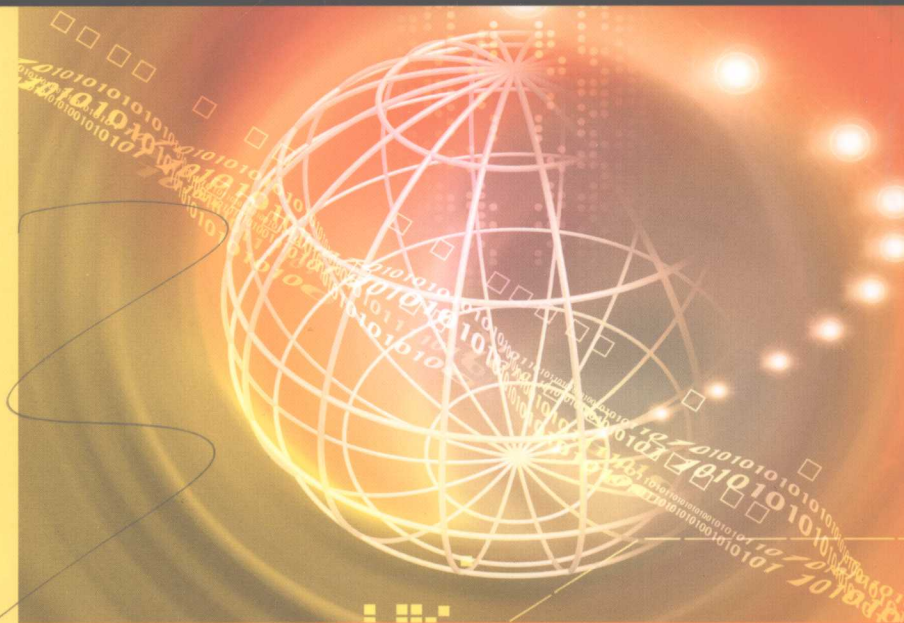




毕硕本 著

矢量图形特征的空间数据 挖掘及其应用

SPATIAL DATA MINING OF VECTOR FEATURES
AND ITS APPLICATION



科学出版社

www.sciencep.com

K878/3

2008

矢量图形特征的空间 数据挖掘及其应用

Spatial Data Mining of Vector Features
and Its Application

毕硕本 著

科学出版社

北 京

内 容 简 介

本书系统地阐述了矢量图形特征的空间数据挖掘这一研究领域内的基本概念、基本过程与相关方法,及其在姜寨遗址一期文化聚落中的应用。全书共包括 12 章,主要内容涉及姜寨一期聚落的空间分析、基于 GIS 数据库的剖面数据挖掘、矢量特征的空间分类挖掘、矢量特征的空间聚类挖掘、基于属性数据库的关联挖掘、基于属性数据库的泛化挖掘、矢量特征的神经网络空间分类、矢量特征的遗传算法空间分类、矢量特征的模糊空间聚类、空间知识推理、矢量特征的模糊空间推理等。

本书内容丰富,组织严谨,逻辑性强,概念、原理、方法和应用结合紧密,图表丰富,可读性强。

本书可作为大专院校计算机、地理信息系统、计算数学专业以及相关专业高年级本科生和研究生参考书,也可供从事计算机、地理信息系统、计算数学专业和数据挖掘相关软件开发的科技人员,以及有关大专院校师生参考。

图书在版编目(CIP)数据

矢量图形特征的空间数据挖掘及其应用 = Spatial Data Mining of Vector Features and Its Application / 毕硕本著. —北京:科学出版社,2007

ISBN 978-7-03-020094-5

I. 矢… II. 毕… III. 地理信息系统-应用-新石器时代文化-文化遗址-考古-西安市 IV. K878-39

中国版本图书馆 CIP 数据核字(2007)第 181794 号

责任编辑:牛宇锋 于宏丽/责任校对:刘小梅

责任印制:刘士平/封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

新蕾印刷厂印刷

科学出版社发行 各地新华书店经销

*

2008 年 1 月第 一 版 开本:B5(720×1000)

2008 年 1 月第一次印刷 印张:18 1/2

印数:1—2 000 字数:357 000

定价:50.00 元

(如有印装质量问题,我社负责调换〈明辉〉)

中国博士后科学基金项目

“面向矢量图形特征的智能空间数据挖掘方法研究”
(2005037743)

江苏省博士后科学基金项目

“面向矢量图形特征的智能空间数据挖掘方法研究”
(0501019B)

江苏省教育厅计划指导性项目

“基于时间序列的聚落考古多源空间数据挖掘方法研究”
(05KJD170109)

资助出版

作者简介



毕硕本,1965年生,山东昌邑人,博士,南京信息工程大学计算机与软件学院教授。1987年7月本科毕业于大庆石油学院计算机与控制工程系计算机软件专业,1996年3月硕士毕业于哈尔滨工业大学计算机科学与技术系计算机软件专业,2004年6月博士毕业于南京师范大学地理科学学院地图学与地理信息系统专业,2006年6月从南京师范大学数学与计算机学院数学博士后流动站出站。中国计算机学会高级会员、中国地理学会会员、中国GIS协会会员。2006年12月被选为江苏省教育厅“青蓝工程”学术带头人培养对象。

目前主要从事计算机软件、地理信息系统、空间数据挖掘、智能计算等方面的研究工作。主持与参加的项目有:中国石油集团总公司科技项目“大庆油田地面建设信息系统”,国家自然科学基金项目“长江三角洲地区地下水开采与地面沉降虚拟调控试验”,863计划项目“虚拟地理环境系统的研究与开发”,南京师范大学“十五”期间“211工程”学科建设项目“田野考古发掘地理信息系统”等。获得国家级“优秀计算机软件金奖”1项,省部级以上科研奖励5项,发表学术论文40多篇,出版著作2部,教材1部,获得软件著作权4项。

前 言

空间数据是人们藉以认识自然的重要数据。空间数据库含有空间数据和非空间数据,非空间数据是除空间数据以外的一切数据。所以,也可以认为空间数据库是通用的数据库,其他数据库是空间数据库的特殊形态。随着计算机信息处理技术的进步,数据和数据库急剧膨胀,而数据库中隐藏的丰富的知识远远没有得到充分的挖掘和利用,数据库急剧增长与人们对数据库处理和理解的困难之间形成了强烈的反差。因此,从空间数据库中自动地挖掘知识,寻找隐藏在空间数据库中的不明确的、隐含的知识,空间关系或其他模式(即进行空间数据挖掘)的研究变得越来越重要。

数据挖掘的目的是把大量的原始的数据转换成有价值的知识,用于描述过去的模式和预测未来的趋势,它可以看成是决策支持过程。智能空间数据挖掘是将不完全、不可靠、不精确、不一致和不确定的数据和信息通过数据挖掘的技术手段逐步改变为完全、可靠、精确、一致和确定的知识和信息的过程和方法。也利用对不精确性、不确定性的容忍来达到问题的可处理性和鲁棒性。智能空间数据挖掘除涉及 GIS、空间数据挖掘技术外,还涉及信息科学的多个领域,是人工神经网络、模糊系统理论、进化计算,同时包括人工智能等理论和方法的综合应用。在当前空间数据飞速增长和对知识强烈需要的情况下,对空间数据挖掘的研究得到越来越广泛的关注,进行空间数据挖掘的研究具有重要的理论意义和应用价值。

将空间数据挖掘的思想和方法引入史前聚落考古的单遗址聚落的空间布局与内部结构的研究,可以增强 GIS 空间分析能力和智能化应用水平,为聚落考古等领域的空间数据研究提供有力的智能数据分析处理工具。把计算智能方法引入针对姜寨聚落遗址资料的空间数据挖掘研究中,进一步挖掘该聚落遗址所蕴涵的知识,不仅可以扩展计算智能的应用研究领域,而且可以提高空间数据挖掘方法的普适性与多样性。

本书是中国博士后科学基金(2005037743)和江苏省博士后科学基金(0501019B)“面向矢量图形特征的智能空间数据挖掘方法研究”的研究成果,也是在作者的博士论文“聚落考古中空间数据挖掘与知识发现的研究”和博士后出站报告“姜寨聚落遗址的智能空间数据挖掘研究”的基础上编写而成的。

本书的编写参阅了大量国内外有关论著、期刊文献,并与相关专家、学者进行了交流。本书系统地阐述了矢量图形特征的空间数据挖掘与智能空间数据挖掘研究领域内的基本概念、基本过程与相关方法,以及在姜寨遗址一期文化聚落中的

应用。

本书在编写和修改过程中,得到了南京信息工程大学科技处处长闵锦忠教授、计算机与软件学院副院长傅德胜教授等许多同志的关心和帮助,在此表示衷心感谢。

由于作者水平有限,加上时间匆促,难免有不妥之处,恳请读者批评指正。

毕硕本

2007年3月

目 录

前言

第 1 章 绪论	1
1.1 引言	1
1.2 数据挖掘概述	12
1.3 空间数据挖掘概述	16
1.4 计算智能	18
1.5 国内外研究现状与发展趋势	26
第 2 章 姜寨一期聚落的空间分析	40
2.1 空间分析的基本概念	40
2.2 空间分布	41
2.3 空间关系的分析方法	43
2.4 姜寨遗址的地理环境	46
2.5 姜寨聚落的三环结构	49
2.6 姜寨遗址的广场问题	56
2.7 姜寨遗址的若干遗迹分布	60
2.8 主要遗迹间空间关系的分析	63
第 3 章 基于 GIS 数据库的剖面数据挖掘	65
3.1 遗址的文化层分布情况	65
3.2 文化层的标准层与剖面图形库结构	66
3.3 文化层的空间数据挖掘	67
3.4 剖面上遗迹的空间数据挖掘	69
第 4 章 矢量特征的空间分类挖掘	71
4.1 空间数据库的设计与图形数据的预处理	71
4.2 分类算法	73
4.3 居住区的分类挖掘	79
4.4 墓葬区的分类挖掘	88
第 5 章 矢量特征的空间聚类挖掘	90
5.1 聚类算法	90
5.2 居住区的聚类挖掘	97
5.3 墓葬区的聚类挖掘	103

5.4	重要遗迹的空间分布挖掘	107
第6章	基于属性数据库的关联挖掘	112
6.1	属性数据库的设计与预处理	112
6.2	关联规则算法	114
6.3	房屋的数据挖掘	121
6.4	墓葬的数据挖掘	123
6.5	窖穴的数据挖掘	128
6.6	灶坑的数据挖掘	129
第7章	基于属性数据库的泛化挖掘	130
7.1	泛化规则算法	130
7.2	遗物的数据挖掘	136
7.3	房屋的分组与时序数据挖掘	146
第8章	矢量特征的神经网络空间分类	153
8.1	引言	153
8.2	神经网络的发展与研究现状	155
8.3	感知器模型	157
8.4	多层感知器	158
8.5	自组织竞争网络	159
8.6	自组织特征映射网络	162
8.7	学习矢量量化网络	164
8.8	支持向量机	166
8.9	姜寨聚落遗址矢量特征的神经网络空间分类研究	174
第9章	矢量特征的遗传算法空间分类	191
9.1	引言	191
9.2	进化计算的研究现状	191
9.3	进化计算概述	194
9.4	遗传算法基本理论	196
9.5	遗传机器学习	205
9.6	姜寨聚落遗址矢量特征的遗传算法空间分类研究	214
第10章	矢量特征的模糊空间聚类	222
10.1	引言	222
10.2	模糊理论研究现状	222
10.3	模糊集合及其表示	224
10.4	模糊关系与模糊矩阵	228
10.5	模糊分类方法	234

10.6 姜寨聚落遗址矢量特征的模糊空间聚类研究.....	234
第 11 章 空间知识推理	244
11.1 空间知识的表示与存储.....	244
11.2 空间知识推理机制的结构.....	246
11.3 空间知识推理.....	248
11.4 考古遗迹的知识推理.....	250
第 12 章 矢量特征的模糊空间推理	257
12.1 引言.....	257
12.2 模糊逻辑研究现状.....	257
12.3 模糊关系.....	258
12.4 模糊逻辑和模糊推理.....	260
12.5 模糊推理的各种模型.....	263
12.6 姜寨聚落遗址的模糊空间推理研究.....	265
参考文献	272

第1章 绪 论

1.1 引 言

随着计算机信息处理技术的进步,数据和数据库急剧膨胀,而数据库中隐藏的丰富的知识远远没有得到充分的挖掘和利用,数据库急剧增长与人们对数据库处理和理解的困难之间形成了强烈的反差。数据挖掘(data mining, DM)和数据库知识发现(knowledge discovery in databases, KDD)技术就是在这种状况下应运而生的,其目的是为数据库理解与应用提供自动化、智能化的手段。空间数据是人们藉以认识自然的重要数据,空间数据库含有空间数据和非空间数据,非空间数据是除空间数据以外的一切数据。所以,也可以认为空间数据库是通用的数据库,其他数据库是空间数据库的特殊形态。由于空间数据的数量和复杂性都在飞快地增长,要求终端用户详细分析这些空间数据,并提取感兴趣的知识或特征是很困难的。因此,从空间数据库中自动地挖掘知识、寻找隐藏在空间数据库中的不明确的和隐含的知识以及空间关系或其他模式(即进行空间数据挖掘)的研究变得越来越重要。在当前,空间数据中矢量图形数据与影像数据相互并存,但矢量图形数据具有许多比影像数据更加优越的特点。以李德仁院士领导的研究群体为代表,对影像数据类型的空间数据挖掘研究,做出了很多富有成效的研究成果,而对面向矢量图形特征进行空间数据挖掘的研究一直很贫乏。

智能空间数据挖掘是将不完全、不可靠、不精确、不一致和不确定的数据和信息通过数据挖掘的技术手段逐步改变为完全、可靠、精确、一致和确定的知识和信息的过程和方法。它利用对不精确性、不确定性的容忍来达到问题的可处理性和鲁棒性。智能空间数据挖掘除涉及地理信息系统、空间数据挖掘技术外,还涉及信息科学的多个领域,是人工神经网络、模糊系统理论、进化计算,同时包括人工智能等理论和方法的综合应用。

1.1.1 研究的目的是与意义

数据挖掘和知识发现(data mining & knowledge discovery, DMKD)的目的是把大量的原始数据转换成有价值的知识,用于描述过去的模式和预测未来的趋势,它可以看成是决策支持过程。同数据库管理系统检索和查询出的信息相比,数据挖掘出的知识是隐含、精炼和高水平的。为实现空间决策支持的目标,一项关键的技术是空间知识库的建立与空间知识的运用。通过对应用领域中的空间数

据进行快速大量获取与预处理的研究,建立空间数据库,研究并实现空间数据挖掘机制,建立应用领域的空间知识库及其空间知识处理机制,对应用领域决策目标的实现提供强有力的支持。

作者利用史前聚落半坡类型姜寨遗址的田野考古发掘资料,进行空间数据挖掘和智能空间数据挖掘的研究,目的在于通过对聚落遗址空间和属性数据的挖掘和智能化挖掘,发现和运用考古学知识,为实现该应用领域的协同决策支持目标进行研究。

针对田野考古中的空间数据进行数据挖掘的研究主要有以下几个方面的意义:

(1) 通过对田野考古资料进行数据挖掘和智能空间数据挖掘,可以更加精确、更加全面地找到古代人类活动的若干规律;通过对田野考古中的智能空间数据挖掘的研究,可以更加有利于考古工作者以更加科学的手段进行田野考古遗存的研究,为考古工作者提供一个智能化的、高效的科研平台,提高考古研究的信息化、智能化程度。

(2) 针对矢量图形特征,对田野考古的聚落遗址的空间数据挖掘的研究,其意义就在于对大量的现存的地下空间数据进行挖掘,找到隐含的空间知识,扩展空间数据挖掘的研究对象和 GIS 的研究领域,推动地理信息科学的智能化应用水平。

(3) 利用神经网络、遗传算法与模糊逻辑等计算智能的新方法对面向矢量图形特征的空间数据进行智能数据提取与智能空间数据挖掘的研究,主要是为了更全面、更真实、更智能化地挖掘原始遗迹空间数据的内在机理和隐含的规律,多角度、多层次、多方法地去研究空间数据隐含的规律性知识,同时克服常规空间数据挖掘方法所存在的弊端,更好地为后期的田野考古学空间分析与决策提供服务。

1.1.2 研究目标

研究面向矢量图形特征的空间数据挖掘方法,拓展空间数据挖掘的研究领域和提高空间数据挖掘的智能化水平,同时把它应用到聚落考古学研究领域,解决以下两个一直长期存在的问题:

(1) 大量现存的聚落考古资料中隐含的知识没有被及时发现和利用、田野考古发掘资料的知识含量和利用率有待提高的问题。采用的解决方案就是对聚落考古学资料进行空间数据挖掘,即运用分类、聚类、关联、泛化等数据挖掘技术方法,以及神经网络、遗传算法和模糊逻辑的计算智能方法,对姜寨一期聚落中的各种遗迹的矢量图形特征进行空间分类与空间聚类的数据挖掘。

(2) 田野考古发掘中如何运用知识指导发掘的问题。田野考古发掘要在丰富

的考古学空间知识的指导下进行,通过该研究成果来提高田野考古发掘的科学性和准确性。采用的解决方案是首先通过空间数据挖掘得到空间知识,并且建立考古空间知识库,然后基于精确推理和模糊逻辑原理,对姜寨一期遗迹中的空间关系进行经典知识推理和模糊知识推理与预测。

为了解决以上两个问题,制定了如下研究目标:

(1) 研究与分析面向矢量图形特征的空间数据,运用数据挖掘方法和计算智能方法,以田野考古平面遗迹图中的矢量图形特征为研究实例,进行空间数据挖掘方法的研究。它涉及:①对图像的数字化的、存储管理,要对图形、图像数据进行分类与建库。②对矢量遗迹平面图进行数据提取等预处理。③根据矢量图形特征数据的要求,以及数据挖掘算法和计算智能算法的特点设计与实现空间数据挖掘算法。④利用数据挖掘方法,以及神经网络、遗传算法和模糊计算这些智能算法,进行针对矢量图形特征的空间数据分类与空间聚类挖掘研究。

(2) 基于经典推理和模糊逻辑,研究与设计空间知识推理机制,对田野考古资料中的房屋等遗迹平剖面图中的矢量图形特征进行精确和模糊空间知识推理。该目标的实现涉及:①研究与设计空间知识表达机制。②研究与设计面向矢量图形特征的田野考古空间知识库。③研究与设计分别基于经典精确知识推理和模糊知识推理的推理机模型。

1.1.3 研究资料

本研究采用的相关实验数据资料说明如下:

资料来源是1988年文物出版社出版,由西安半坡博物馆、陕西考古研究所、临潼县博物馆编著的《姜寨——新石器时代遗址发掘报告》(上、下册)。该资料介绍了姜寨遗址的地理环境及其发现、发掘的经过;并以几个剖面图说明了遗址东南西北中五个部分的地层堆积情况;特别是对出土的遗迹与遗物进行了文化分期,共分为五个时期,并分别对每个时期的遗迹、遗物进行了统计与典型介绍;最后对姜寨遗址的文化特征、文化性质、社会经济状况,以及姜寨一期的村落总体布局和社会组织及社会发展阶段等进行了考古学文化分析与阐释。

姜寨新石器时代遗址位于陕西省西安市临潼区,于1972年发现,至1979年连续8年共进行了11次大面积的科学发掘。此遗址保存较好,内涵丰富,延续时间长,出土的大批珍贵遗物对我国新石器时代考古及原始社会史研究都有相当重要的意义。

本研究进行的空间数据挖掘采用了该资料的遗址发掘探方分布图(1:1000)、一期文化的主要遗迹分布图(1:500)、11幅探方地层剖面图(1:250)、53幅遗迹平剖面图(比例尺多种)、3幅分区墓葬分布图(比例尺多种)。基于属性的

数据挖掘采用正文中的房屋布局情况及其相对层位表,以及附表中的第一期房屋登记表、灶坑登记表、窖穴登记表、土坑墓登记表、瓮棺葬登记表等五类表。另外还利用下册中的与第一期文化相关的遗迹、遗物等照片建立了图像数据库。

1.1.4 研究内容

1.1.4.1 空间数据挖掘对象的层次

空间数据挖掘的研究与应用可分为三个层次:一是针对属性数据的挖掘,是对属性数据的认识和发现的活动,并将挖掘结果通过图形和其他可视化的方式显示出来。二是针对空间数据的挖掘,以遗迹平面分布图、房屋平剖面图等纯矢量空间数据作为操作的对象进行空间数据挖掘。三是联合式,把 GIS 数据库作为操作对象进行空间数据挖掘。

1.1.4.2 空间数据挖掘涉及的内容

(1) 对姜寨遗址文化一期人类活动的空间环境进行分析。基于空间分析技术,研究姜寨遗址一期古代原始环境的空间布局、空间关系等聚落形态。通过建立的文化层 GIS 数据库,挖掘各个文化层和各类剖面上的遗迹现象的空间统计规律,该部分研究内容包括文化层 GIS 数据库的构建与基于 GIS 数据库的空间数据挖掘研究等。

(2) 田野考古探方剖面图、平面遗迹图的空间数据挖掘。田野考古探方剖面图、平面遗迹图的空间数据挖掘涉及对图像的数字化、管理等,首先要对图形、图像数据进行分类与建库,然后对矢量遗迹平面图和矢量文化层剖面图进行预处理,利用多种数据挖掘算法进行空间数据分类与空间聚类挖掘,最后利用房屋等遗迹平剖面图进行空间知识推理。

(3) 遗迹、遗物等遗存文字资料的管理和数据挖掘处理。田野考古文字资料的管理和数据挖掘问题涉及各种遗迹、遗物数据的分类、入库管理、查询统计与基于不同数据类型的数据挖掘,包括遗迹、遗物数据库的逻辑设计和物理设计、数据预处理,以及多种数据挖掘算法的运用等。

1.1.4.3 智能空间数据挖掘涉及的内容

本研究采用神经网络、遗传算法和模糊计算等计算智能方法,对面向矢量图形特征的空间数据挖掘进行智能空间数据挖掘的方法研究,目的在于通过对矢量图形特征的空间数据进行智能挖掘,找到一种智能、新型、高效的面向矢量图形特征的空间数据挖掘方法,并且发现和运用田野考古学知识。基于这个研究目标,本研究所研究的主要内容为:

(1) 针对研究实例中遗迹的矢量图形特征, 研究与设计神经网络、遗传算法和模糊计算三种算法的实现方案。

(2) 研究与设计面向矢量图形特征的空间数据和模糊逻辑方法的模糊知识推理机制; 研究与分析基于模糊逻辑的知识推理模型; 通过对应用实例中模糊知识推理模型的研究与开发, 实现对人类古环境中人类活动空间分布位置等规律的推测。

1.1.5 研究技术路线

本研究采用的技术路线, 主要包含空间数据库建立、空间数据提取与空间数据挖掘, 以及空间知识库建立与空间知识推理三个大的阶段, 见图 1.1。下面说明每个阶段的重要步骤。

1.1.5.1 遗迹几何特征数据库与遗迹、遗物属性数据库的建立

该阶段首先利用姜寨遗址的资料建立图形、属性、图像等三类数据库, 然后进行空间数据挖掘前的数据提取、数据整理、数据转换与系统集成等预处理研究。该阶段的研究内容主要包括:

(1) 遗迹几何特征数据库(或称图形库)的建立。基于聚落考古空间数据挖掘的需求, 通过对图像数据库中遗迹图的数字化, 建立空间实体(主要是房屋、灶坑、窖穴、壕沟、圈栏、牲畜夜宿场、土坑墓、瓮棺墓、道路等遗迹)的空间几何特征数据库。

(2) 设计与建立考古图像数据库及其元数据库。考古图像元数据库主要是对聚落遗址中涉及的各种典型遗迹、遗物等图像数据的分类描述和说明等数据。

(3) 遗迹、遗物属性数据库的设计与遗迹 GIS 数据库的建立。首先设计遗迹、遗物属性数据库的逻辑结构和物理存储结构, 再通过建立拓扑关系和连接各自的遗迹属性数据库, 建立起各个遗迹的 GIS 数据库。

(4) 基于遗迹的 GIS 数据库, 从遗址空间分布、遗迹空间关系等角度, 对姜寨遗址的古环境和遗址内部结构, 以及各个文化层等进行空间分析。运用 GIS 的空间分析功能对遗址与各类遗迹进行空间统计分析、对比分析、缓冲区分析等, 揭示遗址与各类遗迹的区位特征和结构特征。

1.1.5.2 矢量图形特征数据提取与空间数据挖掘

该阶段首先研究如何提取空间实体的矢量图形特征。针对房屋遗迹等面实体而言, 主要研究内容是提取它的面积与方位角特征。然后对遗迹几何特征数据库和遗迹、遗物属性数据库进行空间数据挖掘和智能空间数据挖掘研究。该阶段的

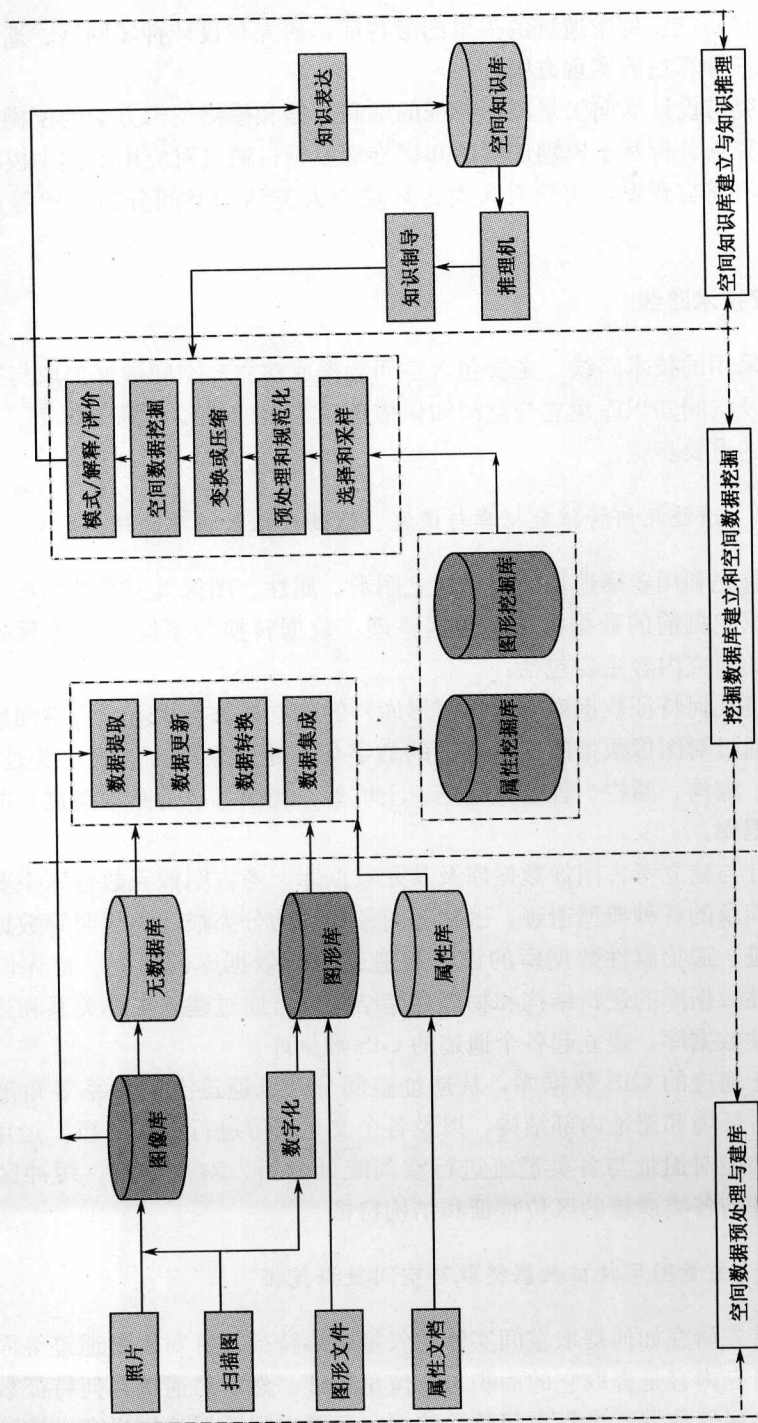


图1.1 研究的技术路线

工作是本研究的核心,它是一个多步骤的处理过程,在该处理过程中会有多次反复,主要包括以下一些处理步骤:

(1) 空间数据挖掘的准备。调查、分析空间数据挖掘相关领域的技术情况,熟悉聚落考古和姜寨遗址的背景知识,弄清需要获取的各类规则的技术要求与实现细节,以及空间数据挖掘的应用需求、数据挖掘算法与智能挖掘算法的各种参数要求。

(2) 空间数据的选择。根据空间数据挖掘的要求,利用数据库投影、选择等操作对空间与属性数据进行处理,从遗迹、遗物属性库与图像元数据库中提取与空间数据挖掘相关的数据。该阶段主要用排除法,将空间数据挖掘不需要的数据从各类数据库中排除,产生用于空间数据挖掘的各类挖掘数据库。

(3) 空间数据预处理。主要是对阶段(2)产生的各类挖掘数据库进行再加工,检查数据的完整性及一致性,根据空间数据挖掘算法的技术要求,测试并确定是否对其中的噪声数据进行剔除处理,对丢失的数据是否利用智能方法进行填补。同时对于遗迹矢量图形数据库中不适合进行空间数据挖掘的图形进行必要的图形处理和初步探测挖掘,并扩充遗迹矢量图形特征数据库的内容。

(4) 空间数据缩减或采样。对经过预处理的挖掘数据库中的大样本数据集,可以根据空间数据挖掘的需要对其进行再处理,主要是通过随机采样、数据投影、数据筛选等数据库操作减少挖掘数据量,采集参与空间数据挖掘过程的样本数据,以及通过数据库连接等数据库操作和建立概念层次树数据库等方法,获取满足空间数据挖掘需要的矢量图形特征空间数据。

(5) 确定空间数据挖掘的目标。根据各类挖掘数据库的数据类型,以及空间数据挖掘算法的技术要求,确定空间数据挖掘可以发现何种类型的知识。按照空间数据挖掘领域的不同专业应用的需要,以及不同类型数据的要求,在空间数据挖掘过程中应采用不同的数据挖掘算法。

(6) 确定空间数据挖掘算法。根据阶段(5)所确定的任务,选择合适的空间数据挖掘算法,包括选取与设计合适的算法模型和运行参数,如选取合适的概念层次树和必要的指导性参数知识、选取合适的分类范围以及聚类个数等指导性参数,使挖掘算法和挖掘的目的相一致。

(7) 进行空间数据挖掘。该阶段中需要采用决策树分类算法、 k -means 聚类算法等进行基于矢量几何特征数据的空间数据挖掘,需要采用 Apriori 关联算法与多维模式的 BUC 算法等进行属性数据的关联规则的挖掘,以及需要建立概念层次树,采用面向属性数据的泛化算法等进行属性数据的泛化规则和时序规则的挖掘。

(8) 进行智能空间数据挖掘。该阶段中需要采用神经网络多种分类算法、模糊聚类算法和基于遗传算法的分类器系统等进行面向矢量图形特征的智能空间数