

ES-SOFM 混合模型及其在水环境评价中的应用

许世刚¹, 高新陵²

(1. 淮阴师范学院经济与法律系, 江苏 淮安 223001 2. 河海大学水利水电工程学院, 江苏 南京 210098)

摘要 提出一种基于进化策略(Evolutionary Strategy, ES)和自组织特征映射(Self-Organizing Feature Map, SOFM)神经网络的混合模型进行水环境评价的新方法. 把 SOFM 网络嵌入到 ES 中, 根据 SOFM 网络的运行状态构造 ES 的适应性函数, 利用 ES 的强搜索能力克服 SOFM 网络对初始权向量和样本输入次序敏感的弱点. 在水环境评价中的应用效果表明, 此方法比模糊聚类法分类更合理.

关键词 自组织特征映射 进化策略 适应性 聚类 水环境评价

中图分类号 X824 **文献标识码** A **文章编号** 1000-198X(2002)05-0053-03

对水质进行科学的评价是保护和改善水环境的前提. 国内外的环境工作者研究了很多的评价方法, 如单指数法、模糊聚类法、灰色聚类法等, 其中模糊聚类法已成为环境综合评价中的重要方法^[1]. 在水环境中各种污染物与环境之间具有复杂多变的联系, 这些方法都很难刻画分类指标空间到类型空间的非线性映射关系^[2]. 自组织特征映射(Self-Organizing Feature Map, 简称 SOFM)神经网络在无师情况下, 通过对输入模式的反复学习, 捕捉各个模式中所含的特征, 并对其进行自组织, 在竞争层将聚类结果反映出来. SOFM 网络模拟了大脑神经细胞对外界信息刺激的反应, 克服了所谓“祖母细胞”现象, 同时具有较强的抗干扰能力. 它已成功地应用于径流预报^[3]、地下水动态分类^[2]等水利水电工程中.

但 SOFM 网络应用中存在着以下弱点: 网络连接权的初始状态对网络的收敛性能有较大影响; 学习模式较少时, 网络的分类效果取决于输入模式的先后次序^[4]. 这些因素制约了 SOFM 网络的更广泛应用. 目前已有一些学者提出的改进方法主要致力于使权向量的初始状态与输入模式易接近. 如: 将连接权向量赋予相同的初值, 增加每一个权向量被选中的机会, 给原学习模式的每个元素加上一个很小的随机值, 随着学习的进行, 逐渐滤掉随机值再进行细调整, 给每一个竞争层神经元增设输出阈值, 目的在于增加所有神经元被选中的机会, 提高连接权向量的利用率. 但这些方法仍难以保证网络收敛至最佳状态, 即难以取得最佳的聚类效果.

进化策略(Evolutionary Strategy, 简称 ES)是模拟生物进化过程的一种随机优化方法, 它能以较大的概率得到全局最优解. 本文研究利用进化策略的强探索能力指导 SOFM 网络的运行, 改善网络的聚类效果, 并将此混合模型应用于丰满电站下游江段的水环境评价.

1 ES-SOFM 混合模型机理

SOFM 网络由输入层和竞争层组成. 输出层所有神经元通过相互竞争和自适应学习, 形成空间上的有序结构, 从而实现输入矢量到输出矢量空间的特征映射^[5]. 网络学习后, 竞争层各神经元的连接权向量的空间分布反映输入模式的空间拓扑结构和概率分布情况, 将需分类的模式提供给网络, 获胜神经元表示模式的分类结果.

进化策略是模拟物种进化过程中的选择、变异等作用, 按“物竞天择, 适者生存”的原则竞争演化来求解优化问题, 它将优化问题的解空间视为若干种群, 让物种模拟进化过程不断进化, 直至满足终止准则^[6].

算法结合的最佳方式是融合. 为了充分发挥 ES 的全局优化能力, 克服 SOFM 网络自身的局限, 实施如下步骤:

步 1 初始化. 随机选取 $w_i \in R^m$, $i = 1, 2, \dots, N$, 构成初始种群, N 为种群规模.

步 2 父代变异. 由于 Schewefel 变异法能显著改善算法的计算效率与优化性能^[7], 本文选用 Schewefel 变异法

$$w_i' = w_i + N(0, \sigma_i')$$

收稿日期 2001-08-15
作者简介: 许世刚(1969—), 男, 江苏灌南人, 博士, 讲师, 主要从事系统优化规划及动能经济研究.

其中 $\sigma_i' = \sigma_i \exp(\xi)$, $\xi = \tau' N(0, 1) + \tau N(0, 1)$ (ξ 服从正态分布 $N(0, \sqrt{\tau'^2 + \tau^2})$), $\tau' = 1/\sqrt{2n}$, $\tau = 1/\sqrt{2\sqrt{n}}$ (n 为进化代数).

步 3 加入 SOFM 神经网络训练. 注意在输入模式对时, 按随机顺序的原则, 消除次序对分类效果的影响.
步 4 计算适应性函数值 $\mathcal{S}(w)$. 在聚类分析中, 类与类之间的边界应清楚, 个体与类之间的隶属关系应明确. 为了反映 SOFM 网络的分类状况, 构造如下适应性函数:

$$\mathcal{S}(w) = \alpha \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m \|w_i - w_j\|_2 - \beta \frac{1}{m} \sum_{j=1}^m \left(\frac{1}{p_j} \sum_{i=1}^{p_j} r_{ji} \|w_j - A_i\|_2 \right)$$

式中: m ——总分类数; p ——样本模式总数; p_j ——属于第 j 类的输入模式数目; w_j ——第 j 类类别中心的特征向量; A_i ——第 i 样本的特征向量.

$\mathcal{S}(w)$ 由两项组成, 第 1 项是为了体现聚类中心之间的差异, $\alpha \in [0, 1]$, α 一般取较小值; 第 2 项强调同类个体与类别中心的最小距离, 它是分类合理性的重要指标. $\beta \in (0, 1]$, 一般取 1 或较大值.

$$r_{ji} = \begin{cases} 1 & \text{第 } i \text{ 样本属于 } j \text{ 类} \\ 0 & \text{第 } i \text{ 样本不属于 } j \text{ 类} \end{cases}$$

步 5 从 $2N$ 个个体中选择 $\mathcal{S}(w)$ 较大的 N 个个体进入下一代.
步 6 收敛性判断. 若 $|(S_{\max} - S_{\min})/S_{\text{avg}}| < \epsilon$
循环终止, 取 $S_{\max}$ 所对应的经过训练的网络对输入模式仿真, 输出结果. 否则转步 2.

其中 $S_{\text{avg}} = \frac{1}{N} \sum_{i=1}^N \mathcal{S}(w_i)$ $S_{\max} = \max_i \{\mathcal{S}(w_i)\}$ $S_{\min} = \min_i \{\mathcal{S}(w_i)\}$

2 ES-SOFM 混合模型在水环境评价中的应用

将此混合模型应用于第二松花江丰满电站下游江段的水质评价. 沿江 8 个断面 19 项污染参数的相对值 (即污染参数实测浓度/污染参数标准浓度) 见表 1^[8].

模糊聚类最大矩阵元法的分类结果见表 2^[8].

表 1 第二松花江 8 个断面 19 项污染参数的相对值
Table 1 Relative values of 19 pollution parameters for 8 sections along the second Songhuajiang River

断面	丰满 1	三道码头 2	哈达湾 3	九 站 4	哨 口 5	红 旗 6	松花江村 7	扶 余 8
pH	0.4124	0.4250	1.1656	0.9703	0.5776	0.3842	0.3755	0.4771
悬浮物	0.1614	0.3242	0.8465	0.5635	0.7930	0.8740	0.3787	0.4567
五日氧	0.7130	0.9625	5.8950	1.7525	1.4250	0.9350	0.4957	0.9143
耗氧量	0.3750	0.7150	2.4875	1.5000	1.6825	1.1233	0.9000	0.7200
溶解氧	0.3880	0.4275	0.9125	0.9455	0.5685	2.7199	0.9963	1.0552
氨 氮	0.1140	0.1937	1.0310	1.2033	1.0788	0.9520	0.2577	0.0290
亚硝酸盐	0.0243	0.0356	0.2854	0.1411	0.2275	0.1836	0.1605	0.0629
硝酸盐	0.6340	0.7987	0.7458	0.6033	0.7845	0.8520	1.9146	0.2319
铜	0.0825	0.0609	0.2001	0.1646	0.1742	0.1195	0.1286	0.1228
锌	0.0370	0.0280	0.0900	0.0585	0.0805	0.0513	0.0263	0.0613
总 汞	0.1920	0.1908	1.1148	0.7588	0.7563	0.6880	0.3780	0.6880
总 铬	0.1850	0.1400	0.1845	0.2290	0.2350	0.2333	0.0100	0.1240
总 铁	1.1916	1.2922	1.7399	1.6326	1.5627	1.2689	1.0833	0.9803
锰	0.0627	0.2097	0.2011	0.1065	0.1031	0.1427	0.0974	0.0523
酚	0.0500	0.3155	2.4788	1.5475	1.2650	1.4230	0.4150	0.5133
氰化物	0.0250	0.0350	0.1455	0.5506	0.2421	0.1234	0.0417	0.0828
硝基苯	0.0100	0.0100	8.0288	0.7379	0.4612	0.6383	0.0100	0.0100
氯 苯	0.0250	0.0250	11.662	1.7427	0.4033	0.0250	0.0250	0.0250
苯 胺	0.0050	0.0050	0.1367	0.0348	0.0416	0.0050	0.0050	0.0050

表 2 模糊聚类最大矩阵元法断面污染分类结果
Table 2 Classification of pollution in each section by the maximum matrix element method of fuzzy cluster

断面代号	3	4	5	6	8	2	7	1
截集水平 λ	1.0000	1.0000	0.6024	0.4917	0.2986	0.2511	0.2360	0.1381
评 语	严重污染	严重污染	中度污染	中度污染	尚清洁	尚清洁	尚清洁	尚清洁

采用 ES-SOFM 混合模型方法 ,参数选取不同的组合 :初始种群 $N = 10\ 20\ 30$,SOFM 训练最大次数 $T = 100\ , 500\ 800$; $\lambda(0) = 0.1\ 0.2\ 1$ ES 循环次数为 $10\ 30\ 50$,都取得了一致的分类结果(表 3) ,说明此方法是稳健的.

对照表 2 ,表 3 ,发现主要区别在于断面 4(即九站)的归属问题.在模糊聚类法中 ,哈达湾和九站的聚类水平 λ 均为 1 ,从而同属严重污染.但在表 1 中两者污染参数的相对值差异很大 ,这种评价是不合理的.产生此现象的根本原因在于建立模糊相似关系时 ,矩阵元的计算采用了“积和”的方式 ,因而易使被研究对象中相差较大的评价单元 ,因截集水平相等而误聚为一类.在大容量样本情况下 ,误判现象会更严重 ,难以保证很强的科学性.而 ES-SOFM 混合模型能够克服这个缺点 ,实现从指标空间到类型空间的非线性映射关系.把哈达湾归为严重污染 ,九站归为中度污染 ,与实际情况相吻合 ,得到最优的评价结论.

3 结 论

- a. ES-SOFM 混合模型克服了 SOFM 网络对样本输入次序和初始权向量敏感的弱点 ,在大概率上保证网络收敛到全局最优解 ,分类更加合理、可靠.ES 的强全局搜索能力 ,对其他聚类分析方法性能的改善也具有指导意义.
- b. ES-SOFM 混合模型在第二松花江的水质评价中取得了理想的效果 ,而它本身是一种通用的聚类方法 ,在其他的水环境评价和水利水电工程中可以进一步推广使用.

参考文献 :

[1] 张美华 ,陈宏.几种模糊聚类法在环境质量综合评价中的应用[J].重庆环境科学 ,1999 ,21(3) :13 ~ 16.
[2] 张翔 ,刘国东 ,丁晶.自组织神经网络在地下水动态分类中的应用[J].工程勘察 ,1998(2) :29 ~ 31.
[3] 胡铁松 ,丁晶.径流长期分级预报的 Kohonen 网络方法[J].水电站设计 ,1997 ,13(2) :24 ~ 29.
[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 ~ 168.
[5] 杨行峻 ,郑君里.人工神经网络[M].北京 :高等教育出版社 ,1992. 226 ~ 277.
[6] Fogel D B. An introduction to simulated evolution optimization[J]. IEEE Trans on Neural Networks. 1994 ,5(1) 3 ~ 14.
[7] 徐志洪 ,卢建刚 ,顾钟文.解全局优化问题的进化策略及其仿真[J].机电工程 ,2000 ,17(2) :87 ~ 90.
[8] 陈守煜.水利水文水资源与环境模糊集分析[M].大连 :大连工学院出版社 ,1987. 98 ~ 106.

ES-SOFM hybrid model and its application to water quality assessment

XU Shi-gang¹ , GAO Xin-ling²

(1. Department of Economics and Law , Huaiyin Teachers College , Huaiian 223001 , China ;

2. College of Water Conservancy and Hydropower Engineering , Hohai Univ. , Nanjing 210098 , China)

Abstract : A new method for water quality assessment is put forward based on Evolutionary Strategy(ES) and Self-Organizing Feature Map (SOFM) neural networks. With SOFM neural networks embedded into ES , the adaptability function of ES is constructed based on the state of SOFM neural networks. The sensitivity of SOFM neural networks to the initial weight matrix and sequence of exemplar input is overcome by the global optimization of ES. The application of the model to water quality assessment shows that the method is superior to the fuzzy cluster method.

Key words : Self-Organizing Feature Map (SOFM); Evolutionary Strategy(ES); adaptability ; cluster ; water environment assessment