

文章编号: 1001-0920(2002)05-0599-05

基于快速、高精度遗传算法神经网络的 薄互储层参数预测

田景文¹, 高美娟²

(1. 大庆石油学院 测控技术与仪器教研室, 黑龙江 安达 151400; 2. 大庆石油学院 勘探系, 黑龙江 安达 151400)

摘要: 针对传统遗传算法(GA)和人工神经网络BP算法各自存在的不足, 引入自适应机制的浮点编码的遗传算法, 并将其与BP网中的梯度下降法相结合, 进行混合交互运算, 形成GA-BP混合算法。该算法使网络具有较快的收敛速度和较高的逼近精度, 能较好地解决综合多种地震信息进行薄互储层参数预测的精度和收敛速度问题, 并通过实例验证了此方法的正确性和实用性。

关键词: 遗传算法; 神经网络; 薄互储层; 地震特征参数

中图分类号: TP 13

文献标识码: A

Thin interbedded reservoir parameters predicting based on high speed and precise genetic algorithm neural network

TIAN Jingwen¹, GAO Meijuan²

(1. Teaching Research Group of Control and Measurement, Daqing Petroleum Institute, Anda 151400, China; 2. Exploration Department, Daqing Petroleum Institute, Anda 151400, China)

Abstract: A genetic algorithm with adaptive and floating-point code is proposed to overcome disadvantages of the genetic algorithm and BP algorithm. This algorithm is combined with BP to give GA-BP mixed algorithm which has higher accuracy and faster convergence speed. The new algorithm also provides improved predict accuracy of thin interbedded reservoir parameters. An example shows the validity and feasibility of this algorithm.

Key words: genetic algorithms; neural networks; thin interbedded reservoir; seismic characteristic parameter

1 引言

基于梯度下降法的BP网络是近年发展起来的进行储层参数预测的新方法, 其逼近非线性函数的能力已在理论和应用方面得到了广泛的研究^[1]。但是, BP算法存在两个突出弱点, 即收敛速度慢和可能收敛到局部极小点。

遗传算法也是近年迅速发展起来的一种全新的优化算法^[2], 其优点是: 1) 并行搜索, 搜索效率高; 2) 不需要目标函数的微分值, 放宽了对目标函数的要求; 3) 本质上属于随机寻优过程, 不存在局部收敛问题。但是, 遗传算法作为一种随机优化算法, 也存在一些不足, 往往只能在短时间内寻找到接近全局最优解的近优解。

收稿日期: 2001-08-02; 修回日期: 2001-10-29

作者简介: 田景文(1962—), 男, 黑龙江克山人, 副教授, 博士, 从事控制理论、人工神经网络模式识别技术的研究; 高美娟(1965—), 女, 北京人, 副教授, 博士生, 从事地震资料处理解释、人工神经网络模式识别技术的研究。

针对传统遗传算法存在的不足,引入自适应机制的浮点数编码的遗传算法,并将其与BP网络混合交互运算,形成快速、高精度的遗传算法神经网络。它克服了BP算法收敛速度慢和可能收敛到局部极小点的缺陷,同时解决了单独利用GA往往只能在短时间内寻找到接近全局最优解的近优解这一问题。

随着油田勘探程度的加深和勘探水平的不断提高,易识别的构造油气藏越来越少,寻找隐蔽的岩性油气藏圈闭,对薄互岩性油气藏进行描述是目前石油勘探的主要目标,而薄互储层参数预测则是薄互岩性油气藏描述的基础。目前,综合多种地震信息进行薄互储层参数预测是一个难题,薄互储层沉积具有储层薄且横向变化剧烈的特点,且地震数据量大,采用传统的BP网络进行的参数预测已达不到所需的精度和速度要求。为此,本文提出改进网络算法,以解决薄互储层参数预测的精度和收敛速度问题。

2 基本原理

2.1 误差回传BP网络

误差回传BP网络中神经元分层排列,网络由一个输入层、一个或多个隐蔽层和一个输出层组成。网络的学习过程主要由输入信息的正向传播和误差的反向传播两个过程组成。在正向传播过程中,输入信息从输入层到隐蔽层再到输出层进行逐层处理。如果输出层的输出与给出的学习样本希望输出不一致,则计算出输出误差,转入误差反向传播过程,通过修改各层神经元之间的权值,使误差达到最小。

2.2 遗传算法

遗传算法(GA)是模仿自然界生物进化思想而得出的一种自适应启发式全局搜索算法,其实质是由复制-交换-变异算子组成的周而复始的循环过程。GA在寻找全局最优的过程中,不需要梯度信息,也不用进行微积分计算,仅通过上述3种算子便可以以很大的概率在解空间搜索到全局最优或次优解,从而有效地减小陷入局部极小点的概率。

复制算子将群体中的个体按与其适应值高低成比例的概率,复制到新群体中。通过复制保留优势个体,淘汰劣势个体,提高了群体的平均适应值,但损失了群体的多样性。复制的作用是实现优胜劣汰,自然选择机制,驱使群体收敛于最优解。交换算子按一定交换概率 P_c 随机选择两个个体,即随机交换部分染色体基因产生两个新个体。遗传算法通过复制和

交换算子可以产生平均适应值更高和更好个体的子代群体,使进化过程朝更优解方向进行。变异算子以很小的概率 P_m 随机地改变染色体串上的某些位,即0变1,1变0。变异算子具有补偿群体多样性损失的重要作用,一方面可以在当前解附近寻找更优解,另一方面可以保持群体的多样性,使群体能够继续进化。

2.3 自适应遗传算法(AGA)

具有比例复制和自适应的交换和变异操作的GA称为AGA。在搜索最优参数过程中,AGA根据个体的具体情况,自适应地改变遗传算法的交换概率 P_c 和变异概率 P_m ,以保持解群体的多样性,防止早熟收敛,提高算法的计算速度和精度^[3]。

$$P_c = \begin{cases} k_1 \frac{f_{\max} - f}{f_{\max} - f_{\text{avg}}}, & f > f_{\text{avg}} \\ k_3, & f < f_{\text{avg}} \end{cases} \quad (1)$$

$$P_m = \begin{cases} k_2 \frac{f_{\max} - f}{f_{\max} - f_{\text{avg}}}, & f > f_{\text{avg}} \\ k_4, & f < f_{\text{avg}} \end{cases} \quad (2)$$

其中, f_{\max} 是群体中的最大适应度, f_{avg} 是群体的平均适应度, f 是用于交换的两个串中较大适应度者, f 是待变异个体的适应度值。

一般推荐值: $k_1 = k_3 = 1, k_2 = k_4 = 0.5$ 。在实际应用中, P_c 常取0.5~1.0, P_m 取较小的值(0.005~0.05)。

2.4 快速、高精度GA-BP混合算法

2.4.1 染色体基因位编码方式

遗传算法的编码方式主要有两种:二进制编码和十进制编码。二进制编码应用较广,但存在如下不足:

1) 算法精度 ϵ 取决于码串的长度及待优化参数的取值范围,即

$$\epsilon = (u - v) / (2^L - 1)$$

其中, u 和 v 是待优化参数的边界值, L 是码串长度。

2) 在具体操作过程中,需要不断地对码串进行进制转换,计算量大。

本文提出的混合算法,首先实现GA中数字串的实数化,即以十进制数字串取代二进制数字串来直接表征参数,于是:

1) 避免了因网络取值范围不明确,难以进行编码的困难;

2) 取消了编码、译码过程,提高了学习算法的速度;

3) 十进制数字串的引入,在数字串长度不变

(GA 的计算量不变) 的情况下, 计算精度可以大为提高。

2.4.2 快速、高精度 GA-BP 混合算法

当 BP 网络陷入局部极小点时, 网络会出现严重的病态。逃离局部极小点一直是 ANN 应用中要解决的关键问题。遗传算法是优化 BP 网络, 逃离局部极小点, 加快网络收敛速度的理想工具^[4]。将引入自适应机制的浮点数编码的遗传算法与 BP 算法相结合, 优化 BP 网络, 可进一步提高网络预测精度和收敛速度。

本文采用的是一个多层前向 BP 网, GA 将网络的一个权重 w 看成一条染色体, 将该网络中所有权重的集合看成一个个体, 在初始化阶段产生大量个体——群体。GA 中的适应度函数, 由

$$E = \frac{1}{2p} \sum_{p,k} (t_{pk} - o_{pk})^2 \quad (3)$$

构造。其中, p 为训练样本数, k 为输出层神经元个数, t_{pk} 为第 k 神经元关于第 p 个样本的多层前向网络的输出, o_{pk} 为第 k 神经元关于第 p 个样本的实际输出。

适应度函数

$$f = \begin{cases} C_{\max} - E, & E < C_{\max} \\ 0, & E \geq C_{\max} \end{cases} \quad (4)$$

式中 C_{\max} 可取进化过程中 E 的最大值。

为保证算法的平稳性和全局收敛性, 在 GA 的选择操作中, 选用最佳保留选择机制, 先按赌轮选择机制进行选择, 然后将当前解中适应度最高的个体复制到下一代群体中, 以保证遗传算法终止时得到的最后结果一定是历代出现过的最高适应度的个体。具体步骤如下:

Step 1: 随机产生 N 组在不同实数区间内取值的初始网络权值, 作为初始群体;

Step 2: 用 BP 算法对这 N 组初始权值分别进行预训练, 若经过预训练后, 这 N 组权值中至少已有一组满足精度要求, 则算法结束, 否则转 Step 3;

Step 3: 分别依据经过预训练的上述 N 组权值所对应的上下限确定取值区间, 在区间内随机生成 $r \times N$ 组新的权值, 连同经过预训练的 N 组权值构成完整的基因群体, 共 $(r+1) \times N$ 组权值;

Step 4: 对这 $(r+1) \times N$ 组权值进行选择、交换、变异等自适应遗传操作;

Step 5: 如果经 Step 4 的操作已至少得到一组满足精度要求的权值, 则算法结束, 否则从经过遗传操作的这 $(r+1) \times N$ 组权值中选出 N 组较好的, 转

Step 2。

算法流程如图 1 所示。

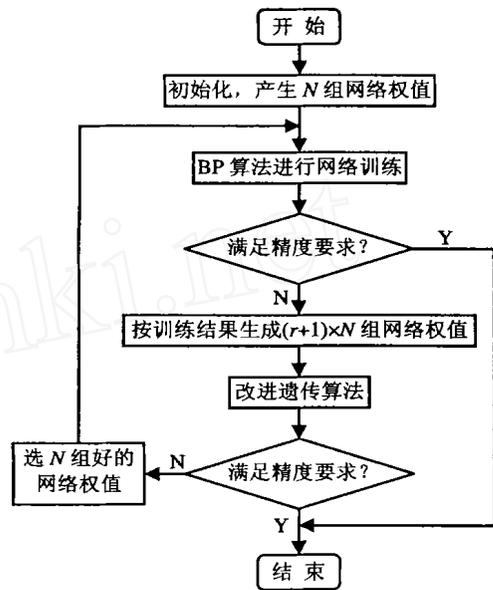


图 1 GA-BP 混合算法流程

3 薄互储层参数预测

3.1 地震特征参数

地震资料提供了地下地质构造和岩石性质的海量信息, 地震属性(地震特征参数)就是从地震数据中产生的几何学、运动学、动力学及统计特征的具体测量。所以, 利用多种地震特征参数和神经网络技术进行薄互储层参数的横向预测, 特征参数的选择非常重要。

进行薄互储层参数预测时, 我们提取了与储层有关的各种地震属性 50 余种, 分为线性参数和非线性参数。线性参数包括瞬时参数、振幅类参数、频谱类参数、相关类参数等; 非线性参数包括分形分维参数和混沌参数等。通过井资料对地震特征参数进行的相关分析, 结合专家经验, 再利用实际地震资料进行检验, 确定出关键参数, 作为网络的输入神经元。

3.2 应用实例

利用本文方法对大庆某工区的实际地震资料进行了砂岩厚度和孔隙度两个参数的预测, 预测目的层是某油层中的 P_1 砂层组。该工区共有 18 口井, 我们首先利用该工区的地震资料提取了 55 种特征参数, 然后利用 18 口井资料对 55 种地震特征参数进行相关分析, 最终选用最大振幅、平均振幅、目的层的主频、目的层的振幅谱能量、瞬时频率、瞬时相位、自相关函数的最大值、分形分维类中的曲线长度

等8个参数作为网络的输入神经元,在15口井及其附近各取6道作为已知样本,共有90个训练样本,另外3口井作为一般化性能测试样本。样本举例如表1。

根据本文研究的薄互储层参数预测的新方法,利用C++编制了相应的应用软件,网络参数的选取灵活性很大。但针对该算法本身所具有的特性,隐层越多,网络连接权值越多,相应的遗传算法的染色体个数也越多,从而影响训练速度。所以,在误差精

度相差不大的情况下,隐层不宜超过4层。

薄互储层参数预测系统在该工区的网络拓扑结构为8-12-12-2,即8个输入神经元,2个隐蔽层,每个隐层的神经元均为12个,输出层有2个神经元(砂岩厚度和孔隙度)。

将训练样本输入网络结构进行训练,当BP网络及改进的遗传算法均完成给定的迭代次数后,整个网络便完成一次迭代。学习参数选取如下:学习速度0.7,惯性系数0.9,系统最大误差0.01,单个样本

表1 样本和地震特征参数

井名	最大振幅	平均振幅	主频	振幅谱能量	瞬时频率	瞬时相位	自相关最大值	曲线长度
Q1	29 636 06	1 709 56	37.11	433 861 984 0	32 16	0 019	103 384 39	4 360 68
Q24	38 407 84	7 575 62	31.25	613 254 592 0	27.90	0 59	194 346 41	3 919 67
Q12	52 361 81	1 829 48	43.95	1 148 697 984 0	38.95	0 18	212 195 40	6 613 60
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

表2 采用本文遗传算法神经网络及BP网络预测砂岩厚度与钻井统计值对比

井号	砂岩厚度统计值/m	本文的遗传算法人工神经网络			BP算法		
		预测值/m	绝对误差/m	相对误差/%	预测值/m	绝对误差/m	相对误差/%
Q1	6.2	6.232 236	0.032 236	0.52	6.111 125	-0.088 87	1.43
Q12	8.0	7.891 461	-0.108 54	1.36	8.028 696	0.028 696	0.36
Q24	4.2	4.216 523	0.016 523	0.39	4.150 432	-0.049 57	1.18
Q262	1.8	1.821 135	0.021 135	1.17	1.868 789	0.068 789	3.82
Q281	6.0	5.953 371	-0.046 63	0.78	5.965 227	-0.034 77	0.58
Q56*	8.0	7.051 205	-0.948 79	11.86	6.826 152	-1.173 85	14.67
Q278	2.8	2.857 336	0.057 336	2.04	2.875 715	0.075 715	2.70
Q29*	5.4	5.858 735	0.458 735	8.49	5.952 022	0.552 022	10.22
Q27	7.0	7.044 447	0.044 447	0.63	7.055 022	0.055 022	0.79
Q320*	7.2	7.699 291	0.499 291	6.93	8.018 740	0.818 740	11.37

表3 采用本文遗传算法神经网络及BP网络预测孔隙度值与钻井统计值对比

井号	孔隙度统计值/%	本文的遗传算法人工神经网络			BP算法		
		预测值/%	绝对误差	相对误差/%	预测值/%	绝对误差	相对误差/%
Q1	21.2	21.225 13	0.019 16	0.12	21.089 37	-0.191 06	0.52
Q12	18.8	18.695 81	-0.104 19	0.55	19.081 09	0.281 09	1.47
Q24	17.8	17.817 08	0.017 08	0.10	18.805 48	0.005 48	0.03
Q262	16.9	16.889 43	-0.010 57	0.06	17.105 98	0.205 98	1.22
Q281	20.6	20.505 58	-0.094 42	0.46	20.630 31	0.030 31	0.15
Q56*	21.7	20.316 41	-1.383 59	6.38	23.478 99	1.778 99	8.19
Q278	19.5	19.399 71	-0.100 29	0.51	19.830 67	0.330 67	1.67
Q29*	25.0	25.983 13	0.983 13	3.93	26.738 87	1.738 87	6.96
Q27	15.6	15.614 43	0.014 43	0.09	15.696 56	0.096 56	0.62
Q320*	24.6	23.716 81	-0.883 19	3.59	23.305 73	-1.294 27	5.26

最大误差 0.001, BP 网络迭代次数 20, 改进的遗传算法迭代次数 20, 整个网络迭代次数 1 000。

表 2 是在上述网络拓扑结构和学习参数下, 对砂岩厚度预测的结果。为便于对比, 同时利用传统 BP 网络(在相同网络拓扑结构和学习参数下)进行了砂岩厚度预测(其中带 * 号的为性能测试样本井)。通过与已知井对比, 本文提出的算法预测的最大绝对误差为 0.95 m, 最大相对误差为 11.86%; BP 网络预测的最大绝对误差为 1.17 m, 最大相对误差为 14.67%, 这说明采用本文算法对砂岩进行的预测精度是较高的, 而且达到相同系统误差时的运算速度比 BP 网络提高 1.5 倍左右。

表 3 是利用本文方法和 BP 网络对该工区的孔隙度进行预测的结果, 通过井处预测值与实际值比较, 表明本文的遗传算法神经网络预测精度高于 BP 网。

4 结 论

快速、高精度遗传算法神经网络, 由于对传统的遗传算法进行了改进, 综合了遗传算法和人工神经网络两者的优点, 克服了 BP 算法收敛速度慢和可能收敛到局部极小点的缺陷, 从而提高了网络的收敛速度和预测精度。对实际资料进行的预测表明, 该方法是进行薄互储层参数预测行之有效的方法, 而且该方法在地震勘探中具有很好的应用前景。

薄互储层参数预测效果既依赖于所采用的网络模型, 也依赖于所选用的地震特征参数, 同时还具有地区性差异。这是薄互储层参数预测的难点。所以, 快速、高精度遗传算法神经网络这一新的薄互储层参数预测方法, 还有待于在实际应用和研究中得到进一步改进和完善。

参考文献(References):

- [1] 焦李成. 神经网络系统理论[M]. 西安: 西安电子科技大学出版社, 1995.
- [2] Grefenstette J. Optimization of control parameters for genetic algorithms[J]. *IEEE Trans on Syst, Man and Cybern*, 1986, 16(1): 122-128.
- [3] 黄秀轩, 朱学峰. 改进的自适应遗传算法[J]. 中国学术期刊文摘, 1998, 14(11): 1415-1417.
(Huang Xiu-xuan, Zhu Xue-feng. Modified adaptive genetic algorithm[J]. *Academic Periodical Abstracts of China*, 1998, 14(11): 1415-1417.)
- [4] 王建成, 高大启, 王静, 等. 改进的遗传和 BP 杂交算法及神经网络经济预警系统设计[J]. 系统工程理论与实践, 1998, 18(4): 136-141.
(Wang Jian-cheng, Gao Da-qi, Wang Jing, et al. Designing of ANN economic early warning system based on improved genetic and BP algorithms[J]. *System Engineering—Theory & Practice*, 1998, 18(4): 136-141.)

(上接第 598 页)

- [8] H Nakanishi, I B Turksen, M Sugeno. A review and comparison of six reasoning methods[J]. *Fuzzy Sets and Systems*, 1993, 57: 257-294.
- [9] G J Wang. On the logic foundation of fuzzy reasoning[J]. *Information Sciences*, 1999, 117: 47-88.
- [10] I B Turksen, Zhao Zhong. An approximate analogical reasoning schema based on similarity measures and interval-valued fuzzy sets[J]. *Fuzzy Sets and Systems*, 1990, 34: 323-346.
- [11] J J Buckley, Yoichi Hayashi. Fuzzy input-output controllers are universal approximators[J]. *Fuzzy Sets and Systems*, 1993, 58: 273-278.
- [12] J J Buckley. Sugeno type controllers are universal controllers[J]. *Fuzzy Sets and Systems*, 1993, 53: 299-303.
- [13] B Kosko. Fuzzy systems as universal approximators[A]. *Proc of IEEE Int Conf on Fuzzy Systems*[C]. San Diego, 1992. 1153-1162.
- [14] L X Wang. Fuzzy systems are universal approximators[A]. *Proc of IEEE Int Conf on Fuzzy Systems*[C]. San Diego, 1992. 1163-1170.
- [15] Laszlo T Koczy, Alessandro Zorat. Fuzzy systems and approximation[J]. *Fuzzy Sets and Systems*, 1997, 85: 203-222.
- [16] Castro J L, Delgado M. Fuzzy systems with defuzzification are universal approximators[J]. *IEEE Trans on Systems, Man and Cybernetics*, 1996, 26(1): 149-152.
- [17] Castro J L. Fuzzy logic controllers are universal approximators[J]. *IEEE Trans on Systems, Man and Cybernetics*, 1995, 25(4): 629-635.