

# 基于样本划分的启发式遗传 BP 算法

文敦伟, 蔡自兴

(中南大学 信息科学与工程学院, 湖南 长沙 410083)

**摘要:** 遗传算法作为一种模拟生物遗传进化过程的随机搜索算法, 具有并行和全局搜索能力、不要求函数可导等特点, 在神经网络学习中得到广泛应用. 合理选择初始群体和控制搜索的盲目性, 有利于提高算法的效率. 为此, 提出了一种新的神经网络学习算法——基于样本划分的启发式遗传 BP 算法. 该方法对神经网络学习样本进行划分, 形成样本子集. 初始群体通过在样本集类上训练神经网络而获得. 这些初始网络中包含了关于样本子集的有用信息, 根据模式定理, 能通过遗传算法保留和加强. 此外, 提出并证明了关于样本集类及其 BP 训练的代数性质, 结合子空间划分进行启发式搜索, 以克服搜索的盲目性. 对上述方法进行仿真实验, 迭代次数和误差较小, 表明本学习算法是切实可行的.

**关键词:** 神经网络; 样本; 学习算法; 遗传算法

**中图分类号:** TP183

**文献标识码:** A

**文章编号:** 1005-9792(2002)01-0092-04

BP 算法是一种重要的神经网络学习算法, 在智能计算和智能控制中得到广泛应用<sup>[1,2]</sup>, 但其存在收敛速度慢, 易陷入局部极小点等缺点, 一般采用进化算法和遗传算法作为神经网络的学习方法<sup>[3-5]</sup>. 遗传算法作为一种模拟生物遗传进化过程的随机搜索算法, 具有并行搜索, 不要求函数可导, 特别是易于获得全局最优解的特点. 遗传算法可以用来优化网络的权值, 实现神经网络的全局优化学习. 用遗传算法优化神经网络连接权的步骤包括编码、交叉变异、选择等<sup>[6,7]</sup>.

用遗传算法优化神经网络, 不需要误差函数的梯度信息, 不需考虑误差函数是否可微, 根据应用要求在该误差函数中增加某些惩罚项, 使网络具有某些特殊的形态, 如连接权值总和约束下的网络训练等. 特别是利用随机全局寻优特点可以避免出现局部极值. 此外, GA 和 BP 优化神经网络各有优缺点. 例如它们都对训练过程中的算法参数敏感, GA 擅长全局搜索, 而 BP 擅长局部搜索. 为进一步结合 GA 和 BP 各自优点, 将两者以一定的方式结合, 形成遗传 BP 算法, 既能克服寻优中的盲目性, 又可避免出现局部收敛情况<sup>[7]</sup>. 在此, 作者结合遗传算法与 BP

算法的特点, 提出一种新的算法——基于样本划分的启发式遗传 BP 算法, 并对该算法进行仿真研究.

## 1 样本集 $S$ 的一个集类及特性

设  $S$  为神经网络训练的样本集. 按问题规模对样本集  $S$  进行划分, 将  $S$  划分为  $n$  个互不相交的子集  $S_1, S_2, \dots, S_n$ .

以  $S$  的划分  $S_1, S_2, \dots, S_n$  为基础, 按如下方法构造集类  $J$ :

$$\begin{aligned} & \emptyset \in J; \\ & S_1 \in J, S_2 \in J, \dots, S_n \in J; \\ & S_1 \cup S_2 \in J, S_1 \cup S_3 \in J, \dots, S_1 \cup S_n \in J; \\ & S_2 \cup S_3 \in J, \dots, S_2 \cup S_n \in J; \\ & \dots \\ & S_1 \cup S_2 \cup S_3 \in J, S_1 \cup S_2 \cup S_4 \in J, \dots, S_1 \cup S_2 \cup S_n \in J; \\ & \dots \\ & S_1 \cup S_2 \cup S_3 \dots \cup S_n \in J. \end{aligned}$$

**性质 1**  $\langle S, J \rangle$  是一个拓扑空间.

**证明** 由上述构造过程, 有以下 3 种情况:

a.  $\emptyset \in J, S \in J$ .

收稿日期: 2001-07-08

基金项目: 国家自然科学基金资助项目(69974043); 国家教育部博士点基金资助项目(99053317); 湖南省自然科学基金资助项目(99JJY20062)

作者简介: 文敦伟(1965-), 男, 湖南华容人, 中南大学副教授, 博士, 从事分布式人工智能、智能控制、计算机应用等研究.

b. 设  $J = (S_1, S_2, \dots, S_n)$ , 对任意  $m$ , 若  $S_1 \in J, S_2 \in J, \dots, S_m \in J$ , 则其交集对任意  $m$ , 显然有

$$\bigcap_{i=1}^m S_i \in J.$$

c. 对  $S_1 \in J, S_2 \in J, \dots, S_n \in J$ , 其并集为

$$\bigcup_{i=1}^{\infty} S_i \in J.$$

故  $\langle S, J \rangle$  是一个拓扑空间.

**性质 2** 对样本集  $S$  上的集类  $J$  上的集合运算  $\cup, \langle J, \cup \rangle$  是独异点.

**证明**

a. 设  $S_i \in J, S_j \in J$ , 由  $J$  的构造方式, 显然有  $S_i \cup S_j \in J$ .

故  $J$  是一个群胚.

b. 对任意  $S_i \in J, S_j \in J, S_k \in J$ , 显然有  $S_i \cup (S_j \cup S_k) = (S_i \cup S_j) \cup S_k$ .

故  $J$  是一个半群.

c. 对任意  $S_i \in J$ , 因

$$S_i \cup \emptyset = \emptyset \cup S_i = S_i, S_i \cup S = S \cup S_i = S,$$

故  $J$  有零元  $S$  和单位元  $\emptyset$ .

可见, 对构造的  $S$  的集类  $J, \langle J, \cup \rangle$  是一个独异点.

显然, 因为  $\langle J, \cup \rangle$  中  $S_i \in J$  关于单位元  $\emptyset$  无逆元, 故  $\langle J, \cup \rangle$  不是一个群,  $J$  的元素个数为

$$C_n^n + C_n^{n-1} + C_n^{n-2} + \dots + C_n^1 + C_n^0 = 2^n.$$

上述性质相关的基本代数概念参见文献[8].

## 2 $\langle J, \cup \rangle$ 上的子样本 BP 训练

**定义 1** 对具有一定结构的神经网络 NN 的子集  $N$ , 令  $N_i \in N$ , 称  $N_i$  满足 (Satisfy)  $S_i$ , 若用  $S_i \in J$  进行自适应变步长 BP 训练  $N_i$  后能达到预期的精度, 记为  $N_i = \text{BP}^+(S_i)$ .

**定义 2** 在由上述  $N_i$  构成的集合  $N$  上定义二元运算“ $+_N$ ”,  $N_i +_N N_j$  为功能上包容  $N_i$  和  $N_j$  的  $N$  上的神经网络, 若  $S_i = \emptyset$ , 令  $N_i = \text{BP}^+(S_i)$  为空集  $\emptyset$ .

根据神经网络逼近定理, 在一般前向多层网络结构下, 满足任意  $S_i \in J$  的  $N_i$  是成立的.

**性质 3**  $\langle N, +_N \rangle$  是 1 个独异点.

**证明** 显然  $\langle N, +_N \rangle$  满足可结合性, 其单位元为满足  $S$  的网络.

**性质 4**  $\text{BP}^+$  是  $\langle J, \cup \rangle$  到  $\langle N, +_N \rangle$  的同态映射.

**证明** 根据上述定义, 有

$$\text{BP}^+(S_i \cup S_j) = \text{BP}^+(S_i) +_N \text{BP}^+(S_j) = N_i +_N N_j.$$

## 3 启发式子样本遗传算法

由于遗传算法是一个随机寻优算法, 这种随机性表现在初始个体的变量随机发生、交叉、变异以及选择(或复制)等. 随机性保证了 GA 能在全局范围内进行搜索. 同时, 随机中又包含了适应值的约束, 从而保证其收敛到最优值. 随机性除具有上述优点外, 又不可避免地存在一定的盲目性. 搜索的盲目性显然增加了 GA 算法耗费的时间. 针对这种盲目性, 对算法进行如下改进:

a. 采用 BP 算法, 按样本子集训练得到初始网络. 显然, 这些初始网络中包含了一些有用的信息, 根据模式定理, 这些有用信息能通过 GA 得到保留和加强.

b. 有目的地引导 GA 探索尚未访问的权值子空间. 这种方法与盲目、随机所产生的新网络相结合, 提高了 GA 搜索效率, 加速了收敛速度. 具体地说, 就是在 GA 的每一个新代中, 在某些子空间内随机产生网络, 与在全范围内产生的网络一道作为新的移民加入新一代.

按照上述思想, 并采用文献[6]中使用的符号, 算法可以表示为:

a. 初始化. 在权值允许的全范围内随机产生  $k_2$  个网络, 连同  $k_1$  个经过样本子集  $\text{BP}^+$  训练的网络一起经过编码后形成初始群体.

b. 度量. 对适应度函数  $\phi(a_k(t)) = \delta(f(\varphi(a_k(t))))$ ,  $\phi$  按网络的误差来确定. 其中:  $\phi(a_k(t))$  为权值  $w_t$  (含阈值), 取  $f(w_t)$  为  $E(w_t, S)$ , 代表网络的误差;  $S$  为样本集;  $\delta(\cdot)$  为比例系数.

c. 搜索. 设神经网络共有  $m$  个权值. 将每个权分为  $d$  个区域, 则总的子空间数为  $d^m$ , 子空间数规模选取要根据网络权值的大小并尽可能根据非线性空间的粒度来选取, 适当控制子空间规模. 使用 1 个  $d^m$  个元素的向量表示子空间状态, 初始化为 0. 若权值在某个子空间, 则标注相应位置为 1, 对于每 1 个新代的  $\lambda$  个个体中,  $\lambda_1$  为上一代的最优个体,  $\lambda_2$  为在某搜索子空间中产生的个体,  $\lambda_3$  为经过上一代交叉变异和选择后产生的个体,  $\lambda = \lambda_1 + \lambda_2 + \lambda_3$ .

算法描述如下:

$$t = 0.$$

初始化:  $N_S = \{N_1(0), N_2(0), \dots, N_{k_1}(0)\}$ ,  
 $N_i = \text{BP}^+(S_i), S_i \in J$ ;  
 $p_{k_1}(0) = \varphi^{-1}(N_S)$  ( $\varphi^{-1}$  为编码);  
 $p_{k_2}(0) = \{a_{k_1+1}(0), \dots, a_{k_1+k_2}(0)\} \in I^{k_2}$ ,  
 $I = \{0, 1\}^l$ ;  
 $p(0) = p_{k_1}(0) \cup p_{k_2}(0), k_1 + k_2 = \mu$ .  
 度量:  $p(0) = \{\phi(a_1(0)), \dots, \phi(a_\mu(0))\}$ ,  
 $\phi(a_k(0)) = \delta(E(\varphi(a_k(0))))$ ;  
 while ( $\wedge(p(t)) \neq T$ ) do.  
 交叉:  $a'_k(t) = r_{(p'_j)}(p(t))$ ,  
 $\forall k \in \{1, 2, \dots, \mu\}$ .  
 变异:  $a''_k(t) = m_{(p'_m)}(a'_k(t))$ ,  
 $\forall k \in \{1, 2, \dots, \mu\}$ .  
 度量:  $p''(t) = \{a''_1(t), \dots, a''_\mu(t)\} :$   
 $\{\phi(a''_1(t)), \dots, \phi(a''_\mu(t))\}$ ;  
 $\phi(a''_k(t)) = \delta(E(\varphi(a''_k(t))))$ .  
 搜索:  $p_{\lambda_2}(t) = \{a_{e1}(0), \dots, a_{e\lambda_2}(0)\}$  ( $a_{ei}(t)$  为  
 在某个未搜索子空间中随机产生的个体).  
 选择:  $\{p_{\lambda_1}(t), p_{\lambda_3}(t)\} = S(p''(t))$ .  
 产生新代:  $p(t+1) = p_{\lambda_1}(t) \cup p_{\lambda_2}(t) \cup p_{\lambda_3}(t)$ .  
 $t = t + 1$ ,  
 end.  
 选择:  $a_k(t) = S(p''(t)), k \in \{1, 2, \dots, \mu\}$ .  
 BP 算法:  $\text{BP}^+(S, \varphi(a_k(t))) \rightarrow \text{NN}$  (由  $\text{BP}^+$  产  
 生最后的神经网络).

### 4 仿真实例

选定下面的非线性系统为仿真对象<sup>[9]</sup>:

$$y = 0.1e^{(2x_1 - 0.5)^2} + 0.1e^{(2x_2 - 0.5)^2} + \sin(2\pi x_1) - \cos(2\pi x_2) - (x_1 - 0.75)^2 - (x_2 - 0.25)^2.$$

$x_1, x_2 \in [0, 1]$ , 其图形如图 1 所示.

采用上述样本划分与启发式遗传 BP 算法, 使用 Matlab5.3 环境编程和 GA 算法<sup>[10]</sup> 进行编程和仿真. 选用的神经网络结构为(2-5-1), 群体规模为 15, 其中按照 BP 算法初始化的神经网络个数为 8, 子样本集合个数为 4, 在按样本训练子集时, 共有 15 个网络获得初步训练(迭代次数为 3 000), 从中获得最好的 4 个网络和对子样本最有代表性的 4 个网络. 其余的神经网络随机确定. 在子空间搜索时, 取  $d = 2$ ,  $\lambda_1 \lambda_2 \lambda_3$  分别取为 2-8-5. 采用实数编码、启发式交叉

和非正态变异方案进行 2 000 代搜索后, 获得的神经网络训练结果如图 2 所示. 可见, 训练所得的误差较小, 表明算法切实可行.

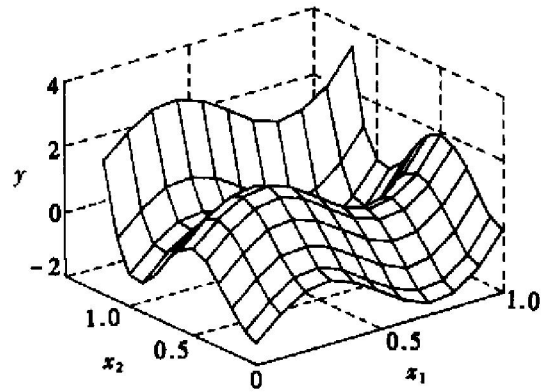
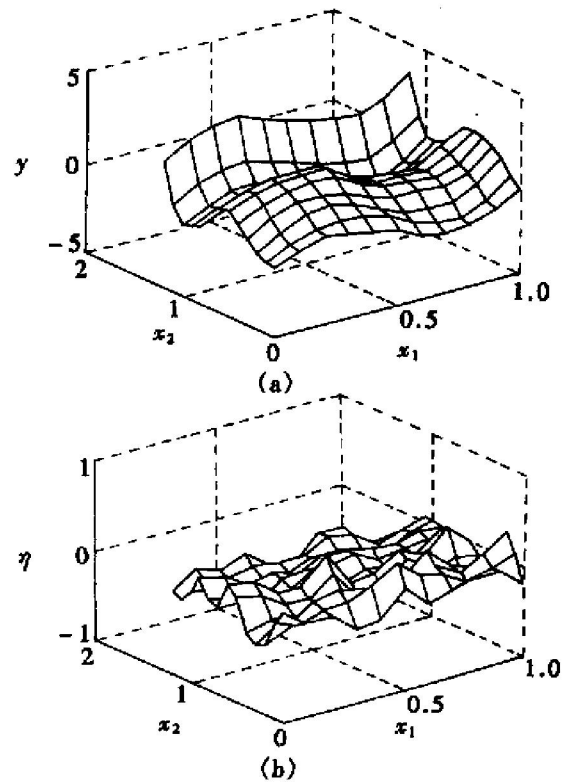


图 1 非线性系统仿真图形



(a) — 网络映射; (b) — 网络误差  $\eta$

图 2 神经网络训练结果

### 5 结 论

- a. 样本子集与受训的神经网络之间存在同态映射关系.
- b. 按照样本划分训练神经网络获得的初始种群, 能较好地反应样本中的有用信息.
- c. 对设计的算法进行仿真实验, 搜索次数较少、误差较小, 表明了算法是切实可行的.

## 参考文献:

- [1] 蔡自兴. 智能控制——基础与应用[M]. 北京: 国防工业出版社, 1998.
- [2] 王耀南. 智能控制系统——模糊逻辑·专家系统·神经网络控制[M]. 长沙: 湖南大学出版社, 1996.
- [3] Park D. Genetic-based new fuzzy reasoning models with application to fuzzy control[J]. IEEE Trans on System, Man and Cybernetics, 1994, 24(1): 39-47.
- [4] Angeline P J. An evolutionary algorithm that constructs recurrent neural network[J]. IEEE Trans on Neural Networks, 1994, 5(1): 54-65.
- [5] YAO Xin, LIU Yong. A new evolutionary systems for evolving artificial neural networks[J]. IEEE Trans on Neural Networks, 1997, 8(3): 694-713.
- [6] 陈国良, 王煦法, 庄镇泉, 等. 遗传算法及其应用[J]. 北京: 人民邮电出版社, 1996.
- [7] 陈荣, 徐用懋, 兰鸿森. 多层前向网络的研究——遗传 BP 算法和结构优化策略[J]. 自动化学报, 1997, 23(1): 43-49.
- [8] 王兵山. 离散数学[M]. 长沙: 国防科技大学出版社, 1985.
- [9] 章云. 非线性系统的神经网络建模与控制[D]. 广州: 华南理工大学电子科学系, 1997.
- [10] Houck C R, Joines J A, Kay M G. Comparison of genetic algorithms, random restart and twoopt switching for solving large location allocation problems[J]. Computers & Operations Research, 1996, 23(6): 587-596.

## Sample division based heuristic genetic BP algorithm

WEN Dunwei, CAI Zixing

(College of Information Science and Engineering, Central South University, Changsha 410083, China)

**Abstract:** As a random search method mimicking heredity and evolution of creature, genetic algorithm has an ability of parallel and global search without the function derivable. It has been widely used in neural network learning. To get high efficiency of the algorithm, the initial population should be chosen properly and the blindness of search should be controlled. In this paper, a novel neural network learning algorithm, sample division based heuristic genetic BP algorithm was proposed. In this algorithm, by dividing the samples into subsets, the power set was formed and the initial population was obtained by training neural networks with BP on the powerset. Using schema theory, useful information about samples, which may be contained in those initial neural networks, can be strengthened and remained. Several algebraic properties of sample sets and their BP training are proposed and proved. Heuristic search is also performed around divided subspaces. The result of simulation shows the validation of this algorithm.

**Key words:** neural networks; sample; learning algorithm; genetic algorithm