G, Mackinlay J D, Card S K. Cone Trees: Animated 3D Visualizations of Hierarchical Information. In: Proc. Human Factors in Computing Systems CHI '91 Conf., New Orleans, LA, 1991. 189~194
- 17 Battista G D, Eades P, Tamassia R, Tollis I. Annotated Bibliography on Graph Drawing Algorithms. Computational Geometry: Theory and Applications, 1994. 4: 235~282
- 18 Buja A, Swayne D F, Cook D. Interactive High-Dimensional Data Visualization. Journal of Computational and Graphical Statistics, 1996, 5(1): 78~99
- 19 Fishkin K, Stone M C. Enhanced Dynamic Queries via Movable Filters. In: Proc. Human Factors in Computing Systems CHI '95 Conf., Denver, CO, 1995. 415~420
- 20 Ahlberg C, Wistrand E. IVEE: An Information Visualization and Exploration Environment. In: Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995. 66~73
- 21 Wilhelm A, Unwin A R, Theus M. Software for Interactive Statistical Graphics - A Review. In: Proc. Int. Softstat '95 Conf., Heidelberg, Germany, 1995
- 22 Leung Y, Apperley M. A Review and Taxonomy of Distortion-oriented Presentation Techniques. In: Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994. 126~160
- 23 Mackinlay J D, Robertson G G, Card S K. The Perspective Wall: Detail and Context Smoothly Integrated. In: Proc. Human Factors in Computing Systems CHI '91 Conf., New Orleans, LA, 1991. 173~179
- 24 Apperley M, Spence I T. A Bifocal Display Technique for Data Presentation. In: Proc. Eurographics, 1982. 27~43
- 25 Rao R, Card S K. The Table Lens: Merging Graphical and Symbolic Representation in an Interactive Focus + Context Visualization for Tabular Information. In: Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994. 318~322
- 26 Sarkar M, Brown M. Graphical Fisheye Views. Communications of the ACM, 1994, 37(12): 73~84
- 27 Munzner T, Burchard P. Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space. In: Proc. VRML '95 Symp., San Diego, CA, 1995. 33~38
- 28 Alpern B, Carter L. Hyperbox. In: Proc. Visualization '91, San Diego, CA, 1991. 133~139
- 29 Keim D A. Designing Pixel-Oriented Visualization Techniques: Theory and Applications. IEEE trans. on visualization and computer graphics. 2000
- 30 Abello J, Korn J. MGVS: A system for visualizing massive multidigraphs. Transactions on Visualization and Computer Graphics, 2001
- 31 Kreuseler M, Lopez N, Schumann H. A Scalable Framework for information Visualization. In: Proc. InfoVis'2000, Salt Lake City, 2000. 27~36
- 32 Stasko J, Zhang E. Focus + Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations. In: Proc. IEEE Information Visualization 2000, Salt Lake City, UT, Oct. 2000. 57~65

(上接第85页)

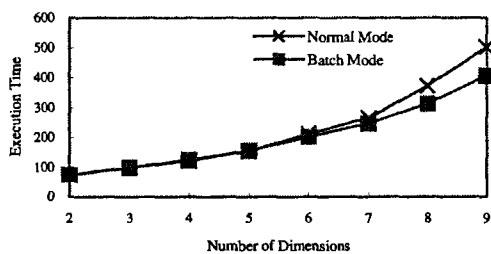


图4 气象数据集

从图中可以看到,在维个数较少的数据集上,一般模式下和批处理模式下计算 PrefixCube 所用的时间相当,随着维个数的增加,批处理模式计算逐渐优于一般模式下的计算。当数据集的维个数为9时,一般模式计算 PrefixCube 需要501秒,而批处理模式计算 PrefixCube 仅需要405秒,比前者少了96秒。这是因为,天气数据集是按维基数从大到小排序的,则在头几维上做划分时,由于维的基数较大,分组中元组的个数就相对较少,因此成批生成的元组个数就少,从而并不能给计算时间的加快带来多少好处;而随着数据集的维个数逐渐增多,由于后几维的基数较小,则在其上做划分后分组中元组的个数要多一些,成批生成的元组个数也会多一些,此时才体现出了批处理模式的优势,降低了 PrefixCube 计算的时间代价。

结论 在文[12]中提出的 PrefixCube 结构,不仅能有效减小数据立方的尺寸大小,而且在立方尺寸压缩比例,立方恢复和更新,以及立方查询三者之间得到了很好的平衡。可是在计算 PrefixCube 时,为了识别元组间的共享前缀而不得不进行大量的元组之间的比较,因此降低了 PrefixCube 的计算时间效率。在本文中,我们提出了以批处理模式计算 PrefixCube 的方法,实现了批量生成一组属于同一小方、且共享某个前缀的元组,从而减少了组内元组之间的比较,有效地缩短

了 PrefixCube 的计算时间。我们的实验结果表明,在大多数数据集上,批处理模式的引入都能有效地降低 PrefixCube 的计算时间代价。在真实气象数据集的实验中,在最大的9维气象数据集上,批处理模式计算 PrefixCube 所需的时间比一般模式下计算 PrefixCube 所需的时间要少得多。

参考文献

- 1 Gray J, Bosworth A, Layman A, Pirahesh H. Data Cube: A Relational Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. In: Proc. of the Int. Conf. on Data Engineering, 1996
- 2 Beyer K, Ramakrishnan R. Bottom-Up Computation of Sparse and Iceberg CUBEs. In: Proc. of the ACM SIGMOD Int. Conf. on Management of Data, 1999
- 3 Poess M, Potapov D. Data Compression in Oracle. In: Proc. of Int. Conf. on Very Large Data Bases, 2003
- 4 Table Compression in Oracle9i Release 2: A Performance Analysis. Oracle Whitepaper, 2003
- 5 Gray J, Sundaresan P, Englert S, Baclawski K, Weinberger P J. Quickly Generating Billion-Record Synthetic Databases. In: Proc. of the ACM SIGMOD Int. Conf. on Management of Data, 1994
- 6 Lakshmanan L V S, Pei J, Han J. Quotient Cube: How to Summarize the Semantics of a Data Cube. In: Proc. of Int. Conf. on Very Large Data Bases, 2002
- 7 Sismanis Y, Deligiannakis A, Roussopoulos N, Kotidis Y. Dwarf: Shrinking the PetaCube. In: Proc. of the ACM SIGMOD Int. Conf. on Management of Data, 2002
- 8 Wang W, Feng J, Lu H, Yu J X. Condensed Cube: An Effective Approach to Reducing Data Cube Size. In: Proc. of the Int. Conf. on Data Engineering, 2002
- 9 Lakshmanan L V S, Pei J, Zhao Y. QC-Trees: An Efficient Summary Structure for Semantic OLAP. In: Proc. of the ACM SIGMOD Int. Conf. on Management of Data, 2003
- 10 Feng J, Si H, Feng Y. Indexing and Incremental Updating Condensed Data Cube. In: Proc. of the Int. Conf. on Scientific and Statistical Database Management, 2003
- 11 Hahn C, Warren S, London J. Edited synoptic cloud report from ships and land stations over the globe. <http://cdiac.esd.ornl.gov/cdiac/ndps/ndp026b.html>. <http://cdiac.esd.ornl.gov/ftp/ndp026b/SEP85L.DAT.Z>
- 12 Feng J, Fang Q, Ding H. PrefixCube: Prefix-sharing Condensed Data Cube. To appear: ACM International Workshop on Data Warehousing and OLAP, 2004