

# 基于改进 GA 神经网络的可持续发展水平研究

程兆麟<sup>1</sup>, 曾孟佳<sup>2</sup>

(1.西华师范大学 区域经济研究所;2.西华师范大学 计算机学院,四川 南充 637002)

**摘要:**提出了基于改进遗传算法神经网络的可持续发展水平评估模型,该模型利用遗传算法辅助网络训练,克服了传统评价体系中权重确定带有主观性、评价结果缺乏公允等缺陷。建立了适合我国国情的可持续发展评价指标体系,并在此基础上利用样本数据对该模型的评分效果进行了实证研究,结果表明该模型具有良好的应用性。

**关键词:**遗传算法;神经网络;可持续发展能力;指标体系

**中图分类号:**F061.3

**文献标识码:**A

**文章编号:**1001-7348(2005)04-0074-03

## 0 前言

面对当今全球的环境问题,人们终于认识到了为了适应当今和未来发展需要,必须努力寻求一条经济、社会与环境、资源相协调的可持续发展道路。可持续发展理论、概念及实现途径和方法的研究已成为当前国内外研究的前沿和各国政府及公众关注的热点<sup>[1-3]</sup>。不过至今仍未建立一个能被一致认可的可持续发展的评价指标体系和评价模型。

目前,国内外学术界和决策部门对区域可持续发展的研究评价体系大多是在建立在对不同层次指标进行加权评判的基础范式之上的,这类范式虽然简便易算,可操作性较强,但是权重的确定却存在着极大的主观性,评估结果缺乏公允。

基于此,本文提出了一种基于 GA-NN 的可持续发展评估模型和算法,在学习并且展现评价专家经验的同时,减少主观因素影响。

## 1 可持续发展水平指标体系

### 1.1 指标体系的建立

1987年,世界环境与发展委员会(WCED)

在其《我们共同的未来》报告中提出:“可持续发展是满足当代人的需要而又不损害后代人满足其需要的能力的发展”<sup>[4]</sup>。可见,可持续发展是一个复杂的系统,根据定义,利用层次分析法将可持续发展能力划分为3个子系统:经济系统发展能力、社会系统支撑能力和环境与资源系统支撑能力<sup>[5,6]</sup>。

本文基于层次性与结构性,全面性与代表性,可操作性与可靠性,动态性与静态性等诸原则构建可持续发展指标体系(见表1)。

### 1.2 可持续发展指标主成分预处理

为了简化输入指标,并排除部分指标之间存在的相关性,本文采用了主成分分析对数据进行了预处理(数据来源于1999年、2000年《中国统计年鉴》<sup>[7]</sup>)。预处理后的结果如下:

**经济系统:**提取4个主成分,累计方差贡献率达到77.952%,各主成分分别表征地区经济发展总量水平、经济均量水平、产业结构、经济发展速度;

**社会系统:**提取4个主成分,累计方差贡献率达到78.992%,各主成分分别表征社会教育与安定水平、人口密度与社会文明、基

础设施与生活质量状况、社会科技发展水平;

**环境系统:**提取4个主成分,累计方差贡献率达到73.078%,各主成分分别表征自然灾害抵抗力与生态系统质量状况、社会资源状况、环境系统质量状况、自然资源状况。

## 2 基于遗传算法的神经网络

### 2.1 前向人工神经网络

神经网络(Neural Networks, NN)是以大量的具有相同结构的简单单元联接来模拟人类大脑的结构和思维方式的一种可实现的复杂网络系统,它以其并行分布处理、自组织、自适应、自学习、具有鲁棒性(Robustness)和容错性等独特的优良性质引起广泛的关注,尤其是在住处不完全、样本缺损和参数漂移的情况下,仍能保证稳定的输出。

神经网络模型中,在分类差别方面应用最广泛的就是前向反馈网络(BP网络),其一般由一个输入层、一个输出层和一个或多个隐含层组成。输入神经元接受环境的信息,位于网络最底层;输出神经元把信息传递到环境中去位于网络最高层;隐含神经元不与环境发生交互作用。神经元I接收其它神经元传递来的输入信息,根据和函数 $net$ ,进行加

表1 可持续发展指标体系

经济系统	经济总量	经济总量 GDP,全社会固定资产投资总额,社会消费品零售额;
	经济均量	人均 GDP, GDP 密度,人均实际利用外资额;
	经济结构	非农产业产品比重,城乡经济结构(乡城人均纯收入比),投资结构(固定资产投资中非国有比重),产值结构(工业 GDP 中非国有比重);
	经济效益	工业全员劳动生产率,工业产值利税率,社会消费品零售额增长率;
	经济速度	GDP 增长率,全社会固定资产投资增长率,社会消费品零售额增长率;
	经济外向度	进出口额占 GDP 百分比,FDI 占 GDP 百分比,单位面积货运周转量;
社会系统	人口发展	人口密度,社会老龄化指数,平均受教育年限;
	基础设施	交通密度(公路+铁路),100 户电话机数,人均邮电业务量,城市人日生活用水量;
	科技教育	从业人员万人拥有国企专业技术人员,教育科技经费占 GDP 比重,万人中学生数,专利授权率,大专以上文化程度人员比重;
	生活质量	恩格尔系数(农、城),收入水平(农、城),城市人均居住面积;
	城市化	城镇密度,非农人口比重;
资源与环境	安定与文明	社会保障覆盖率(以三保人口参保率为例),城镇失业率,万人交通事故率,万人拥有医生数,成人识字率;
	政府能力	财政收入占 GDP 比重;
	自然资源	人均水资源保有储量,人均耕地,人均土地,人均煤保有储量,人均铁矿石保有储量;
	社会资源	劳动力比重,旅游资源(以年接待国际旅游人数替代),负担系数;
环境系统	自然灾害	受灾面积占幅员比重;
	森林覆盖率	森林覆盖率,年造林面积占幅员比重,保护区占辖区面积比重;
	生态系统	工业三废处理达标率,除涝治碱率

选择。选择过小,则收敛速度很慢,反之则可能回馈过度而产生震荡。③在有随机扰动下不能达到最佳效果,学习过程容易陷入局部极小值。

对于 BP 算法存在的缺陷,本文采用以下措施加以改进:①单元激活函数选择 log-sigmoid 函数,即  $f_i = \frac{1}{1+e^{-x_i}}$ ,并对数据进行归一化处理;②运用 Levenberg-Marquardt(L-M)算法训练前向神经网络。L-M 算法在内慧足够的情况下,具有其它算法不可比拟的优势<sup>[8,9]</sup>;③增加进化层,进化层利用遗传算法修正 BP 网络权值与偏差,以克服网络训练过程中的局部极小化缺陷。

本文中设计的 BP 网络采用两层结构,即一个输入、一个输出层和一个隐含层。输入层共有 M 个神经元,隐含层有 N 个隐含节点,输出层有 R 个输出神经元,p 为输入向量(如图 1 所示)

2.2 改进遗传算法

遗传算法(Genetic Algorithm, GA)由自然界的遗传进化理论发展而来,是一种基于自然选择和群体遗传机理的搜索算法,它模拟了自然选择和遗传过程中的繁殖、杂交和突变的现象。目前,遗传算法已经成功地解决了许多复杂的优化问题,它最大的优点是:即使对多态的和

非连续性的函数,也能获得全局最优解。因此,许多学者都尝试将遗传算法与神经网络结合在一起,利用遗传算法优化神经网络的结构和权重系数。本文中采用的遗传算法通过对传统遗传算法中的交叉、变异过程进行自适应改进,提高其优化效率。

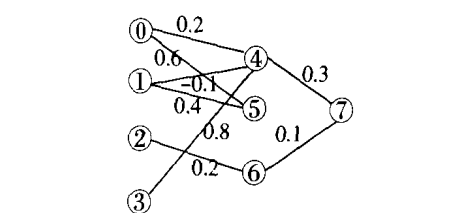


图2 遗传算法编码图

熵值。如果单纯采用二进制编码,会造成编码串太长,且需要再解码为实数,影响网络学习的精度。本文采用二值编码和实数值的混合编码方法:神经网络的结构采用二值编码,而神经网络的连接权重和阈值系数采用实数编码(如图 2 所示)。

编码结构中,二值码“1”表示有相联,二值码“0”表示无相联。起始的群体中的每个码串是随机产生的,但实值编码被限定在可能的连接权重系数取值范围之内。起始码串采用均匀分布的随机分布函数产生,并成为下一次遗传搜索过程中个体的父母集合。

(2)目标函数及适应度函数。将由码串表示的每个个体反编码为相应的神经网络,然后输入所有训练样本,计算神经网络的输出与期望输出之间的误差平方和,因此,目标函数采用改进的 De Jong 函数,即求  $\min f(x)$ :

$$f(x) = \sum_{j=1}^{N_{var}} \sum_{i=1}^{N_{ind}} (x_{ij}^{sample} - x_{ij}^{chrom})^2$$

理论上,De Jong 函数在 0 处有全局最优解 0,表示遗传优化后的权重与输入误差平方和为 0,即最相似;随着 f(x)增大,相似性变差。实际上,最好的权值并不一定是最小的距离,而应该是适当的距离。

将同一代群体中的 L 个染色体按目标函数值从小到大进行排序,排序后染色体的序号记为 s。构建适应度函数:

$$f_s = 2 \left[ \frac{1}{L} - \frac{s}{L(L+1)} \right]$$

则 s 越小,适应度越大,染色体越好。

(3)遗传算法的交叉、变异。在传统遗传算法下,对所有个体的交叉概率  $P_c$  和变异概率  $P_m$  选用某一固定参数,在整个遗传算法过程中保持不变。在这种情况下,如果群体中个体彼此非常接近,则个体的选择概率相当,交叉后不能产生新的个体,这样搜索过程就不能有效继续进行,进化过程陷于停滞

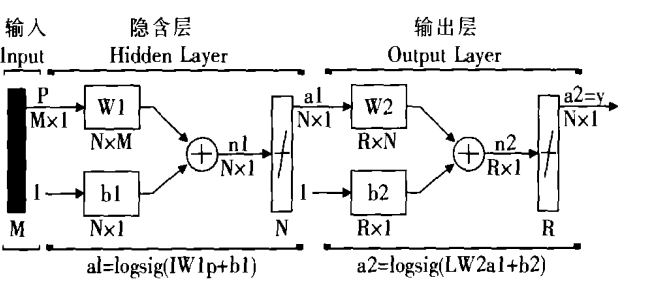


图1 BP 网络结构图

非连续性的函数,也能获得全局最优解。因此,许多学者都尝试将遗传算法与神经网络结合在一起,利用遗传算法优化神经网络的结构和权重系数。本文中采用的遗传算法通过对传统遗传算法中的交叉、变异过程进行自适应改进,提高其优化效率。

(1)遗传算法的编码。神经网络的权重学习是一个复杂的连续参数优化问题;每个遗传码串代表一种神经网络的结构和权重/

权平均,然后根据激励函数  $f_i$  产生输出信息,输出信息又按照网络的拓扑结构传递到下一个神经元。

在运用 BP 算法时要注意:①初始值的选择。BP 网络的连接模式是用权值向量表示的,权值向量决定着网络如何对环境中的任意输入作出反应。权值应当是随机产生的小数,如果神经元的输入数值为 R,则初始权值区间可估计为  $[-2/R, 2/R]$ ;②学习率的

评价与预测  
中国科学院评价研究中心合办

状态(stagnation),这就是通常所说的早熟性收敛。在这种情况下,交叉操作作用不明显,而且增加了目标函数值的计算时间,只能依靠变异操作产生新的个体。引入变异算子后搜索范围扩大,有利于找到全局最优解,但由于搜索范围广,收敛速度将受到影响。相反,在进化过程中如果个体彼此差异很大,交叉操作可以使有效模式相互组合,繁殖大量“好”的后代,提高进化速度。变异时,都采用随机选取一位进行翻转变值的方法,不能有效提高二进制编码的搜索性能及稳定性。本文采用一种交叉和变异概率的自适应调整规则,使得每个个体按其适应度大小选择不同的交叉概率  $P_c$  和变异概率  $P_m$ 。而且,在遗传过程中根据适应度的变化自动调节这两个控制参数。这样,群体中每个个体对环境的变化就具有自适应调节能力<sup>[10]</sup>。

在优化过程中,设  $P_c$  表示交叉率;  $P_m$  为变异率;  $f_{\max}$  表示种群最大适应度值;  $\bar{f}$  表示种群平均适应度值;  $f'$  表示在要交叉的两个个体中较大的适应度值;  $f$  表示要变异的个体适应度值。这里,  $k_1, k_2, k_3, k_4$  是在 0 和 1 之间取值的常数,  $k_3$  和  $k_4$  较大。则遗传算法的自适应交叉和变异概率由下式表示:

$$P_c = \begin{cases} k_1(f_{\max} - f') / (f_{\max} - f'), & f' > \bar{f} \\ k_3, & f' \leq \bar{f} \end{cases}$$

$$P_m = \begin{cases} k_2(f_{\max} - f) / (f_{\max} - f'), & f > \bar{f} \\ k_4, & f \leq \bar{f} \end{cases}$$

交叉概率和变异概率可以自适应调整,对适应度小的个体采用较大的交叉概率,以加速其更新速度;对适应度大的个体采用较小的交叉概率,使这些较优的解不易丢失。若群体中的个体趋于相同时,个体的适应度就接近群体的平均适应度,由上式可知,每个个体的变异概率也较大,因此很快得到新个体,从而保证了群体的多样性,并克服了算法的早熟现象。

### 2.3 改进的 CA-NN 网络流程图

结合 BP 网络和遗传算法的特点,本文设计了如图 3 所示的算法。

## 3 可持续发展能力验证

### 3.1 训练样本及 GA-NN 模型建立

神经网络模型有效与否与样本的设置重要关系,在组织学习样本时,应该考虑各方情况,尽可能多地为网络提供必要

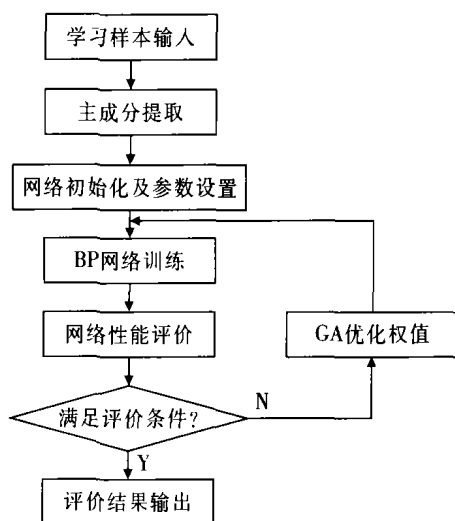


图3 GA-NN 网络流程图

的信息。最初的样本输出不可避免地包含了专家的主观意见,一方面可以通过增加专家数量和专家水平来加以修正,另一方面可以通过改进指标体系来进行完善。

区域经济可持续发展分析标准见表 2。

本文建立的网络模型采用两层结构,根据训练样本和指标体系,建立 4 个样本,4 个输入节点,10 个隐含层节点和 5 个输出点的可持续发展能力评价 GA-NN 模型。4 个训练样本的期望输出值列于表 2 右端。网络训练经过 514 个时间单位(epoch)后误差陷入局部极小(见图 4),从而转入 GA 进化。应用

表 2 经济可持续发展能力分析标准和期望输出值

主成分	F1	F2	F3	F4	期望输出值				
I 级	0.045	0.214	-0.993	0.823	0.99	0.01	0.01	0.01	0.01
II 级	-0.899	0.318	1.115	1.269	0.01	0.99	0.01	0.01	0.01
III 级	-0.912	0.327	-1.353	1.295	0.01	0.01	0.99	0.01	0.01
IV 级	-0.721	0.293	-1.478	-0.117	0.01	0.01	0.01	0.99	0.01
V 级	-0.365	0.193	-1.034	-0.386	0.01	0.01	0.01	0.01	0.99

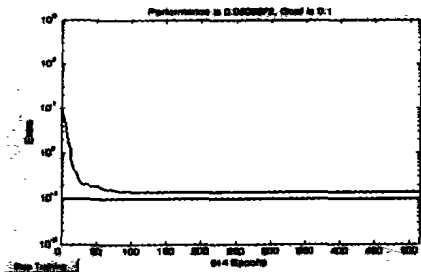


图 4 网络误差变化图

表 4 西部部分地区经济可持续发展能力综合评价结果

区域	F1	F2	F3	F4	综合评价结果和等级					专家评级	
重庆	0.001185	-0.00062	0.00018	0.000426	0.000	0.000	0.995	0.035	0.018	III	I
四川	0.00359	-0.00058	0.00011	-4.1E-05	0.918	0.000	0.026	0.074	0.018	I	I
贵州	0.00101	-0.00076	0.000122	-0.00103	0.000	0.000	0.012	0.062	0.860	V	V
云南	-0.00036	0.000944	-0.00056	-0.00207	0.000	0.057	0.135	0.658	0.000	IV	IV
陕西	0.001208	-0.006	0.000151	6.5E-06	0.212	0.753	0.000	0.030	0.000	II	II
青海	-0.00085	-0.00021	3.21E-05	1.36E-05	0.000	0.000	0.682	0.187	0.000	III	IV

GA 算法,通过 500 代优化,得到全局最优网络权重。经过训练,样本拟合度达到 100%。网络学习结果见表 3。

### 3.2 模型的计算机实现

本文提出的 GA-NN 算法是在 WindowsXP 平台上基于 Matla6.5 实现的<sup>[11]</sup>。

### 3.3 区域可持续发展能力综合评价

待评价的经济系统的检测样本见表 4。应用经过上述训练后的网络权重和阈值,检测样本及评价结果见表 4 (表 4 输入的数据经过归一化处理)。

由表 4 评价结果可以得知,GA-NN 综合评价结果与专家评价基本一致。被误判的地区为重庆和青海,其原因笔者认为,在建立的综合评价指标体系中,经济均量水平占权重较大,而重庆在 1997 年建立直辖市后,包括了大量的国家贫困地区在内,使其均量水平大大降低,而青海则是由其人口较少而使得得分上升,从而造成与专家评价不一致。

## 4 结论

GA-NN 网络模型具有全局寻优、自适应的能力,该网络模型中的 GA-BP 算法由于采用了 GA 进化层,克服了传统 BP 算法的缺点,具有较强的适应性。利用该模型来进行区域可持续发展能力综合评价,在理论和实

表 3 GA-NN 网络训练结果

等级	训练结果				
I 级	0.9788	0.0236	0.0000	0.0000	0.0000
II 级	0.0487	0.9325	0.0268	0.0000	0.0000
III 级	0.0000	0.0369	0.9524	0.0214	0.0000
IV 级	0.0000	0.0000	0.0245	0.9632	0.0125
V 级	0.0000	0.0000	0.0000	0.0486	0.9264

# 中国轿车企业规模经济效益实证分析

胡洪力

(浙江工商大学 工商管理学院, 浙江 杭州 310035)

**摘要:**在数据包络分析的  $CR^2$ 、 $C^2GS^2$ 、NIRS 模型基础上, 建立了一个评价中国轿车企业规模经济的数学模型, 并对 1992~2001 年中国轿车企业规模经济效益进行了实证分析, 发现其中的 7 年处于规模递减阶段, 其余的 3 年规模收益不变。这说明我国平均的轿车生产企业规模偏大了。这是由于我国轿车企业的投入产出效率太低, 加之市场需求有限, 致使我国轿车企业 10 年中有 7 年的时间处于规模不经济的状态。

**关键词:**DEA; 轿车企业;  $CR^2$  模型;  $C^2GS^2$  模型; NIRS 模型; 规模经济; 实证分析

**中图分类号:**F426.471

**文献标识码:**A

**文章编号:**1001-7348(2005)04-0077-03

## 0 前言

目前, 对我国轿车企业规模经济问题讨论时, 国内专家一般都认为我国轿车企业没有达到最小有效规模 (Minimum Efficient Scale, 简称 MES), 并存在着严重的规模不经济。例如, 张仁琪、高汉初(2001)的研究则认为我国轿车厂的 MES 为年产 20~30 万辆<sup>[1]</sup>。那么, 到底我国轿车企业的规模经济状况如何呢? 本文采用数据包络分析方法, 通过建立评价中国轿车企业规模经济的数学模型,

对 1992~2001 年我国轿车企业规模经济状况进行了实证分析。

## 1 企业规模经济效益评估的数学模型

数据包络分析(Data Envelopment Analysis 简称 DEA)是由美国运筹学家 Charnes 和 Cooper 等学者于 1978 年发展起来的<sup>[2]</sup>。它以相对效率概念为基础, 用于评价具有相同类型的多投入、多产出的决策单元是否技术有效和规模有效的一种非参数统计方法。基于

DEA 的基本原理, 我们先建立不变报酬模型, 首先测量综合效率  $\theta_0$ 、技术效率和规模效应的共同影响, 然后建立可变报酬条件下测量纯技术效率  $\theta_0$  的模型, 然后再从中分离出规模效率, 并与不变报酬模型下的效率指数  $\theta_0$  相比较, 从而得出决策单元的规模报酬类型。

### 1.1 测量综合效率的不变报酬模型

假设有  $n$  个决策单元 (Decision Making Units, 简称 DMU), 每个决策单元都有  $m$  种类型的输入指标和  $g$  种类型的输出指标。设  $X_{ij}$  是第  $j$  个决策单元对第  $i$  种输入的投入量,

实际应用上都具有重要意义。较高的评分准确率也证明了本文指标选择的合理性。

## 参考文献:

- [1] Maclaren V W. Urban sustainability reporting[J]. Journal of the Americal Planning Association, 1996, 62(2): 185-202.
- [2] Cocklin C R. Methodological problems in sevaluating sustainability[J]. Environment Conservation, 1989, 16(4): 27-32.
- [3] 赵景柱, 梁秀英, 张旭乐. 可持续发展概念的系统分析[J]. 生态学报, 1999, (3): 393-398.
- [4] WCED(World Commision on Environment and

Development, 1987). Our common future[M]. Oxford University Press.

- [5] Shi Xiaoqing. Dynamic analysis and assessment for sustainable development[J]. Journal of Environmental Sciences, 2002, 14(1): 88-94.
- [6] Xhao Jingzhu, Johannes Baptist Opschoor. Indicator system and evaluation framework for sustainable development[J]. Journal of Environmental Sciences, 1999, 11(4): 492-497.
- [7] 国家统计局城市社会经济调查总队. 中国城市统计年鉴 2000[M]. 北京: 中国统计出版社, 2001.
- [8] 吕俊, 张兴华. 几种快速 BP 算法的比较研究[J]. 现代电子技术, 2003, (24): 96-99.

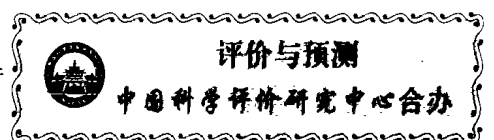
- [9] Yarlalagadda P K D V. Prediction of processing parameters for injection moulding by using a hybrid neural network[J]. Proceedings of the Institution of Mechanical Engineers. Part B. Journal of engineering manufacture, 2001, 215(10): 1465-1470.
- [10] Srinivas M, Patnaik L M. Adaptive probabilities of crossover and mutation in genetic algorithms [J]. IEEE Transactions on System, Man and Cybernetics, 1994, 24(4): 656-667.
- [11] 飞思科技产品研发中心. MATLAB6.5 辅助神经网络分析与设计[M]. 北京: 电子工业出版社, 2003.

(责任编辑: 慧 超)

收稿日期: 2004-07-06

基金项目: 国家自然科学基金资助项目 (70273009)

作者简介: 胡洪力 (1973-), 男, 河南永城人, 浙江工商大学工商管理学院讲师, 经济学博士, 主要从事汽车产业发展、企业成长方面的研究。



评价与预测

中国科学评价研究中心合办