

# GA-BP 神经网络对含油饱和度的动态预测研究

潘少伟<sup>1</sup>, 梁鸿军<sup>2</sup>, 李 良<sup>2</sup>, 杨少春<sup>3</sup>

(1. 西安石油大学 计算机学院, 陕西 西安 710065;

2. 中国石油长庆油田勘探开发研究院, 陕西 西安 710021;

3. 中国石油大学 地球科学与技术学院, 山东 青岛 266580)

**摘 要:**为克服传统 BP 神经网络收敛速度慢、易陷入局部极小值等不足,采用遗传算法对其进行了优化。之后利用遗传算法优化的 BP 神经网络—GA-BP 实现了对江苏油田庄 2 断块某小层含油饱和度的动态预测;首先确定 GA-BP 神经网络的输入、输出层神经元,接着建立经验公式,实现了输入神经元时间  $T$  的量化,然后选取训练样本对 GA-BP 神经网络进行训练,最终建立起含油饱和度的动态预测模型并利用该模型对 5 年以后的含油饱和度进行了预测。该预测结果对油田现场下一步的生产实践具有重要的指导意义。

**关键词:**BP 神经网络;GA-BP 神经网络;含油饱和度;动态;预测

**中图分类号:**TP39

**文献标识码:**A

**文章编号:**1673-629X(2012)12-0157-04

## Research on Dynamic Prediction of Oil Saturation with GA-BP Neural Network

PAN Shao-wei<sup>1</sup>, LIANG Hong-jun<sup>2</sup>, LI Liang<sup>2</sup>, YANG Shao-chun<sup>3</sup>

(1. School of Computer, Xi'an Shiyou University, Xi'an 710065, China;

2. Research Institute of Petroleum Exploration & Development, Changqing Oilfield, Xi'an 710021, China;

3. School of Geosciences and Technology, China University of Petroleum, Qingdao 266580, China)

**Abstract:** In order to overcome the slower convergence rate and falling into local minimal value easily of traditional BP neural network, the genetic algorithm is used for optimization. Then the oil saturation of a certain layer of Z2 fault-block in Jiangsu oilfield is predicted. Firstly, the input and output layer neurons are established. Secondly, an empirical formula is given that realizes the quantification of the input neuron-time. Lastly, the dynamic prediction model on oil saturation of the GA-BP neural network is established after the training with the selected samples. And the oil saturation of 5 years later is predicted with the dynamic prediction model. The prediction result of oil saturation has an important guiding significance to the production practice.

**Key words:** BP neural network; GA-BP neural network; oil saturation; the dynamic; prediction

## 0 引言

基于人工神经网络方法的储层参数预测,目前已取得了大量研究成果<sup>[1-9]</sup>,但也暴露了一些不足。例如用 BP(Back Propagation)神经网络建立储层参数预测模型时,就存在两点不足:一是收敛速度慢,二是目标函数易陷入局部极小值。遗传算法(Genetic Algorithm, GA)是模拟达尔文遗传选择和自然淘汰的全局性概率搜索算法,具有较强的解决问题的能力。笔者

就是通过遗传算法优化 BP 神经网络的连接权值和网络结构,并在进化过程中采取保留最佳个体的方法,解决了 BP 神经网络上述的两点不足,从而建立了江苏油田庄 2 断块某小层含油饱和度的神经网络预测模型。并利用该模型对 5 年以后的含油饱和度进行了预测,其结果对该油田下一步生产实践具有重要的指导意义。

## 1 GA-BP 神经网络

文中,GA-BP 神经网络是指遗传算法优化的 BP 神经网络。

### 1.1 BP 神经网络

BP 神经网络通常是指基于误差反向传播算法的多层前向神经网络,由 D. E. Rumelhart 和 J. L. McCel-

收稿日期:2012-05-30;修回日期:2012-09-02

基金项目:国家科技重大专项子课题(2011ZX05032-001);陕西省自然科学基金项目(2010JM8032,2012JQ8040);陕西省教育科学计划项目(2010JK772,11JK1071)

作者简介:潘少伟(1979-),男,山东平度人,博士,主要从事油藏描述方法研究。

land 及其研究小组于 1986 年研究设计出来<sup>[5]</sup>。BP 算法为目前应用最广泛的神经网络学习算法,绝大部分的神经网络模型采用 BP 算法或它的变化形式,同时它也是前向神经网络的核心部分。BP 神经网络的主要学习过程是:将输入模式从输入层经隐含层单元逐层处理,并传向输出层,每一层神经元的状态只影响下一层神经元的状态。如果在输出层不能得到期望的输出,则将误差信号沿原来的连接通路反向传播,通过修改各神经元的权值,使误差信号最小<sup>[10]</sup>。通常采用的方法是梯度下降法,即扩展的  $\delta$  规则,因此在算法上就要求神经元的特性函数必须是可导的。而且,由于该算法两次迭代的搜索方向总是正交的,即  $[-\nabla J(a_{k+1})] \cdot [-\nabla J(a_k)] = 0$ ,这说明,该方法在寻优过程中极易陷入局部极小值,并且收敛速度较慢,因此要对其进行改进。

## 1.2 遗传算法

遗传算法由美国 Michigan 大学的 J. Holland 教授于 1975 年提出。它的主要特点是群体搜索策略和群体中个体之间的信息交换,搜索不依赖梯度信息,也不需要求解函数可微,只需要该函数在约束条件下可解,因此该方法适合解决各种复杂的非线性问题。遗传算法提出了与生物进化相似的步骤,即:自然选择、交叉、变异、优胜劣汰;同时,它还具有 4 个基本要素,分别是:基因编码、初始群体设定、适应度函数和遗传算子。

### (1) 基因编码。

文中直接采用了实数编码(通常,标准的遗传算法都采用二进制编码,但二进制编码会造成编码串过长,从而降低学习效率,另外在编码和解码的过程中存在量化误差,会影响神经网络的学习精度<sup>[11]</sup>,因此,在含油饱和度预测过程中涉及的数据直接采用了实数编码)。

### (2) 初始群体设定。

遗传算法是对群体的迭代操作,必须准备一个初始群体,文中初始群体是随机选取的。由初始群体(祖先)经过遗传操作,形成一代代子代,逐步搜索到最优解。

### (3) 适应度函数。

适应度函数  $F(P_i)$  的计算公式如下:

$$F(P_i) = 1/E(W_i) \quad (1)$$

$$E(W_i) = 1/2 \times \sum (O^{(P)}(W)_i - t^P)^2 \quad (2)$$

其中  $W_i$  为群体中的第  $i$  个个体,是需要优化的神经网络权值,  $P_i$  为  $W_i$  对应的染色体。  $E(W_i)$  为误差函数,  $O^{(P)}(W)_i$  是第  $P$  个样本输入时神经网络的输出值,  $t^P$  为第  $P$  个样本的期望输出值。

### (4) 遗传算子。

遗传算子包括选择算子、交叉算子和变异算子。

选择算子是遗传算法的关键,它体现了自然界适者生存的思想。从群体中选择染色体作为下一代的父染色体,可采用圆盘选择方法计算。

遗传算法中必须设定一个交叉算子,用交叉概率  $P_c$  表示,通过交叉概率  $P_c$  做算术交叉来完成实数编码方案的交叉操作,其计算公式如下<sup>[7]</sup>:

$$w_{yp_i} = w_{yp_i}(1 - P_c) + w_{yp_n}P_c \quad (3)$$

$$w_{yp_n} = w_{yp_n}(1 - P_c) + w_{yp_i}P_c \quad (4)$$

式中:  $w_{yp_i}$ 、 $w_{yp_n}$  分别是第  $l$ 、 $m$  个染色体个体中第  $i$  个神经元到第  $j$  个神经元的权值;  $w_{yp_i}$ 、 $w_{yp_n}$  分别表示第  $l$ 、 $m$  个染色体个体产生 2 个子代的第  $i$  个神经元到第  $j$  个神经元的权值。交叉算子通过(3)式和(4)式来完成交叉操作。

变异算子的作用就是改变字符串某个位置上的字符,通过变异操作可以弥补遗传物质的丢失。变异保证了遗传算法的全局收敛性,变异概率  $P_m$  通常取值较小。

## 1.3 GA-BP 神经网络

GA-BP 神经网络的基本思想就是将遗传算法和 BP 神经网络结合起来,利用遗传算法的种群搜索方式对给定的网络权值、阈值进行最优配置。其目的是寻找最容易获得全局最优的网络参数,从而改善 BP 神经网络过分依赖梯度信息,实现误差全局最小的最大可能性。

## 2 GA-BP 神经网络对含油饱和度动态预测

GA-BP 神经网络对含油饱和度的动态预测就是首先把时间  $T$  量化,然后将其作为一项参数输入到神经网络中,进而建立起相应的神经网络预测模型,利用该模型实现对未来含油饱和度的预测。江苏油田庄 2 断块于 1995 年 10 月开始注水开发,到目前为止,积累了大量的动静态资料。在动静态开发资料的基础上,采用 GA-BP 神经网络建立含油饱和度的动态预测模型,指导该油田下一步的生产实践。

### 2.1 GA-BP 神经网络的结构设计

GA-BP 神经网络输入层和输出层的神经元个数分别为 6 和 1,其结构如图 1 所示。利用如下公式来计算隐含层的神经元个数  $n_1$ <sup>[12]</sup>:

$$n_1 = a + \sqrt{n + m} \quad (5)$$

上式中,  $m$ 、 $n$  分别为神经网络输入层、输出层的神经元个数,  $a$  为 1~10 之间的常数,取  $n_1$  的值为 6。

### 2.2 输入神经元 $T$ 的量化及训练样本的选取

时间是一个抽象的概念,如何将其量化是一项全新的课题。依据江苏油田庄 2 断块的具体开发历

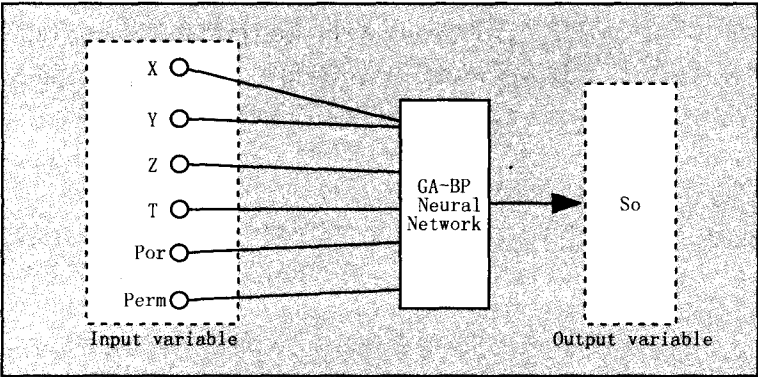


图 1 GA-BP 神经网络结构图

程,得出时间  $T$  定量化的经验公式为:  $T = (X - 1994) * 12 + b$ 。其中,  $X$  代表了年份,  $b$  代表了月份。规定 1993 年 12 月对应的  $T$  值为 0,把每口井的测井日期(××年××月)代入到上述公式即可求得每口井对应的  $T$  值(见表 1)。

表 1 输入神经元—时间  $T$  的量化列表

时间	量化的时间变量 $T$	时间	量化的时间变量 $T$
1993 年 12 月	0	...	...
1994 年 1 月	1	2002 年 10 月	106
1994 年 2 月	2	2002 年 11 月	107
1994 年 3 月	3	2002 年 12 月	108
1994 年 4 月	4	...	...

把每口井的测井解释结果和 值作为样本来对 GA-BP 神经网络进行训练。同时为保证所建的神经网络预测模型具有较高精度,并且能较精确的反映地下储层参数的实际变化规律,对样本数据还须进行归一化处理。

2.3 GA-BP 神经网络算法

GA-BP 神经网络算法的实现过程如下<sup>[7]</sup>:

在此之前,需初始化相应的参数:群体规模  $S = 120$ ,染色体  $P_i(0)$ ,  $i = 0, 1, \dots, 119$ ,  $t = -1$ ,交叉概率  $P_c = 0.8$ ,变异概率  $P_m = 0.2$ ,终止条件,  $\varepsilon_1 = 0.03$ ,  $\varepsilon_2 = 0.001$ ,显然  $\varepsilon_2 < 0.05\varepsilon_1$ ,BP 算法的计算次数  $j = 2000$ 。

- (1)  $t++$ , 计算适应度  $F(P_i(t))$ ;
- (2) 选择,利用  $P_c$ 、 $P_m$  进行交叉、变异的遗传操作;
- (3) 如果满足  $F(P_i) = F_{\max}(P_i)$  的  $P_i$  使得  $E(W_i(t)) < \varepsilon_2$  则执行下一步,否则转到(1);
- (4) 如果满足  $F(P_i) = F_{\max}(P_i)$  的  $P_i$  使得  $E(W_i(t)) < \varepsilon_1$  则转到(8),否则,以  $W_i(t) \mid F(P_i) = F_{\max}(P_i)$  作为网络初始化权值,  $i = -1$ ;
- (5)  $t++$ ,  $j++$ , BP 算法的前向计算;
- (6) BP 算法的后向计算;
- (7) 如果  $E(W_i(t)) < \varepsilon_1$  则执行下一步;否则,如果  $i < j$ ,则用  $W_i(t)$  替代 GA 群体中最差的个体,转到

- (5),否则修改  $\varepsilon_2$ ,  $\varepsilon_2 = 0.05E(W_i(t))$ ,并转到(1);
- (8)结束。

训练结束后,得到  $E(W_i(t)) = 0.009 < \varepsilon_1 = 0.03$ ,即训练后得到的总误差小于预定误差。可见,通过训练得到的 GA-BP 神经网络预测模型达到了预定的要求。

2.4 模型检验

选取 60 个样本数据来测试所建的 GA-BP 神经网络预测模型,样本中含油饱和度的测井解释值和 GA-BP 神经网络的预测值如下(表 2)。由表 2 可知,GA-BP 神经网络的预测值和测井解释值比较接近,相对误差基本上在  $\pm 13\%$  以内。

2.5 性能比较

在相同条件下,比较了 GA-BP 算法与 BP 算法的训练误差(图 2),这里纵坐标表示每次训练时的平均误差,横坐标表示训练次数。由图 2(a)可以看出,在

表 2 样本的测井解释值和 GA-BP 神经网络的预测值列表

属性	测井解释值	神经网络预测值	绝对误差	相对误差
含油饱和度(%)	58.090	56.324	1.766	3.0%
	30.220	30.001	0.219	0.7%
	5.840	6.358	0.518	8.9%
	0.000	0.000	0.000	0.0%
	53.269	46.390	6.879	12.9%
	8.250	7.920	0.330	4.0%
...	...	...	...	...

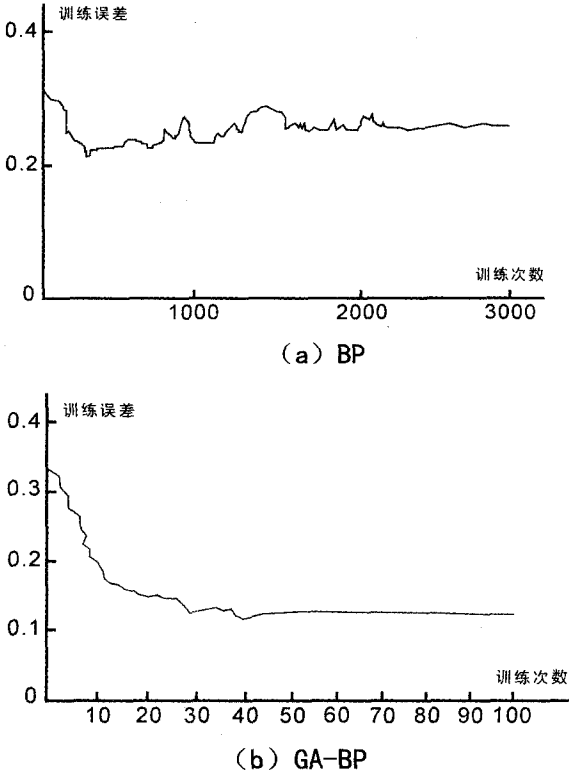


图 2 神经网络学习过程误差曲线

前 2000 次的迭代过程中, BP 算法的训练误差波动较大, 在迭代到 2000 次以后趋于收敛, 但仍然存在“波动现象”; 由图 2(b) 可看出, GA-BP 算法在迭代生成 40 代以后就已经收敛, 而且不存在“波动现象”, 显而易见, 在收敛性方面 GA-BP 算法优于 BP 算法。

## 2.6 未来含油饱和度的预测

利用已获得的 GA-BP 神经网络预测模型, 对 5 年以后的含油饱和度进行了预测, 预测结果见表 3。通过对预测结果进行总结, 得到这样一个结论: 伴随着注水开发, 含油饱和度在不断的减小; 在含油饱和度高值区, 含油饱和度减小的幅度较大, 在含油饱和度低值区, 在含油饱和度减小的幅度较小。

表 3 GA-BP 神经网络对 5 年以后  
含油饱和度的预测值列表

属性	深度(m)	测井解释		GA-BP 神经网络预测	
		T 值	测井解释值	T 值	预测值(5 年之后)
含油饱和度(%)	1601.000	3	58.090	63	46.332
	1602.625	105	30.220	165	25.013
	1603.500	105	5.840	165	4.328
	1605.000	76	0.000	136	0.000
	1606.125	22	53.269	82	46.361
	1607.000	22	8.250	82	6.310
	...	...	...	...	...

## 3 结束语

(1) 针对 BP 神经网络存在的两点不足, 文中引入了 GA-BP 神经网络, 并分析了 GA-BP 神经网络算法的迭代求解过程。

(2) 利用 GA-BP 神经网络实现了对江苏油田庄 2 断块某小层含油饱和度的动态预测: 首先确定 GA-BP 神经网络的输入、输出层神经元, 接着建立经验公式, 实现了输入神经元时间  $T$  的定量化, 然后选取训练样本对 GA-BP 神经网络进行了训练, 建立起含油饱和度的动态预测模型。并在同等条件下对比分析了 GA-BP 算法和 BP 算法的训练误差, 结果表明 GA-BP 算法的收敛性大大优于 BP 算法。

(3) 利用 GA-BP 神经网络实现了 5 年以后含油饱和度的动态预测, 其结果对油田下一步的生产实践

具有一定的指导意义。但地下储层参数的动态变化异常复杂, 仅仅应用 GA-BP 神经网络来对其进行预测是远远不够的。文中只是提出了一种含油饱和度的动态预测方法, 以期达到抛砖引玉的作用。

## 参考文献:

- [1] 徐守余, 王艳红. 利用神经网络建立储层宏观参数动态模型—以胜利油田二区为例[J]. 油气地质与采收率, 2005, 12(6): 10-12.
- [2] 邵才瑞, 印兴耀, 李洪奇, 等. 储层属性的遗传神经克里金插值方法及其应用[J]. 中国石油大学学报(自然科学版), 2007, 31(5): 35-40.
- [3] 单敬福, 纪友亮, 柳成志. 改进人工神经网络原理对储层渗透率的预测—以北部湾盆地涠西南凹陷为例[J]. 石油与天然气地质, 2007, 28(1): 106-109.
- [4] 王文娟, 曹俊兴, 张元标, 等. 基于微粒群算法的神经网络储层物性参数预测[J]. 西南石油大学学报, 2007, 29(6): 31-33.
- [5] 陈蓉, 王峰. 基于 MATLAB 的 BP 神经网络在储层物性预测中的应用[J]. 测井技术, 2009, 33(1): 75-78.
- [6] 陈钢花, 董维武. 遗传神经网络在煤质测井评价中的应用[J]. 测井技术, 2011, 35(2): 171-175.
- [7] 周金应, 桂碧雯, 李茂, 等. 基于岩控的人工神经网络在渗透率预测中的应用[J]. 石油学报, 2010, 31(6): 985-988.
- [8] 吴秋波, 吴元, 王允诚. 混合学习法前向网络多属性储层参数预测[J]. 西南石油大学学报(自然科学版), 2011, 33(2): 68-72.
- [9] Vander B M. Neural networks in geophysical applications[J]. Geophysics, 2000, 65(4): 1032-1047.
- [10] 杨宁, 史仪凯, 袁小庆, 等. 基于 BP 网络横向磁场永磁电机调速系统的设计[J]. 西北工业大学学报, 2011, 29(5): 824-828.
- [11] 邓一兵, 胡伟, 高峰, 等. 遗传神经网络在载人飞船环控决策系统中的应用研究[J]. 西安交通大学学报, 2010, 44(7): 64-69.
- [12] 张立明. 人工神经网络的模型及其应用[M]. 南京: 复旦大学出版社, 1994.

(上接第 156 页)

- [5] 行平台研究与实现[J]. 北京航空航天大学学报, 2003, 29(10): 897-900.
- [5] 赵会群, 孙品. 面向服务的可信软件体系结构代数模型[J]. 计算机学报, 2010, 33(5): 890-899.
- [6] 高勇, 刘瑜, 谢昆青, 等. 一个基于 Petri 网的 Web 服务组合模型[J]. 计算机工程, 2006, 32(6): 17-18.
- [7] 辜希武, 卢王鼎. 基于 Pi-演算的 BPEL4WS Web 服务组合形式化模型[J]. 计算机科学, 2007, 34(3): 69-74.
- [8] 黎升洪. 基于进程代数的 Web 服务编排方法及其类型理

论研究[D]. 上海: 上海大学, 2009.

- [9] 赵春娟, 肖迎元. 一种基于语义的 Web 服务组合方法[J]. 天津理工大学学报, 2010, 26(5): 29-33.
- [10] 梁晟. 基于语义 Web 的服务自动组合技术的研究[D]. 北京: 中国科学院, 2000.
- [11] 杨鑫, 陈俊亮. WSC/ADL: Web Services 组合系统体系结构描述语言[J]. 软件学报, 2006, 17(5): 1182-1194.
- [12] 骆翔宇, 谭征, 苏开乐, 等. 一种基于认知模型检测的 Web 服务组合验证方法[J]. 计算机学报, 2011, 34(6): 1041-1061.