

空间智能: 地理信息科学的新进展

黄 波^{1,2}, 吴 波^{1,3}, 刘 彪², 曹 凯¹

(1. 香港中文大学 地理与资源管理系; 2. 香港中文大学 太空与地球信息科学研究所;
3. 福州大学 空间数据挖掘与信息共享教育部重点实验室, 福建 福州 350002)

摘 要: 在总结多年来研究 GIS 智能计算的理论与实践基础上, 提出地理信息科学发展的新方向: 空间智能。空间智能强调发现与应用空间模式, 以增强 GIS 处理复杂数据和解决复杂问题的能力。空间智能主要的技术体系由空间分析、空间优化和空间模拟三大模块构成, 其技术基础包括空间统计与索引、智能代理、高级启发式, 以及数学规划等系列智能技术。由于空间智能融合了机器学习、统计分析和人工智能等多个学科理论, 面向解决实际工程需求中大量存在的复杂时空问题, 因此理论上具有广阔的发展空间, 实践上也有重大的应用需求。随着空间智能体系的完善和技术的进一步成熟, 它将在实际应用中具有巨大的价值。

关键词: 空间智能; 空间认知; 自学习; 空间优化

中图分类号: P208 **文献标识码:** A

1 引 言

地理信息科学是 GIS 技术及其应用发展到一定水平后的必然要求, 它不是只满足于利用计算机技术来对地理信息进行可视化表达及其空间查询, 而是强调了地理信息系统的空间分析能力。目前, 地理信息科学在信息处理、存贮、提取以及管理和分析等一系列基本问题上取得了长足的进展, 并呈现出网络化、开放性、虚拟现实、集成化和空间多维性等多元化的发展趋势, 然而它的发展仍然面临着两大主要难题: 一是在规模宏大、复杂多样的数据中如何发现和应用空间模式。在现实生活中, 很多与地理相关的数据是一个庞大的数据集, 以区域和城市规划管理工作中城市人口研究为例, 除了使用有关于人口的 GIS 统计型数据外, 还使用与之密切关联的国内生产总值、工农业生产总值等社会经济统计型数据。这些数据之间具有多层次结构和逻辑关系, 在这样的复杂情况下如何发现和应用的空間模式? 二是缺乏处理复杂空间问题的功能。GIS 的主要功能还处于数据采集、管理与分析阶段, 缺乏复杂的空間决策支持能力。如空間组合定位

问题一直没能得到很好地解决。

地理信息科学的目的在于揭示地理信息的产生机理, 从而寻找出空間信息处理与分析的机制。因此, 开展空間分析、模拟和再现地理现象的研究, 进而在众多的空間数据中发现对地理现象起主导作用和具有重要影响的空間模式和规律是极为重要的。然而, 目前的 GIS 普遍缺少自我学习的能力, 不能自我纠错校正; 不能通过经验改善自身的性能; 不能自动获取和发现所需要的知识、模式和规则等。随着 GIS 应用需求的多源化、复杂化、综合化和智能化, 传统 GIS 技术的局限性表现得愈加明显。因此, 开展空間信息的认知、时空过程表达、分析、模拟与决策支持等一些基本地理问题研究, 探讨智能化的 GIS 技术的基本理论、技术方法和领域应用迫在眉睫^[1]。地理信息科学是多学科交叉融合的系统科学, 它的发展与一系列的现代信息技术的进步息息相关。随着计算技术、人工智能与运筹学等领域的快速发展, 这些领域里所取得的一些最新进展给我们带来了新的思考方法。将它们与地理信息科学的基本问题有效结合起来, 可能在很大程度上能揭示地理信息的发生机理, 极大地提高空間信息认知、时空过程分析与模拟的能力, 从而产生新的研究领域: 空間智能。

收稿日期: 2007-07-28 修订日期: 2008-03-30

基金项目: 香港政府研究基金 (RGC) 资助。

作者简介: 黄 波 (1968—), 男, 香港中文大学地理与资源管理系副教授, 研究方向为时空统计模型、空間优化、移动 GIS 以及图像处理等。已发表论文 100 多篇, 其中 SCI 检索近 50 篇。

2 空间智能的框架

空间智能的基本概念来源于心理学,它主要指形象思维的智能,具有在复杂环境下准确感觉视觉空间的能力,并且能把所知觉到的表现出来,对色彩、线条、形状、形式和空间关系很敏感,有辨别空间方位的能力。借鉴这个概念,本文提出地理信息科学的空间智能理论与方法,它的突出的特点是需要具备空间智能,即一种能够发现与应用空间模式能力,并能通过外部环境和经验不断进行学习的能力。它包含两层涵义:其一,空间认知能力;其二,自学习能力。所谓空间认知能力是指个体对客体或空间图形在头脑中进行识别、编码、贮存、表征、分解组合和抽象概括的能力,主要包括空间观察、空间记忆、空间想象和空间思维等能力因素。自学习能力包含以下几种功能:(1)强化学习功能。使得系统在下次完成同样或类似的任务时更为有效。(2)自适应功能。自适应功能要求能够对所经历事物重新构造模式、规则并修改相应的记录。(3)知识获取功能。即能够在复杂环境下挖掘空间模式、发展趋势和规律。空间智能是人们应用 GIS 进行更高层次的分析和应用的必然结果,表现为寻找未知模式、规律或趋势而在复杂时空数据中进行主动搜索和智能化的探索过程。

图 1 是作者提出的空间智能的基本框架图。框架的底层是空间统计、运筹规划及智能算法等近现代分析计算工具,它们为空间分析、优化和模拟提供技术基础。中间层是空间智能,自学习与空间认知,它们主要为地学现象和地学过程分析、模拟、预测和决策提供一种可能的技术支撑。顶层是地学规律,在于揭示地理现象的发生机理。空间智能与空间认知和空间自学习之间表现为相互促进和依存的关系。空间智能在对问题的分析、模拟及优化的基础上,从复杂多变的数据中寻找时空数据中隐含相关性或关系,形成比较具体化和系统化的空间模式,并在地学规律支持下发现问题和提出假设;空间认知可以看成是一个自下而上的精炼过程和自上而下的加工过程,它使用归纳方法发现规律,使用演绎法评估所得出的规律,反过来又能够通过空间智能技术指导空间分析、优化和模拟等基本问题。空间智能主动探索未知模式、规律或趋势的本身就是一个自学习和自适应的过程;反过来自学习能够增强空间智能在复杂环境下挖掘空间模

式和发展趋势的能力。图 1 还表明地学规律对空间智能过程具有巨大的影响,空间模式的发现与应用总是在先验的地学规律约束下进行的;空间智能把地理空间作为模型空间进行研究,通过智能计算、优化以及模式发现,形成具体化和系统化的知识,从而发现地学规律。因此,空间智能可以看成是联系空间分析与地学规律之间的桥梁。

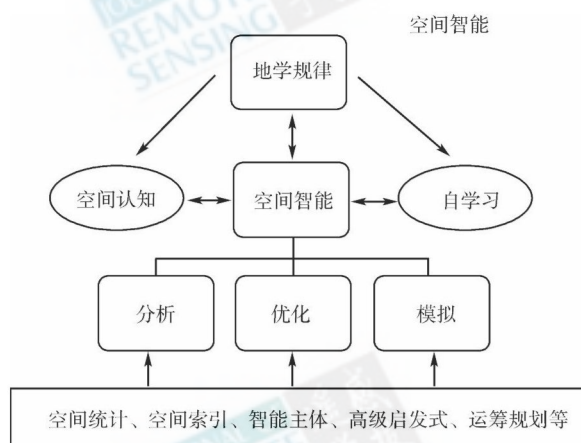


图 1 空间智能框架图

Fig 1 The framework of spatial intelligence

在技术层面,空间智能技术体系基础是软件化的知识和智能化的技术。因此,需要强调开发 GIS 主动“学习”的技术,本质上,GIS 学习是创建数据集分析程序的方法。由于很多推论问题属于无程序可循,所以有一部分 GIS 学习能力的开发强调采用相对容易处理的近似算法,使用综合归纳分析而不是演绎推理的方法。

空间智能与机器学习,统计分析和人工智能等各学科之间既有联系,又有所区别。机器学习是通过学习和训练获得或重现模式,侧重于设计新方法从数据中提取知识的一般技术行为。空间智能的模式发现则是针对特定应用领域以及终端用户的理解为目标。相对于空间统计分析,空间智能强调在隐含的、复杂的条件下对空间数据分析客观规律的挖掘,发现与应用空间模式。尽管人工智能一定程度上量化地反映类别间的数学分布的特征,但纯粹的智能计算方法,由于没有地学知识和空间特征的支持而难以真实反映一些特殊的地学分布,特别是处理复杂空间信息时难以确定其统计参数。空间智能与空间数据挖掘也有所区别,空间数据挖掘是指利用数据挖掘方法,按照一定的度量值和临界值从空间数据库中抽取知识以及与之相关的预处理、抽样和数据变换的一个多步骤相互链接、反

复进行的人机交互过程^[2 3]。相比之下,空间智能注重的是如何从空间数据库中应用分析和模拟方法来抽取知识与模式,并强调空间优化。

3 空间智能技术与应用

3.1 数据建模与分析

尽管 GIS 技术与众多的数据分析技术已经很好地融合在一起,但与空间统计学理论的结合却显得相对滞后,导致 GIS 大多缺乏鲁棒的空间分析功能。由于统计学习具有一定的模式发现和信息挖掘能力,因此,将统计分析技术与 GIS 相结合可以使 GIS 的智能分析功能得到拓展。时空模型是 GIS

领域近年来一个比较活跃的研究内容,并在空间数据(例如土地利用)变化预测方面得到了很好的应用及推广^[3]。如何构建时空模型(即在描述时空属性的同时顾及时空相关性)是该领域的重要研究内容之一。数据建模是 GIS 的基础,时空数据分析则是 GIS 成功应用与发展的关键。在空间智能的研究中,时空数据的分析与预测将是一个至关重要的方面。尽管时间序列的分析方法相对比较成熟,如回归分析、马尔可夫链方法等。但是,这些方法在强调时间序列数据分析的同时,却忽略了空间数据的分析。因此,建立统一的时空模型,同时兼顾时间和空间的数据的分析是空间智能研究的一个重要内容。

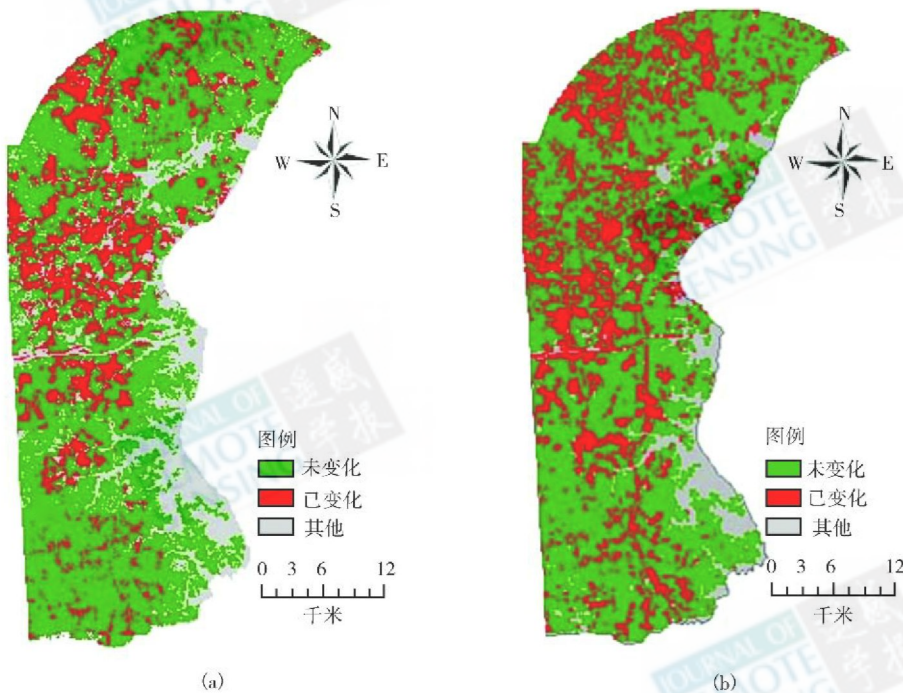


图 2 美国 New Castle 市城市土地利用扩展预测与实际变化的对比图

(a) 2002 年土地利用变化的预测; (b) 2002 年实际土地利用变化

Fig 2 Predicted land use change in comparison with actual change in New Castle, USA

(a) Predicted land use change in 2002 (b) Actual land use change in 2002

一种较好的时空数据分析与预测方法是扩展的 logistic 回归模型, logistic 回归在因变量、自变量、正态假设以及模型的可解释性等方面具有很大优势。笔者研究表明利用二元 logistic 模型来预测城乡土地空间格局变化, 利用多元 logistic 模型结合时空数据模型预测乡村用地到商业用地, 乡村用地到工业用地, 以及乡村用地到其他用地类型的多类转换, 均取得了较好的分析与预测效果。图 2 是利用空间因子, 如到城市中心的距离、到主干道的距离、人口密度等, 采用 logistic 回归模型对美国 New

Castle 市在 1984、1992、1997、2002 四年的相关数据进行分析, 对该市 2002 年的城市扩展模型进行预测的结果, 精度高达 80% 以上^[3]。

作者今后将进一步尝试其他 logistic 模型, 如 Nested Logit Model、Multinomial Probit Model 等, 并考察它们在土地利用变化分析与预测中的适用性。在模型分析的基础上, 建立一个 logistic 回归分析函数库, 满足用户的多种分析需求。利用这些模型对不同地区, 不同数据源的土地转换进行预测与评估有待于进一步研究。Logistic 模型的系数估计与样本的数量

与类型相关, 采取系统抽样和随机抽样相结合的方式消除空间相关性, 以及优选最佳样本数量等的统计方法也值得深入探讨。此外, 指数光滑的方法也将用于调节两个或者更多的连续 logistic 模型之间的模型的系数。这项研究也将着眼于处理空间相关性, 时间相关性以及时空相关性的方法。

3.2 空间优化

空间优化是指在一定区域, 一定资源和位置相关性的约束条件下, 极大或者极小空间目标函数^[4]。GIS 应用中的很多实际问题最终能够归结为最优化问题, 如资源分配、最优选址、土地利用优化布局等。由于这些问题涉及大量的非同类数据, 如经济的、政治的和自然的等因子, 数据类型上则表现为浮点型, 整型, 排序型, 甚至亚元变量等。这些复杂的数据综合交错在一起, 所建立的模型涉及复杂的非线性组合优化问题, 相比普通的优化问题更为复杂。由于大多数问题是非多项式的, 因而很难通过闭合形式或解析的求解方法来处理, 使用穷举法等方法也难以找到问题的最优解决方案。在 GIS 现有的功能里, 能够满足 GIS 需要的 (大量的, 规模的, 实时的) 有效计算方法的发展已经落后于 GIS 的发展。解决这类问题需要引进新的理论和方法, 如结合最优算法 GIS 可以处理很多大而复杂的空间寻址问题。目前, 作者已经设计了一系列具有启发性算法来解决道路和选址问题, 如遗传算法, 蚁群算法, 取得了理想的效果。蚂蚁算法是一种自适应的启发式方法, 它是建立在转换规则和荷尔蒙激素矩阵基础上的; 而遗传算法则是在自然选择基础上的染色体交叉和变异的系列操作。如果对这些方法做进一步改进, 相信可以解决更复杂的空间优化问题。图 3 是在消防站点选址的多目标规划中, 作者分别采取蚁群算法和遗传算法进行选址计算的效率比较图。

图 3 表明蚁群算法与遗传算法能够很快达到目的, 但蚁群算法比遗传算法效率更高。由于遗传算法与蚁群算法是基于群体行为的启发式方法, 如何将两者有效结合以加快收敛速度是值得深入研究的。此外, 局部搜寻或者全局搜寻经常需要找到一个相邻的目标或者是满足一定条件的目标。然而, 很少有将空间索引技术用于提高搜索效率, 因此在这方面也需要进行探讨。人工免疫系统是近期出现的一种新的启发式方法, 这也是值得关注的一个方向。

在增强以上启发性算法的同时, 我们的研究也

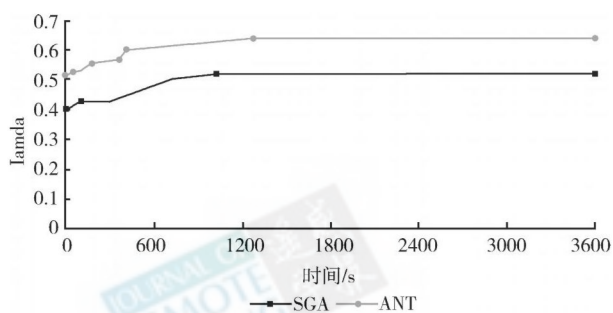


图 3 蚁群算法和遗传算法效率对比图

Fig 3 Comparison of the efficiency between ant algorithm and genetic algorithm

将聚焦在多目标优化上。空间决策中经常使用多目标最优来寻求一个合理的解决方案。传统的赋权相加法常把多目标最优问题转化为单目标问题, 虽然方法简单易行, 却找不到累托前沿的凹点。一些改进的启发式算法, 比如非支配顺序的遗传算法 (NSGA), 细分帕累托遗传算法 (NPGA) 和微遗传算法 (MicroGA) 也常用来确定累托前沿点, 但是它们有其特定的应用范围, 并且不能保证所寻求的解是最优的。因此, 需要更高级, 更通用的启发式方法, 以期将这些启发式和 GIS 相结合而精确快速地找到非统治解。除此之外, 复杂的应用研究也值得关注, 如多次覆盖问题。竞争和协作选址模型和需求不定选址模型也是很有意思的专题。

3.3 基于智能主体的模拟

复杂性科学理论指出, 复杂系统中大量的微观智能主体之间随时间的推移相互作用能够在系统宏观尺度上突现出新的结构和功能。这种思想强调局部的规则将导致系统宏观全局的变化, 关注数量众多的各种智能主体聚集在一起而产生的宏观现象, 假如为不同的智能主体赋予不同的行为频率和时空跨度, 那么基于智能主体的模型就在一定程度上表现出传统数学模型几乎无法描述的多尺度特性。因此智能主体模型内在地表达了系统的层次性。智能主体模型是一种自下而上的方法, 用来模拟主体与主体以及主体与外界环境之间相互作用的全局结果。主体在不同的环境里通常表示不同的对象, 比如, 在生态系统里, 它表示植物或者动物; 在交通中, 它就表示交通工具; 在人群中, 它就表示人。这些模型通常设定一个环境或者一个框架, 在这个环境或者框架里相互影响发生, 并根据不同的行为 (程序规则) 或者特征参数来定义主体。

近年来,人们开始将智能主体技术作为一种模拟工具,用来模拟复杂的人类和社会活动,以及微观领域的相互作用,并在市场和传播模拟等方面得到了成功的应用。但是目前智能主体模拟受限于简单的行为定义,并不能反映真实的群体行为。因此,需要能整合更复杂行为的高级智能主体模型。例如在交通模拟中,虚拟世界的旅行者模拟的是某个地区的真实的人的旅行和驾驶行为。单个交通工具的相互影响产生了现实的交通动态,分析者通过交通动态可以估计汽车的废气排放量,并评估这个运输系统的表现。作者曾用智能主体模型对实时的交通状况进行了模拟研究,在这个模型中,驱动这些模拟工具的思想主要借助于智能主体模拟来完成。举例来说,在车辆及行人流量的模拟中,使用者最优的交通分配、等候模型可以作为全局约束来管理个体行为。

复杂的行为模拟还可以通过整合智能主体模型和分析模型来实现高级智能主体模型。此外,群体智能模型也可以得到合适的应用。作者将进一步把智能代理研究用于疏散计划的分析模型相结合。由恐慌引起的人群惊跑常常带来诸多不幸或者严重的受伤。非理性的群体行为通常导致过度拥挤和撤退缓慢,从而带来更大的损失。在智能主体模型中,集体恐慌是一种自然现象。这种现象是由相关的复杂个体级别的行为和个体间的相互影响造成的(催眠效用,原始本能的相互激发)。在我们的研究中,将用智能主体模型来模拟群体恐慌行为,以了解并预防恐慌。不同的撤离计划将会通过智能主体模拟进行测试。智能主体技术在检验不同假设中的应用还很少见,但随着时间的推移,这一技术将得到广泛的应用。

4 结论与讨论

本文提出空间智能的框架并分析在现有的一些技术支持下,如空间索引、统计分析、智能代理、启发式、以及数学规划等,可以找到一些解决复杂空间问题的有效方法。今后我们将进一步补充完善、系统化这些方法,建立起空间智能的原型系统。这个系统包含分析、模拟、优化三个紧密组成模块,这三者相互作用,共同促进空间智能的发展。优化模块可以推动分析与模拟的发展,如在分析或模拟过程中,可能会出现数量众多的参数,最优化的方法有助于快速找到最合适的参数集;同样,分析和模拟可以给优化问题提供良好的初始值或者搜索范围,从而提高计算效率。因此,它们的结合能够更好的处理一些复杂问题。

参考文献 (References)

- [1] XU G H. Application and Prospect on Remote Sensing and Resources and Environment Information System [J]. *Remote Sensing of Environment* 1994 9(4): 241—246 [徐冠华. 遥感与资源环境信息系统应用与展望 [J]. 环境遥感, 1994 9(4): 241—246]
- [2] Li D R, Wang S L, Li D Y. Spatial Data Mining Theories and Applications [M]. Science Press 2006 [李德仁, 王树良, 李德毅. 空间数据挖掘理论与应用 [M], 科学出版社, 2006]
- [3] Huang B, Xie C L, Tay R, et al. Land Use Change Modeling Using Unbalanced Support Vector Machines [J]. *Accepted for publication in Environment and Planning B*.
- [4] Huang B, Liu N, Chandramouli M. A GIS Supported Ant Algorithm for the Linear Feature Covering Problem with Distance Constraints [J]. *Decision Support Systems* 2006 42 (2): 1063—1075

Spatial Intelligence: Advancement of Geographic Information Science

HUANG Bo^{1, 2}, WU Bo^{1, 3}, LIU Biao², CAO Kai¹

(1. Department of Geography and Resource Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China;

2. Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China;

3. Key Lab of Spatial Data Mining and Information Sharing of Ministry of Education, Fuzhou University, Fujian Fuzhou 350002, China)

Abstract Notwithstanding the astounding growth achieved by Geographic Information Systems (GIS) in recent decades, some critical bottlenecks still continue to pose challenges to the advancement of this technology. The inability to handle large and diverse datasets and the lack of functionalities to solve complex spatial problems such as combinatorial location problems is essential one among these hurdles. Recent advances in Computational Intelligence (CI) and Operations Research have, however, opened up new avenues to overcome these obstacles to the development of GIScience.

Judicious integration of the aforementioned techniques into GIScience could possibly lead to an innovative discipline namely spatial intelligence. This paper presents our preliminary investigation on this subject including its framework, major acquisition methods and sample applications. The basic concept of spatial intelligence derived from psychology refers to the mental process associated with the brain's attempts to interpret certain types of information received. This information basically includes any kind of mental input such as visual pictures, maps and plans. Based on this concept we propose that the introduction of spatial intelligence within the domain of GIScience is an ability to discover and apply spatial patterns which is usually elicited through analysis/mining, optimization and simulation. Two characteristics of spatial intelligence are highlighted here. One is the ability of spatial cognition and the other is the self learning capability. Spatial cognition refers to the process of recognizing, encoding, saving, expressing, decomposing, constructing and generalizing spatial objects which can be obtained from spatial observation, spatial perception, spatial indexing and spatial deductive inference. Self learning includes enforcing learning, adaptive learning and knowledge acquiring abilities to actively dig up knowledge from observation data. Promoting spatial intelligence is a logical requirement for higher level analysis and application of GIScience. Through active learning and searching process in complex spatiotemporal data, spatial intelligence discovers unknown spatial patterns, trends and regularities.

From the technical perspective, we emphasize the use of meta-heuristic and other intelligent algorithms to address complex geospatial problems. A three-tiered structure is proposed for the spatial intelligence framework. At the bottom of the framework, spatial statistics, programming and intelligence computation are used to provide the foundation of spatial analysis, simulation and optimization. The middle level consists of spatial intelligence, self learning and spatial cognition for analyzing, simulating, interpreting and decision making for geospatial processes and phenomena. At the top of the framework, GIScience laws and regularities are used to mine unknown patterns. To realize the goal of spatial intelligence, a solid research that integrates key topics with the concepts and modeling approaches derived from Information Science and Operations Research to advance GIS theories have been developed. The core supporting methods and techniques pertinent to the proposed framework include spatial analysis, simulation and optimization. The development of spatial analytical models to represent spatial and temporal features and their relationships forms a vital aspect of this research. Spatial optimization is employed to maximize or minimize a planning objective, given the limited area, finite resources and spatial relationships for a location-specific problem, once spatiotemporal patterns have been discovered. Simulation is an important tool to evaluate and improve models and spatial patterns.

Some successful applications of spatial analysis, optimization and simulation are also reported in this paper. Logistic regression models, e.g., binomial, multinomial and Nested Logit are applied and examined to predict various spatiotemporal changes including rural to commercial, rural to recreational and rural to other land uses. A range of heuristic algorithms such as Genetic Algorithms (GA) and Ant Colony Systems (Ant), to troubleshoot complex routing and location problems and multiobjective optimization for spatial decision making are studied. A case study of integrating agent-based modeling with analytical models by drawing upon microscopic traffic simulations for emulating real-time traffic conditions is also conducted.

Key words spatial intelligence; spatial cognition; self learning; spatial optimization