空间数据挖掘知识的地图可视化表达*

王佐成[1,2]，薛丽霞[1,2]，李永树[2]，徐京华[2]
(1.重庆邮电学院 软件学院，重庆 400065；2.西南交通大学 土木学院，四川 成都 610031)

摘　要：通过分析空间数据挖掘所能发掘的知识类型,对发掘出的知识表达方式进行了研究,提出地图是空间数据挖掘规则和知识的可视化表达的优秀和成熟的表示方法,并对如何表达进行了探讨。

关键词：空间数据挖掘；可视化；地图
中图法分类号：TP317.4　　**文献标识码**：A　　**文章编号**：1001-3695(2006)02-0253-03

Mapping Knowledge of Spatial Data Mining

WANG Zuo-cheng[1,2], XUE Li-xia[1,2], LI Yong-shu[2], XU Jing-hua[2]
(1. *College of Software, Chongqing University of Posts & Telecommunications, Chongqing* 400065, *China*; 2. *College of Civil Engineering, Southwest Jiaotong University, Chengdu Sichuan* 610031, *China*)

Abstract: This paper proposes that map is the proper visual method of spatial knowledge or rules from spatial data mining. Map can represent static and dynamic things, from the point of view of map expressing, spatial knowledge or rules from spatial data mining can be classified three categories, spatial characteristic rules, spatial relation rules and temporal and spatial evolution rules. The visual representing method for each category is proposed by virtue of maps.

Key words: Spatial Data Mining; Visualization; Map

1　引言

数据挖掘与知识自提出后得到迅速的发展,在挖掘的理论方法和技术上已经取得很大进展。然而,随着挖掘出的知识的积累,出现了知识的发现与人类可认知和理解的瓶颈,也即所谓的“知识爆炸”。例如在数据库关联规则发现中,一个10 000条记录的商业数据库可能发现几十上百条关联规则,对于空间数据挖掘尤其如此。虽然通过基于限制的数据挖掘方法可以缩减规则规模,然而对用户来说,理解这些规则将是艰巨的工作,此时,知识可视化表达被提出。研究表明,在人所获取的各种信息中,通过视觉而得到的占60%以上。

地图作为可视化表达具有空间属性空间对象的理想的载体,它是人类空间形象思维的再现。作为空间科学中时空分析与表达的手段,地图被称为“第二语言”;作为沟通的艺术形象,地图又是“国际化”的符号[1]。它作为空间数据挖掘的一个重要数据源,在空间数据挖掘中进行了大量的研究,挖掘出的空间知识借鉴非空间数据挖掘系统的知识表达方法来进行表达,如规则、表、交叉表、饼图或条图、判定树、数据立方体或其他可视化表示[2]。但是,空间数据挖掘出的知识大多数是带有空间属性,使得空间数据挖掘出的知识表达过程更加复杂,表达方式更加多样,地图正符合这种可视化表达需要,然而,在对空间知识进行表达时却没有被充分利用,实际上地图在将空间数据挖掘出的知识进行可视化表达方面具有其他表达方法不可比拟的优势。

收稿日期：2005-02-06；修返日期：2005-03-25
基金项目：国家自然科学基金资助项目(40371098)

2　空间数据挖掘所能够发掘的知识

空间数据挖掘所能够发现的知识主要包括空间特征规则,空间区分规则,空间分布规律,空间分类规则,空间聚类规则,空间关联规则,空间演变规律,面向对象的知识,空间偏差型知识[3]。一般表现为一组概念、规则、法则、规律、模式、方程和约束等形式的集合,是对数据库中数据属性、模式、频度和对象簇集等的描述[4]。这些知识中有的属于“浅层知识”,如某区域有无高速公路、河流的长度和最大宽度等,这些知识一般通过 GIS 的查询功能就能提取出来;还有一些属于“深层知识”,如空间位置分布规律、空间关联规则、形态特征区分规则、空间演变规律等,它们没有直接存储于空间数据库中,必须通过运算和挖掘才能发现。对这些知识可用特征表、谓词逻辑、产生式规则、语义网络、面向对象的表达方法和可视化等来表达[5],并且目前在数据挖掘领域已经在应用,但是,由于空间数据挖掘发现的知识绝大多数具有空间属性,传统的知识表达方法在表达空间属性上具有很大的局限性。

3　地图表现理论

地图所表现的各种复杂的空间和非空间对象,是通过特有的符号系统——包括点、线、面状符号、色彩以及文字、声音、动画视频等所构成的地图语言来实现的。这种符号系统不仅能表现静态的空间结构特征,如制图对象的地理位置、范围、质量特征、数量指标等,而且能够直观地显示各种制图对象的动态信息,如空间分布变化及其相互关系[1](图 1)。

静态信息表现方式主要包括采用点状、线状、面状与体状的方式和采用虚拟现实技术表现三维以及多维的分布信息。

针对历史、现状及未来的发展动态信息的表现方式较为多样，包括动画/视频、电子沙盘的空间漫游以及热区的声音、文本超链接，同时地学图谱在表现空间形态结构与时空变化规律方面具有独特的优势，在知识可视化表达上是一种重要方式。

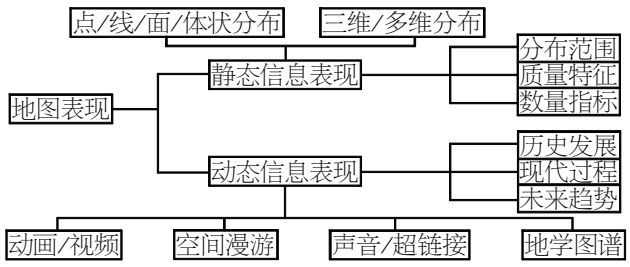


图 1 地图表现的方式和内容

静态信息表现主要通过符号的视觉变量来表现，包括形状、尺寸、色彩、亮度、图案和纹理等六种。其中点状符号和线状符号的主要视觉变量是形状、尺寸、色彩和图案，面状符号的主要视觉变量是色彩、图案、亮度和纹理，三维或多维虚拟现实对六种视觉变量都采用。动态信息表现除了传统的图形、图表、静态图像、注记等传统地图所具有的上述六种视觉变量外，还有动画、视频等动态地图符号以及文本、声音介绍和背景音乐等语言类符号。

4 对发掘的空间知识进行地图可视化表达

根据空间数据挖掘的知识类型，可以采用不同的地图表现方式，运用多种视觉变量来表达。空间挖掘的知识类型多样，但从知识的地图可视化表达角度看可分为三大类：空间特征规则、空间关系规则和空间演变规律，不同类型的知识选用的地图可视化表达不同。下面分别就空间数据挖掘知识类型来选用适合的地图可视化表达方式进行论述。

4.1 空间特征规则的地图可视化表达

空间特征规则(Spatial Characteristic Rules) 是某类或几类空间实体的几何和属性的共性特征。空间几何特征是指目标的位置、形态特征、走向、连通性、坡度等普遍的特征；空间属性特征指目标的数量、大小、面积、周长、名称等定量或定性的非集合特征。空间特征可视化表现在地图上即呈现分布规律，表现为实体在空间的垂直、水平或垂直——水平分布规律，如高山植被的垂直分布、公用设施的城乡差异、异域地物的坡度坡向分布等规律。

数据挖掘发现的空间特征规则是用户感兴趣的知识，实际上是空间特征的精练和升华。规则中的谓词(用 O 表示) 作为地图可视化表示的内容限定，规则谓词的属性集$\{S_1, S_2, \ldots, S_n\}$即为特征，对所需要表示的特征赋给视觉变量 V 来可视化表达。空间位置作为谓词和谓词属性的隐含限制条件，决定特征的可视化空间编码。图 2(a) 描述了在空间 I 中空间特征规则地图可视化表达，由空间谓词 O 限定地图内容，统一的视觉变量 V 保证地图表达的可读性，A_i 表示空间谓词 O 的子集，其所对应的 V 取值由其特征值决定。如“中国从沿海到内地人口密度逐步降低”是一条描述人口地理分布的特征规则，其中人口密度作为地图的主题限定，可以采用面状图式选用纹理视觉变量来表现，如图 2(b) 所示。

4.2 空间关系规则的地图可视化表达

空间关系规则是一类表现空间对象之间关系的知识，包括空间分类、空间聚类以及空间关联规则。空间分类规则根据空间区分规则把空间对象数据集中的数据映射到某个给定的类上用于数据预测。空间聚类规则把特征相近的空间实体数据划分到不同的组中，组之间的差别尽可能大，组内的差别尽可能小，可用于空间实体信息的概括和综合。与分类规则不同，进行聚类前并不知道将要划分成几个组和什么样的组，也不知道根据哪些空间区分规则来定义组。空间关联规则是空间实体之间的拓扑关系、距离关系、方位关系。

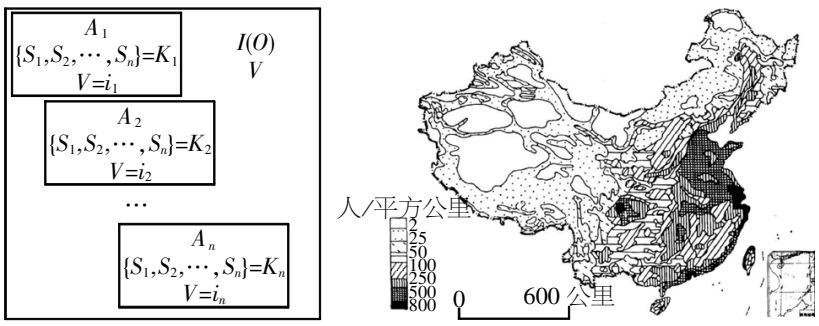


图 2 空间特征规则地图表达

空间分类和聚类规则主要在同类空间对象之间按照相同的特征指标进行挖掘产生。空间分类和空间聚类在地图表达上可以表示为

$$M = C(T_i,\ V,\ O, \{S_1, S_2, \ldots, S_n\})\ (i = 1, 2, 3\ldots)$$

其中，M 为地图对象，C 表示分类或聚类操作方法，T_i 是类别标志，V 代表地图视觉变量，O 代表对象集，作为地图内容限定，$\{S_1, S_2, \ldots, S_n\}$ 代表 O 的属性集。分类时在$\{S_1, S_2, \ldots, S_n\}$中给定指标，而在聚类时在$\{S_1, S_2, \ldots, S_n\}$中没有指标存在，通过操作 C 聚集后产生指标。在对空间分类知识的地图可视化表达中，将属性集$\{S_1, S_2, \ldots, S_n\}$与视觉变量 V 按照可视化表达要求进行映射。图 3(a) 表达了这种可视化方法，按照指标将 I 空间分为四个子空间，每个子空间通过视觉变量的变化来可视化表达子空间的特征，空间位置关系通过显式表达出来。一条典型的空间分类规则如“我国水系流域分为外流区和内流区”，地图可视化表达如图 3(b) 所示。采用色彩视觉变量来表达内流区和外流区。

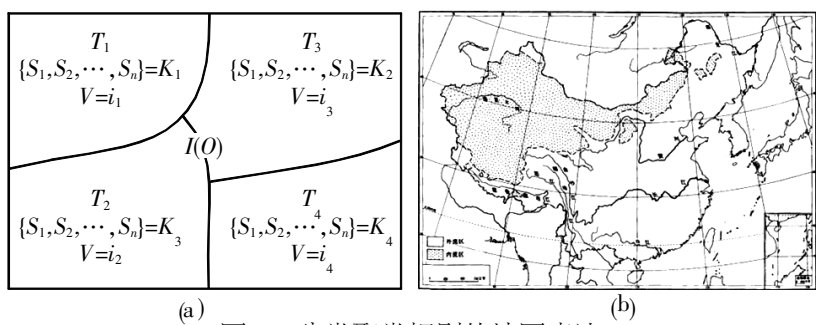


图 3 分类聚类规则的地图表达

空间关联规则可以表达为 $A \rightarrow B(s\%, c\%)$，在知识的地图可视化表达时，谓词 A 和 B 作为地图内容限定，s 和 c 是规则的支持度和置信度，将其赋给视觉变量来表现，如图 4(a) 所示。如果谓词带有数值属性，也将其赋给视觉变量来表现。实例如“is_公寓 ∧ adjacent to_公园 → is_expensive(30%, 70%)”，地图可视化如图 4(b) 所示，通过图案视觉变量来表现规则谓词，色彩视觉变量表现规则的兴趣度。

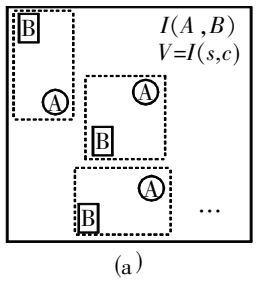
4.3 时空演变规律的地图可视化表达

时空演变规律是对时序空间数据库进行空间数据挖掘发现的空间变化规则。时空演变规则是指空间目标依时间的变化规则，即哪些地区易变，哪些地区不易变，哪些目标易变，怎

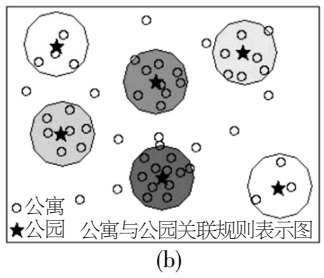
么变,哪些目标固定不变[6]。例如土地利用演变、世界人口流动、全球气候变迁等时空变化的知识。空间演变规则在地图表达上可以表示为

$$M = C(O, V\{S_1, S_2, \ldots, S_n\})$$

其中,M 为地图对象,C 表示时空演变轴,可以取时间 t 或者空间 r,取时间时地图表达的是对象集 O 随时间演变规律,取空间时地图表达的是对象集 O 随空间演变规律,V 代表地图视觉变量,视觉变量在与属性集$\{S_1, S_2, \ldots, S_n\}$按照可视化表达要求进行映射后,主要采用动态信息表现视觉变量,如地图动画/视频、声音、超链接、地学图谱以及传统的运动符号等。$\{S_1, S_2, \ldots, S_n\}$代表 O 的属性集,也是演变规律的数值特征。典型的实例如"地球板块构造学说的地图动画","物种进化图谱"。图 5 用运动符号法表达了"影响我国的台风从我国东南沿海登陆影响我国"时空演变规律。



(a)



(b)

图 4　空间关联规则地图表达

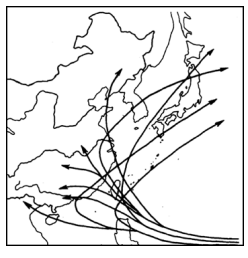


图 5　台风时空演变规则地图表达

5　总结

空间数据挖掘能够从存放在空间数据库、数据仓库或其他信息库中的大量数据中挖掘有趣的模式和知识,要使数据挖掘变得有效,空间数据挖掘系统就应当能够提供多种形式来显示和表达这些发现的模式和知识,地图作为有效的可视化表达空间知识的方式,当然并不排斥数据挖掘其他表达方法,如规则、表、交叉表、饼图或条图等。读图者对地图的理解能力会影响到地图对空间知识的可视化表达能力,因此结合数据挖掘知识的其他表达方法,地图将能够更好地表达空间知识。

参考文献:

[1] 廖克. 现代地图学[M]. 北京:科学出版社,2003.

[2] Jiawei Han, Micheline Kamber. Data Mining Concepts and Techniques[M]. Beijing, High Education Press, 2002.

[3] 周海燕. 空间数据挖掘的研究[D]. 郑州:中国人民解放军信息工程大学,2002.

[4] 李德仁,王树良,史文中,等. 论空间数据挖掘和知识发现[J]. 武汉大学学报(信息科学版),2001,26(6):491-499.

[5] 邸凯昌,李德仁,李德毅. 空间数据发掘和知识发现的框架[J]. 武汉测绘科技大学学报,1997,22(4):328-332

[6] 邸凯昌. 空间数据发掘与知识发现[M]. 武汉: 武汉大学出版社,2000.

作者简介:

王佐成(1973-),男,博士研究生,研究方向为空间数据库、数据挖掘;薛丽霞(1976-),博士研究生,研究方向为空间数据库、数据挖掘;李永树(1957-),教授,博导,研究方向为空间数据结构、测量工程;徐京华(1958-),教授,研究方向为地图学与地理信息系统应用。

(上接第 212 页)精度要求,新算法的收敛速度远远高于基本 BP 算法的收敛速度,体现了较强的优越性。

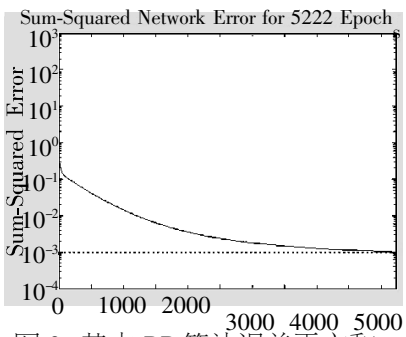


图 2　基本 BP 算法误差平方和随训练步数变化图

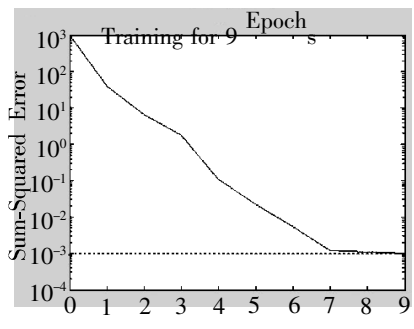


图 3　GOBp 算法误差平方和随训练步数变化图

但 L-M 算法必须存储矩阵 J^TJ,它是一个 $n \times n$ 的矩阵,其中 n 为网络权值。这取决于计算机内存的大小,一般情况下,参数数目不超过几千个。所以当网络很大时,用 L-M 算法计算神经网络是不可能的。另外全局优化的理论和算法远不如局部优化理论完善,很难判断以找到的局部极小点是否为全局极小点。

5　总结

标准 BP 算法的收敛速度慢以及易陷入局部收敛是影响其广泛应用的主要原因。本文利用 L-M 算法提高其收敛速度,利用填充函数法增强其全局寻优的能力,提出了基于 L-M 算法的全局优化 BP 网络,并实现 GOBP 算法。经实验验证,该网络解决贷款问题,其收敛速度及收敛能力远远优于基本 BP 算法。

参考文献:

[1] AN J Y. An Approach to Fault Diagnosis of Chemical Processes via Neural Networks[J]. Journal of AICHE, 1993, 39(1):82-87.

[2] Horinik K, Stinchcombe M, White H. Multilayer Feedforward Networks are Universal Approximators[J]. Neural Networks, 1989, 2(5):359-366.

[3] Ge R, Qin Y. A Class of Filled Functions for Finding Global Minimizers of a Function of Several Variables[J]. Journal of Optimization Theory and Applications, 1987, 54(2):241-252.

[4] Ge R. A Filled Function Method for Finding a Global Minimizer of a Function of Several Variables[J]. Mathematical Programming, 1990, 46:191-204.

[5] Ge R, Qin Y. The Globally Convexized Filled Functions for Global Optimization[J]. Applied Mathematics and Computation, 1990, 35:131-158.

[6] Ge R. The Filled Function Transformations for Constrained Global Optimization[J]. Applied Mathematics and Computation, 1990, 39:1-20.

[7] Simon Haykin. Neural Networks. 神经网络原理[M]. 叶世伟,史忠植. 北京:机械工业出版社,2004.

[8] 李换琴,万百五. 训练前向神经网络的全局优化新算法及其应用[J]. 系统工程理论与实践,2003,(8):42-47.

[9] 胡守仁. 神经网络导论[M]. 长沙:国防科技大学出版社,1993.

作者简介:

盛立(1982-),男,山东烟台人,硕士,主要研究方向为人工神经网络、数据挖掘;刘希玉(1964-),男,山东济南人,教授,博士,主要研究方向为人工神经网络、进化计算;高明(1981-),女,山东滨州人,硕士,主要研究方向为数据挖掘、神经网络。