

自组织特征映射神经网络原理和应用研究*

李春华 李 宁 史培军

(环境演变与自然灾害教育部重点实验室;北京师范大学资源学院:100875,北京)

摘要 根据耕地利用在经济、社会和生态方面面临的压力状况,以压力-状态-响应框架为基础,设计一套压力大小量化的指标体系,引入具有较强的聚类 and 容错能力的自组织特征映射(self-organizing mapping, SOM)神经网络模型,在说明 SOM 网络模型和算法的基础上,应用 SOM 的聚类功能,以 MATLAB 语言构建 SOM 网络模型,对我国的 31 省市自治区耕地利用压力大小进行了分类,并结合相关文献的研究成果阐述耕地压力的地域差异原因。结果显示我国耕地压力的区域差异与经济地域差异有高度的一致性,表明经济发展是耕地压力的主要来源。选取大样本的神经网络训练得到的结果和现实的一致也表明, SOM 模型是一种适用的耕地压力区域分类新方法。

关键词 自组织特征映射网络;耕地压力分类;地域差异;中国

随着我国经济和人口的快速增长,耕地资源大量减少,耕地保护压力成为学术界研究的热点,许多学者^[1]从不同的角度探讨耕地减少的因素,有的以不同类型的自然地理区和经济地理区为单位分析耕地变动的驱动力,有的学者以县、市、省级行政区为单位,分别选取影响耕地变动的社会、经济、自然因素作分析,这可能揭示地域性的影响因素,但在地域经济联系密切的条件下,地域分工的经济规律的作用愈加明显,耕地利用会受到其他地区的影响,使这种分析有其局限性。他们使用得较多的是相关、聚类、回归、因子和主成分分析等统计方法,由于土地利用的变化过程与其驱动因素之间具有很强的非线性关系,选取常规变量的因子分析法和逐步回归法等也多是建立在线性模型基础上的,所以不适用土地利用变化的复杂性问题的分析。由于大多数经济、人文方面的研究很难获得完备的数据集,信息来源可能既不完整又含有假象,甚至存在特例和反例,传统的一些聚类方法对所需数据的要求比较严格,因此也难以胜任这些工作。自组织特征映射(self-organizing mapping, SOM)模型判定方法则能够根据学习获得的知识经验对复杂问题进行合理的判断决策而突破以上障碍^[2]。本文以我国耕地压力分类为例,通过 SOM 的聚类功能,描述我国省级耕地压力差异在空间上的变化,进一步探索形成这种差异的原因,找出减轻耕地压力的方法。

1 自组织特征映射模型的原理

1.1 SOM 网络的拓扑结构 SOM 网络一个典型的特性就是可以在一维或者二维的处理单元阵列上形成

输入信号的特征拓扑分布, SOM 网络具有抽取输入信号模式特征的能力。其中应用较多的二维阵列模型由 4 部分组成: 1) 处理单元阵列,用于接受事件输入,并且形成对这些信号的判别函数; 2) 比较选择机制,用于比较判别函数,并选择一个具有最大函数输出值的处理单元; 3) 局部互连作用,用于同时激励被选择的处理单元及其最邻近的处理单元; 4) 自适应过程,用于修正被激励的处理单元的参数,以增加其对应于特定输入的判别函数的输出值^[3-4](图 1)。

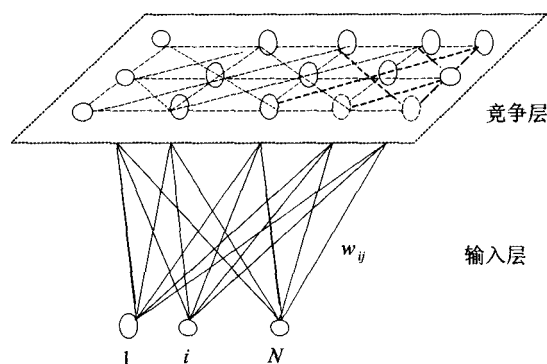


图 1 SOM 网络的拓扑结构示意图

假如网络输入为 $X \in R^n$, 神经元 i 与输入单元的连接权值 $W_i \in R^n$, 则输入神经元 i 的输出 O_i 为:

$$O_i = W_i X.$$

网络实际具有响应的输出单元 k , 该神经元的确定是通过“赢者通吃”的竞争机制得到的, 输出为:

$$O_k = \max_i \{ o_i \}.$$

以上 2 式可修正为:

$$O_i = (i + r_k o_i - o_k) / s_i,$$

* 北京市自然科学基金资助项目(8062020); 国家自然科学基金杰出青年基金资助项目(40425008); 国家自然科学基金资助项目(40541001)

收稿日期:2006-03-01

$$i = \sum_{j=1}^m w_{ij} x_j, O_k = \max_i \{o_i\} - ,$$

式中 w_{ij} 为输入神经元 i 和输出神经元 j 之间的连接权值, x_i 为输入神经元 i 的输出, 为一个很小的正数, r_k 为系数, 它与权值及横向连接有关, s_i 为与处理单元 i 相关的处理单元集合, o_k 称为浮动阈值函数. (t) 为非线性函数, 即:

$$(t) = \begin{cases} 0, & t < 0, \\ (t), & 0 \leq t \leq A, \\ A, & t > A. \end{cases}$$

1.2 SOM 网络学习算法 SOM 网络的学习算法过程如下:

1) 初始化. 赋予 N 个输入神经元到输出神经元较小的连接权值. 选取输出神经元 j 临接神经元的集合 s_j . 其中, $s_j(0)$ 表示时刻 $t=0$ 的神经元 j 的临接神经元的集合, $s_j(t)$ 表示 t 时刻 j 的临接神经元的集合. $s_j(t)$ 随着时间的增长而不断缩小.

2) 提供新的输入模式 X .

3) 计算欧式距离 d_j . 即输入样本与每个输出神经元 j 之间的距离

$$d_j = \|X - W_j\| = \sqrt{\sum_{i=1}^N [x_i(t) - w_{ij}(t)]^2},$$

并计算出一个具有最小距离的神经元 j^* , 即确定某个单元 k , 使得对于任意的 j , 都有

$$d_k = \min_j (d_j).$$

4) 给出一个周围的临域 $s_k(t)$.

5) 按照下式修正输出神经元 j^* 及其临接神经元的权值

$$w_{ij}(t+1) = w_{ij}(t) + (t) [x_i(t) - w_{ij}(t)],$$

其中, (t) 为一个增益项, 并随时间变化逐渐下降到零, 一般取

$$(t) = \frac{1}{t} \text{ 或者 } (t) = 0.2(1 - \frac{t}{10\,000}).$$

6) 计算输出 o_k .

$$O_k = f(\min_j \|X - W_j\|),$$

式中 $f(\cdot)$ 一般为 $0 \sim 1$ 函数或其他非线性函数.

7) 提供新的学习样本来重复上述学习过程.

2 SOM 模型在我国耕地压力程度分类中的应用

2.1 耕地压力来源的 PSR 框架分析 本文样本指标根据压力-状态-响应框架提取. 就耕地压力而言, 其来源主要有 2 个: 一是直接因素, 即农业结构调整、城市化的发展、生态退耕、后备耕地资源的限制及自然灾害毁损; 二是间接因素, 它包括农地产权制度的缺陷、现

行土地征用制度的缺陷、现行土地管理体制的缺陷、土地市场不健全及耕地价值认识不清等^[5]. 在以上双层压力左右下, 耕地的质量和数量都在发生变化, 就质量而言, 耕地的本底质量、结构质量和经济质量发生着变化, 而耕地数量变化是指耕地变为其他农用地以及变为建设用地从而数量减少和未利用地开垦为耕地而导致耕地数量的增加. 耕地利用的主体(国家、集体、农户和其他用地者)会对耕地利用的质量和数量状况作出响应而采取措施, 正确的措施会有利于耕地的利用和保护, 错误办法和措施会给耕地造成一种新的压力.

2.2 样本的选取和指标的意义 首先利用频度统计, 即利用 CNKI 数据库对 1998—2005 年有关耕地可持续利用评价的 36 篇文章和土地利用 PRS 分析方法的 48 篇文章进行频度统计, 筛选出使用频率较高的指标, 然后, 按照压力-状态-响应框架(图 2), 提取压力层的指标, 从而建立压力状态的量化备选指标体系, 最后为使指标体系具有可操作性, 考虑指标数据的可得性并征询专家意见, 得到具体指标体系(表 1)

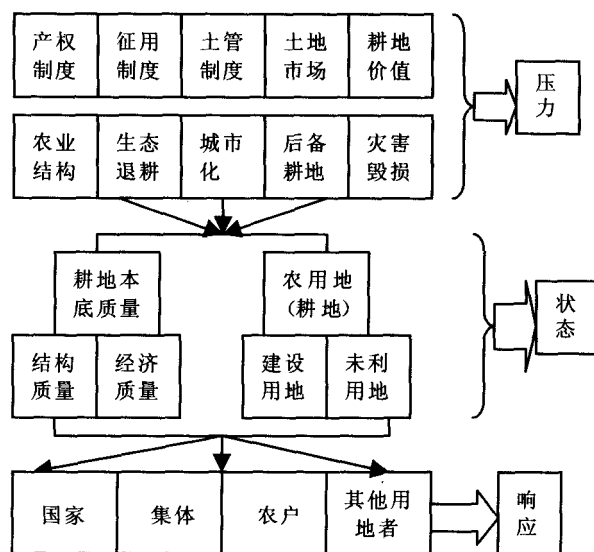


图 2 我国耕地利用的压力来源框图

3 网络样本设计和网络计算^[6-7]

3.1 网络样本设计 使用的 SOM 网络模型是以 MATLAB 语言构建, 利用其神经网络工具箱提供的新建、训练、模拟等函数可以方便地完成整个学习过程. 根据《新中国五十年统计资料汇编》、1986—2004 年的《中国统计年鉴》(因为 1986 年新《土地管理法》通过, 我国开始实行集中的土地管理制度)、各省同期的统计年鉴、同期的《中国工业经济统计年鉴》、《中国农业经济统计年鉴》、《中国人口统计年鉴》、《中国环境年鉴》、《中国土地年鉴》、《中国国土资源年鉴》等获取原始数据资料. 原始数据经过人均和地均计算后, 再作归

表 1 我国耕地压力状态的量化指标体系

指 标	指 示 意 义
城市化水平 (p_1) ,人均居住用地 (p_2)	反映城市化发展和耕地数量之间的关系
人均耕地面积 (p_3) ,复种指数 (p_4)	反映耕地利用现状
人口密度 (p_5) ,城市人口密度 (p_6)	反映现状耕地的人口压力
文盲率 (p_7) ,每万人科技人员数 (p_8)	反映人地系统中人的素质的地域差异
后备耕地资源率 (p_9) ,土地垦殖率 (p_{10})	表示地区耕地资源禀赋
地均投资 (p_{11}) ,地均耕地用电量 (p_{12}) ,地均耕地化肥量 (p_{13})	反映耕地投入程度和投入量对耕地质量的影响
城市居民食品消费 (p_{14}) ,农村居民食品消费 (p_{15})	反映保证粮食安全对耕地最低数量的需求压力
工业废水处置率 (p_{16}) 和废物处置百分比 (p_{17})	表示地区环境质量对耕地质量的影响
客运量 (p_{18}) ,路网密度 (p_{19})	反映交通建设占用对耕地面积变化的影响
生态退耕面积占耕地总面积的比例 (p_{20})	表示耕地面临的生态退耕的压力
成灾与受灾比 (p_{21}) ,水土保持系数 (p_{22})	反映耕地利用环境脆弱性程度
作物种植面积与总播种面积比 (p_{23})	反映农业结构调整和耕地数量变化的关系
城乡收入差别 (p_{24})	反映由于城乡收入差别引起的农户土地利用行为与耕地数量之间的关系
城市划拨用地比重 (p_{25}) ,城市经营性用地出让中招标拍卖比重 (p_{26}) ,耕地征用价格 (p_{27})	反映土地市场化程度和耕地征用制度对耕地的影响
户均耕地面积 (p_{28}) ,土地利用中耕地案件的结案率 (p_{29}) ,农产品价格指数 (p_{30})	反映土地产权制度和土地管理状况 ,农产品价格变化对耕地数量变化的影响

一化处理 ,导入 SOM 网络 ,网络的输入模式为 :

$P_k = (p_1^k , p_2^k , \dots , p_n^k) , k = 1 , 2 , \dots , 31 , n = 30$.
即一共有 31 组样本向量 ,每个样本中包含 30 个元素 .
3.2 网络创建和训练及计算 首先利用 new 函数创建一个 SOM 网络

net = new som (minmax (p) , [5 , 3])

式中 p 为输入向量 , minmax (p) 指定了输入向量元素的最大值和最小值 , [5 , 3] 表示创建网络的竞争层为 5 × 3 的结构 ,它是通过不断调整网络结构得出的 ,然后利用函数 train 和仿真函数 sim 对网络进行训练并仿真 .由于训练步数影响网络的聚类性能 ,所以这里设计 10 , 100 和 1 000 这 3 个步长分别观察其分类性能 .经过尝试发现 ,其中第 1 000 步的聚类结果是 : [7 4 5 10 8 10 10 7 8 8 4 8 4 8 4 4 4 8 5 5 5 4 5 5 10 4 10 10 10 10] ,它跟实际情况相符 ,所以选用这一分类结果 (表 2) .

表 2 基于 SOM 模型的中国耕地压力程度分类结果

类别	所含地区
1	北京 ,天津 ,上海
2	江苏 ,浙江 ,山东 ,广东 ,福建 ,辽宁
3	河北 ,河南 ,安徽 ,湖北 ,湖南 ,江西 ,陕西 ,四川
4	山西 ,海南 ,贵州 ,云南 ,广西 ,重庆
5	西藏 ,青海 ,宁夏 ,新疆 ,甘肃 ,内蒙古 ,吉林 ,黑龙江

4 结果分析

根据 SOM 网络聚类分析的结果 ,结合我国经济发展的地域差异 ,可以把以上 5 种类型 (1 ~ 5 类) 归结为 3 类地区 :

第 1 类 经济发达地区 .其范围大致包括上述第

类和第 2 类 ,它们位于环渤海 ,长江三角洲和东南沿海地区 .该区耕地承受的压力主要来自 2 个方面 :

1) 经济发展、工业化和城市化使耕地数量减少 .该区城市化和耕地保护的矛盾主要表现在 :非农建设用地导致的耕地面积减少主要集中于经济发达地区 ,耕地递减的速率与经济发达程度高度相关 .另外 ,人口由农村向城镇、由经济落后地区向经济发达地区流速的加快 ,将使该区人地资源短缺的矛盾进一步加剧 .未来半个世纪 ,该区将进入高速城市化阶段 ,城镇用地将成为建设用地占耕地的主体 .据“全国土地利用总体规划”预计 ,1985 — 2000 年全部新增建设用地占耕地数中 ,城镇占地达到 43 % ,居各主要建设用地类型之首 [8] .

2) 经济发展、工业化和城市化使耕地质量下降 .一方面 ,农业内部用地因单纯追求经济利益而导致结构不合理 ,化肥普遍代替传统的绿肥、畜肥以及河泥肥 ,导致耕地质量下降 .工业化和城市化使经济取得快速增长的同时引起的环境污染 ,也致使耕地生态环境日渐恶化 .另一方面 ,该区耕地无论是开发程度还是利用强度 ,都已达到了较高水平 ,农业用地后备资源不足 ,进一步的建设依然可能将大量占用平整肥沃的高质量耕地和园地 .因此 ,工业化、城市化不可逆转的趋势与土地资源不足的矛盾使这些地区耕地承受很大压力 .

第 2 类 经济中等发达地区 .其范围包括上述的第 3 类 ,主要是我国中部省份 ,大部分地区也是国土资源部公布的粮食主产区 (河北、河南、黑龙江、吉林、辽宁、湖北、湖南、江苏、江西、内蒙古、山东、四川、安徽) ,该区耕地因为自身自然条件和承担我国粮食安全任务

而承受压力,其来源主要是:农业结构调整、基础设施建设、城乡居民居住、灾毁,其中农业结构调整占用耕地占耕地减少量的一半以上.1990—2000 年粮食主产区耕地减少 354.2 万 hm^2 ,1996—2000 年在我国实施严格耕地保护制度的情况下,粮食主产区耕地仍然减少 66.7 万 hm^2 ,年均减少 13.3 万 hm^2 .另外,城镇村和工矿用地面积比重和人均占有量均高于全国平均水平.建设用地增长主要是交通和农田水利设施用地,约占全国同类用地的一半左右,该区受盐碱、渍涝危害的耕地面积大,造成中低产田分布广,2002 年粮食主产区成灾与受灾面积分别占全国的 66%和 67%.而且粮食主产区多为传统农区,工业化技术起点低,环保措施不力,工业生产废物给农业生态环境带来严重影响,这不仅说明粮食主产区生态环境恶化较为严重,而且表明粮食主产区抗灾防灾的能力也比较弱^[9].但是,该区农业开发历史久远,土地开发利用程度较高,人口稠密,城镇密集,部分地区的光热资源较好,耕地资源相对丰富,人均耕地高于全国平均水平,耕地自然和社会经济条件一般较好,后备土地资源有一定开发潜力,再加上国家对粮食主产区的保护政策,使耕地的压力有所减轻.

第 3 类 经济欠发达地区.其范围包括第 1 类和第 2 类,该区除了吉林、黑龙江、山西、海南等省外,是经济欠发达地区,其耕地利用压力主要来自区域自然环境和社会经济不利条件:

1) 社会经济方面.人均国民生产总值低于全国平均水平,工农业产值之比远远低于发达地区,城镇功能不完善,人力资源较丰富,但素质较低,资金严重不足,自然资源较丰富,但匹配性差,科学技术滞后,技术人才贫乏,这些都不利于耕地资源的可持续利用.

2) 自然环境方面.主要属山地、丘陵和高原地区,生态环境复杂多样,生态系统破坏难恢复,生态退耕的压力大,自然灾害较频繁,自然资源有退化趋势,土地质量较差.我国主要的沙漠、戈壁和“三化”土地(退化、沙化和碱化)都分布在该区,它们受自然灾害影响,灾毁的面积较大,耕地成片性差,规模经营不足.

因此,尽管该区耕地面积比例较高,但是耕地利用的自然条件限制性因素较多,在同样的消费水平条件下,东部和西部地带耕地承受的压力相当,只是驱动这种变化的动因不同,东部地带主要是由于耕地资源流失严重,导致人均耕地面积剧烈降低,西部由于自然、经济、技术条件的限制,耕地生产力水平较低且提高缓慢,人均消费所需耕地面积较大,致使耕地仍然承受了与东部相当的压力^[10].

5 结论和讨论

1) 耕地压力程度的区域差异明显.我国耕地利用压力的地域差异和经济发展程度密切相关,说明我国经济发展是耕地压力的主要来源.为了证明这一点,笔者用径向基函数神经网络模型(RBF)对耕地压力的影响因素进行识别和对耕地压力大小进行排序,结果和 SOM 网络计算结果基本一致,这种结果也和蔡运龙等^[10]用最小人均耕地面积法计算的耕地压力区域差异也是一致的.本研究的政策含义是耕地的区域平衡不能只限定在省域范围,而应根据不同地区耕地利用承受压力状况求得不同地域尺度的平衡.耕地压力程度的神经网络(SOM)方法测定的结果为耕地的区域调控提供了可供操作的依据.

2) SOM 模型是一种耕地压力区域分类的新方法.SOM 模型能把高维的信息数据以有序方式映射到低维空间上,形成一种拓扑意义上的有序图,能实现实时学习、网络具有自稳定性、无须外界给出评价函数、能够识别向量空间中最有意义的特征、抗噪音能力强等优点,它是一种无导师的聚类方法,网络的学习过程无需提供理想输出模式,在分类过程中可避免人为因素,这些都有利于耕地压力的聚类分析,它比传统的有监督分类法更有优势.本文把它引入耕地压力区域分类,得到满意的结果.

3) 有待进一步改进方面.一是网络的设计和训练.SOM 是一种人工神经网络模型,其权值的调整仅考虑了学习率及输入模式与获胜神经元及其邻域,却忽略了输入模式分量与全体参与竞争的神经元权值向量间的某种相关关系.当网络的连接过多,结点数目庞大时,其计算量大,需要较长的学习时间,所以要提高网络计算的性能就必须改进网络的设计和训练方法.二是耕地质量指标的问题.我国耕地压力状态的量化指标体系的建立是耕地利用压力程度量化结果科学性的又一个关键.本文耕地质量状况指标主要从耕地本底质量、结构质量和经济质量.笔者认为耕地压力状态的质量量化指标还应具体细化,以耕地的环境条件与土壤条件为基础建立耕地质量评价指标体系,耕地质量标准应当是指耕地的生产能力,而不仅仅是耕地本身的相关指标.

6 参考文献

- [1] 刘旭华,王劲峰.国家尺度耕地变化驱动力的定量分析方法[J].农业工程学报,2005,21(4):57
- [2] 徐建华.现代地理学中的数学方法[M].北京:高等教育出版社,2002,448-452

- [3] 丛爽. 面向 MATLAB 工具箱的神经网络理论与应用 [M]. 合肥: 中国科学技术大学出版社, 1998, 125-128
- [4] 王伟. 人工神经网络原理—入门与应用 [M]. 北京: 北京航空航天大学出版社, 1995, 157-165
- [5] 杨瑞珍. 我国耕地资源流失原因的深层剖析与政策建议 [J]. 中国人口·资源与环境, 2005, 15(3): 108
- [6] 付强, 王志良, 梁川. 自组织竞争人工神经网络在土壤分类中的应用 [J]. 水土保持通报, 2002, 22(1): 39
- [7] 刘兴久. 模糊聚类分析在土壤分类中的应用 [J]. 东北农学院学报, 1988(2): 12
- [8] 王卫. 我国耕地保护的经济分析及对策研究 [J]. 自然资源学报, 1997, 12(4): 383
- [9] 王大伟, 刘彦随, 卢艳霞. 农业结构调整对全国粮食安全的影响分析—以粮食主产区为例 [J]. 中国人口·资源与环境, 2005, 15(2): 65
- [10] 蔡运龙, 傅泽强, 戴尔阜. 区域最小人均耕地面积与耕地资源调控 [J]. 地理学报, 2002, 57(2): 127

THE PRINCIPLE AND APPLICATION OF SELF-ORGANIZING MAPPING NETWORK IN THE CLUSTER OF CULTIVATED LAND USE PRESSURE IN CHINA

Li chunhua Li ning Shi peijun

(Key Laboratory of Environmental Change and Natural Disaster, Ministry of Education of China;
College of Resource Science and Technology: Beijing Normal University, 100875, Beijing, China)

Abstract Cultivated land is one of the most important natural resources in China. In order to find the way for the sustainable cultivated land use and to overcome the default of regular cultivated land use pressure evaluation, First a logging theological identification technology is introduced and the principle of SOM (self-organizing mapping) network is summarized. Then a set of cultivated land-use pressure evaluation indicators based on "Press-State-Response" framework is set up according to the cultivated land current condition in China in terms of economy, society and ecology. A program is written and MATLAB 6.1 software is employed to measure the cluster of regional cultivated land use pressure. The results of SOM show that 31 provinces (cities) or autonomous regions are classified into 5 categories. On the basis of economic development level, they are grouped into 3 groups: developed-area, middle-developed area and less-developed area. The regional cultivated land pressure level disparities are obvious. Cultivated land pressure status estimated on SOM can explore distinctively the reason and the results of cultivated land use pressure changes, which will help administrator to adopt suitable land policy and management measures to alleviate cultivated land protection pressure and improve cultivated land quality. The evaluation results also indicate that the application of SOM neural network to assessing regional differentiations of cultivated land use pressure level without assuming parametric relationship is convenient, precise and feasible which can be an alternative approach of assessing regional differentiations of cultivated land use pressure level.

Key words self-organizing mapping network; cultivated land-use pressure cluster; regional difference; China