

林业数据挖掘与可视化的应用分析

王占刚², 庄大方¹, 邱冬生¹, 明 涛^{1,2}

(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: 面对海量的林业空间与非空间数据, 提取隐含的信息是“数字林业”急需解决的关键问题之一。数据挖掘和可视化是处理和分析海量空间数据的有力工具, 利用它们可以有效地剖析隐藏在数据背后的信息与知识, 为我林业规划与决策提供服务。本文将空间数据挖掘和可视化技术整合为一个有机的整体, 集成两者的优势, 阐述了可视化技术在空间数据挖掘研究中的意义。并讨论了空间数据可视化表达模式和可视化与空间数据挖掘的整合。针对林业资源数据的特点, 提出在可视化表达空间数据的基础上对林业数据的多维信息进行综合分析。最后, 开发了原型系统, 初步实现了林业数据可视化挖掘。

关键词: 空间数据挖掘; 可视化; 可视化挖掘; 林业

1 引言

随着国家“数字林业”概念的提出, 林业部门采集和更新数据的能力大幅度提高, 面对海量的林业空间与非空间数据, 如何从中提取隐含的信息、空间关系或有意义的特征或模式, 揭示各种空间规律、关系和趋势, 就成为“数字林业”急需解决的关键问题之一^[1]。由于林业资源数据自身的复杂性以及不确定性(具有多维、动态、多源、多尺度、空间性等特点), 导致利用传统的数据查询检索、空间分析统计等方法难以迅速有效地处理日益丰富的数据, 于是出现了“数据丰富、但信息贫乏”的现象^[2]。

数据挖掘 (Data Mining, DM) 与科学计算可视化 (Visualization in Scientific Computing, ViSC) 是近年来在应用需求推动下发展较为迅速的技术。如果将可视化与数据挖掘技术有机地集成, 形成可视化数据挖掘 (Visual Data Mining, 简称 VDM) 技术, 针对复杂的林业资源数据, 从多维、多角度分析^[3], 在表达原始数据、数据挖掘结果的过程中, 发现林业数据中的复杂关系和综合影响, 挖掘林业信息机理知识, 达到描述数据特征和推断预测的目的, 为规划决策提供依据。

2 空间数据挖掘与可视化应用的关键技术

2.1 数据挖掘与可视化应用分析

空间数据挖掘 (Spatial Data Mining, SDM) 是在数据挖掘的基础之上, 结合地理信息系统、遥感图像处理、全球定位系统等相关研究领域而形成的一个分支学科^[4,5]。由于空间数据本身具有海量、多尺度、多维和非结构化等特征, 使得它比一般的数据挖掘更加复杂。空间数据挖掘可以从空间数据库中提取隐含在其中的、但又是潜在有用的空间知识、关系和特征, 它可以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、趋势和模式等^[5]。

可视化技术能够使人直观地理解多维动态空间数据中的各种复杂模式——空间数据结构特征和复杂关系, 以及其时空变化规律等。可视化挖掘是应用可视化技术从海量数据中发现潜在有用知识的过程, 它利用人类的形象思维能力较强这一特点, 以图形图像的简明形式向用户表达抽象的挖掘信息, 使用户能够交互地观察数据、挖掘过程及结果, 加深了用户对数据含义、数据之间的相互关系

收稿日期: 2006-12-12; 修回日期: 2007-03-27.

基金项目: 中国科学院知识创新工程 (KZCX3-SW-357)

作者简介: 王占刚 (1977-), 博士研究生, 专业方向为空间数据挖掘、三维建模和 GIS. E-mail: wangzg@reis.ac.cn

和发展趋势的理解^[6-9]。可视化数据挖掘技术为用户提供了一个高效的可视化分析环境,在有效表达空间信息的基础上,通过人机交互融入用户自身的专家知识和经验,达到深层次剖析数据的目的,从而不断挖掘新的信息,探寻系统的内在联系和演化规律。

2.2 数据挖掘与可视化在林业应用中的关键技术

可视化为人们提供高度的交互功能与丰富的可视化表现功能,并充分调动其主观能动性,利用已有的经验与知识对研究对象进行分析。针对可视化在空间数据挖掘中应用的特点,本文从以下两个方面对其进行探讨。

(1) 空间数据可视化表达模式

空间数据本身具有多维、动态和海量等特征,因此,其可视化较为复杂。在可视化过程中,不仅要展现空间数据的外部形态结构,还要表达其内在的关系特征,通常应针对具体的研究对象与目的选取合适的可视化方法进行表达。

基于不同的分类方法,可以将空间数据可视化方法划分为不同的类别。依据数据对象的维数可以将其划分为二维可视化、三维可视化以及多维可视化技术。其中,二维可视化的主要表现形式是电子地图以及各种统计图、专题图等,这是一种以抽象形式来表达丰富内涵的方法;三维可视化包括 2.5 维和真三维表达,其着重于表达现实三维空间中实体的形态结构以及空间关系;多维可视化在展示空间位置的同时,还可以描述对象的其他属性信息。它是目前空间信息可视化领域的研究热点之一。多维数据的可视化大都采用降维的方法来实现^[10]。通常需要把高维的信息转换到人类视觉能够感知的二维或者三维空间。在空间多维信息表达的过程中,应充分利用图形的形态、色彩、符号、叠置、布局、聚焦、顺序、亮度、透明等特殊表达模式,以易于理解的方式向使用者展示复杂的空间信息。

林业领域的相关数据较为丰富,如森林类型数据、立地条件、林地使用状况、地形地貌数据、土壤类型、植被分布特征等多种数据。由于这些数据具有多维、动态以及空间的特性,使得面向林业数据的可视化不仅要展现森林数据的空间形态,而且还要表达其多维的特性,这给我们的研究提出了较高的要求。例如在研究森林空间分

布的过程中,可以针对不同的森林因子(树种、蓄积量、龄组)选取合适的可视化模式进行表达,充分借助于图形图像的色彩、形状、角度等的变化以及三维场景、动画等技术^[11],从多尺度、多观察模式来分析各个森林因子与地形和位置的关系。森林资源的动态变化为我们提供了大量的时间序列数据,利用三维场景、动画等可视化技术可以充分展现林业动态数据的发展演变规律。如利用动画技术表达森林的生长演变,描述森林资源的动态变化过程。

另外,虚拟地理环境是近年来发展较为迅速的技术,它通过图形、声音、触觉乃至气味等人类知觉,从多通道为用户营造一个真实的虚拟环境,使其产生身临其境的感觉,增强对空间信息的理解^[11]。

(2) 空间数据挖掘与可视化整合

将空间数据挖掘与可视化有效整合是实现可视化挖掘的核心。可视化挖掘是在对空间数据进行可视化表达的基础上,通过人机交互将人们敏锐的观察能力与丰富的专家知识融入挖掘过程,并将自己的经验与知识反馈给挖掘过程,使用户在不断地改进中反复进行挖掘,以深入剖析空间数据的结构特征与复杂关系,达到最大限度挖掘信息与知识的目的^[12]。

人机交互技术可以使用户与计算机之间进行信息的交流与反馈。本文将人机交互的作用归纳为以下几个方面:原始数据可视化操作、挖掘算法选择、信息可视化操作、循环往复等,其对应于整个空间数据挖掘的过程。如图 1 所示,在对原始数据处理的基础上,结合用户观察的结果进行数据选择、聚焦、放缩、旋转等可视化操作,形成预处理数据集,为用户提供直观的视觉感知;接着用户借助于自身的专家知识,根据挖掘意图选择合适的挖掘算法进行处理,生成中间结果;然后根据经验,对获得的信息进行可视化操作并生成挖掘结果;凭借人自身的形象思维能力,不断交互改进挖掘的方法与表达方式,循环往复直到获得所需的空間知识为止^[13],真正实现空间数据探索分析的目的。在人机交互过程中,不仅仅包含用户对数据的可视化操作,而且包括对整个数据挖掘过程的监控与干预,用户可以通过取消、中断、恢复等功能有效地控制挖掘的过程。

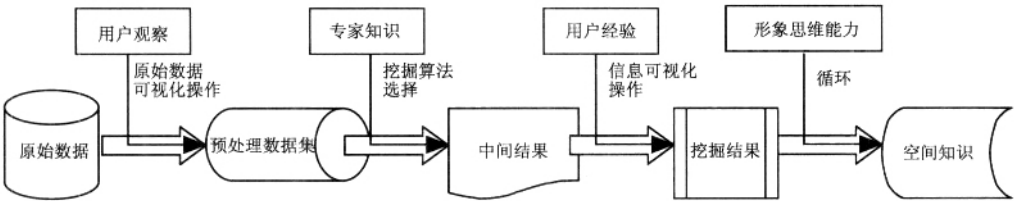


图 1 人机交互在空间数据挖掘过程中的作用

Fig.1 The application of the human-machine interaction in the spatial data mining

3 林业数据挖掘与可视化的应用

林业信息数据量大、关系复杂,利用可视化数据挖掘技术对森林资源的多维信息进行综合分析,可以发现多维信息间的复杂关系和综合影响,模拟预测森林资源的生长状况和生态环境态势。充分挖掘隐藏在森林资源数据内部的有效信息,为规划决策提供依据。

利用丰富的林业资源数据、基础地理数据以及遥感图像和地面调查数据,通过人机交互融入用户自身的专家知识和经验选择合适的挖掘方法,经过聚类、分类、统计、空间变换、叠加、融合等空间分析功能的处理,并反复选用适当的可视化模式将结果清晰美观地表达出来,为用户展示森林资源的空间分布规律、动态变化趋势以及实际经营管理情况。例如表达树种、蓄积、龄组、林地面积和土壤分布等特征因子,从而揭示森林资源的结构、组成以及相互间的联系、影响、制约,并从时空两种尺度展示森林资源消长的特点与演变规律,为森林经营与管理提供科学的依据^[14]。

本文以东北长白山森林数据为例,将可视化与数据挖掘集成为一个整体,以 Visual Basic 6.0 为编程工具,基于 ESRI ArcObjects 9.0 开发了原型系统。系统针对不同空间对象采用不同的可视化表达模式,利用叠置、颜色、线型、符号来综合表达树种、年龄、树高、郁闭度、蓄积量等特征。在表达基本空间信息的基础上,通过用户交互不断改进可视化表达方式,利用各种专题统计图,真正实现可视化分析的目的。如图 2 所示,首先选定东北长白山地区森林资源图层,然后针对图层的不同字段通过人机交互选择多种表达模式:将优势树种以不同颜色的小班边界线表示;小班蓄积量以不同高度的柱状图表示;区域地形的坡度以不同颜色的多边形表示。

该功能可以综合表达同一林业图层的多种信息,用户在此基础上通过形象思维能力进一步获取森林资源信息,详见图版 ,彩图 4。

图中所示为长白山东北部地区的森林资源分布图,图中优势树种信息由具有不同颜色的中空的线框表达,坡度信息由颜色渐变的实体多边形展示,每个林班的蓄积量值由柱状图表达。通过此图,用户可以有效地获取树种、坡度因子以及森林蓄积量之间的空间关系。可以看出,该地区优势树种主要包括阔叶林、混交林、白桦林、椴树林、杨树林、果树林、柞树林、灌木林等等,其中阔叶林与混交林所占比例较大;另外,该地区地形较为平缓,起伏不大,中部地区尤为平缓,坡度大于 9 度的地区均分布在周边地区;蓄积量分布的特点是北部低,南部高。阔叶林、混交林与白桦林等优势树种绝大部分分布在地势平缓的地区;混交林地区与南部阔叶林的平均蓄积量较大,北部地区蓄积量均较小;蓄积量大的地区地势基本平稳。

4 结语

本文将可视化技术与空间数据挖掘整合为一个有机的整体,集成两者的优势,提出利用可视化数据挖掘的方法来对林业资源数据进行综合分析,并在此基础上开发了原型系统,初步实现了林业数据的可视化挖掘。可视化数据挖掘技术可以为人们提供丰富的可视化表达能力与高度的交互性,充分调动人的形象思维能力,实现探索性分析。通过该方法,有助于探寻林业系统中的内在联系和演化趋势,不断发现新的林业规律与模式。

林业领域在长期的生产实践过程中积累了丰富的数据。但目前,国内外在林业数据挖掘方面的研究还很少,大部分处理分析仍是基于简单的数据

统计与空间分析, 无法满足林业领域日益增长的研究和应用需求。随着国家“数字林业”工程的不断发展, 相关部门对于林业信息与知识的需求必将增大。因此, 面向林业领域应用的数据挖掘技术, 探讨具有多维、动态、多尺度、多源等诸多特性的林业数据挖掘算法与模型, 对于进一步推动“数字林业”的发展具有重要的意义。

参考文献

- [1] 徐爱俊, 方陆明, 唐丽华. “数字林业”及其若干关键技术分析与探讨. 内蒙古林业调查设计, 2004, 27 (4): 21~24.
- [2] Jiawei Han, Kamber M. 数据挖掘: 概念与技术. 北京: 机械工业出版社, 2001.
- [3] 王英杰, 袁勘省, 余卓渊. 多维动态地学信息可视化. 北京: 科学出版社, 2003.
- [4] 李德仁, 王树良 等. 论空间数据挖掘和知识发现. 武汉大学学报(信息科学版), 2001, (6): 491~499
- [5] 邱凯昌著. 空间数据发掘与知识发现. 武汉: 武汉大学出版社, 2001.
- [6] Keim D A. Information visualization and visual data mining. IEEE Transactions on Visualization and Computer Graphics, 2002, 8(1): 1~8.
- [7] Ferreira de Oliveira, M C Levkowitz H. From visual data exploration to visual data mining: A survey. IEEE Transactions on Visualization and Computer Graphics, 2003, 9 (3): 378~384.
- [8] 芮小平, 张彦敏. 空间信息可视化挖掘研究. 测绘科学, 2005, 30 (2): 64~66.
- [9] 段晓军, 杜小勇, 易东云. 可视化数据挖掘技术及其应用. 计算机应用, 2000, 20 (1): 54~56.
- [10] 芮小平, 赵 扬, 杨崇俊 等. 空间信息多维可视化技术综述. 小型微型计算机系统, 2004, 25(9): 1636~1640.
- [11] 龚建华, 林琨. 虚拟地理环境-在线虚拟现实的地理学透视. 北京: 高等教育出版社, 2001.
- [12] 王占刚. 基于改进三棱柱的三维地质模型及其可视化研究. “硕士论文”中国矿业大学(北京), 2003.
- [13] 贾泽露, 刘耀林, 张彤. 可视化交互空间数据挖掘技术的探讨. 测绘科学, 2004, 25 (9): 34~37.
- [14] 高连有, 崔岩. 浅析林业可视化技术在林业信息化中的应用. 内蒙古林业调查设计, 2004, (6): 33~35.

Application Analysis of Data Mining and Visualization in Forestry

WANG Zhangang^{1,2}, ZHUANG Dafang¹, QIU Dongsheng¹, MING Tao¹

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

Abstract: How to extract the implicit information from massive forestry data has been an urgent key task for us. Data Mining and Visualization are powerful tools for processing and analyzing massive spatial data. They can help us effectively discover information and knowledge from data for serving forestry planning and decision-making. This paper integrates data mining and visualization into visual data mining with their advantages. The application of visualization in spatial data mining is presented. Visualization mode of spatial data and integration of spatial data mining and visualization are discussed. Aimed at the characteristics of data in forestry resource, this paper proposes that we can comprehensively analyze multidimensional information about forestry resource. Finally, a prototype was developed and visual mining in forestry was realized.

Key words: spatial data mining; visualization; visual mining; forestry

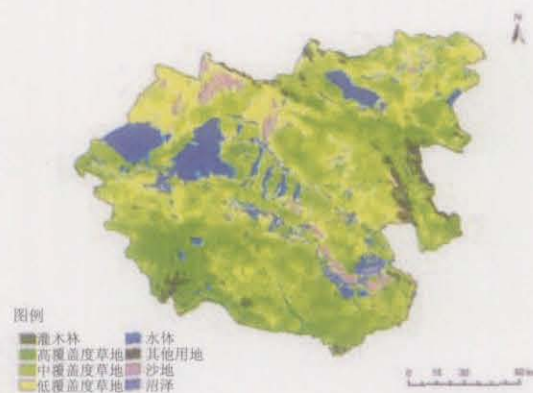
张 帅 等： 黄河源区玛多县土地利用/覆被及
景观格局变化的遥感分析（彩图 1~3）

图版 I

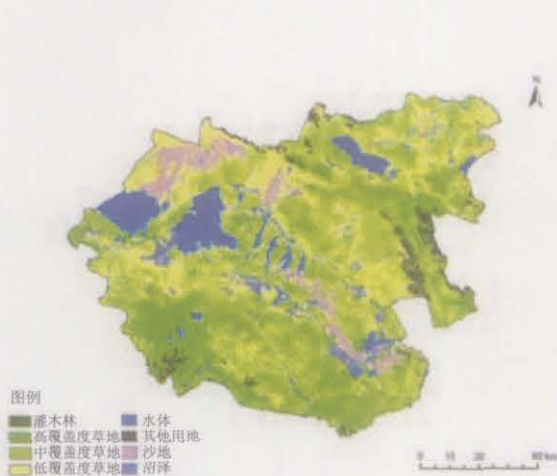
王占刚 等： 林业数据挖掘与可视化的应用分析（彩图 4）



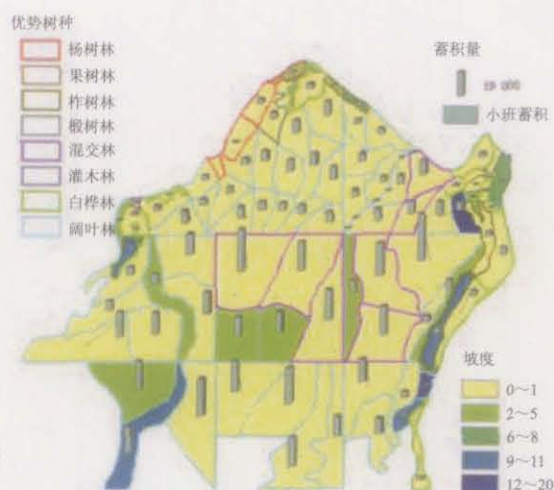
彩图 1 1977 年土地利用/覆被状况图
Fig.1 Land use and land cover map of
Madoi county in 1977



彩图 2 1990 年土地利用/覆被状况图
Fig.2 Land use and land cover map of
Madoi county in 1990



彩图 3 2003 年土地利用/覆被状况图
Fig.3 Land use and land cover map of
Madoi county in 2004



彩图 4 优势树种、坡度、蓄积量综合表达图
Fig.4 Synthesis chart of predominant tree
species, grade and cumulation