



全国高校地理信息科学教学丛书

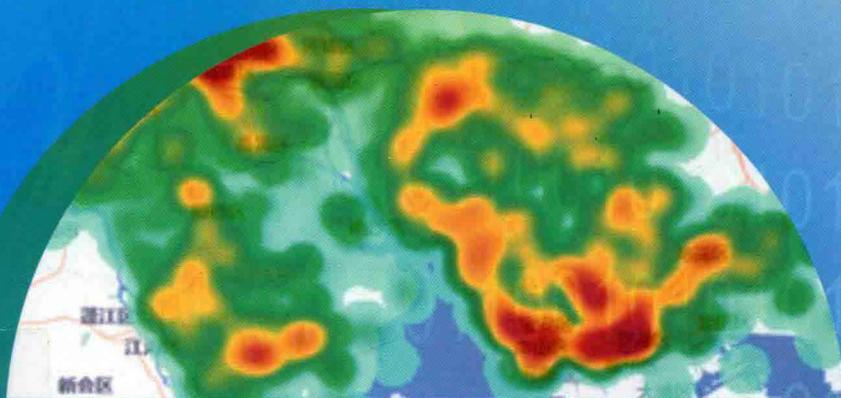


普通高等教育“十一五”国家级规划教材

GIS空间分析

(第三版)

刘湘南 王平 关丽 卢浩 张春晓 编著



科学出版社

8.1.2 地理空间大数据的特征	214
8.1.3 地理空间大数据分析与传统空间数据分析的区别	214
8.1.4 数据驱动的科学	215
8.2 大规模地理空间数据高效分析	215
8.2.1 空间分析并行计算	215
8.2.2 基于 GPU 计算的空间分析	218
8.2.3 分布式空间分析	220
8.2.4 空间分析基础设施 CyberGIS	226
8.3 基于 GIS 的社会感知计算	227
8.3.1 时空行为模式分析	228
8.3.2 公众情感分析	230
8.3.3 空间交互分析	231
复习思考题	233
第 9 章 地理计算与空间建模	234
9.1 GIS 与专业模型的耦合	234
9.1.1 松散耦合	234
9.1.2 紧密耦合	235
9.2 地理计算智能	237
9.2.1 智能计算概述	237
9.2.2 人工神经网络	239
9.2.3 遗传算法	242
9.2.4 元胞自动机	243
9.2.5 深度学习	246
9.3 地理空间模型	247
9.3.1 地理相互作用模型	247
9.3.2 地理过程模拟	249
9.3.3 地理空间优化	250
9.4 地理设计与空间决策	254
9.4.1 地理设计	254
9.4.2 空间决策支持系统	256
9.4.3 空间决策分析	256
复习思考题	258
主要参考文献	259

(3) 可执行程序集成方法。可执行程序集成方法是指 GIS 与应用分析模型以可执行文件的方式独立存在, 二者的内部、外部结构均不改变, 相互之间独立存在。二者的交互通过文件、命名管道、匿名管道或数据库以约定的数据格式进行。可执行程序集成方式可分为独立方式和内嵌方式。

(4) DDE 与 OLE 集成方法。DDE 或 OLE 与内嵌的可执行程序的集成方法相似, 属于松散的集成方法, 系统的数据交换使用了操作系统内在的数据交换支持, 使程序的运行更加流畅。

(5) 模型库的集成方法。模型库指按一定的组织结构存储模型的集合体。模型库可以有效地管理和使用模型, 实现模型的重用。模型库符合客户机/服务器 (C/S) 工作模式, 当需要模型时, 模型被动态地调入内存, 按照预先定义好的调用接口来实现模型与 GIS 的交互操作。

(6) 基于组件的集成方法。组件技术是系统集成最流行的方法之一。应用软件模块和支持组件编程的语言有很多, 包括 VB、Delphi、VC 等, 都可开发 GIS 与应用分析模型的集成系统。目前组件技术有 Sun 的 JavaBeans、OMG 的 Corba 技术、Microsoft 的 Com 及 Microsoft 对 Com 技术的发展 Com+ 和 .Net 组件技术。

9.2 地理计算智能

随着 GIS 应用水平的不断提高, 越来越多的复杂应用问题对 GIS 空间分析功能提出了更高的要求。因此, 把数学、计算科学、统计学和信息科学等领域的理论和方法引入地学研究, 将神经网络、遗传算法、元胞自动机、深度计算等人工智能技术与 GIS 相结合, 可增强 GIS 空间分析功能, 提高解决实际应用问题的能力。

地理计算是地理信息科学的核心内容之一, 主要研究地理信息科学的方法学问题。它应用数学计算方法和技术描述空间特征、解释地理现象、解决地理问题, 本质上可认为是对地理学时间与空间问题所进行的基于计算机的定量化分析。地理计算和智能挖掘结合产生了地理计算智能。GIS 为地理计算提供数据库支持, 人工智能技术和智能计算技术提供计算原理和工具, 高性能计算服务系统提供动力。

近年来, 智能计算技术中的神经网络模型、遗传算法模型、元胞自动机模型及深度计算等不断被引入并成为地理计算的核心, 用于解决复杂地理问题。目前, 已成功应用于遥感分类、土地覆被变化模拟、景观生态评价、城市动态模拟等领域。

9.2.1 智能计算概述

智能计算 (computational intelligence, CI), 也称为“软计算”。传统人工智能 (artificial intelligence, AI) 使用的是知识, 在面对许多涉及识别、认知、理解、学习、决策等方面的问题时, 特别是人类仅凭自身的经验、直觉就能解决的问题时, 方法上就存在知识表达或建模的困难。智能计算则基于操作者提供的数据, 不依赖于知识, 以数据为基础, 通过训练建立联系, 进行问题求解。Bezdek 认为智能计算具有适应性运算能力、计算的容错能力、人脑的计算速度, 以及与人脑类似的决策与思维的正确率。智能计算的本质与传统硬计算不同, 其目的在于适应现实世界遍布的不确定性, 因此智能计算的指导原则是扩展对客观世界不确

定性的容忍,以达到对不确定性问题的可处理性、鲁棒性、低成本求解等目标。

1. 智能计算定义

(1) 智能计算是受自然界(生物界)规律的启迪,根据模仿其原理设计求解问题的算法,如人工神经网络技术、遗传算法、进化规划、模拟退火技术和群集智能技术等。

(2) 智能计算(如神经网络、进化、遗传、免疫、生态、人工生命、主体理论等)作为第二代人工智能方法,是连接主义、分布式人工智能和自组织系统理论等共同发展的结果。

(3) 智能计算是用计算机模拟和再现人类的某些智能行为。从方法论的角度,智能计算大致可分为三种基本类型:以符号操作为基本特征的符号机制;以人工神经网络为代表的联结机制;以遗传算法为代表的进化机制(进化论)。符号机制从抽象层次模拟和再现人类的某些智能行为,演绎方法构成其主要的逻辑框架;联结机制从神经元相互作用的层次模拟再现人类的某些智能行为,归纳法,尤其是不完全归纳法构成其主要的逻辑框架;进化机制从自然进化的角度探寻智能的形成方式,基于试探和反馈的自适应奖罚策略构成其主要的逻辑框架。

(4) 智能计算广义地讲就是借鉴仿生学思想,基于生物体系的生物进化、细胞免疫、神经细胞网络等某些机制,用数学语言抽象描述的计算方法,用以模仿生物体系和人类的智能机制。从方法论的角度和目前的研究状况来看,智能计算有五种基本类型:①适用于处理不确定信息的模糊数学和粗集理论;②再现人类某些智能行为的神经网络;③以模拟生物进化规律为特征的进化算法;④以免疫操作为基本特征的免疫算法;⑤以脱氧核糖核酸(DNA)复制为基本特征的DNA计算。

总的看来,智能计算技术是从模拟自然界生物体系和人类智能现象发展而来的,可以在人们改造自然的各种工程实践中取得实际效果。

关于人工智能和计算智能的关系,不同学者持有不同的观点。Bezdek 等认为智能具有三个层次:第一层次是生物智能(biological intelligence, BI),它是对智能的产生、形成和工作机理的直接研究,主要是生理学和心理学研究者所从事的工作,大脑是其物质基础;第二层次是人工智能,是非生物的,它以符号系统及其处理为基础,来源于人的知识和有关数据,主要目标是应用符号逻辑的方法模拟人的问题求解、推理、学习等方面的能力;第三层次是计算智能(computational intelligence, CI),由数学方法和计算机实现,来源是数值计算及传感器所得到的数据。他们认为计算智能是人工智能的子集,人工智能是计算智能到生物智能的过渡。另一些学者认为人工智能和计算智能是不同的范畴。Eberhart 将计算智能定义为一种包含计算的方法,显示出有学习或处理新情况的能力,从而使系统具有如泛化、恢复、联想和抽象等一种或几种推理功能。计算智能系统通常包括多种方法的混合,如神经网络、模糊系统、进化计算系统及知识元件等。事实上,无论是CI还是AI都各具特点、问题、潜力和局限,只能相互补充而不能相互取代。

智能理论与技术的研究无止境,智能技术的新概念、新名词将不断出现,智能技术的最高层次是在结构和功能上接近人的大脑。一般认为,计算智能是神经网络、模糊计算、进化计算及其融合技术的总称,是基于数值计算和结构演化的智能,是智能理论发展的高级阶段。

2. 智能计算特点

(1) 智能性。智能计算技术的智能性包括自适应、自组织和自学习性等,这种自组织、

自适应特征赋予该技术具有根据环境的变化自动发现环境的特性和规律的能力。

(2) 稳健性。智能计算的稳健性是指在不同环境和条件下算法的适用性和有效性,利用智能计算技术求解不同问题时,只需设计相应的适应性评价函数,而无需修改算法的其他部分。

(3) 不确定性。智能计算技术的不确定性是伴随其随机性而来的,其主要操作都含有随机因子,从而在算法的进化过程中,事件发生与否带有较大的不确定性。

(4) 强化计算。智能计算不需要很多待求解问题的背景知识,而主要依赖于大量快速的运算从数据集中寻找规则或规律。这是智能计算领域的普遍特征。

(5) 容错性。神经网络和模糊推理系统都有很好的容错性。从神经网络中删除一个神经元,或是从模糊推理系统中去掉一条规则,并不会破坏整个系统。由于具有并行和冗余的结构,系统可以继续工作。

(6) 全局优化。传统的优化方法一般采用的是梯度下降的爬山策略,遇到多峰函数时容易陷入局部最优。遗传算法能在解空间的多个区域内同时进行搜索,并且能够以较大的概率跳出局部最优以找出整体最优解。

智能计算以连接主义的思想为主,并与模糊数学和迭代函数系统等数学方法相交叉,形成众多的发展方向。智能计算主要包括模糊逻辑(fuzzy logic)、神经网络(neural network)和概率推理(probabilistic reasoning),随后还增加了进化计算[evolutionary computation,包括遗传算法(genetic algorithms)、进化策略(evolutionary strategies)和进化规划(evolutionary programming)等三个分支]、学习理论(learning theory)、置信网络(belief network)和混沌理论(chao theory)内容。一些研究认为智能计算还应包括非线性科学中的小波分析、混沌动力学、分形几何理论、免疫算法(immune algorithm)、DNA计算、模拟退火技术(simulated annealing algorithm)、多智能体(multi-agent)系统及粗糙集理论(rough sets)和云理论(cloud theory)等。

智能计算并不是单一的方法,而是众多方法和技术的集合。大体而言,模糊逻辑(fuzzy logic)、神经计算(neural computation)和遗传算法(genetic algorithms)是智能计算技术的核心,这些技术是互补关系,而不是竞争关系。模糊集合理论借助隶属度来刻画模糊事物的亦此亦彼性,考虑模糊性,重在处理不精确的概率。而粗糙集以自己的上近似集和下近似集为基础,笼统考虑随机性和模糊性,具有很强的定性分析能力,可用于不确定影像分类、模糊边界划分等。云理论是一个分析不确定信息的新理论,由云模型、不确定性推理和云变换三大部分构成。云理论把定性分析和定量计算结合起来,适于处理GIS中随机性和模糊性为一体的属性不确定性。空间统计学可估计模拟决策分析的不确定性范围,分析空间模型的误差传播规律,改善GIS对随机过程的处理等。神经网络反映大脑思维的高层次结构,善于直接从数据中进行学习;模糊计算模仿低层次的大脑结构,推理能力较强;进化计算模拟生物体种群的进化过程,实现优胜劣汰,很适合于求解全局最优问题,但其学习的精度不如神经网络,推理能力不如模糊系统。

9.2.2 人工神经网络

人工神经网络(artificial neural network, ANN)是一种非线性分析模型,它对于模拟和揭示具有非线性、自组织性、开放性等特征的地理复杂系统及其规律具有明显的优势。

人工神经网络从积累的工程实例中训练、学习建立各影响因素间的非线性映射关系，对于非线性系统数据具有较高的拟合能力及预测精度。另外，它对于残缺不全或不确定信息具有较强的容错能力，因此 ANN 模型更适用于研究多因素的非线性地理问题，在一定程度上可以避免传统数学方法建立模型时所遇到的尴尬。将 GIS 与人工神经网络技术相结合处理复杂地理问题可以提高 GIS 的“智商”，具有理论上的可行性和实践价值。

1. 人工神经网络模型基本概念

人工神经网络又称为人工神经系统 (ANS)、神经网络 (NN)、自适应系统 (adaptive systems)、自适应网 (adaptive networks)、联结模型 (connectionism) 等。它是一个并行的分布处理结构，由多个处理单元及其联结的无向信号通道互连而成。这些处理单元具有局部

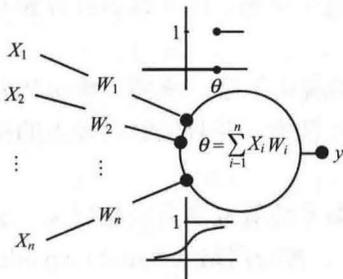


图 9.6 MP 神经元模型

内存，并可以完成局部操作。每个处理单元有一个单一的输出联结，这个输出可以根据需要被分支成希望个数的许多并行联结，且这些并行联结都输出相同的信号，即相应处理单元的信号大小不因分支的多少而改变。处理单元的输出信号可以是任何需要的数学模型，每个处理单元中进行的操作必须是完全局部的。也就是说，它必须仅仅依赖于经过输入联结到达处理单元的所有输入信号的当前值和存储在处理单元局部内存中的值。

神经网络的基本处理单元为神经元。神经元由分离的多条输入纤维和一条输出纤维构成，利用轴突上活动电位的电脉冲信息，电脉冲通过突触的这种连接体转换成化学信号，并且传递到下一个神经元。1943 年，McCulloch 和 Pitts 定义了简单的人工神经元模型，称为 MP 模型。它的一般模型可以用图 9.6 来描述。

2. 典型的人工神经网络模型

1) 层次型神经网络

层次型神经网络起源于 20 世纪 60 年代出现的感知器。80 年代中期，Rumelhart 等发表了反向传播 (back propagation, BP) 算法的学习算法。由于层次型神经网络能够学习输入模式和对应的输出模式之间的关系，在模式识别和控制等方面都有非常广泛的应用。BP 网络是其中较典型的一类模型，也是各种人工神经网络中发展和应用最为瞩目的一类模型。它是一种非线性变换单元的前馈式网络，具有大量可供调节的参数和很强的非线性模拟运算能力，已成为一种重要的信息处理方法，特别是在分析预测非线性系统未来行为上具有巨大潜力。

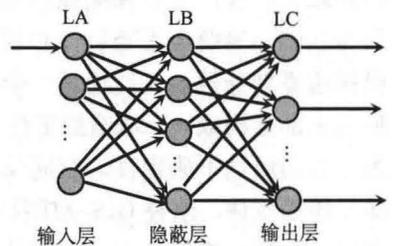


图 9.7 三层前馈型 BP 网络

一个典型的三层前馈型 BP 网络的拓扑结构如图 9.7 所示，从结构上分为输入层 LA、隐蔽层 LB 和输出层 LC。同层结点间无关联，异层神经元间前向连接。其中，LA 层含 m 个结点，对应于 BP 网络可感知的 m 个输入；LC 层含有 n 个结点，与 BP 网络的 n 种输出响应相对应；LB 层结点的数目 u 可根据需要设置。

三层前馈型 BP 网络存储知识 (即调整网络连接权值及结点阈值) 时采用的 BP 方法，即误差逆传播学习方法，是一种典型的误差修正方法。其基本思路是：把网络学习时输出层出

现的与“事实”不符的误差，归结为连接层中各结点间连接权及阈值（有时将阈值作为特殊的连接权并入连接权）的“过错”，通过把输出层结点的误差逐层向输入层逆向传播以“分摊”给各连接结点，可算出各连接结点的参考误差，并据此对各连接权进行相应的调整，使网络适应要求的映射。

2) 互联型神经网络

互联型神经网络可以分为联想存储模型和用于模式识别及优化的网络。联想存储模型主要包括福岛的时空间模式联想存储模型、霍普菲尔德网络（Hopfield network）、HASP（human associative processor）、BAM（bidirectional associative memory）及MAM（multidirectional associative memory）等形式。层次型神经网络的信号流为单向，主要用于模式识别，而互联型神经网络为双向，用于联想存储及最优化。计算机通过设定有序的地址来存储记忆的内容。要将内容取出时，可以通过地址读取内容。对要存储的内容，进行符号化的变形，基本上以二进制信号的形式存储。

互联型神经网络中存在反馈回路，它在时间序列数据上表现良好，较为流行的互联型递归神经网络包括 Elman 神经网络、Jordan 神经网络和 Hopfield 神经网络等。Hopfield 神经网络是较典型的互联型神经网络模型。它是美国物理学家霍普菲尔德（Hopfield）在 1982 年提出的，引入了物理学能量函数的思想，对稳定性问题给出解决方案。该模型是由 N 个结点全部互连而构成的一个反馈型动态网络，可实现联想记忆，并能进行优化问题求解（图 9.8）。

Boltzmann 神经网络是最早的全连接神经网络之一，它能够学习内部表征、解决组合数学问题。从结构上讲，它是 Hopfield 神经网络的推广或变形。Boltzmann 神经网络不仅可以解决优化问题，还可以通过学习，模拟外界所给的概率分布，实现联想记忆。

人工神经网络是一个高度的非线性映射模型，许多自然资源的动态变化过程就是一个非线性映射过程。因此，利用 BP 模型经学习训练后，建立资源分布动态模型，可以达到理想的收敛效果，将为资源管理模拟仿真提供一种新的、高精度分析方法。例如，可应用人工神经网络建立输入量（林龄）与输出量（单位面积蓄积量）之间的非线性映射关系，对森林资源分布格局进行模拟。

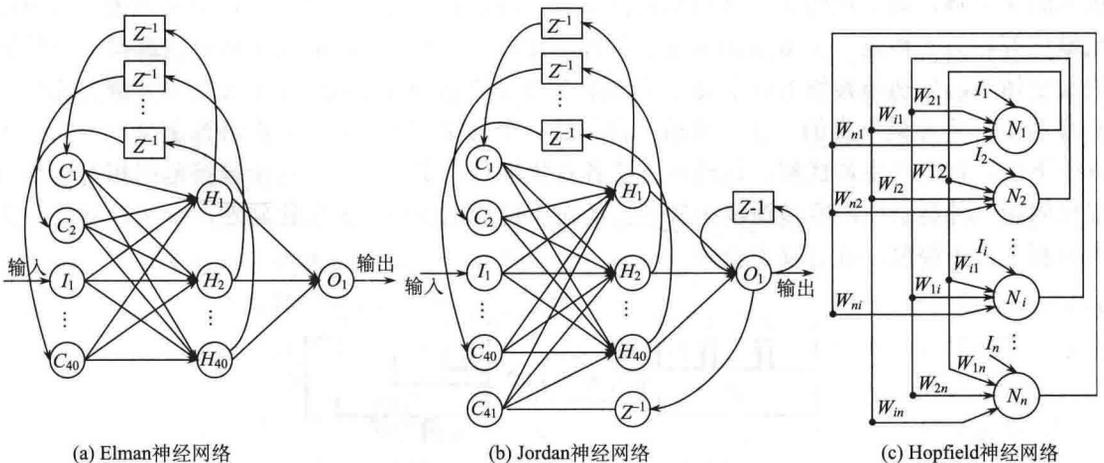


图 9.8 互联型递归神经网络