

# 基于遗传算法的 BP 神经网络在油田措施 规划预测中的应用

## Application of BP Neural Network Based On Genetic Algorithm in Oil Field Measure Program

龚安 王霞 (中国石油大学(华东)计算机与通信工程学院 257061)  
姜焕军 (泰山学院信息科学技术系 山东泰安 271000)

**摘要:**如何有效地确定神经网络的结构和参数,一直是神经网络研究中的一个难点。遗传算法是一种基于自然选择和生物遗传机理的全局搜索算法,本文提出了一种改进的遗传算法来优化 BP 神经网络,并将其应用于油田措施规划预测模型中。结果表明,该方法具有收敛速度快和预测精度高的特点。

**关键词:**BP 神经网络 措施规划 遗传算法

### 1 引言

随着我国油田开发进入中后期,开发对象逐步转向难开采的低品位和高含水油藏,为确保企业在稳产的前提下获得最大效益,现场采取了多种措施,如压裂、酸化、大修、补孔、转抽等。如何构建高精度的油田措施规划预测模型,优化措施投入量,以获得最佳措施产量成为目前油田现场最为关注的问题。基于遗传算法的 BP 神经网络以其强大的容错能力和对非线性数据的处理能力可构建科学合理的油田措施规划预测模型,依据该模型可实施措施效果的预测和优化,使企业效益最大化。

### 2 BP 神经网络构建措施规划预测模型

BP 网络是基于误差反向传播神经网络 (Back Propagation Neural Network) 的简称,是实现映射变换的前馈神经网络中最常用的一种网络。从结构上讲,三层 BP 神经网络是一个典型的半线性前馈神经网络,它被分成输入层、中间层(或隐含层)和输出层。同层节点间无关联,异层节点间前后相连接。其中,输入层节点数对应于 BP 网络可感知的输入个数,输出层节点数与 BP 网络的输出个数相对应,中间层节点的数

目可根据需要设置。

三层前馈型 BP 网络存储知识时采用的误差反传学习算法是一种典型的误差修正方法。首先通过网络向前发送输入模式,然后计算实际输出与理想输出之间的误差,并将其归结为连接层中各节点连接权及阈值的“过错”,通过输出层节点的误差逐层向输入层反向传播以“分摊”给各连接节点,从而可算出各连接节点的参考误差,并据此对各连接权进行相应的调整,使网络趋向于要求的映射。通过不断反复上述“输入模式→误差计算→误差反传→权值调整”过程,网络实际输出与理想输出之间的误差不断地减少,直到最终获得期望的输出,网络学习(或网络训练)便告结束。此时,该网络就可以应用于实际预测工作中了。

实现油田措施规划的预测,首先要根据已有的措施数据和措施产量构建预测模型。人工神经网络(ANN)以其在对非确定性、非规则性数据,特别是带噪音的、杂乱的非线性数据处理方面强大的处理功能,成为构建油田措施规划预测模型理想的技术手段。ANN 通过对历年措施数据样本的学习,自动获得最佳逼近样本数据规律的函数,无需数学物理模型和人工干预即可自动建立预测模型并精确映射任意高度非线性的

输入输出关系,且容错性和自适应性良好。

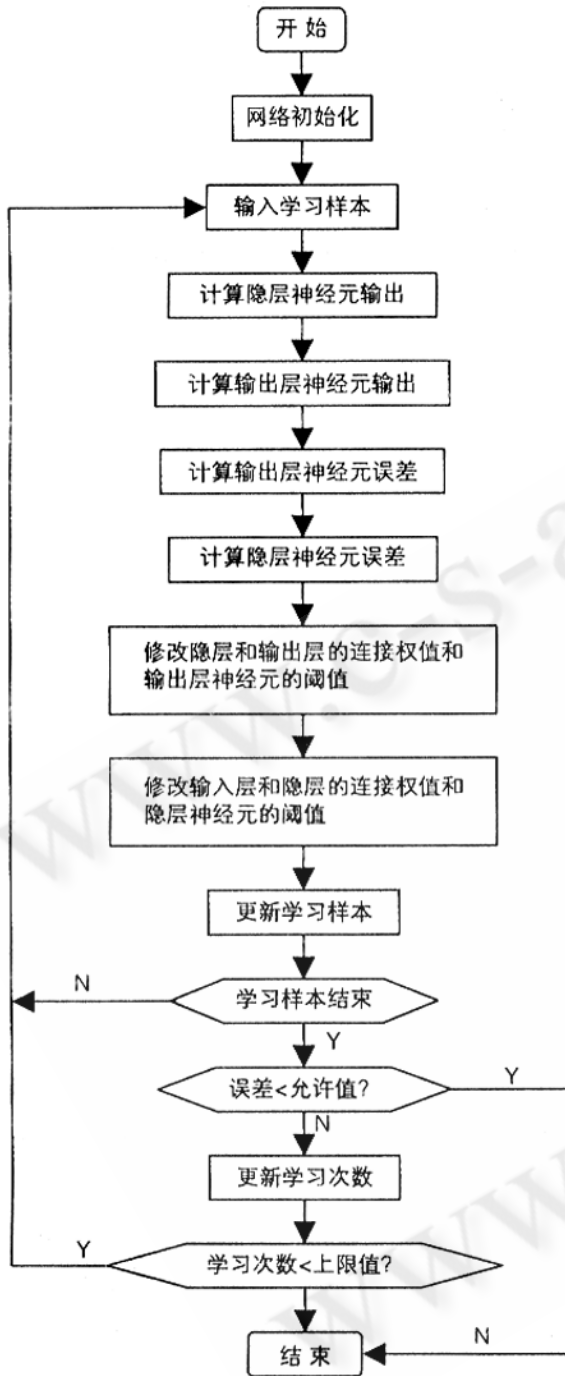


图 1 BP 算法程序流程图

如图 2 所示,利用人工神经网络从大量的历史数据中学习训练,提取油田措施规划领域的相关知识,并将知识表示为网络连接权值的大小与分布,即可建立油田措施规划预测模型。BP 神经网络具有大规模的并行处理能力,计算速度快,且由于知识信息分布存储在网络的连接权与阈值上,因而具有很强的容错性和

自学习、自适应能力以及异域联想功能,对复杂的非线性建模非常容易,所以应用广泛。但经典 BP 网络仍存在收敛速度慢、稳定性差,易陷入局部极小的缺陷,限制了 BP 神经网络的进一步应用,而用遗传算法来优化 BP 神经网络可克服上述缺点达到理想的效果。

### 3 遗传算法优化预测模型

遗传算法是模拟达尔文的遗传选择和自然淘汰的生物进化过程的计算模型,是一种基于自然选择和遗传变异等生物机制的全局性概率搜索算法。主要特点是采用整体搜索策略,搜索不依赖梯度信息和求解函数可微,只需函数约束条件下可解,因此具有自适应、全局优化和隐含并行性等优点。

遗传算法擅长全局搜索,而 BP 人工神经网络应用于局部搜索时比较有效,因此首先用遗传算法对 BP 神经网络的初始连接权进行优化,在解空间中定位出一个较好的搜索空间,然后再采用 BP 算法在这个小空间中搜索出最优解。具体步骤如下:

#### 3.1 编码

用遗传算法训练 BP 神经网络,首先需要将网络连接权和阈值进行编码。常用的编码方式有二进制编码和实数编码。当网络的规模稍大,采用二进制进行编码染色体的长度就可能很长,从而影响遗传算法的效率,因此应当采用实数编码的方式。编码时,每个连接权值(染色体的一个基因位)用一个实数表示,连接网络的所有权值用一组实数表示,遗传操作在任意两组实数上进行,以缩短了染色体长度,提高计算效率。

#### 3.2 群体初始化

群体初始化的关键是设定群体的规模,即权值组合数目的确定。群体规模作为遗传算法的主要控制参数之一,对遗传算法效能的发挥影响较大。如果种群规模大,种群中个体的多样性增加,容易找到最优解,但是会延长收敛的时间;如果种群数目较少,收敛速度提高了,但容易陷入局部极小。一般的遗传算法总是凭借经验确定种群数目。针对这种情况,提出一种自适应确定种群数目的遗传算法,在算法初期采用较大的种群数目,随着算法的深入,相应减小种群数目。这样,既保证了个体的多样性,避免算法陷入局部极小值,又提高了算法的收敛速度,使算法很快找到最优解。

### 3.3 适应度函数的计算

在遗传算法中,对适应度函数的唯一要求就是对输入可以计算出能加以比较的非负结果,因此这里采用网络的误差函数作为适应度函数,并认为误差大的个体其适应度小。具体表示如下:

$$F = C - E \quad (1)$$

其中 C——常数

E——网络的误差函数

$$E = \frac{1}{2} \sum_m \sum_k (Y_{m,k} - \bar{Y}_{m,k})^2 \quad (2)$$

这里,  $Y_{m,k}$  和  $\bar{Y}_{m,k}$  分别是网络第  $m$  个训练样本的第  $k$  个输出节点的实际输出和期望输出。

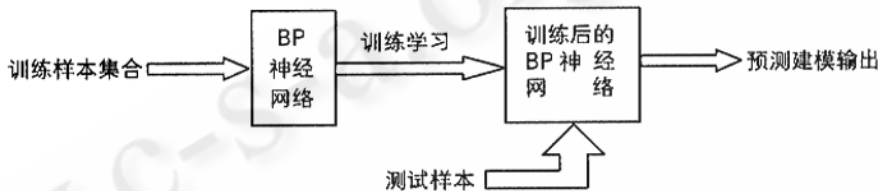


图 2 基于 BP 网络的预测建模框图

### 3.4 遗传操作

(1) 选择算子。计算完各个体的适应度后,选择适应度大的个体遗传到下一代,从而使问题的解越来越接近于最优解空间。目前最常用的是选择算子适应度比例法。在该方法中,各个个体的选择概率和其适应度成比例。对于适应度为  $f_i$  的个体,则其被选中概率为:

$$p_i = f_i / \sum_{i=1}^N f_i \quad (3)$$

其中,  $N$  为群体的规模,在实际应用中,一般将适应度最大的个体无条件地遗传给下一代。

(2) 交叉操作。交叉就是将父代两个权值个体的部分结构加以替换重组而生成新个体的操作。选择合理的交叉算子是本步骤的关键,由于前面编码采用的是实数编码的方式,因此与之对应,交叉方法采用如下方式:

假设要参与交叉的操作的两个基因链分别为  $X_i$  和  $X_j$ , 两个基因链上所对应的基因位分别是  $x_i$  和  $x_j$ , 并且  $X_i$  的适应度高于  $X_j$  的适应度,定义如下所示的两个中间变量

$$\Delta_i^1 = \begin{cases} \min \{ x_i + \frac{1+p_c}{2} (x_i - x_j), x^{\max} \} & x_i \geq x_j \\ \max \{ x_i + \frac{1+p_c}{2} (x_i - x_j), x^{\min} \} & x_i < x_j \end{cases} \quad (4)$$

$$\Delta_j^1 = \begin{cases} \max \{ x_j + \frac{1+p_c}{2} (x_j - x_i), x^{\min} \} & x_j \geq x_i \\ \min \{ x_j + \frac{1+p_c}{2} (x_j - x_i), x^{\max} \} & x_j < x_i \end{cases} \quad (5)$$

其中,  $x^{\max}$  和  $x^{\min}$  分别为  $x_i$  ( $x_j$ ) 的取值上限和下限,  $p_c$  为交叉率,这样,交叉后所产生的两个新个体  $y_i$  和  $y_j$ , 可以用下式确定:

$$y_i = \frac{1+p_c}{2} * \Delta_i^1 + \frac{1-p_c}{2} * \Delta_j^1 \quad (6)$$

$$y_j = ax_i + (1-a)x_j \quad (7)$$

其中  $a$  为  $[0,1]$  之间的随机数,这种交叉算子可以保证产生的两个个体,一个在父代所在的区域搜索,另一个向适应度高的父代个体的方向搜索。

(3) 变异。变异是指模仿生物突变现象使种群中少数个体突变而增加种群的多样性,获得更宽的搜索范围。变异主要是发生在未发生交叉的权值个体中进行。在选择变异区间的时候,对适应度大的个体变异区间取的较小,适应度小的个体的变异区间较大。这样,既能减少变异操作对好的个体的破坏,又能够保证遗传算法的搜索能力。

当采用遗传算法学习 BP 网络权值时,要有效地配合使用交叉和变异算子,在遗传算法中,交叉算子因其全局搜索能力强而作为主算子,变异算子因其搜索能力强而作为辅助算子。一般来讲,交叉概率  $P_c$  要取较大的值,而变异概率  $P_m$  相对要小的多,这样,群体在即将寻到最优值时,不会因变量变异而破坏此进程。

遗传操作完成之后,按适值选取最后一代群体中  $N$  个可能具有全局性的进化解,分别以这些解为初始权值,用 BP 神经网络进行求解,比较  $N$  个由神经网络求得的最优解,从而获得全局最优解。

## 4 应用实例

把本文提出的预测模型应用于某油田的油田措施规划的预测,取该油田 1992~2001 年的历史数据作为训练样本。以连续  $n$  年的数据预测第  $n+1$  年的数据。以生产井数、注水井数、注水量、总措施量作为输入,即可得到产油量的预测输出模型。取神经网络的输入层节点数  $r=4$ ,隐层节点数  $s_1=5$ ,输出层节点数  $s_2=1$ ,得到如下表所示的训练样本:将表 1 种的 10 个输入样本送入 BP 神经网络的输入层,将遗传算法和 BP 算法相结合进行神经网络的训练。取遗传算法初始种群大小为 60,最大遗传代数为 100,交叉概率为 0.9,变异概率为 0.01,初始权值的取值范围为  $-1\sim 1$ ,取神经网络的学习速率为 0.01,最大迭代次数为 5000 次,误差目标值为 0.005,得到隐层和输出层权值。将 98 年~01 年的输入作为已知,98 年~01 年的输出作为预测与实际值比较,分别用 BP 算法和基于遗传算法的 BP 算法进行预测(预测值 1 和误差 1 是用基本的 BP 神经网络进行预测的结果和与实际值产生的误差,预测值 2 和误差 2 是用基于遗传算法的 BP 神经网络进行预测的结果和与实际值产生的误差),训练结果如表 2 所示。由表 2 可以看出,基于遗传算法的 BP 算法的预测效果比纯 BP 算法的预测效果好,相对误差比较低,精度比较高,此预测模型是有效的。

表 1

序号	输入样本				输出样本
	生产井 (口)	注水井 (口)	注水量 ( $10^4$ 吨)	总措施量 (口)	产油量 ( $10^4$ 吨)
1	33	2	40.2	0	52.7
2	35	4	60.7	0	63.9
3	38	8	80.5	3	74.4
4	41	10	90.7	5	90.6
5	38	18	114.2	9	89.7
6	37	19	120.7	12	78.2
7	34	21	130.3	18	70.4
8	28	26	136.7	24	63.3
9	24	30	142.3	27	58.7
10	17	28	147.6	30	50.3

表 2

	1998	1999	2000	2001
实际值	70.4	63.3	58.7	50.3
预测值 1	69.2	64.3	60.8	52.5
误差 1	1.7%	1.58%	3.58%	4.37%
预测值 2	69.8	63.9	59.6	51.7
误差 2	0.85%	0.94%	1.5%	2.78%

## 5 结论与建议

本文提出了一种将遗传算法和神经网络 BP 学习算法相结合的油田措施规划预测模型,用遗传算法良好的全局搜索能力来弥补 BP 神经网络在全局搜索上的缺陷,实例应用证明这种改进的油田措施规划预测模型比一般的预测模型收敛性能好,训练结果精度高。因此,该模型是合理的、可行的、在油田措施规划中有一定的实用价值。

### 参考文献

- 1 王保中、康立山、何巍,基于实数编码遗传算法的多层神经网络 BP 算法[J],武汉大学学报,1998。
- 2 徐丽娜,神经网络控制[M],电子工业出版社,2003。
- 3 吴建生、金龙、农吉夫,遗传算法 BP 神经网络的预报研究和应用[J],数学的实践与认识,2005。
- 4 景广军、梁雪梅、范训礼,遗传神经网络预测模型的设计及应用[J],计算机工程与应用,2001。
- 5 J H Holland, Genetic Algorithms, Scientific American, 1992
- 6 Yao X, Liu Y, A new evolutionary system for evolving artificial neural networks[J], IEEE Trans on Neural Network, 1997
- 7 陈玲、冯其红、于红军, BP 网络方法在油田措施规划中的应用[J],石油勘探与开发,2002。
- 8 田旭光、宋彤、刘宇新,结合遗传算法优化 BP 神经网络的结构和参数[J],计算机应用与软件,2004。