

可视化技术在空间数据挖掘中的应用

王占刚^{1,2}, 庄大方¹, 邱冬生¹

(1. 中国科学院地理科学与资源研究所资源环境数据中心, 北京 100101; 2. 中国科学院研究生院, 北京 100039)

摘要: 在阐述可视化与空间数据挖掘关系的基础上, 探讨了可视化在空间数据挖掘过程中应用的各个环节, 提出了将具体应用划分为概念层、逻辑层和基础层 3 个层次。以地质模型数据挖掘为例, 对应 3 个层次阐述了可视化应用的关键技术: 地质模型可视化, 交互式挖掘与探索性可视化分析。开发了一个原型系统, 初步实现了可视化挖掘功能。

关键词: 可视化; 空间数据挖掘; 人机交互; 探索性可视化分析; 地质模型挖掘

Application of Visualization Technology in Spatial Data Mining

WANG Zhan-gang^{1,2}, ZHUANG Da-fang¹, QIU Dong-sheng¹

(1. Data Center for Resources and Environment Sciences, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101; 2. Graduate School, Chinese Academy of Sciences, Beijing 100039)

【Abstract】 Based on the relation between visualization and spatial data mining, every step about the application of visualization in visual data mining is discussed. Application is divided into three levels: concept level, logistic level and base level. Taking geological model mining as an example, pivotal techniques are brought forward in accordance with the three levels: geological model visualization, interactive data mining and exploratory visual analysis. A prototype system is developed and visual data mining is realized.

【Key words】 visualization; spatial data mining; man-machine interaction; exploratory visual analysis; geological model mining

面对日益丰富的空间数据, 如何从中提取隐含的知识、空间关系和特征、发现人们未知的各种空间规律和趋势, 是空间分析领域急待解决的关键问题之一^[1]。数据挖掘(data mining, DM)是从大量数据中发现知识的有效途径, 但其算法本身非常复杂, 过程较为抽象, 挖掘的结果通常难以理解^[2]。可视化技术可以从复杂的多维数据中产生图形展示客观事物及其内在的联系, 从而丰富科学发现的过程^[3,4]。通过丰富的图形表现能力, 可视化技术能够准确地表达原始数据、挖掘结果、挖掘过程, 使用户深入地理解问题并选择更适当的数据挖掘算法, 达到深入剖析数据的目的。尤其面向空间数据, 可视化对于揭示空间实体的结构特征、相互关系及演变规律具有十分重要的意义。

1 可视化与空间数据挖掘的关系

地学可视化是科学计算可视化在地球科学领域的一个研究分支, 它的目的是在可视化表达、分析地学数据的基础上, 为用户提供一种信息交流与反馈的机制, 使其在观察数据外观的基础上更深层次地剖析其内在的信息和知识。空间数据挖掘从空间数据中提取隐含的知识、关系、模式和特征等, 帮助人们理解空间数据, 揭示空间特征, 发现人们未知的各种空间规律、关系和趋势^[5]。

可视化技术能够形象地展示空间信息及其内在属性, 从而挖掘数据间的规律性和发展趋势, 给予人们意想不到的洞察力。(1)可视化技术可以有效地揭示空间实体的结构特征、相互关系及演变规律, 使用户对数据的剖析更清晰、理解更深刻。(2)可视化技术能够为挖掘过程提供人机交互的手段, 实现用户与计算机之间交流与反馈的机制, 是将专家知识与经验作用于计算机的桥梁。(3)可视化技术可以为人们营造一个高效的空

间信息表达与分析环境, 为数据挖掘提供导航作

用, 辅助人们进行空间认知^[6]。通过该环境, 可以激发人的形象思维, 帮助人们寻找数据中的模式和特征, 加深用户对数据含义、数据之间的相互关系和发展趋势的理解^[7,8]。可见, 如果能充分利用可视化技术丰富的图形表达能力, 在准确表达原始数据的基础上, 通过人机交互融入用户自身的专家知识与经验, 实时地选择数据挖掘方法, 并将每一个过程清楚地表达出来, 就能让用户深入地理解问题并根据需求不断地改进挖掘方法, 使得整个空间知识发现经历一个逐步求精的优化过程^[6], 实现探索性数据分析的目的。

2 可视化在空间数据挖掘中的应用

2.1 应用环节

数据挖掘的步骤可分为 3 步: 数据准备, 数据挖掘和结果的解释评估^[2]。其中, 数据准备是对多种数据源进行清理、集成、选择和变换的预处理工作; 数据挖掘步骤是使用某种方法提取感兴趣的数据模式; 结果的解释评估是用符合用户习惯的表达方式向用户提供挖掘的知识, 并且依据某种兴趣度对模式进行度量。从空间数据挖掘过程和自身特点来看, 其每一步都需要可视化的支持。

结合传统数据挖掘的步骤, 本文将可视化技术在空间数据挖掘过程中的应用归纳为以下几个环节: 原始数据可视化, 人机交互操作, 数据挖掘过程表达以及挖掘结果的可视化。其对应于整个空间数据挖掘的过程, 如图 1 所示, 在对原始数据可视化的基础上进行相应的预处理工作, 形成预处理数

基金项目: 中国科学院知识创新工程基金资助项目(KZCX3-SW-357)

作者简介: 王占刚(1977 -), 男, 博士研究生, 主研方向: 地学可视化, 数据挖掘, GIS; 庄大方, 研究员、博士生导师; 邱冬生, 博士、副研究员

收稿日期: 2006-09-30 **E-mail:** wangzg@lreis.ac.cn

数据集,为用户提供直观的视觉感知;然后利用人机交互操作,将用户的挖掘意图直接作用于数据集生成中间结果;在数据挖掘过程中进行实时的可视化表达,同时借助于人形象思维能力进行探索性数据分析,循环改进挖掘方法与途径,获得挖掘结果;最终将数据挖掘结果可视化,获得所需的空间知识。

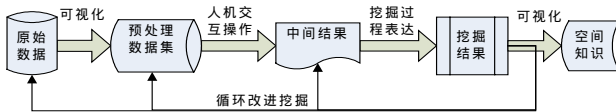


图1 可视化在空间数据挖掘中的应用环节

2.2 应用层次

依据挖掘目标的深度和广度可以将可视化在空间数据挖掘中的应用分为若干层次。MacEachen等人从概念层、操作层和实现层3个层次论述了可视化在数据挖掘中的应用^[9]:概念层的数据挖掘主要强调空间模式分析,辅助理解时空过程,注重于剖析隐含的时空特征。可视化在本层次的作用是在进行探索性可视化分析。操作层挖掘是将概念层的挖掘目标细分为一些子目标,并且每一个目标可以通过一些具体的操作实现,其重点在于挖掘过程的确定。实现层挖掘是在概念层与操作层的基础上实现挖掘目标的具体过程,是用户实际操作的挖掘工具。可视化在本层次中的作用是提供一个可视化的分析环境。

本文借鉴 MacEachen 的挖掘层次划分方法,根据可视化在空间数据挖掘中的用途及表现形式,将具体应用划分为概念层、逻辑层、基础层3个层次,如表1所示。可视化在概念层以动画、漫游、联合表达等形式启发人的形象思维能力,在不断改进挖掘的过程中实现探索性的可视化分析目的,帮助用户发现隐藏的时空模式、规律等;在逻辑层,可视化通过联合表达、图形叠加、高亮法、聚焦等技术,在图形表达的基础上为用户提供动态显示与交互操作机制,用户通过这种机制将基本可视化工具集成,形成空间数据可视化挖掘模型,挖掘其内在空间特征与关系;基础层可以为用户提供实际可用的数据挖掘环境和工具,它以图形显示、统计图、专题图、符号等形式表达外部空间形态与结构,给用户以初步的印象并激发其形象思维。

表1 可视化在空间数据挖掘中的应用层次

	数据挖掘目的	可视化用途	表现形式
概念层	挖掘时空模式、规律	实现探索性可视化分析,结合人的形象思维发现空间知识	动画、漫游、联合表达等
逻辑层	挖掘内在空间特征与关系	在图形表达的基础上提供动态显示与交互操作机制,将基本可视化工具集成形成空间数据可视化挖掘模型	联合表达、图形叠加、高亮法、聚焦等
基础层	挖掘外部空间形态、结构	提供基本的可视化工具与环境	图形显示、统计图、专题图、符号等

3 可视化的地质模型数据挖掘

3.1 关键技术

对应前面提出的可视化在空间数据挖掘中应用的环节和层次,本文从以下几个关键技术对可视化的地质模型挖掘进行研究探讨。

(1)地质模型可视化

可视化技术能够使人直观地理解空间数据中的各种复杂模式——空间结构形态结构和复杂关系等。地质体、地质构造等空间现象在现实世界中都是以三维形式客观存在的,因此以真三维的形式来

表达模型,可以更好地挖掘模型中空间形态结构及内部属性信息。地质模型具有岩性、深度、粒度等多维属性,因此,可以采用降维的方法来实现多维数据可视化。通常的做法是把高维信息转换到人类视觉能够感知的二维或三维空间^[10]。

在可视化过程中,还可以运用统计图、专题图、符号图等形式来丰富展现的内容,这是以抽象形式来表达丰富内涵的方法,是对空间数据可视化的一种有益的补充。另外,还可以利用虚拟现实等前沿技术来模拟真实的环境,为用户营造一个可视化的虚拟地理环境^[11],进一步促进数据挖掘与地理知识的发现。

(2)交互式挖掘

人机交互是实现地质模型挖掘的重要技术,它是用户与计算机之间进行信息交流与反馈的有效途径^[12]。人机交互过程不仅包含用户对空间数据的操作,而且包括用户对整个数据挖掘过程的干预与监控,例如通过取消、中断、恢复、功能组合等有效控制挖掘过程。在地质模型可视化处理之后,可以通过直接三维交互的形式对模型进行一些特殊的操作,如切片、虚拟钻探、填挖方分析、透明切片与透明钻探等,以剖析模型内在的特征与关系。

交互式挖掘是将基本可视化工具集成所形成的一种特殊的空间数据挖掘模型。它利用基本可视化工具对空间数据进行表达,然后结合用户敏锐的观察能力与丰富的专业知识对其进行分析检验,并通过一系列特殊的人机交互操作将自身丰富的经验知识实时反馈到挖掘过程中,同时运用联合表达、图形叠加、高亮法、聚焦等技术进行多角度表达,实时地显示动态挖掘结果。用户不断选择合适的可视化工具,直到获得所需的内在空间特征与关系为止。

(3)探索性可视化分析

探索性可视化分析是在空间数据可视化与交互式挖掘的基础上形成的一种较高层次的挖掘方式,它着重于挖掘隐藏的时空模式、规律等深层次信息,强调借助于人的形象思维能力,不断改进挖掘方法与表达方式,循环往复直到获得用户需要的信息。这种方法比较符合人类认识客观事物的规律,其最大特点是没有具体的挖掘目标。用户通过人机交互将自己的经验与知识反馈给挖掘过程,并不断检验分析结果、改进可视化方法。整个挖掘过程经历了一个不断改进、逐步求精的循环过程。探索性可视化分析为地质模型挖掘营造了一个可视化环境,用户可以自由地观察与剖析模型,寻找其蕴涵的知识机理,真正实现探索性知识发现。

3.2 系统实现

目前,地质模型研究的传统方法大多注重其外在形态的表达,对内部结构与属性的分析做得不够;没有形成深度挖掘地质模型的功能,用户还无法根据自己真正的需求来剖析模型。

本文借鉴数据挖掘的思想,将可视化本身作为一种信息挖掘工具用于地质模型的信息剖析。在已有三维地质模型的基础上^[12],以Visual C++6.0为开发工具、OpenGL为三维图形开发包,开发了一个软件原型系统GeoVisMiner 1.0。该系统实现了从3个层次上对地质模型进行可视化挖掘:在基础层实现了地层三维显示、分层着色、多模式表达(点、线、面模式等)、半透明显示等功能,用户可以通过直接三维交互的形式进行模型几何变换与切割操作;在逻辑层实现了切片、虚拟钻探、填挖方分析、透明切片与透明钻探等功能;在概念层实现了虚拟漫游功能。基础层的功能是基本的可视化及人机交互工具;逻辑层在基础层之上将一些工具进行组合实现了相应功能;概念层中的虚拟漫游功能实现了探索性可视化分析的功能,用户可以在对模型进行了多种操作之后定义漫游空间,并以第一人称的方式进行自主的实时漫游,身临其境地体会模型的内部信息,实现数据体验。图2、图3为系统效果图。

(下转第71页)