

## 基于遗传算法和 BP 算法的多层感知机杂交训练算法<sup>1</sup>

穆文全 廖晓峰 虞厥邦

(电子科技大学光电子技术系 成都 610054)

**摘 要** 基于梯度下降的神经网络训练算法易于陷入局部最小,从而使网络不能对输入模式进行准确分类。本文提出综合遗传算法和 BP 算法的杂交算法 GA-QP,它结合遗传算法的全局搜索特性和 BP 的局部收敛特性,实现对神经网络的有效训练。实验表明该算法优于 BP 算法,实验结果令人满意。

**关键词** 神经网络,杂交训练算法,遗传算法,QP 算法, BP 算法

**中图分类号** TN-052

### 1 引 言

神经网络已被广泛地应用到信号处理,模式识别等领域, BP 算法是迄今应用最广泛的神经网络训练方法<sup>[1]</sup>。BP 算法通过修改连接权和阈值来逼近任何映射,因而 BP 算法学习过程实际上是一个优化过程。BP 算法及其变种已得到人们深入的研究<sup>[1-4]</sup>,文献 [2] 提出基于共轭梯度的神经网络训练算法, Fahlman 提出 BP 的快速变种 (Quickprop,简称 QP)<sup>[4]</sup>,它们在局部搜索时比较成功。然而,网络训练仍然面临以下问题:(1)网络训练结果对权向量初始化值异常敏感,不同的初始化可能导致截然不同的结果。(2)输出误差函数曲面包含许多远离全局最小点的局部最小点,梯度下降算法可能收敛到这些局部最小而使网络不能对输入模式进行准确分类。避免这类问题的一种方法是一旦训练陷入局部最小,就加入随机量,使其跳出局部最小,并继续训练;然而这种作法不能保证同一个局部最小不会重新出现。(3)输出误差曲面有许多局部最小,即使为寻找一个稍好的局部最小, BP 算法也显得效率低下。模拟退火<sup>[3]</sup>解决某些问题取得一定成功,然而需要大量时间。因此,如何提高网络训练的效率和质量是目前重要的研究问题。文献 [5,6] 提出利用遗传算法 (Genetic Algorithms,简称 GA) 训练前馈神经网络,取得一定成果,然而 GA 较弱的局部搜索能力致使网络训练精度不能太高。对此本文提出了一种神经网络杂交学习算法 GA-QP。该算法把遗传算法与 QP 算法<sup>[4]</sup>结合起来。遗传算法用于全局搜索,使其搜索到全局最优或近似全局最优的附近, QP 用于局部搜索,使其迅速地收敛到最终的优化值。实验表明,该算法训练神经网络优于 BP 算法,网络训练质量和效率大大提高。

### 2 遗传算法

遗传算法是基于自然选择和自然遗传系统的、高度并行和自适应的、搜索和机器学习过程<sup>[7]</sup>。遗传算法同时处理空间的若干点而不像梯度下降算法那样只处理单点,从而有助于搜索全局最优点,而免于陷入局部最小。遗传算法只需要适值函数值 (fitness function value) 而不需要任何辅助信息 (如微分值等)。

<sup>1</sup> 1995-07-18 收到, 1996-01-04 定稿

国家教委博士点专项基金 9361403 号和中科院自动化所国家模式识别重点实验室基金资助

遗传算法包含 5 个基本组成部分: (1) 染色体 (即串) 的编码方法; (2) 适值函数的定义; (3) 群体的初始化; (4) 定义于染色体上的算子; (5) 参数选择。这里, 我们定义遗传算法为 9-元实体

$$GA = (P^0, I, \lambda, L, f, s, c, m, T),$$

其中  $P^0 = (a_1^0, \dots, a_\lambda^0) = \begin{bmatrix} a_{11} \\ \vdots \\ a_{1,L} \end{bmatrix} \cdots \begin{bmatrix} a_{\lambda 1} \\ \vdots \\ a_{\lambda,L} \end{bmatrix} \in I^\lambda$  为初始群体;  $I = \{0, 1\}^L$  为染色体编码;

$\lambda \in N$  为群体个数;  $L \in N$  为染色体长度;  $f: I \rightarrow R$  为适值函数;  $s: I^\lambda \rightarrow I$  为父染色体选择操作;  $c: I^2 \rightarrow I^2$  为交叉操作;  $m: I \rightarrow I$  为变异操作;  $T: I^\lambda \rightarrow \{0, 1\}$  为结束判决。

在一个群体中有  $\lambda$  个染色体, 初始染色体群体  $P^0$  是随机产生的, 实体  $a_k^t$  是在第  $t$  代群体  $P^t$  中第  $k$  个染色体, 染色体  $I$  是  $L$  长的二进制串,  $N$  和  $R$  分别是正整数集合和实数集合, 遗传算法的进化过程 (在  $T=0$  时继续) 直到结束判决满足 ( $T=1$ )。

父染色体选择操作  $s$  一般根据圆盘选择方法 (Roulette wheel selection), 它从群体  $P^t = (a_1^t, \dots, a_\lambda^t)$  中选择染色体作为下一代的父染色体  $a_i^t$ ,  $a_i^t = s(P^t) = a_q^t$ , 而  $q$  则由下式产生:

$$q = \min \left\{ k | \forall k \in \{1, \dots, \lambda\}, a_i \leq \sum_{j=1}^k f(a_j^t) \right\}, \quad (1)$$

其中,  $a_i$  是随机产生实数, 满足  $0 \leq a_i \leq \sum_{j=1}^{\lambda} f(a_j^t)$ 。

交叉操作  $c$  在遗传进化中起了最主要的作用<sup>[7]</sup>。交叉操作分两步: 首先用于交叉的染色体, 根据其适值随机选择出来, 然后, 任意选择一个整数  $q \in [1, L-1]$ , 根据 (2) 式产生子染色体

$$\begin{aligned} \left\{ \begin{array}{l} a_i^t \\ a_{i+1}^t \end{array} \right\} &= c \left( \left\{ \begin{array}{l} a_i^t \\ a_{i+1}^t \end{array} \right\} \right) \\ &= \left\{ \begin{array}{l} [a_{i+1,1}^t, a_{i+1,2}^t, \dots, a_{i+1,q}^t, a_{i,q+1}^t, \dots, a_{i,L}^t]^T \\ [a_{i,1}^t, a_{i,2}^t, \dots, a_{i,q}^t, a_{i+1,q+1}^t, \dots, a_{i+1,L}^t]^T \end{array} \right\}, \\ \forall i &\in \{1, 3, \dots, 2K+1, \dots, \lambda-1\}. \end{aligned} \quad (2)$$

变异操作  $m$  用于弥补遗传中遗传物质丢失, 它从群体  $p^t$  中产生中间群体  $p'^t$ , 其定义如下:

$$a_i'^t = m(a_i^t), \forall i \in (1, \dots, \lambda); \quad (3)$$

$$a_{i,k}' = \begin{cases} a_{i,k}, & \text{for } k \in \{1, 2, \dots, p-1, p+1, \dots, L\}; \\ \bar{a}_{i,k}, & \text{for } k = p. \end{cases} \quad (4)$$

这里  $p$  是根据变异概率得出的, 且  $1 \leq p < L$ 。

标准 GA 中, 按适值比例进行选择操作存在一些问题。一个数学上的问题便是这个过程不能保证渐近收敛到全局最优。问题的实质是对于任何问题, GA 没有收敛保证: 任何代群体中最好的染色体可能丢失, 而又没有任何安全措施使其在后代中得以恢复。我们克服这个问题的途径是采用一种称为精英选择的启发式算法, 按这种算法, 每一代中最好的染色体将无条件保留至下一代, 这种方式保证了渐近收敛。这样, GA 不仅改善了群体的平均适值, 同时也提高了群体中最佳适值。

适值适化 (Fitness Scaling) 是我们用于训练多层感知机 (MLP) 的另一种启发式算法, 目的在于避免早熟收敛和提高进化质量。所谓早熟收敛是指群体中的染色体过早达到这样一种状

态:任何遗传算子不再能产生比其父辈更好的子孙。让我们回顾一下 GA 整个运作过程,在 GA 运作的初期,几个表现突出的染色体和一群平庸的染色体组成整个群体是经常的,如果仅用标准 GA 的选择规则  $p_{\text{select}}(i) = f_i / \sum_j f_j$  (其中  $p_{\text{select}}(i)$  为第  $i$  条染色体的选择概率),表现突出的个体将占据下一代有限个数染色体的大部分,这样将导致早熟收敛,从而易于陷入局部最小,而失去了 GA 的全局收敛特性。另一方面,我们考虑一下 GA 运行的后期,在某一代中可能有各式各样的个体,然而,群体的平均适值可能非常接近于群体中最佳个体的适值,表现一般的个体与表现最佳的个体在下一代中可能获得相同拷贝,这样,用于改善优化质量的适者生存原则,便退化成为一种随机过程。这里,我们采用启发式线性适化方法,对个体适值进行适当的调整。线性适化定义如下:

$$f^L = af + b, \quad (5)$$

其中  $f^L$  是适化后的适值,  $f$  为适化前的适值,  $a$ ,  $b$  根据采取的具体适化措施来确定。总之,启发式适值适化防止表现突出个体在早期占据排他的主导性作用,同时在后期鼓励竞争,提高个体最优解的质量。我们下面所说遗传算法指的均是采用这种启发式策略的遗传算法。

### 3 杂交算法 GA-QP

遗传算法具有极强的全局搜索能力,已广泛地应用于函数优化<sup>[6]</sup>,复杂组合优化<sup>[9,10]</sup>,然而它的局部搜索能力较弱;QP 是一种基于 Newton 方法的 BP 快速变种<sup>[4]</sup>,其计算非常简单,它只利用被修改权值的局部信息:

$$\Delta w(t) = s(t) \cdot \Delta w(t-1) / [s(t-1) - s(t)], \quad (6)$$

其中

$$s(t) = \partial E / \partial w(t). \quad (7)$$

的确,这个新值是对权值的最佳值的较粗的逼近。实验表明,QP 具有极快的收敛速度及极强的局部搜索能力;然而它缺乏全局搜索能力,而且解决平坦点 (flat spot) 时显得力不从心。为此,我们提出 GA-QP 杂交算法,综合上述启发式遗传算法和 QP 的优点而摒弃其缺点。在 GA-QP 中,神经网络的训练分为两步:首先,遗传算法训练神经网络,使其定位于权空间全局最优点或近似全局最优点附近,然后用遗传算法中最佳适值的染色体初始化网络连接权。QP 在此基础上进行局部优化。其算法描述如下:

- 步骤 1 初始化有限个数群体的染色体,并作为当前代  $P(0), t = 0$ 。
- 步骤 2 计算  $p(t)$  中染色体的适值,根据适值选择交配池内的父母染色体。
- 步骤 3 把交叉操作和变异操作根据交叉概率和变异概率应用到父母染色体,并产生下一代  $P(t+1), t = t + 1$ 。
- 步骤 4 如果  $P(t)$  中的染色体不满足终止条件 1, GO 步骤 2, 否则 GO 步骤 5。
- 步骤 5 如果满足终止条件 2, 则结束。否则,用群体中适值最大的染色体初始化网络权值,  $t = 0$ 。
- 步骤 6 按  $\Delta w(t) = -\epsilon s(t)$ 。反向传播并修改网络值。
- 步骤 7 前向传播,  $t = t + 1$ , 计算  $s(t)$ 。
- 步骤 8 如果  $s(t) = 0$ ,  $\Delta w(t) = \alpha \Delta w(t-1)$ , GO 步骤 10, 否则 GO 步骤 9。
- 步骤 9  $\Delta w(t) = s(t) \Delta w(t-1) / [s(t-1) - s(t)]$ 。
- 步骤 10 如果满足终止条件 2, 则结束, 否则 GO 步骤 7。

这里,终止条件 1 指的是遗传算法训练终止的条件,终止条件 2 指整个杂交算法终止的条件。

## 4 实验结果

### 4.1 “异或”问题

异或问题实质就是映射问题:  $(0,0) \rightarrow 0, (0,1) \rightarrow 1, (1,1) \rightarrow 0, (1,0) \rightarrow 1$ 。我们建立了一个 2 个输入结点, 2 个隐含层结点和 1 个输出结点的多层感知机, 其中有 9 个权值需进行调整, 包括阈值。我们参数设置如下: 群体个数为 50, 交叉概率为 0.8, 变异概率为 0.01。我们按二进制编码方法, 13 位二进制码表示一个权变量, 则一条染色体长为  $13 \times 9$  位, 适值函数  $fitness = 1 - \sum_{s=1}^4 (T_s - A_s)^2$ , 其中  $T_s$  为  $s$  模式的理想输出,  $A_s$  为  $s$  模式的实际输出; 终止条件 1 定义为差平方和小于 0.01, 即当差平方和小于 0.01 时, 由 QP 继续进行局部优化; 终止条件 2 定义为差平方和小于 0.0001。QP 参数分别设置为学习率  $\epsilon = 0.7$ , 惯性量  $\alpha = 1.2$ 。由于不同的 CPU 处理速度和浮点计算能力会使同样的算法具有不同的执行时间, 因此, 我们对算法的时间复杂性用迭代次数作为考察的标准。实验表明, 我们的算法优于 QP 算法。图 1(a) 是我们按 GA-QP 和 QP 分别执行 20 次, 取其平均结果而得的收敛图形比较。

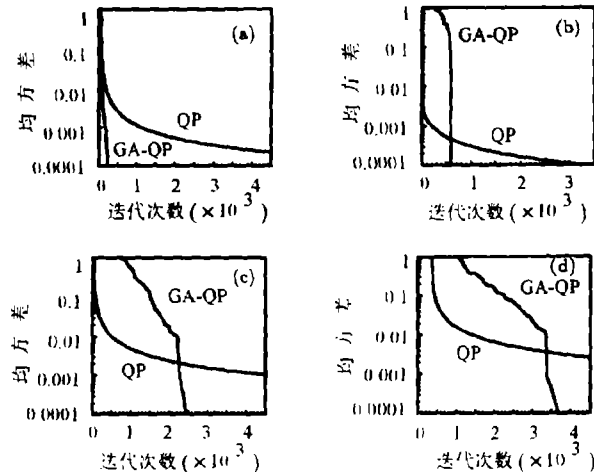


图 1

- (a) “异或”问题基于 GA-QP 和基于 QP 训练的收敛过程  
 (b) 编码解码器 3-2-3 基于 GA-QP 和基于 QP 训练的收敛过程  
 (c) 编码解码器 4-2-4 基于 GA-QP 和基于 QP 训练的收敛过程  
 (d) 编码解码器 10-5-10 基于 GA-QP 和基于 QP 训练的收敛过程

### 4.2 编码解码 (Encoder/Decoder) 问题

编码 / 解码问题是神经网络训练的典型基准点之一。一个  $N-M-N$  编码 / 解码器指的是一个  $N$  个输入层和输出层单元,  $M$  个隐含单元的 3 层神经网络, 给出  $N$  个不同输入模式, 每个模式只有一位为 1, 其余为 0。网络训练是指给定任一模式, 输出值为其输入模式。S.E.Fahlman 认为它是一个比“异或”问题更好的基准点<sup>[4]</sup>, 它在模式识别上具有更好的推广性。

图 1(b), 图 1(c), 图 1(d) 分别是编码 / 解码器 3-2-3、4-2-4、10-5-10 基于 GA-QP 算法和基于 QP 训练的收敛过程比较, 其中 QP 算法的学习率  $\epsilon = 0.8$ , 惯性因子  $\alpha = 1.2$ , 其它参数定义如同“异或”问题。实验表明, 基于遗传算法的神经网络训练优于基于 BP 网络训练。

## 5 结 论

本文提出了基于遗传算法和 QP 方法的杂交神经网络训练方法,并将其应用于异或问题和编码/解码问题,以检验此算法的有效性。由于遗传算法自身的全局收敛特性,使得我们的算法能最大限度地进行搜索,避免了陷入局部最小,同时加上 QP 有效的局部收敛特性,使得此算法迅速进行局部搜索,从而提高了算法的收敛速度。总之,算法的收敛速度比 BP 快,收敛质量和效率大大提高,因此,该算法有望在实际中得到应用。

## 参 考 文 献

- [1] Rumelhart D, McClelland J. Parallel distributed processing: Exploration in the microstructure of cognition. Cambridge: MIT Press, 1986, 423-443, 472-486.
- [2] Madyastha R, et al. An algorithm for training multilayer perceptrons for data classification and function interpolation. IEEE Trans. on CAS, 1994, CAS-41(12): 866-875.
- [3] Szu H. Non convex optimization. Proc. SPIE. San Diego: 1986, Real Time Signal Processing IX, 698, 59-65.
- [4] Fahlman S E. Fast-learning variation on back-propagation: An empirical study. Proceedings of 1988 Connectionist Models Summer School. San Mateo, CA: 1988, Morgan Kaufmann Publishers, 38-51.
- [5] Yoon B, et al. Efficient genetic algorithms for training layered feedforward neural network. Information Science, 1994, 76(1): 67-85.
- [6] Kitano H. Empirical studies on the speed of convergence of neural network training using genetic algorithms. Proc. AAAI-90, 1990, MIT press, 789-795.
- [7] Goldberg D. Genetic Algorithms in Search, Optimization and Machine Learning. Reading: Addison Wesley Publishing Company, 1989, Chapter 1-3.
- [8] DeJong K A. Genetic algorithms: A 10 year perspective. Proceedings of an International Conference on Genetic Algorithms and Their Applications. Pittsburgh: 1985, Lawrence Erlbaum Associates Publishers, 169-177.
- [9] Bac F Q, Perov V L. Optimization problems. Biological Cybernetics, 1993, 69(3): 229-234.
- [10] Anbati B, et al. Heuristic combinatorial optimization by simulated Darwinian evolution: a polynomial time algorithms for traveling salesman problem, Biological cybernetics, 1991, 65(1): 31-35.

## A HYBRID NEURAL NETWORK TRAINING ALGORITHM BASED ON GENETIC ALGORITHM AND BP ALGORITHM

Mu Wenquan    Liao Xiaofeng    Yu Juebang

(University of Electronic Science and Technology of China, Chengdu 610054)

**Abstract** Neural network BP training algorithm based on gradient descend technique may lead to entrapment in local optimum so that the network inaccurately classifies input patterns. This paper presents a hybrid training algorithm GA-QP combined to genetic algorithm with BP algorithm. Experiments show that the hybrid algorithm outperforms BP algorithm. Satisfactory experimental results are obtained.

**Key words** Neural network, Hybrid training algorithm, Genetic algorithm, Quickprop algorithm, BP algorithm

穆文全: 男, 1969年生, 博士生, 电路与系统专业。

廖晓峰: 男, 1964年生, 博士生, 电路与系统专业。

虞厥邦: 男, 1934年生, 教授, 电路与系统专业。