

数据挖掘技术在地理信息系统中的应用

The Application of Data Mining in Geographic Information System

黄红兵, 蒋望东

Huang Hongbing, Jiang Wangdong

(广西师范大学计算机科学系, 广西桂林 541004)

(Dept. of Comp. Sci., Guangxi Normal Univ., Guilin, Guangxi, 541004, China)

摘要:地理信息系统(GIS)可提供大量信息,但却无法发现数据中存在的关系和规则,利用数据挖掘技术可以从GIS数据库中挖掘出隐含的知识和空间关系,从而增强GIS的功能。

关键词:地理信息系统 数据挖掘 数据库 知识发现

中图分类号: TP311.13

Abstract: The concept and characters of data mining technique are explained. The necessity and possibility of application of data mining in geographic information system(GIS) is discussed. The contribution of data mining to GIS is summarized.

Key words: geographic information system, data mining, database, knowledge discovery from database

数据挖掘(Data Mining, 简称DM)是指从存储的大量的、不完全的、有噪声的、模糊的、随机的数据中,提取隐含在其中的、事先不为人知的、但又是潜在在有用的信息和知识的过程^[1]。数据挖掘主要是探索隐藏在数据中一些可能的性质,具有发掘与预测功能。

地理信息系统(Geographic Information System, 简称GIS)具有采集、查询、检索与统计、计算、管理、处理、分析建模和显示空间数据的功能,它以地理空间数据库为基础,采用空间模型分析方法,适时提供多种空间的和动态的有用信息,广泛用于综合研究、资源开发、城市规划、市政设施、房地产、交通管理、区域发展规划、环境保护、灾害防治、投资环境评价和决策等领域^[2]。

GIS的技术优势在于它的数据综合、模拟与分析评价能力,可以得到常规方法或普通信息系统难以得到的重要信息,实现地理空间过程演化的模拟和预测。空间分析功能是地理系统的核心功能,也是地理信息系统区别于其它计算机系统的根本点^[2]。

目前,GIS系统无法有效地发现大量的数据中存在的关系和规则,很难把握数据背后隐藏的重要

规律,无法根据现有的数据对未来做出精确的预测,大量的地理数据还没有发挥应有的作用,在一定程度上阻碍了GIS的进一步发展。

随着计算机技术的飞速发展,人们对地理信息系统的要求不断提高,不再满足于地理信息系统简单的绘图功能和空间分析功能,人们期望从地理信息系统的大量空间数据中获取更多的知识,数据挖掘技术有望解决这一问题,即空间数据挖掘。从GIS数据库发现空间关系及其它知识,是从数据库中发现知识(KDD)研究在GIS数据库中的一个具体应用。

1 GIS 中的数据挖掘技术

GIS是空间数据库发展的主体,空间数据库包含大量的属性数据和空间数据,其语义信息远比一般的关系数据库、事务数据库丰富和复杂,隐藏着丰富的知识。空间数据挖掘和知识发现技术的引入,使得系统能够自动进行学习,自动获取知识。

由于GIS数据库中不仅包含了大量的属性数据,而且包含了大量的空间数据。如何整理和解释这些数据,尽可能挖掘出对人们有用的知识和信息,总结出规律,可以说是目前GIS数据处理中急需解决的重要问题。数据挖掘技术在GIS中的应用,将使GIS原有的查询和空间分析功能提高到一个新的阶

段,将使 GIS 原有的查询和空间分析功能得到提高。数据挖掘技术可从大量的地理信息数据中发现知识,新发现的知识加入到原来的知识库,使原有的知识库不断得以扩充。数据挖掘技术在 GIS 中的应用为建立智能化的 GIS 奠定基础。

数据挖掘技术可以对地理信息数据进行深层次分析,以更好地利用这些数据。在 GIS 数据库中挖掘有用的隐含知识、更多的地理知识和地理规律,在 GIS 数据库中挖掘有用的隐含知识,这将会提高 GIS 的应用水平。

2 GIS 中的空间数据挖掘

空间数据库是一类重要的、特殊的数据库, GIS 是空间数据库发展的主体。空间数据是占据一定空间的数据对象,空间数据库系统存储的是不同数据类的空间对象,以及这些空间对象之间的空间关系。由于空间数据量巨大,空间数据类型、空间存储结构复杂,空间数据挖掘的关键是空间数据挖掘算法的效率。空间数据挖掘和知识发现技术的发展一方面可使 GIS 查询和分析技术提高到发现知识的新阶段,另一方面,从中发现的知识可构成知识库用于建立智能化的 GIS 系统。

空间数据挖掘是指从空间数据库中提取用户感兴趣的空间模式与特征、空间与非空间数据的普遍关系及其它一些隐含在数据库中的普遍的数据特征,它是数据挖掘技术在空间数据库方面应用的延伸^[3]。

GIS 是一个对地球表面及空间物体描述的信息系统,利用数据挖掘技术可从 GIS 的空间数据和属性数据中找出有关规律,为 GIS 自身的发展提供更广阔的空间。

3 GIS 数据库中可发现的知识

GIS 数据库中隐藏着大量的知识,如某地区的森林资源分布等。这些知识中有些属于“浅层知识”,如某区域的矿藏分布等;对于“深层知识”,如空间关联规则、形态特征区分等规则,无法直接从空间数据库中获取,必须通过一定的加工才能提炼出来。GIS 数据之间不是相互孤立的,在解决实际问题时,经常要同时使用多种挖掘算法。

GIS 数据库是空间数据库的主要类型,从地理信息系统数据库中可以挖掘到的知识,主要类型有^[4]:

3.1 空间分布规律

空间分布规律是指物体在地理空间的分布规律,包括垂直分布规律、水平分布规律以及垂直水平方向的联合分布规律^[4]。垂直分布规律如在高山地区,不同海拔高度的地方往往分布着不同类型的植被,随着海拔的升高植被的种类逐渐减少;水平分布规律包括经度地带性分布规律和纬度地带性分布规律,如在同一纬度上从沿海到内陆降水逐渐减少,同一经度上从低纬度到高纬度气温逐渐下降。根据事物的几何性质与属性之间的关系,将 GIS 的空间数据与属性数据对应起来,可以发现地理事物的几何与属性的对应关系。

3.2 空间关联规则

空间关联规则是指空间目标间相邻、相连(大城市大多与大江大河相连)、共生(沙漠地区中人口较多的居民点一般都和绿洲相关)、包含等空间关联规则。

3.3 空间聚类规则

空间聚类规则,是指特征相近的空间目标聚类成上一级类的规则,可用于 GIS 的空间概括和综合。

3.4 空间特征规则

空间特征规则是指某类或几类空间目标的几何的和属性的普遍特征。

3.5 空间区分规则

空间区分规则是指两类或多类目标间几何的或属性的不同特征。

3.6 空间演变规则

如果 GIS 数据库是时空数据库或 GIS 数据库中存在同一地区在多个不同时间的数据,则可以发现空间演变规则。空间演变规则是指空间目标随时间的变化规则。

3.7 有关目标的几何信息^[5]

从 GIS 数据库中,不难得到有关目标的位置、形状大小、分布等有关信息,通过归纳和演绎的方法就可以获得该类目标的规律性几何信息知识。如世界上许多著名渔场的形成与洋流有密切关系,大都位于寒暖流交汇处。

3.8 目标的几何性质与属性之间的关系^[5]

从 GIS 的图形和属性数据库中,不难发现目标间的相连(如城市大多与大江大河相连),及共生关系(沙漠地区的居民点一般都和绿洲相关)。

3.9 目标的几何性质与属性之间的关系的知识

将 GIS 中的空间数据与属性数据对应起来,可

发现目标的几何与属性之间的对应关系。

4 数据挖掘技术在 GIS 中的应用

数据挖掘技术可以从 GIS 数据库中挖掘出更多的隐含知识、地理知识和地理规律,目前主要应用于以下几个方面^[5]:

(1)精练 GIS 中的数据。虽然现有的 GIS 数据库中存储大量的数据,但是其中只有部分数据是必需的,而有部分数据是冗余的。利用数据挖掘技术,可以有效地发现数据间的相互关联,如此一来, GIS 数据库中就可只保存那些必要的数据,而不必存储其它的冗余数据,从而达到对 GIS 数据库进行精练的目的。

(2)更新 GIS 的数据。现有的 GIS 数据库中存储大量的数据,其中有相当一部分的数据随时间推移不断发生变化,如何对这些变化了的数据及时进行处理,是一个迫切需要解决的问题。用数据挖掘中的空间分析方法可以有效的解决这一问题,它通过对不同时域的数据进行比较,得到事物随时间变化的规律,并找到影响变化的主要因素。这样,在以后的分析中,只要检查这些主要因素是否变化,若有变化,就及时进行数据更新,否则就不予考虑。

(3)扩充和完善 GIS 数据库。虽然 GIS 数据库中包含大量的数据,但可直接从中获取的知识毕竟是有限的。而数据挖掘技术可以大量数据中发现新的知识和规则,并将新发现的知识和规则补充到原有的知识库中去,从而获得更新的知识,这对许多应用来说是至关重要的。

(4)GIS 空间信息系统智能化。专家系统与 GIS 的结合,使 GIS 成为一个空间咨询和决策支持系统,而数据挖掘与 GIS 的结合,可使 GIS 成为真正

的智能空间信息系统。

5 结束语

从 GIS 数据库中挖掘知识需要用到多学科的知识,学科交叉、知识综合性较强。目前完全自动地从 GIS 数据库中发现深层次的知识难度较大,从 GIS 数据库中发现知识需要专家系统的支持,发展方向是实现 GIS、ES(专家系统)、DM 的完全集成,建立智能化的 GIS。

数据挖掘在 GIS 中的应用必将促进数据挖掘自身的发展,在 GIS 中进行数据挖掘时会碰到很多困难,如在处理非常大的数据库时挖掘的效率将会急剧下降;数据变化快使早期发现的模式无效^[6]。这些问题都有待进一步深入研究后加以解决。

参考文献:

- 1 黄晓霞,萧蕴诗.数据挖掘应用研究及展望.计算机辅助工程,2001,(4):23~28.
- 2 吴信才.地理信息系统原理与方法.北京:电子工业出版社,2002.
- 3 Ester M, Kriegel H, Xu X. Knowledge Discovery in Large Spatial Databases: Focusing Techniques for Efficient Class Identification. In Proc. Of 4 Int L Symp. On Large Spatial Databases(Ssd 95), Portland, Maine, 1995. 67~82.
- 4 赵鹏.数据挖掘在规范推理及地理信息系统的研究: [硕士论文].合肥:安徽大学,2003.
- 5 蒋良孝,蔡之华. GIS 数据库的数据挖掘.计算机工程与应用,2003,(18):202~204.
- 6 朱晓强,王行风.数据挖掘在 GIS 中的应用研究.计算机工程与应用,2003,(28):208~213.

(责任编辑:黎贞崇)

海威公司大屏幕显示系统推广应用面大

桂林海威电子有限公司开发的全彩、双基色大屏幕体育/证券/信息广告显示系统,2003年3月22日在桂林市通过了自治区科技厅主持的技术鉴定。

系统的全彩系统采用了灰度分配器控制方式,双基色显示系统应用了多页面组合灰度发生器原理。它还采用了非线性级差显示技术,全彩色技术应用色度空间校正技术等。它的各基色灰度达 2048 级,非线性校正后各基色灰度达 256 级。

该系统已在自治区内外的体育、税务、邮电、会展、证券、宣传、广告等行业安装了 1100 多套,销售额达 5000 多万元。它的灰度分配器技术、相位像素多组合处理技术、模糊专家调灰系统及多页面组合灰度发生器技术,达到了国内领先水平。

(广西科学院 罗海鹏)