

基于细胞状演化神经网络的黄河三角洲区域预测

颜颖¹, 曾文华^{2*}, 姜青山²

(1. 厦门大学计算机科学系, 2. 厦门大学软件学院, 福建 厦门 361005)

摘要: 将细胞状遗传算法 (cGA) 与反向传播 (BP) 神经网络相结合, 构建了一个细胞状演化神经网络时间序列模型, 以 1984~2000 年间的多时相遥感影像为主要数据源, 应用该模型对黄河口地区陆地面积进行预测, 并分析了其将来的演变趋势。结果表明, 在 2001~2010 年间研究区域陆地面积将呈现增长与蚀退交替演变、增长高峰逐渐下降的趋势。

关键词: 演化神经网络; 细胞状遗传算法; BP 神经网络; 时间序列模型; 遥感图像

中图分类号: TP 183

文献标志码: A

文章编号: 0438-0479(2006)S-0296-05

黄河三角洲位于鲁东北, 经济发达、人口稠密, 并于上世纪六七十年代发现丰厚的油气资源, 成为我国第二大石油生产基地。不幸的是全球变暖、海平面上升及人类活动等对其生态环境造成了一系列的严重后果, 其中最直接、最敏感的反映就是海岸侵蚀和三角洲面积减少。自上世纪 70 年代以来, 黄河入海水沙量大幅度减少, 甚至出现严重的断流现象, 导致尾间河道和河口严重淤积, 给河口的防洪带来严重威胁。黄河三角洲延伸造陆速度明显减慢, 海岸带上出现严重蚀退, 增加了油田变浅海开采为陆上开采的难度^[1]。若今后入海水沙量进一步减少, 必将给黄河三角洲的生态环境及工农业生产带来严峻挑战, 因此研究和预测三角洲的演变情况具有非常重要的现实意义。

传统预测一般采用分析要素的方法进行研究。首先列出影响预测的各种可能因素, 然后对这些因素逐一进行分析, 最后由模型模拟或经验推断得出预测的结论。然而, 如果将这种方法应用到黄河三角洲的演化预测中, 就存在较大缺陷。主要原因是影响预测的因素过于复杂, 无法完整地罗列, 即使将因素完全罗列出来, 也很难确定各因素对预测结果影响的强弱程度, 从而难以准确地抽取主导因素进行处理。因此, 本文构建了一个细胞状演化神经网络时间序列模型, 以 1984~2000 年间的多时相遥感影像为主要数据源, 选取黄河口地区作为研究区域, 对该区域 2001~2010 年的陆地面积进行预测, 并分析了其将来的演变趋势。

1 研究区域概况

黄河口海岸是指北起五号桩 (东营港附近), 南至

宋春荣沟之间的一段岸线^[2]。由于东营港至孤东油田之间岸线、宋春荣沟附近及南侧岸线基本稳定^[3], 因此本文选择孤东油田至宋春荣沟之间的一段岸线作为研究区域, 如图 1 所示。

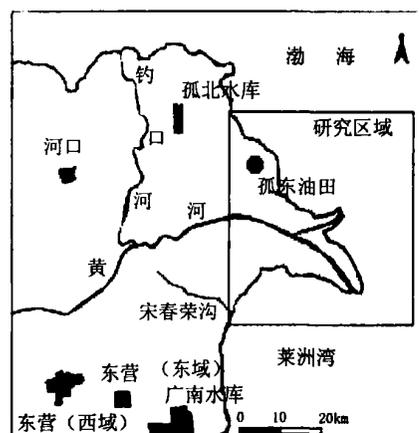


图 1 研究区域位置

Fig 1 Location of the study region

根据历年遥感影像综合分析和黄河口地区海岸线形状, 黄河口海岸线的动态变化可以分为 4 个阶段^[4]。

(1) 1976~1987 年, 黄河口呈独特的“棉絮”状逐年向外伸展。此阶段, 河口内海岸线整体发生变化, 淤进蚀退交错进行;

(2) 1987~1992 年, 黄河口逐渐近似于“楔形”向海推进, 主河道位置稳定在东南方向, 沙嘴稳定延伸, 但顶部尚未完全盈满, 黄河口岸线不断向东南方推进;

(3) 1992~1996 年, 黄河口逐渐形成了一个饱满的“楔子”向海洋突伸, 发育进入晚期, 造陆速率减缓, 黄河来水来沙量减少, 海岸线动态变幅不大;

(4) 1996~2000 年, 人工引黄河从北汊入海, 改变了泥沙淤积条件, 使北汊造陆速率加快, 迅速向海淤出

收稿日期: 2005-11-10

基金项目: 福建省青年科技人才基金项目 (2002J005) 资助

作者简介: 颜颖 (1980-), 女, 硕士研究生。

* 通讯作者: whzen@xmu.edu.cn

一个小沙嘴,而原大沙嘴顶端则开始蚀退,此时段大沙嘴顶端蚀退了约 2.1 km,出现了造陆速度减缓,岸线逐渐蚀退的现象.由于研究区域陆地面积变化迅速,尾间迁徙频繁,应用传统的调查手段已难以对其进行及时、准确的监测与研究,因此本文采用 1984~2000 年间的 Landsat 多时相 TM、ETM+ 系列数据作为主要研究数据,构建了一个细胞状演化神经网络时间序列模型对该区域的面积变化情况进行预测.

2 细胞状演化神经网络时间序列模型

2.1 模型的引入

时间序列预测法^[5]是通过找出符合系统变化规律的函数 F ,以过去的一个观测值序列 $x^{(0)}, x^{(1)}, \dots, x^{(t)}$ 作为输入,预测出未来的值 $x^{(t+1)}$,且 $x^{(t+1)} = F(x^{(0)}, x^{(1)}, \dots, x^{(t)})$.但是由于黄河三角洲海陆变迁是河流径流、泥沙、海洋动力作用等众多因素综合作用的结果,且泥沙来源丰富、数量巨大、组成较细,三角洲岸线演变迅速频繁,整个系统的发展变化高度非线性,无法直接找到准确描述其规律的函数.而神经网络具有高度的自学习能力,能够对非线性函数以任意精度逼近,非常适合模拟三角洲区域形态演化系统的发展变化规律.同时,为避免神经网络易陷入局部最优的缺陷,本文构建的模型还引入了细胞状遗传算法.该算法的种群具有特定的空间结构,能够通过更多的迭代来获取适应性和基因型多样性.研究证实细胞状遗传算法在处理复杂优化任务时优于其它演化模型,如训练神经网络^[6-8].因此本文将细胞状遗传算法和神经网络相结合,构建了一个细胞状演化神经网络时间序列模型,并将它应用于黄河三角洲的区域预测中.

2.2 模型的构建

细胞状演化神经网络时间序列模型如图 2 所示.该模型包含 2 个子模块,即“细胞状遗传算法”和“BP 神经网络模型”.其中“细胞状遗传算法”用于优化 BP 神经网络的初始权值/阈值,并定位出一个较好的搜索空间.算法中适应度函数由神经网络对一组训练数据的一次训练误差计算得出,而“BP 神经网络模型”在获得细胞状遗传算法给出的权值/阈值初值后,即输入训练数据,以训练神经网络,使其误差达到所要求的精度要求;最后,“BP 神经网络模型”在获得最终权值/阈值后,输入观测值序列,通过模型模拟,得出预测值.

(1) 细胞状遗传算法描述

在细胞状遗传算法中,个体分布具有特定的拓扑结构,并只与各自的邻居发生相互作用.最常采用的种群拓扑结构是二维超环面栅格,即所有个体都分布在这些栅格中,并各自占据一个栅格.而在其上定义的邻

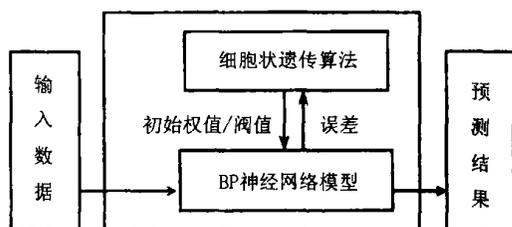


图 2 细胞状演化神经网络时间序列模型

Fig 2 Cellular evolutionary neural networks time series model

域一般包括 5 个个体,即它自身,及其东、西、南、北 4 个方向上与之相邻的个体.这种邻居关系模型称为 NEWS 模型或 LINEAR^[9],如图 3 所示.

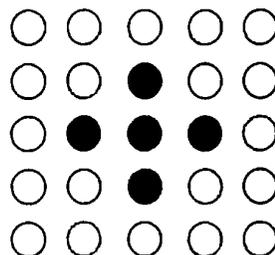


图 3 NEWS 模型

Fig 3 NEWS model

设 x, y 为种群区域的横坐标和纵坐标, $Width$ 、 $Height$ 分别表示该种群区域的宽度和高度,则细胞状遗传算法用伪代码描述如下^[10]:

```

参数初始化
生成随机初始化种群
Do While 未达到最大演化代数
  For  $x = 1$  to  $Width$ 
    For  $y = 1$  to  $Height$ 
      计算个体  $i$  (位于坐标  $(x, y)$ ) 的邻域
      从个体  $i$  的邻域中选择一个个体  $j$ 
      个体  $i$  和个体  $j$  进行交叉, 生成子代
      对子代进行变异
      计算适应度
      将新个体插入种群
    End For
  End For
  更新种群
Loop

```

(2) BP 神经网络模型

在时间序列的预测中,前馈网络是最常使用的神经网络,因此在本文构建的模型中选用了最为传统的 BP 神经网络来作为时间序列的预测模型.本文采用的 BP 网络使用三层结构,包括输入层、隐含层和输出层,

如图 4所示. 输入层由 3 个神经元构成, 隐含层由 15 个神经元构成, 输出层由 1 个神经元构成. 第一层的传输函数是 bg-sigmoid 函数 $f^1(x) = 1/(1 + e^{-x})$, 第二层的传输函数是线性函数 $f^2(x) = x$. BP 神经网络的训练过程描述如下:

- (a) 参数初始化;
- (b) 导入训练集, 计算 BP 神经网络的实际输出;
- (c) 计算目标值与实际输出的均方误差;
- (d) 根据均方误差更新权值阈值;
- (e) 判断误差精度是否满足终止条件. 如果满足, 则算法停止; 否则重复执行 (b) ~ (e).

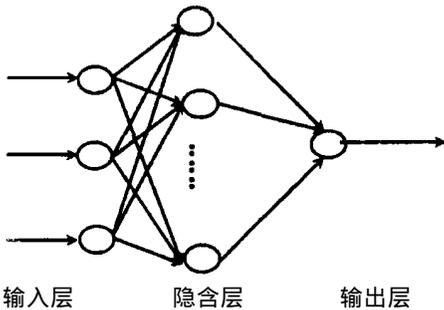


图 4 BP 神经网络

Fig 4 BP neural network

2 3 模型详细设计

(1) 编码设计

神经网络的学习是一个复杂的连续参数优化问题, 这里采用实数编码^[11]. 权值 阈值的取值范围设定为 (-10, 10), 精确到 0.001, 因此每个权值 阈值的编码长度都是 5 位. 第一位为 0 或 1, 表示权值 阈值的正负, 第二位为整数部分, 第三位到第五位为小数部分. 按照“输入隐含节点间连接权 → 隐含节点阈值 → 隐含输出节点间连接权 → 输出节点阈值”的顺序把各权值 阈值连成一个编码串, 串上的每一个部分对应着一个网络权值 阈值.

(2) 适应函数设计

在演化算法中适应函数是用来判断群体中个体优劣、对个体进行选择的标准. 本模型采用的适应函数是 $fitness(i) = \frac{1}{1 + E(net_i)}$, 其中 $E(net_i)$ 是神经网络一次训练的误差, 即对每个个体进行解码, 得到一个 BP 神经网络, 对网络输入一组训练样本进行一个周期的训练, 从而计算出网络的输出误差.

(3) 遗传算子设计

(a) 选择算子设计. 本模型采用标准轮转法选择策略, 即基于相对适应值的选择. 设随机数 $0 \leq r \leq 1$

$$p_i = fitness(i) / \sum_{j=1}^N fitness(j), p_0 = 0 \text{ 若 } p_0 + p_1 + \dots + p_{i-1} < r \leq p_1 + p_2 + \dots + p_i, \text{ 则选择个体 } i$$

(b) 交叉算子设计. 由于采取了实数编码, 每个基因值表示神经网络的一个权值 阈值, 因此交叉操作可以直接以权值 阈值为单位, 随机交换两父体的若干个权值 阈值, 以生成交叉后的两个后代.

(c) 变异算子设计. 首先按轮转法选择父体, 然后对权值 阈值进行自适应变异. 自适应变异就是将变异算子和个体适应值联系起来, 使适应值大的个体在较小范围内搜索, 而适应值较小的个体在较大范围内搜索. 该算子引入了“解变异温度”, 这类类似于模拟退火算法中温度 T 的概念. 设对权值 阈值 v 进行变异, $v \in (a, b)$; v' 为变异后的权值 阈值; r 是 $[0, 1]$ 上的随机数; λ 是决定自适应程度的参数, 用于调整局部搜索区域, 一般取值 $2 \sim 5$; r_2 表示一个随机均匀产生的正整数模 2 所得的结果. 自适应变异描述如下:

$$T = 1 - fitness(i) \tag{1}$$

$$v' = \begin{cases} v + \Delta(T, b - v), & r_2 = 0 \\ v - \Delta(T, v - a), & r_2 = 1 \end{cases} \tag{2}$$

$$\Delta(T, y) = y(1 - r^\lambda), \Delta(T, y) \in (0, y) \tag{3}$$

式 (1) ~ (3) 表明, 当 T 减小时, $\Delta(T, y)$ 接近 0 的概率增加. 这样定义的变异算子将保护较好的解, 使其在较小区域内进行搜索; 而适应度不好的解, 则需搜索较大区域. 这使得变异过程能够根据解的质量自适应地调整搜索区域, 从而明显地提高搜索效率.

3 模型在研究区域的应用研究

为了应用该模型对研究区域的陆地面积进行预测, 我们开发了一个黄河三角洲区域面积预测系统对其进行仿真. 该系统构建的平台是 Windows 2000 Professional 编程环境是 Visual Studio Net 编程语言是 C#, 采用的数据库是 SQL Server 2000 整个预测过程分为两大部分:

(1) 对遥感图像进行处理, 提取海岸线, 计算研究区域的陆地面积;

(2) 应用上述细胞状演化神经网络时间序列模型对研究区域在 2001 ~ 2010 年间的陆地面积进行预测.

3 1 研究区域陆地面积计算

计算研究区域陆地面积^[12]主要有以下 3 个步骤:

(1) 遥感图像预处理. 这包括遥感影像的几何精校正、影像配准和裁取研究区域三个方面的工作. 处理后的研究区域遥感图像如图 5 所示.

(2) 研究区域海岸线的提取. 首先用模糊 Kohonen 聚类网络拓扑结构自探索算法^[13]对裁取的子区影像

进行预处理, 然后以此为基础, 选取训练样区并运用监督分类方法进行分类, 最后提取出平均高潮线作为海岸线, 如图 6 所示。

(3) 研究区域陆地面积计算. 对提取了海岸线的图像进行颜色填充, 陆地填充为黑色, 背景填充为白色, 计算颜色为黑色的像素点个数, 从而得到以像素为单位的陆地面积, 如表 1 所示。

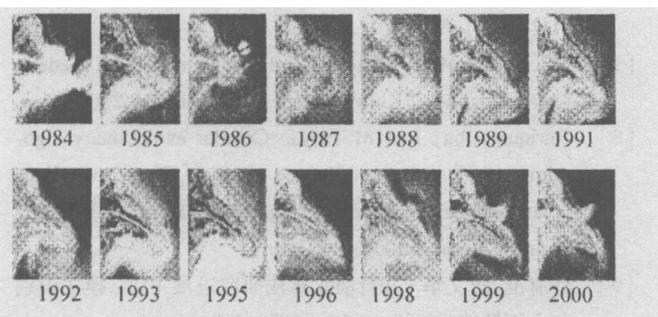


图 5 研究区域遥感图像

Fig 5 RS images in the study region

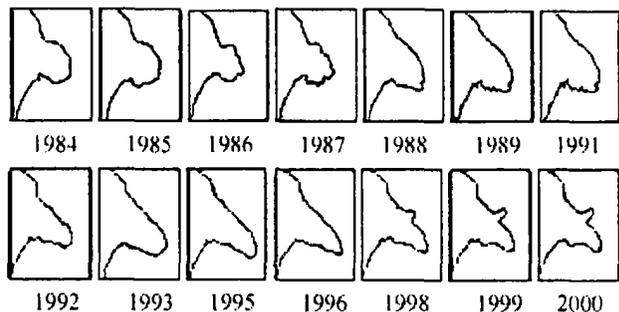


图 6 研究区域海岸线提取图像

Fig 6 Coastline extracted from the RS images in Fig 5

3.2 研究区域陆地面积预测

为预测研究区域 2001~2010 年的陆地面积, 模型参数设定如表 2 所示. 首先将预测训练数据 (表 1) 导入模型, 输入模式设定为 $(x^{(t-1)}, x^{(t-2)}, x^{(t-3)})$, 则输出模式为 $x^{(t)} = f(x^{(t-1)}, x^{(t-2)}, x^{(t-3)})$; 然后使用细胞状遗传算法对 BP 神经网络进行训练和优化; 最后通过模型中已训练好的 BP 神经网络, 模拟研究区域陆地面积从 2001 年到 2010 年的演化过程, 从而得出预测结果, 如表 3 图 7 所示。

3.3 预测结果分析

预测结果表明, 2001~2010 年 10 年间, 研究区域的陆地面积将呈现增长与蚀退交替演变的趋势, 以 3~4 年为一个周期, 每个周期都出现一个增长高峰, 随着时间的推移, 增长高峰呈现逐渐下降的趋势, 陆地面积变化趋向缓和, 如图 7 所示。

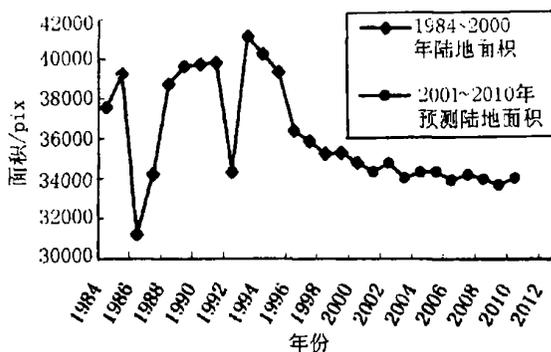


图 7 研究区域陆地面积预测与分析结果

Fig 7 Analysis and prediction of the land area in the study region

表 1 研究区域 1984~2000 年的陆地面积

Tab 1 The area of our study region from 1984 to 2000

日期	面积 /pix
1984 / 10 / 05	37548
1985 / 03 / 04	39232
1986 / 06 / 05	31216
1987 / 05 / 07	34219
1988 / 12 / 03	38753
1989 / 02 / 13	39646
1991 / 01 / 26	39854
1992 / 04 / 02	34355
1993 / 04 / 02	41164
1995 / 03 / 10	39373
1996 / 05 / 31	36413
1998 / 05 / 05	35282
1999 / 06 / 25	35330
2000 / 05 / 02	34837

表 2 模型参数

Tab 2 Parameters of the model

参数名称	参数值
种群规模	100
种群区域宽度	10
种群区域高度	10
最大演化代数	100
交叉率	0.800
变异率	0.010
适应性变异参数	2.000
学习率	0.060
最大迭代次数	20000
最小误差	0.010

表 3 2001~2010年的预测陆地面积

Tab 3 Predicted land area from 2001 to 2010

年份	面积 /pik
2001	34390
2002	34809
2003	34100
2004	34394
2005	34418
2006	33988
2007	34266
2008	34059
2009	33781
2010	34125

4 结 论

本文分析了黄河三角洲黄河口海岸历年的演变情况,以 1984~2000年的遥感图像作为数据源,构建了演化神经网络时间序列模型,并开发了一个黄河三角洲区域面积预测系统对研究区域 2001~2010年间的陆地面积变化情况进行预测,发现该区域陆地面积按照上述模型的预测,将呈现增长与蚀退交替演变、增长高峰逐渐下降的趋势,这将引起我们的关注。

参考文献:

[1] 刘曙光,李从先,丁坚,等. 黄河三角洲整体充淤平衡及其地质意义 [J]. 海洋地质与第四纪地质, 2001, 21(4): 14-18

[2] 杨虎,郭华东,王长林. TM-SAR 数据融合在黄河口沙嘴动态监测中的应用研究 [J]. 地理学与国土研究, 2001,

17(4): 15-19.

[3] 常军,刘高焕,刘庆生. 黄河口岸线演变时空特征及其与黄河来水来沙关系 [J]. 地理研究, 2004, 23(1): 339-346

[4] 常军,刘高焕,刘庆生. 黄河三角洲海岸线遥感动态监测 [J]. 地球信息科学, 2004, 6(1): 94-98

[5] 钟颖,汪秉文. 基于遗传算法的 BP神经网络时间序列预测模型 [J]. 系统工程与电子技术, 2002, 24(4): 9-11.

[6] 许有准. 并行演化算法及其在神经网络中的应用研究 [D]. 厦门:厦门大学, 2005.

[7] Erick Cantú-Paz Efficient and Accurate Parallel Genetic Algorithms [M]. Kluwer Academic Publishers, 2000

[8] Enrique Alba, José M Troya Cellular evolutionary algorithms: evaluating the influence of ratio [C]// Proceedings of the Parallel Problem Solving from Nature VI Berlin Springer-Verlag, 2000, 29-38

[9] 许有准,曾文华. 并行演化算法研究进展 [J]. 模式识别与人工智能, 2005, 18(2): 183-193

[10] Enrique Alba, Bernabé Dorronsoro The exploration/exploitation tradeoff in dynamic cellular genetic algorithm, [J]. IEEE Transactions on Evolutionary Computation, 2005, 9(2): 126-142

[11] 于漫,朱岩. 集中式粗粒度并行模型和并行进化神经网络 [J]. 系统工程理论与实践, 2003, 23(6): 75-79.

[12] Yan Ying, Xu Youzhun, Zeng Wenhua, et al. An evolutionary neural networks time series model for predicting the yellow river estuary area [C]// Environmental Informatics Archives Xiamen University Press, 2005: 98-103

[13] 颜颖,曾文华,姜青山. 模糊 Kohonen聚类网络拓扑结构自探测算法 [C]// 模糊逻辑与计算智能研究进展上册. 深圳:中国科技大学出版社, 2005: 202-207.

Prediction of the Yellow River Estuary Area Based on Cellular Evolutionary Neural Networks

YAN Ying¹, ZENG W enhua^{2*}, JIANG Q ing-shan²

(1 Department of Computer Science Xiamen University, 2 School of Software Xiamen University, Xiamen 361005, China)

Abstract An cellular evolutionary neural networks time series model was constructed to predict the change in the coastline of the Yellow River Estuary area. Firstly, the location and evolution of our research region was described. Secondly, an evolutionary neural networks time series model was constructed. Finally, the model was applied to predict the evolution of our research region using Remote Sensing (RS) images between 1984 and 2000 as data source. In the end of this paper, the results were analyzed. It indicates that the increase and decrease in the area of our research region appears alternately from 2001 to 2010 and the change tends to slow down.

Key words evolutionary neural networks; cellular genetic algorithm; BP neural networks; time series model; RS images