

Spatial Data

Mining Theories and
Methods for
Land Use Zoning

面向土地用途分区的 空间数据挖掘理论与方法

► 牛继强 著



非
外
借



WUHAN UNIVERSITY PRESS
武汉大学出版社

国家自然科学基金项目 (41771438, 41201387)

信阳师范学院学术著作出版基金

联合资助出版

信阳师范学院第七批重点学科——地理学

Spatial Data
Mining Theories and
Methods for
Land Use Zoning

面向土地用途分区的空间数据挖掘理论与方法

► 牛继强 著



WUHAN UNIVERSITY PRESS

武汉大学出版社

图书在版编目(CIP)数据

面向土地用途分区的空间数据挖掘理论与方法/牛继强著. —武汉: 武汉大学出版社, 2017. 10

ISBN 978-7-307-19751-0

I. 面… II. 牛… III. 数据处理—应用—土地利用—研究
IV. F301.2—39

中国版本图书馆 CIP 数据核字(2017)第 241923 号

责任编辑:王金龙

责任校对:汪欣怡

版式设计:汪冰滢

出版发行: **武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件: cbs22@whu.edu.cn 网址: www.wdp.com.cn)

印刷: 虎彩印艺股份有限公司

开本: 787 × 1092 1/16 印张: 11.75 字数: 262 千字 插页: 4

版次: 2017 年 10 月第 1 版 2017 年 10 月第 1 次印刷

ISBN 978-7-307-19751-0 定价: 45.00 元

版权所有, 不得翻印; 凡购我社的图书, 如有质量问题, 请与当地图书销售部门联系调换。

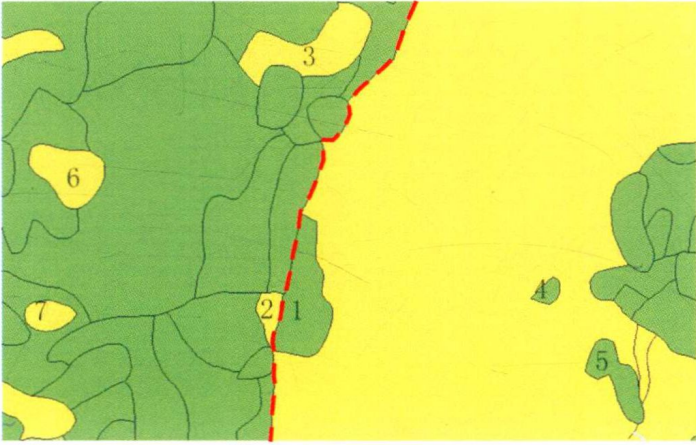


图 4-7 权重取值对聚类结果的影响



图 5-5 图斑的组成



图 5-6 宜城市区位图

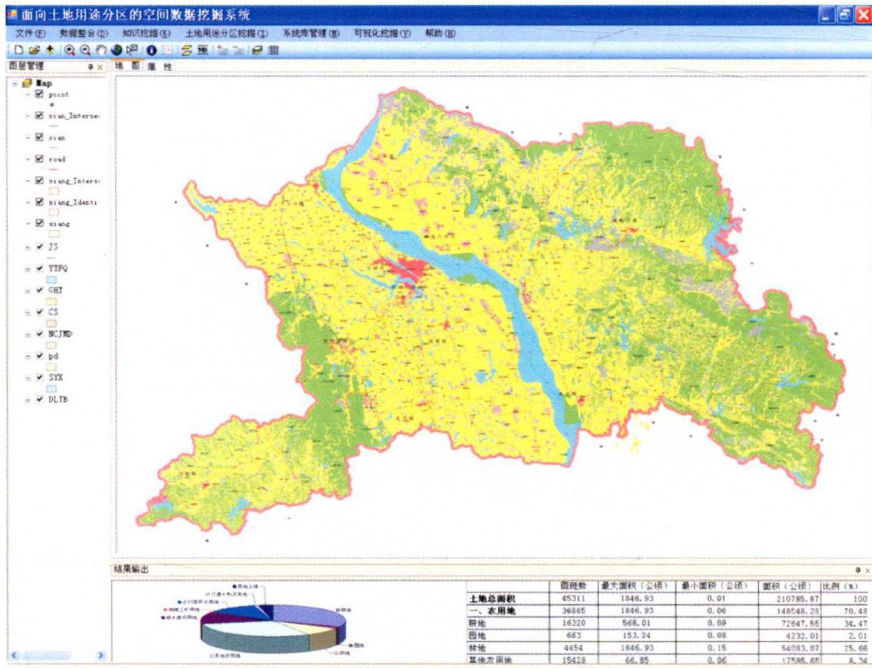


图 5-7 宜城市土地利用现状图

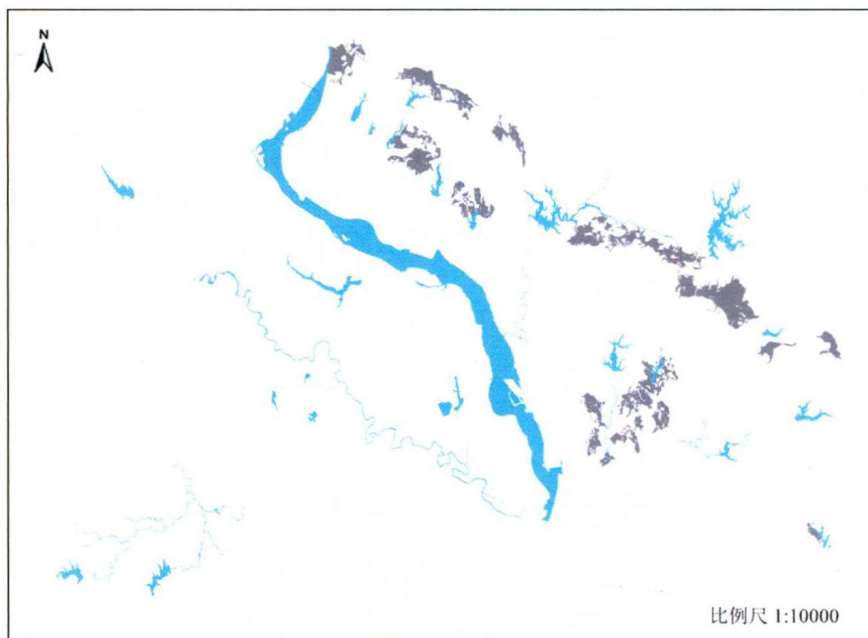


图 5-11 宜城市土地用途分区中非分区图斑的提取

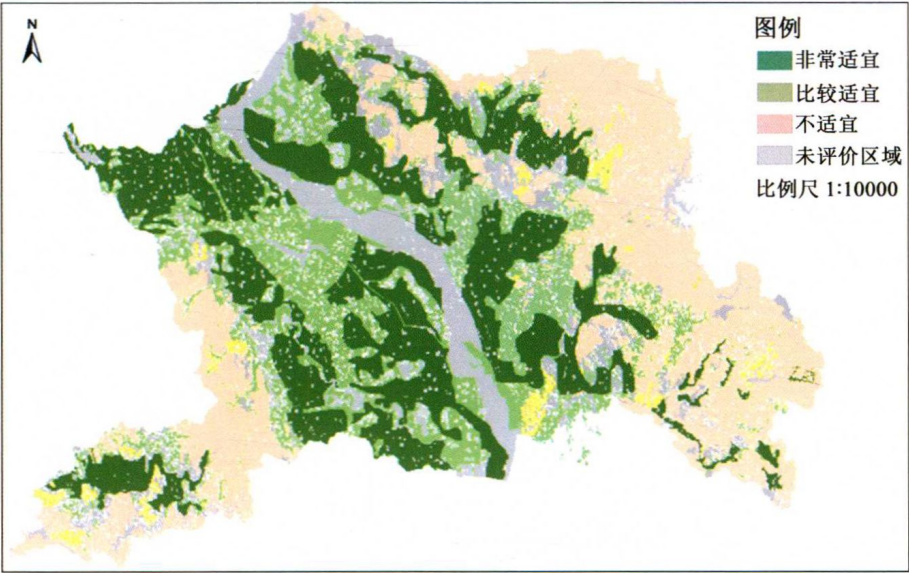


图 5-12 宜城市土地适宜性评价

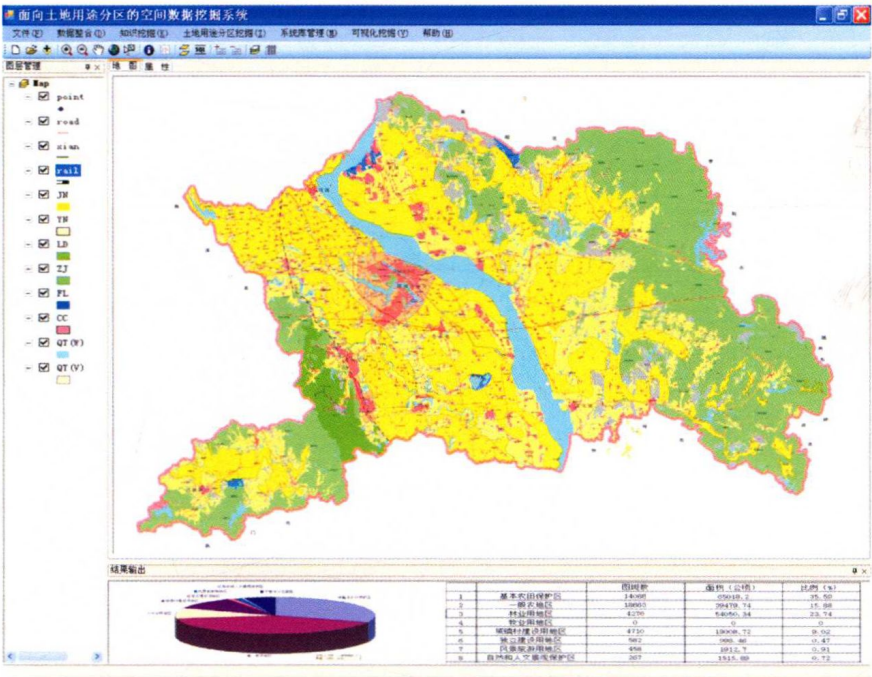


图 5-13 宜城市土地用途分区图

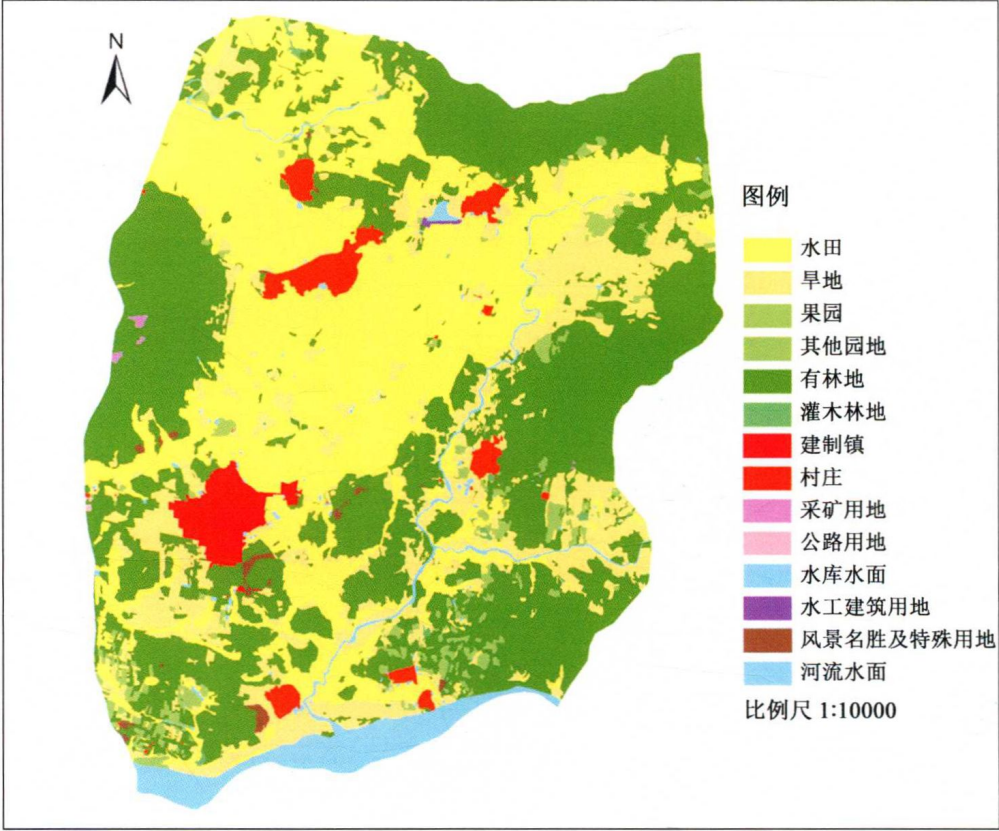
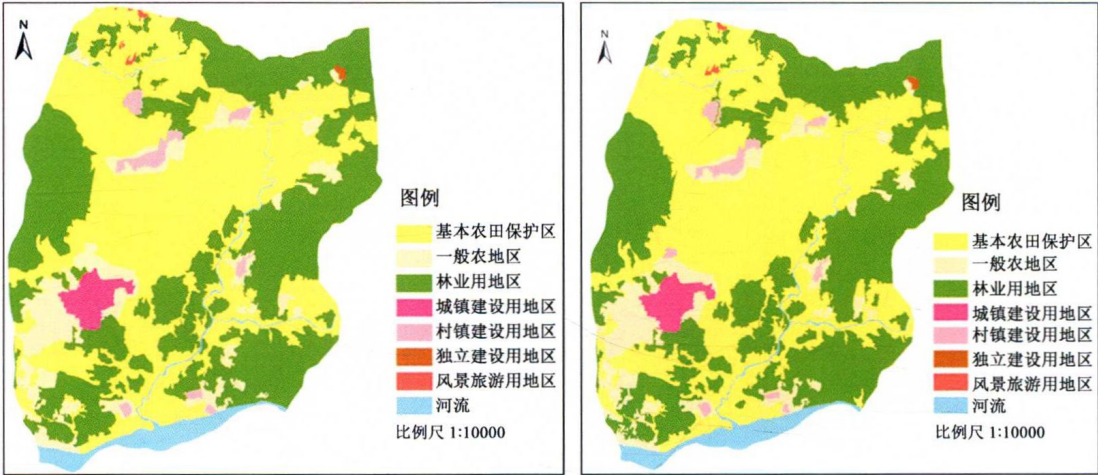


图 5-17 海南省昌江县乌烈镇土地利用现状图



(a) 克隆选择模型 (b) 遗传 K-均值算法

图 5-18 海南省昌江县乌烈镇土地用途分区图

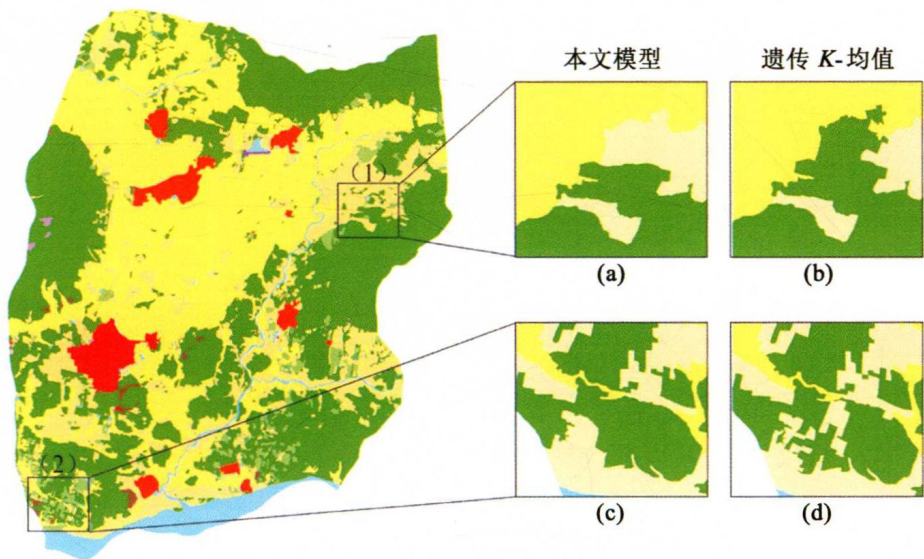


图 5-20 土地用途分区方案的对比分析

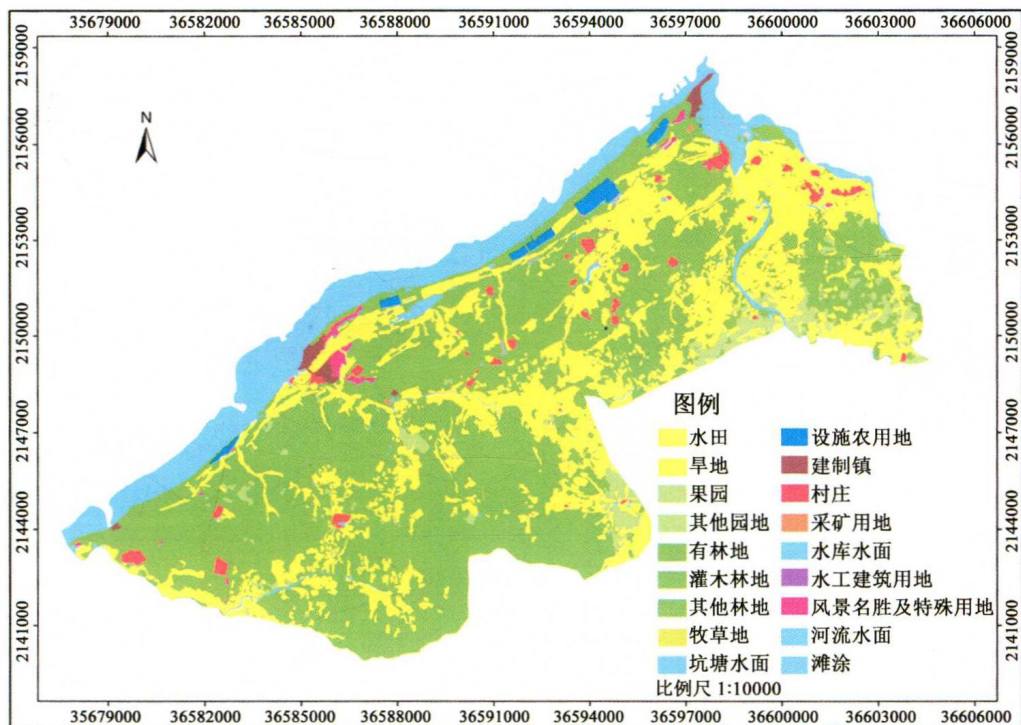


图 5-21 研究区土地利用现状图

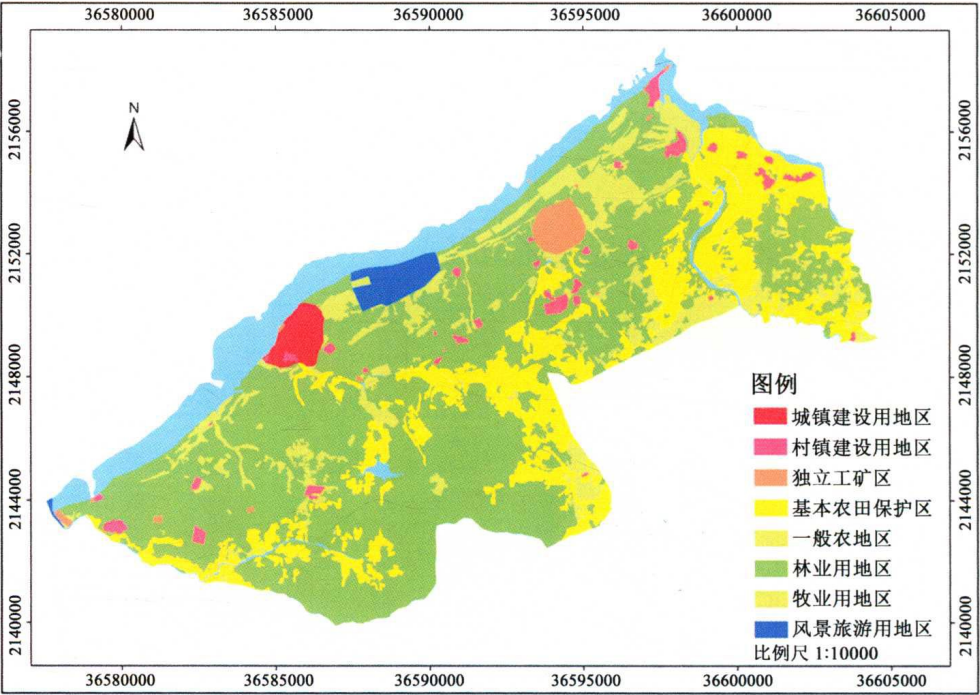


图 5-22 顾及地理实体语义相似度的土地用途分区图

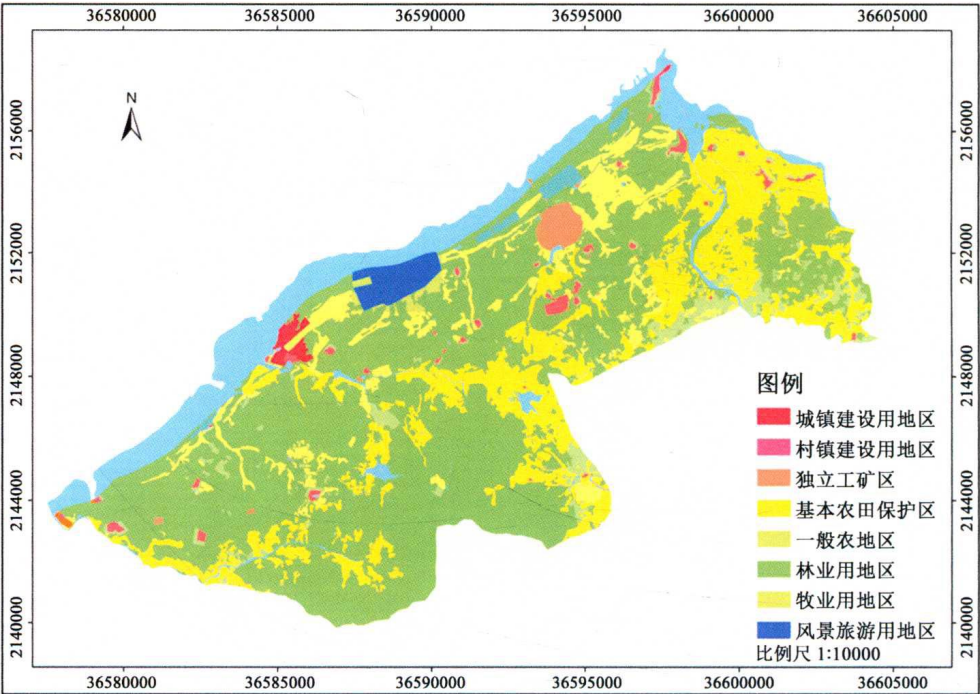


图 5-23 基于叠置分析的土地用途分区图

前 言

近年来,由于空间信息技术领域内对地观测技术、数据库技术、网络技术等的飞速发展,使得土地利用数据的获取与管理变得更为便利,我国已经实施的农用地分等定级、更新调查和“全国第二次土地大调查”等工程获得了大量的数据和资料,并建设了土地利用数据库。这些数据的复杂程度和数量远远超出人脑的分析能力,如何快速、定量地从这些大型时空数据库中挖掘有用的特征和知识已经成为土地利用数据库利用的瓶颈问题。空间数据挖掘可以从时空数据库中获取用户感兴趣的空间模式与特征、数据的关联关系以及其他一些隐含在空间数据中的规律和特征,目前已经成为国内外研究的热点。土地用途分区是土地利用规划的核心问题,但是目前还缺乏系统深入的研究,特别是在土地用途分区的智能化方面。因此,针对目前土地用途分区中存在的问题,发展面向领域的空间数据挖掘模型是时空数据不断积累过程中所提出的迫切要求。界定了面向土地用途分区的空间数据挖掘的研究内容和体系,并系统研究了该问题的理论方法和应用。

在分析国内外对土地用途分区和空间数据挖掘的研究进展的基础上,建立起面向土地信息的空间数据挖掘的基础理论和技术框架,进一步完善了空间数据挖掘的理论和方法。从土地用途分区、空间数据挖掘的定义出发,定义了面向土地利用分区数据挖掘的概念、特征和内容;提出了一种包括数据层、知识层、挖掘层和人机交互层四层结构的空间数据挖掘体系结构;阐述领域空间数据挖掘的基本步骤和从土地利用数据库中能发现的知识类型;探讨了土地用途分区数据挖掘的基本方法,主要包括空间计算模型——空间关系度量的方法、空间数据关联规则的挖掘方法——模糊概念格、空间数据聚类分析的方法——人工免疫系统的聚类算法。在对土地用途分区的问题进行描述的基础上,分析了土地用途分区知识体系,并构建了基于领域知识的土地用途分区模型。

概念格是用数学的形式化的方法对从数据中产生概念的过程进行分析的有力工具。这与数据挖掘是从大量数据中产生知识的过程是一致的,因此,概念格理论经过改进,可用于对空间数据库进行数据挖掘。本书针对概念格难以表达空间概念的问题,研究了多值背景下概念格的构建方法,并对形式概念分析理论进行了扩展,研究了基于模糊概念格的土地利用数据空间关联知识的挖掘,构建了面向土地利用的模糊概念格渐

进式算法和 Hasse 图绘制算法；针对土地利用空间数据海量的特征，引入了基于辞典序索引树算法，提出了土地利用空间关联规则的提取方法，为土地用途分区提供指导。在地理信息科学中，已有用本体的概念属性和语义距离来测度地理实体的语义相似度的相关研究。结合概念的属性、语义距离等影响语义相似度的因素，提出了一种基于本体结构的地理实体语义相似度测度模型，同时将基于语义距离的方法与基于属性的方法相结合，从而更加全面地量化本体结构中实体之间的语义相似度。

土地用途分区是在综合考虑影响土地质量与土地利用方式的各类因素（包括自然、社会、经济方面的因素）的基础上，将研究区域划分为若干均质区片的方法。土地用途分区是一个非常复杂的多目标优化问题，而聚类分析是一种典型的解决组合优化问题的方法。在分析传统的克隆选择算法的基础上，通过引入混沌理论对其进行扩展，使用 Logistic 方程改进克隆选择算法，并提出两种算法的三种结合方式，构建了混沌免疫克隆选择算法模型（CICSA）。传统聚类方法存在过分依赖数据集聚类原型的问题，为了解决这一问题，基于混沌免疫克隆选择算法提出了一种基于知识的多目标优化聚类模型。该模型是用混沌免疫克隆选择算法进行聚类，借助混沌免疫克隆选择算子的优势，将进化搜索与随机搜索、全局搜索和局部搜索相结合，通过对候选解进行操作，能够快速得到全局最优解，而不受到样本集方差分布的影响。因此，使用混沌免疫克隆选择算法能同时处理多类原型的数据聚类问题，并可以在聚类的过程中获得类数信息。提出兼顾空间特征和语义特征的综合距离计算方法，并构建顾及语义相似度的土地用途分区模型，选择典型地区开展应用研究，以期获得一种高效的土地用途分区方法。

在面向土地用途分区的空间数据挖掘的相关理论与技术研究的基础上，研究并开发了原型系统，该软件原型系统包括以下功能模块：土地利用数据管理模块、土地利用知识挖掘模块、土地用途分区挖掘模块、系统库管理模块和可视化表达模块。通过原型系统的开发，进一步明确了面向土地用途分区的空间数据挖掘的功能，解释了土地用途分区的具体过程。选择湖北宜城城市的土地利用数据库和相关数据，进行数据整合，形成可用于挖掘的整合数据库，并以此数据库进行实验研究：（1）使用模糊概念格获取土地利用的空间关联规则，并将这些规则和其他领域知识用于混沌免疫克隆选择算法抗体的编码，使用混沌免疫克隆选择算法进行基于多目标的土地用途分区聚类实验。实验结果证明笔者所研究的基于知识的土地用途分区聚类挖掘模型是智能、高效、准确的分区工具。（2）将克隆选择算法与 GIS 进行有效集成，采用空间聚类的思想，实现了利用克隆选择算法的土地用途分区方法。实验结果表明，该模型能够在多约束下进行土地用途分区，可以实现全局优化，具有稳定、结果可靠等优点，提供的土地用途分区方案科学合理，可以为土地利用规划和土地利用调控和管理提供支持。（3）通过科学设定参数，考虑空间邻近、几何形状等因素对次要图斑进行合并，获得土地用途分区方案。实验结果表明，该模型所获得的土地用途分区方案较好地反映了

人类的认知特性，概括了土地利用的特征，可以满足土地利用规划中的需求，并且分区效率有所提升。

本书是在国家自然科学基金项目“领域知识驱动的空间聚类及其人工免疫优化算法研究”（No. 41201387）、“领域知识驱动的土地利用空间优化配置与多情景模拟”（No. 41771438）的资助下取得的研究成果之一，并得到信阳师范学院学术著作出版基金、信阳师范学院第六批一级重点学科——地理学一级重点学科的联合资助。

笔者拟定提纲并撰写了本书。信阳师范学院的李卓凡和杨娜娜在整理资料、绘制图表和校对方面做了大量的工作。在本书的研究和撰写过程中，得到了信阳师范学院领导的关怀和同事的支持，在此表示诚挚的谢意。

由于研究问题的复杂性和研究时间的限制，本书的研究内容仍是初步的，在一些方面尚不成熟，错误和疏漏之处在所难免，敬请读者不吝指正。

牛继强

2017年9月

目 录

第1章 绪论	1
1.1 问题的提出	1
1.2 国内外研究进展	3
1.2.1 土地用途分区	3
1.2.2 领域知识的获取与表达	6
1.2.3 人工免疫系统算法及其改进	8
1.2.4 空间数据挖掘	9
1.2.5 存在的问题	15
1.3 研究目标与方案	16
1.3.1 研究目标	16
1.3.2 研究内容	16
1.3.3 研究思路	17
1.4 相关学科及其关系	18
1.4.1 数据挖掘	19
1.4.2 人工智能	20
1.4.3 地学数据分析	20
1.4.4 空间推理	20
1.4.5 空间数据库系统	21
1.4.6 土地利用规划	22
1.5 本章小结	22
第2章 面向土地用途分区的空间数据挖掘理论	23
2.1 有关概念	23
2.1.1 土地用途分区	23
2.1.2 空间数据挖掘	27
2.2 土地领域的空间数据挖掘内容与特征	28
2.2.1 空间数据挖掘的内容	28

2.2.2 空间数据挖掘的特征	29
2.3 土地领域空间数据挖掘的体系与过程	30
2.3.1 空间数据挖掘体系	30
2.3.2 空间数据挖掘过程	32
2.4 空间数据挖掘的方法	34
2.4.1 空间关系	34
2.4.2 空间统计学	36
2.4.3 聚类分析	37
2.4.4 Apriori 算法	39
2.4.5 概念格	41
2.4.6 决策树	43
2.4.7 人工智能	44
2.5 土地用途分区知识与挖掘模型	48
2.5.1 土地用途分区知识体系	48
2.5.2 土地用途分区挖掘模型	55
2.6 本章小结	57
第3章 土地利用知识挖掘模型	58
3.1 空间关联规则挖掘	58
3.1.1 空间关联规则的定义	58
3.1.2 空间关联规则挖掘的步骤	59
3.1.3 已有挖掘算法	60
3.2 土地利用数据的模糊概念格	61
3.2.1 土地利用的概念格表示	61
3.2.2 土地利用的模糊概念格表示	65
3.3 模糊概念格构建算法的关键技术	66
3.3.1 模糊概念格构建	67
3.3.2 渐进式算法的改进	70
3.3.3 Hasse 图的绘制	73
3.3.4 复杂空间系统的处理	75
3.4 土地利用空间关联规则提取	77
3.4.1 相关概念	77
3.4.2 生成非冗余规则	78
3.4.3 空间关联规则的直接提取法	80
3.5 基于本体的地理实体语义相似度模型	81

3.5.1 语义相似度与地理空间实体语义相似度	81
3.5.2 改进的地理实体语义相似度测度模型	83
3.6 本章小结	86
第4章 领域知识驱动的土地用途分区挖掘模型	87
4.1 土地用途分区的聚类问题	87
4.1.1 土地用途分区的问题	87
4.1.2 土地用途分区聚类问题的描述	90
4.2 混沌免疫克隆选择算法	93
4.2.1 克隆选择理论	94
4.2.2 混沌免疫克隆选择算法	98
4.3 混沌免疫克隆选择空间聚类	100
4.3.1 编码方案	102
4.3.2 抗体-抗原亲和度函数改进	103
4.3.3 混沌免疫克隆选择聚类算子	106
4.3.4 终止条件	109
4.3.5 混沌免疫克隆选择算法描述	109
4.4 顾及语义相似度的土地用途分区挖掘模型	111
4.4.1 小图斑模式的划分	111
4.4.2 土地用途分区策略	112
4.4.3 土地用途分区算法	113
4.4.4 土地用途分区结果的评价	114
4.5 本章小结	115
第5章 土地用途分区案例研究	116
5.1 面向土地用途分区的空间数据挖掘原型系统设计	116
5.1.1 系统结构设计	117
5.1.2 系统功能设计	118
5.1.3 系统数据库设计	118
5.1.4 系统模型库实现	120
5.1.5 系统知识库实现	122
5.2 土地用途分区数据整合	124
5.2.1 分区单元处理	124
5.2.2 数据预处理	125
5.3 湖北省宜城市土地用途分区	129

5.3.1	研究区概况	129
5.3.2	社会经济条件	130
5.3.3	土地利用现状与空间格局分析	131
5.3.4	土地用途分区知识获取	133
5.3.5	非分区图斑的提取	137
5.3.6	目标函数的确定	137
5.3.7	分区知识	139
5.3.8	实验结果	140
5.3.9	土地用途分区方案评价	142
5.4	海南省昌江县乌烈镇土地用途分区	143
5.4.1	研究区概况	143
5.4.2	模型总体框架	144
5.4.3	土地用途分区过程	144
5.4.4	实验结果与分析	147
5.5	海南省昌江县海尾镇土地用途分区	150
5.5.1	研究区概况	150
5.5.2	土地利用类型的语义相似度实验	151
5.5.3	土地用途分区实验	151
5.6	本章小结	154
第6章	结语	155
6.1	总结	155
6.2	创新点与研究特色	156
6.3	后续研究问题及展望	157
参考文献		159

第1章 绪 论

土地用途分区是土地利用总体规划的核心问题，但是目前还缺乏系统深入的研究，特别是在土地用途分区的智能化方面。由于近年来空间信息技术领域内对地观测技术、数据库技术、网络技术等的飞速发展，大量空间数据从 RS、GIS、GPS 多种应用中收集出来，这也为土地利用数据的获取提供了便利。我国已经实施的农用地分等定级和“全国第二次土地大调查”也因此获得了大量的数据和资料，并建设了土地利用数据库。这些数据的复杂程度和数量远远超出人脑的分析能力，虽然空间数据库具有保存这些由空间数据类型和对象的空间关系来表示空间对象的能力，但用户不可能详细地分析所有数据并提取感兴趣的空間知识。空间数据挖掘将是一个有效的工具，为解决土地用途分区这个问题提供了契机。

1.1 问题的提出

土地利用规划是对一定区域未来土地利用超前性的计划和安排，是依据区域社会经济发展和土地的自然历史特性在时空上进行土地资源合理分配和土地利用协调组织的综合措施（严金明，2001）。土地利用总体规划也是我国土地资源管理工作的“龙头”，因此编制土地利用总体规划是现阶段土地管理的核心任务。实践表明，编制土地利用总体规划是解决各种土地利用矛盾的重要手段，也是保证国民经济顺利发展的重要措施。土地利用总体规划的核心是调整土地利用结构和确定土地利用布局，一般采用土地利用分区和用地指标相结合的方式加以解决。土地利用分区是将规划区内土地划分为不同的土地利用区域，为土地利用的调控和管理提供依据。土地利用分区一般可以分为土地利用地域分区和土地用途分区。目前我国开展的土地利用总体规划修编在市级以上主要进行土地利用地域分区，也称为土地利用功能分区；在县、乡（镇）级主要进行土地利用用途分区。根据土地利用总体规划修编的工作要求，新一轮规划将要建设规划数据库和规划管理信息系统。

传统的土地利用空间分区是综合考虑影响土地质量与土地利用方式的各类因素（包括自然、社会、经济方面的因素）的基础上，将研究区域划分为若干均质区片的方法，以建立微观土地利用单元与区域土地利用管制之间的桥梁，以实现土地资源的