

免疫进化子波网络及其学习算法

王 磊, 焦李成, 刘 芳, 张艳宁

(西安电子科技大学雷达信号处理国家重点实验室, 陕西西安 710071)

摘 要: 本文首次对子波神经网络的分类机理进行了详细地研究, 在此基础上将免疫进化算法与子波神经网络相集成, 并提出了免疫进化子波网络模型及其学习算法. 其中, 免疫进化算法是一种集免疫机制和进化机制于一体的新的全局并行算法, 它可以通过对进化环境的自适应和自学习, 有针对性地抑制由原进化算子操作的盲目性而引起的退化现象. 理论分析和用于双螺旋线分类的仿真结果表明, 免疫进化子波网络不仅是可行的, 也是十分有效的. 由于免疫进化算法本身所固有的并行运算规则、智能搜索方式和概率判断准则, 从而有利于弱化子波网络的应用条件, 加强其联想记忆和信息处理的能力.

关键词: 进化算法; 人工神经网络; 子波分析; 免疫; 双螺旋线

中图分类号: TN957 文献标识码: A 文章编号: 0372-2112(2001)12A-1878-08

An Immune Evolutionary Wavelet Neural Network and Its Learning Algorithm

WANG Lei, JIAO Li cheng, LIU Fang, ZHANG Yarr ning

(Key Lab for Radar Signal Processing, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: The classifiable function of the wavelet neural network (WNN) is studied in this paper, and an integration of the immune evolutionary algorithm and WNN is proposed, based on the model of the immune evolutionary wavelet neural network (IEWNN). In this model, the immune evolutionary algorithm (IEA) used for training the network is a globally parallel algorithm which integrates the immune and the evolutionary mechanisms. The IEA is conducive to alleviating the undulating phenomenon produced by the blind fold behaviors of the original operators in the existent evolutionary algorithms through adaptively learning form the evolutionary environment. The analysis in theory and simulations show that WNN based on IEA is not only feasible but also effective. Because IEA has the parallel operation rules, the intelligent searching behavior and the criterion based on probability, it is conducive to relaxing the conditions of applying WNN and improving the ability of association memory and information processing.

Key words: evolutionary algorithm; artificial neural network; wavelet; immunity; twir spiral problem

1 引言

众所周知, 人们最初提出人工神经网络模型(ANN)是基于对生物神经处理系统的观察, 力图用一种网络结构去模仿并实现生物系统的神经信息处理能力. 随着这种模型在理论与应用研究方面的不断深入与发展, 人们逐渐认识到该网络模型中隐层节点函数的选取与网络连接权值的学习算法, 是影响网络的逼近性能和应用便利性的两个重要方面, 所以, 也是我们进一步研究和改进 ANN 性能所要针对的两个主要对象^[1-3].

在对隐层节点函数的选取方面, 我们注意到近年来由信号处理领域发展起来的子波分析理论是由一族函数去表示或逼近一个信号或函数. 其中, 这一族函数由一个基本子波函数经平移或伸缩构成, 从而也称子波函数系, 并用其变换系数来描述原始信号. 子波函数系表示的突出特点在于时宽乘积很小, 且在时间和频率轴上都集中. 子波变换具有可调的时频

窗口, 具有放大、缩小和平移等功能, 故具有“数学显微镜”的功能^[4-6]. 基于这种认识, 人们对原有前向神经网络逼近连续函数的研究过程中提出了子波神经网络的一些模型, 其中包括: 时间一尺度子波神经网络、紧支撑正交子波神经网络以及自适应斜交子波神经网络等. 这些模型基本上是将子波变换与前向神经网络相结合, 同时将子波参数做为网络参数的一部分, 通过网络学习与修正使误差能量函数最小化, 以达到最佳逼近和分类的目的. 由此又引出了学习算法的问题.

在进行网络学习并对其参数进行修正的算法方面, 我们不妨以自适应子波网络为例, 该模型能够对目标进行特征提取与分类, 但其子波参数的选择需要依靠操作人员的经验或尝试, 而且训练过程易于陷入局部极小, 收敛速度也相对较慢. 有鉴于此, 人们提出了着眼于全局优化的遗传算法来克服上述算法所存在的固有缺陷. 众所周知, 以遗传算法为代表的进化算法都是一种具有“生成+检测”的迭代搜索算法, 具有

广泛的适应性和鲁棒性^[7-9]. 这些算法以体现群体搜索和群体中个体之间信息交换的两大策略的交叉和变异算子(进化规划只强调变异算子),为每个个体提供了优化的机会,从而使整个群体在优胜劣汰的选择机制下保证了进化的趋势.在对算法的实施过程中不难发现上述的两个遗传算子都是在一定发生概率的条件下,随机地、没有指导地迭代搜索,因此它们在为群体中的个体提供了进化机会的同时,也无可避免地产生了退化的可能.在某些情况下,这种退化现象还相当明显.另一方面,每一个待求的实际问题都会有自身一些基本的、显而易见的特征信息或知识.然而遗传算法的交叉和变异算子却相对固定,在求解问题时,可变的灵活程度较小.这无疑对算法的通用性是有益的,但却忽视了问题的特征信息对求解问题时的帮助作用,特别是在求解一些复杂问题时,这种“忽视”所带来的损失往往就比较明显了.

针对以上问题,人们将生物学中免疫概念及其理论应用于进化算法,在保留原算法优良特性的前提下,力图有选择、有目的地利用待求问题中的一些特征信息或知识来抑制其优化过程中出现的退化现象,这些算法即称为基于免疫策略的进化算法(Evolutionary algorithms based on immune strategy)^[10,11].这些算法利用局部特征信息以一定的强度干预全局并行的搜索进程,抑制或避免求解过程中的一些重复和无效的工作,以克服原进化策略算法中交叉和变异算子操作的盲目性.算法在执行时,可以有针对性地抑制群体进化过程中出现的一些退化现象,理论分析和仿真计算都表明免疫进化算法不仅是有效的而且也是可行的,并较好地解决了原进化算法中的退化问题.

基于以上对 ANN 模型中隐层节点函数的选取与网络连接权值的学习算法这两方面的考虑,我们在分析子波神经网络内在机理的基础上,提出了采用免疫进化算法对子波网络进行学习训练模型与方法.由于免疫进化算法本身所固有的并行运算规则、智能搜索方式和概率判断准则,从而有利于弱化子波网络的应用条件,加强其联想记忆和信息处理的能力.

2 子波神经网络模型的机理分析

子波神经网络与一般的人工神经网络模型在结构上具有很大的相似性,例如,一个典型的三层网络结构如下图 1 所示^[12].

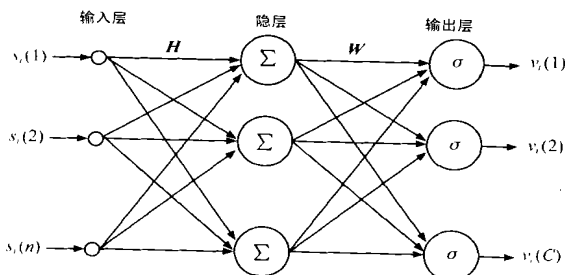


图 1 子波神经网络模型

在上面的结构中,我们设输入层神经元的个数为 n ; 隐层

神经元个数为 N ; 输出层神经元个数为 C ; 训练样本集的容量为 M ; 输入信号 S 的矩阵维数 $\dim S = n \times M$; 输出信号 V 的矩阵维数 $\dim V = C \times M$; 子波 H 的矩阵维数 $\dim H = N \times n$; 子波伸缩、平移参数矩阵 A, B 的维数 $\dim A = \dim B = N \times 1$; 输出层连接权值 W 的矩阵维数 $\dim W = C \times N$; 输出层可调阈值 θ 的矩阵维数 $\dim \theta = C \times 1$. 对于多输入多输出函数 $f: R^n \rightarrow R^m$, 上述网络方程可表示成如下形式:

$$v_i = \sigma(u_i) = \sigma \left[\sum_{k=1}^N W_k \sum_{t=1}^n s_t(t) h \left(\frac{t - b_k}{a_k} \right) \right] \quad (1)$$

我们设目标集合为: $\{\omega_1, \omega_2, \dots, \omega_c\}$, 且各类目标互不相容. 因为每个输出单元均可以看成是一个二类识别问题, 故对第 i 维而言: 若目标为 ω_i , 则 $P(\omega_i) = P_i \approx n_i/M$, 其中 $P(\omega_i)$ 表示目标出现的概率, 且有: $\sum_{i=1}^C P_i = 1$, n_i 表示目标为 ω_i 类的学习样本集的容量; 若目标不为 ω_i (此时记为 $\bar{\omega}_i$), 则 $P(\bar{\omega}_i) = 1 - P_i \approx (M - n_i)/M$, 其中 $P(\bar{\omega}_i)$ 表示目标非 ω_i 类出现的概率.

对于输出层单元:

$$V = \sigma(O)$$

其中 σ 为 Sigmoid 函数, O 为实际输出矩阵. 由于 $\sigma(\cdot)$ 是单调可逆的, 所以 V 与 O 之间存在确定的一一对应关系. 设 D 为理想输出, 则:

$$\dim O = \dim D = C \times M$$

其中

$$b_k = \begin{cases} b_1, & S_k \in \omega_i \\ -b_2, & S_k \in \bar{\omega}_i \end{cases} \quad (2)$$

式中, S_k 为 S 矩阵中第 k 列列矢量, 即网络输出的第 k 个学习样本. b_1, b_2 均为常数, 且 $b_1 > 0, b_2 > 0$.

我们一般定义网络训练的目标函数为实际输出的均方误差, 即:

$$\min: E = \frac{1}{MC} \sum_{k=1}^M \sum_{j=1}^C (d_{jk} - O_{jk})^2$$

式中, O_{jk} 表示第 k 个学习样本的输出矢量 O_k 的第 j 个分量, d_{jk} 代表第 k 个学习样本的期望输出矢量 d_k 的第 j 个分量. E 的矩阵矢量形式如下所示:

$$E = \frac{1}{MC} \sum_{k=1}^M \|d_{jk} - O_{jk}\|^2 = \frac{1}{MC} \|D - O\|^2 \quad (3)$$

由于

$$O = WH = \theta I_M^T$$

其中 $H = HS$ 为子波特征矢量矩阵, $\dim H = N \times M$; I_M 为每个分量为 1 的 M 维列矢量.

将上式代入式(3)得:

$$E = \frac{1}{MC} \|D - WH - \theta I_M^T\|^2 \quad (4)$$

我们首先通过对输出层神经元的分析, 来讨论网络调节 W, θ 以及子波参数 A, B , 以使误差 E 达到最小.

由式(4)可得:

$$E = \frac{1}{MC} \text{tr}[(D - WH - \theta I_M^T)(D - WH - \theta I_M^T)^T]$$

其中, $\text{tr}(\cdot)$ 表示矩阵求迹运算. 另外, 对 θ 求极小化 E , 令 ∇E

$(\theta) = 0$, 得:

$$\theta = \frac{1}{M}DI_M - W \left(\frac{1}{M}HI_M \right) \quad (5)$$

令

$$m_d = \frac{1}{M}DI_M = \frac{1}{M} \sum_{k=1}^M d_k \quad (6)$$

$$m_h = \frac{1}{M}HI_M = \frac{1}{M} \sum_{k=1}^M h_k \quad (7)$$

上式中的 h_k 为 H 中第 k 个列矢量, 即第 k 个学习样本的子波特征矢量. 显然, m_d 为平均期望输出矢量, m_h 为特征提取层子波特征平均矢量. 将式(6)和式(7)代入式(5)可得:

$$\theta = m_d - Wm_h \quad (8)$$

再将式(8)代入式(4)得:

$$E = \frac{1}{MC} \| (D - m_d I_M) - W(H - m_h I_M) \|^2 \quad (9)$$

我们这里定义:

$$\bar{D} = D - m_d I_M \quad (10)$$

$$\bar{H} = H - m_h I_M \quad (11)$$

将式(10)和式(11)代入式(9)即得:

$$E = \frac{1}{MC} \| \bar{D} - W\bar{H} \|^2 \quad (12)$$

对 W 极小化 E , 令 $\nabla E(W) = 0$, 则:

$$W = \bar{D}\bar{H}^+ \quad (13)$$

其中, \bar{H}^+ 表示矩阵 \bar{H} 的 Moore Penrose 伪逆.

将式(13)代入式(12), 并利用伪逆的性质可得:

$$\begin{aligned} E &= \frac{1}{MC} \| \bar{D} - \bar{D}\bar{H}^+ \|^2 = \frac{1}{MC} \text{tr} [(\bar{D} - \bar{D}\bar{H}^+) (\bar{D} - \bar{D}\bar{H}^+)^T] \\ &= \frac{1}{MC} \text{tr} [\bar{D}\bar{D}^T - \bar{D}(\bar{H}^+ \bar{H})^T \bar{D}^T - (\bar{D}\bar{H}^+) \bar{H}\bar{D}^T \\ &\quad + (\bar{D}\bar{H}^+) \bar{H}\bar{H}^+ (\bar{D}\bar{H}^+)^T] \\ &= \frac{1}{MC} \text{tr} [\bar{D}\bar{D}^T - \bar{D}\bar{H}^T (\bar{H}\bar{H}^T)^+ \bar{H}\bar{D}^T] \\ &= \frac{1}{MC} \{ \text{tr}(\bar{D}\bar{D}^T) - \text{tr}[\bar{D}\bar{H}^T (\bar{H}\bar{H}^T)^+ \bar{H}\bar{D}^T] \} \end{aligned} \quad (14)$$

由前面的式(2)可得:

$$\frac{1}{M} \text{tr}(\bar{D}\bar{D}^T) = (b_1 + b_2)^2 \sum_{j=1}^C P_j (1 - P_j) = (b_1 + b_2)^2 \left[1 - \sum_{j=1}^C P_j^2 \right] \quad (15)$$

将上式代入式(14)得:

$$\begin{aligned} E &= (b_1 + b_2)^2 \left\{ \frac{1 - \sum_{j=1}^C P_j^2}{C} - \frac{1}{(b_1 + b_2)^2} \text{tr} [\bar{D}\bar{H}^T (\bar{H}\bar{H}^T)^+ \bar{H}\bar{D}^T] \right\} \\ &= (b_1 + b_2)^2 \left\{ \frac{1 - \sum_{j=1}^C P_j^2}{C} - \frac{1}{(b_1 + b_2)^2} \text{tr} [\bar{H}\bar{H}^T (\bar{H}\bar{D}^T)^+ \bar{D}\bar{H}^T] \right\} \end{aligned} \quad (16)$$

由于 P_j 为常数, 故对 θ 、 W 按最小平方误差准则极小化误差 E 等价于极大化由下面定义的可分性测度函数.

我们定义可分性测度函数 J 为如下所示的形式:

$$J = \text{tr} [S_T^+ S_B] \quad (17)$$

其中

$$S_T = \frac{1}{M} \overline{HH}^T = \frac{1}{M} \sum_{k=1}^M (\bar{h}_k - m_h) (\bar{h}_k - m_h)^T$$

$$S_B = \frac{1}{(b_1 + b_2)^2} \frac{1}{M^2 C} \overline{HD}^T \bar{D}\bar{H}^T = \sum_{j=1}^C P_j^2 (m_h^j - m_h) (m_h^j - m_h)^T$$

上式中的 m_h^j 为第 j 类学习样在网络中获得的子波特征平均矢量. 将上面两式和式(17)一起代入式(16)可得权值点 θ 、 W 处的输出均方误差值为:

$$\begin{aligned} E_{\min|\theta=0, W=W} &= (b_1 + b_2)^2 \left[\frac{1 - \sum_{j=1}^C P_j^2}{C} - \text{tr} (S_T^+ S_B) \right] \\ &= (b_1 + b_2)^2 \left[\frac{1 - \sum_{j=1}^C P_j^2}{C} - J \right] \end{aligned} \quad (18)$$

上式中的 S_T 可以看成是在学习样本集基础上的, 对定义在网络特征提取层子波特征空间内的样本混合散布矩阵的一个估计, 其物理意义可以理解为样本在子波特征空间的聚类程度; S_B 可以看成是在学习样本集基础上对定义在网络特征提取层子波特征空间内的加权的样本类间散布矩阵的一个估计, 而且加权的权重正比于各类的先验概率. 其物理意义可理解为样本在子波特征空间的类间散布.

由以上分析可知, J 反映了在特征空间任一目标类区统计中心点, 与其它目标总的类区中心之间距离的平方, 与所有类区构成的总的“云团”的半径的平方之间的最大比率统计均值. 其中, “云团”是指特征矢量顶端点在空间的分布所构成的图象. 因为 J 是一个比率统计均值, 所以显然有: $0 \leq J \leq 1$. 其中, 当 $J = 0$ 时, 各类区基本重合, 这说明网络特征提取层的子波特征类间相隔太近; 当 $J = 1$ 时, 则表明子波特征类间散布最大.

另一方面, 可分性测度函数 J 与学习样本集定义在子波神经网络分类器的特征提取层上的子波特征有关, 而与输出层的权重 W 和阈值 θ 无关. 寻求分类均方误差最小的过程, 对于输出层来说, 就是寻求学习样本集定义在子波神经网络分类器特征提取层上的子波特征最大可分性的过程. 这里需要指出的是: 输出层并不能改变学习样本集定义在特征提取层子波空间内的可分性.

接下来, 我们分析网络特征提取层的功能. 对 A 、 B 极小化 E (式(4)) 且满足对 W 极小化 E (式(4)) 等价于对 A 、 B 极小化 $E_{\min|\theta=0, W=W}$ (式(4)). 因此, 对 A 、 B 极小化 E 就等价于对 A 、 B 极大化 J :

$$\max J(A, B) = \text{tr} [S_T^+ S_B]$$

求解上述优化问题得 $J_{\max} \geq J$ (当 A 、 B 初始值为极值点时等号成立). 再由式(2)得:

$$E_{\min} = (b_1 + b_2)^2 \left[\frac{1 - \sum_{j=1}^C P_j^2}{C} - J_{\max} \right]$$

再由式(3)可以看出, 寻求可分性测度函数最大化的过程, 即为寻求极值点 A 、 B , 使学习样本集定义在由 A 、 B 所确定的一组子波基(即网络的特征提取层)上的子波特征空间内的可分性最大化. 与输出层作用不同, 特征提取层通过参数 A 、 B 的调节从根本上改变了 J 的大小. 进一步讲, 寻求可分性测度函数最大化的过程也就是寻求学习样本类别可分性最大的子波

特征的过程.

当 $J_{\max} = \frac{1}{C} \left(1 - \sum_{j=1}^C P_j^2 \right)$ 时, 可分最小平方误差可以下降到 0.

通过以上分析, 我们可以得出以下两点:

(1) 对于任意多类识别问题的子波神经网络分类器, 寻求最小均方差的过程也就是寻求可分性最大的子波特征以及对子波特征进行最大可分性划分的过程;

(2) 对于特征提取层, 就是寻求 $J = \text{tr} [S_T^+ S_B]$ 最大, 使可分性测度 $J = J_{\max}$; 对于输出层, 就是对于特征提取层的输出 (即提取的信号子波特征) 进行最大可分性的划分.

3 免疫进化子波网络及其学习算法

在前面的分析中我们曾提到过, 原有子波神经网络在对子波参数的选取方面需要依靠操作人员的经验或尝试, 而且训练过程易于陷入局部极小, 收敛速度也相对较慢. 因此, 人们提出了利用进化算法来克服上述算法所存在的缺陷. 而进化算法中的交叉和变异算子都是在一定发生概率的条件下, 随机地、没有指导地迭代搜索, 因此它们在为群体中的个体提供了进化机会的同时, 也无可避免地产生了退化的可能. 同时, 每一个待求的实际问题都会有自身一些基本的、显而易见的特征信息或应用背景. 但原有的各种进化算子却相对固定, 在求解问题时, 可变的灵活程度较小. 这无疑对算法的通用性是有益的, 但却忽视了问题的特征信息或应用背景对求解问题时的辅助作用.

在对原有进化算法进行改进过程中所提出的免疫概念及其举措, 主要是试图利用局部特征信息以一定的强度干预全局并行的搜索进程, 抑制或避免求解过程中的一些重复和无效的工作, 以克服原算子操作方面的盲目性. 也就是说在保留原算法优良特性的前提下, 力图有选择、有目的地利用待求问题中的一些特征信息或知识来抑制其优化过程中出现的退化现象. 由于免疫进化算法本身所固有的并行运算规则、智能搜索方式和概率判断准则, 从而有利于弱子波网络的应用条件, 加强其联想记忆和信息处理的能力. 下面我们就对这种算法的基本特点和主要性能做一简单的介绍.

3.1 免疫进化算法

免疫概念的提出是受生物自然科学的启发. 在生命科学中, 免疫功能主要是由参与免疫反应的细胞或者说由其构成的器官完成的. 这种免疫细胞主要有两大类^[13]: 一类为淋巴细胞. 这类细胞因为对抗原的反应有明显的专一性, 所以是特异性免疫 (Specific Immunity) 反应的主要细胞; 第二类细胞则具有摄取抗原、处理抗原并将处理后的抗原以某种方式提供给淋巴细胞的作用, 其重要特征是在参与各种非特异性免疫反应 (Nonspecific Immunity) 的同时, 也能积极地参与特异性免疫反应.

基于上述认识, 本文在标准进化规划算法中引入一个新的算子, 即免疫算子 (Immune Operator). 同生命科学中的免疫理论类似, 免疫算子也分两种类型: 全免疫 (Full Immunity) 和目标免疫 (Target Immunity), 二者分别对应于生命科学中的非

特异性免疫和特异性免疫. 其中, 全免疫是指群体中每个个体在变异操作后, 对其每一环节都进行一次免疫操作的免疫类型; 目标免疫则指个体在进行变异操作后, 经过一定判断, 个体仅在作用点处发生免疫反应的一种类型. 前者主要应用于个体进化的初始阶段, 而在进化过程中基本上不发生作用, 否则将很有可能产生通常意义上所说的“同化现象” (Assimilative Phenomenon); 后者一般而言将伴随群体进化的全部过程, 也是免疫操作的一个常用算子.

在本文中, 抗原 (Antigen) 指的是所有可能错误的基因, 即非最佳个体的基因; 疫苗 (Vaccine) 是根据进化环境或待求问题的先验知识, 所得到的对最佳个体基因的估计; 抗体 (Antibody) 是指根据疫苗修正某个个体的基因所得到的新个体. 其中, 根据疫苗修正个体基因的过程即为接种疫苗 (Vaccination), 其目的是消除抗原在新个体产生时所带来的负面影响.

在实际的操作过程中, 首先, 对所求解的问题进行具体分析, 从中提取出最基本的特征信息; 其次, 对此特征信息进行处理, 以将其转化为局部环境下求解问题的一种方案; 最后, 将此方案以适当的形式转化成免疫算子并用来产生新的个体. 这里需要说明的是: 一方面, 待求问题的特征信息往往不止一个, 也就是说针对不同特征信息所能提取的疫苗也可能不止一种, 那么在接种疫苗过程中可以随机地选取一种疫苗进行接种, 也可以将多个或所有的疫苗按照一定的逻辑次序予以接种; 另一方面, 疫苗实质上是对最优个体 x_{\max} 在某一分量上的估计, 而且这种估计的正确与否还有待于其后的选择机制做进一步的判断. 这也就是说, 疫苗只会影响到算法的搜索效率, 而不至于涉及到算法的一些核心性质, 如收敛性等.

综上所述, 为了有效地利用人脑对待求问题的先验知识, 避免完全随机的寻优算法中出现的退化现象, 本文提出了免疫规划的方案. 其思想主要是在合理提取免疫疫苗的基础上, 通过接种疫苗和免疫选择两个操作步骤来完成的. 前者是为了提高个体的适应度, 后者则为了防止群体的退化. 具体而言, 它们分别是:

接种疫苗: 设个体 $b \in R^n$, $b = (b_1, \dots, b_n)$. 给其接种疫苗是指按照先验知识修改 b 的某些分量, 使所得个体以较大的概率具有更高的适应度. 这些先验知识可以是最优解某些分量的大概取值范围, 也可以是一些分量之间的一定制约关系. 而且, 接种疫苗操作应该满足如下条件: 若 b 已经是最佳个体, 则 b 以概率 1 转移为 b . 设有群体 $B = (b^1, \dots, b^N)$, 对 B 接种疫苗是指在 B 中按比例 α 随机抽取 $N_\alpha = \alpha N$ 个个体而进行的一种操作. 疫苗是从问题的先验知识中提炼出来的, 其中所包含的信息量及其准确性对算法的执行效率起着重要的作用.

免疫选择: 这一操作分两步完成. 第一步是免疫检测, 即对抗体进行检测, 若其适应度仍不如父代, 说明在变异过程中出现了严重的退化现象. 这时, 由父代个体代替其参加选择竞争; 第二步是退火选择, 即在目前的子代群体 $D_k = (d^1, \dots, d^N)$ 中以概率:

$$P(d^i) = \frac{f(d^i) e^{f(d^i)/T_k}}{\sum_{i=1}^N f(d^i) e^{f(d^i)/T_k}} \quad (19)$$

选择个体 d^i 进入新的父代群体, 其中 T_k 是单调递减趋于 0 的退火温度控制序列. 我们首先来考虑如下所示一类问题的求解方法.

不失一般性, 我们将一般的优化问题描述成如下所示的形式: 设 $\Omega \subset R^n$ 是紧集, 适应度函数 $f(x)$ 在 Ω 上连续且大于零. 求准最大值点 $x^* \in \Omega$, 并使 x^* 与理论最大值点的误差满足预先给定的误差精度要求. 由于固有的计算误差, 我们不能要求准最大值点 x^* 就是理论上的最大值点, 故设所要求的误差精度 Δ 为 10^{-l} . 在执行了算法的第一步

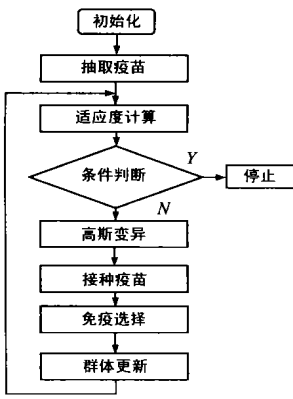


图 2 免疫进化算法流程图

之后, 准最优解 x^* 的搜索空间也就随之确定, 我们设其为 X . 也就是说, 搜索空间 X 由 R^n 中的网格点所构成, 这些网格点的每一个分量均为 $\Delta = 10^{-l}$ 的整数倍.

算法 1 免疫进化算法

(1) 初始化: 首先, 根据要求确定解的精度; 其次, 在 Ω 中随机产生 N 个个体, 这些个体的每一分量均为 l 位小数, 并由此构成初始的父代种群 A_0 ;

(2) 根据先验知识抽取疫苗 H ;

(3) 计算当前第 k 代父本种群 A_k 所有个体的适应度, 并进行停机条件的判断. 若条件满足, 则停止运行并输出结果; 否则, 继续;

(4) 对当前的父代群体 A_k 进行变异操作, 生成子代群体 B_k . 设 $A_k = (a^1, \dots, a^N)$, $a^i \in \Omega \subset R^n$, $a^i = (a^i_1, \dots, a^i_n)$, $i = 1, \dots, N$. 对个体 a^i 进行变异是指按如下方式产生新的个体 b^i :

$$b^i_j = a^i_j + \xi^i_j, j = 1, \dots, n$$

其中 ξ^i_j 为不超过随机变量 $\eta^i_j \sim N(0, \sigma^2_{ij})$ 的最大 l 位小数, 方差 σ^2_{ij} 根据下式进行计算:

$$\sigma_{ij} = \frac{T_k M_j + m_j}{T_k + f(a^i)}$$

其中 T_k 即为退火选择中所用到的序列, 而且 T_0 应充分地大. 由上式可以看出, 当 T_0 充分大时, $\sigma_{ij} \approx M_j$, 而 M_j 是一个与搜索空间的形状和大小有关的量, 所以算法在搜索初期对所有的个体都有较大的进化机会; 当进化程度较高, 即 k 较大时, T_k 趋近于 0 而 $f(a^i)$ 趋近于一个常数, 因此 σ_{ij} 正比于 m_j , 这时子代个体变异后与父代个体的距离由 m_j 控制, 故 $m_j \neq 0$. 另外, 对群体进行变异操作指的是对群体中的每一个个体均进行变异.

(5) 对群体 B_k 进行接种疫苗操作, 得到种群 C_k ;

(6) 对群体 C_k 进行免疫选择操作, 得到新一代父本 A_{k+1} , 并转至第 3 步.

该算法的运行流程如图 1 所示, 其中算法执行的终止条件可以设定为最大迭代次数或最佳个体不被更新的最大次

数.

3.2 免疫进化算法的收敛性

上述算法的状态转移情况可用如下的随机过程来表述:

$$A_k \xrightarrow{\text{高斯变异}} B_k \xrightarrow{\text{接种疫苗}} C_k \xrightarrow{\text{免疫选择}} A_{k+1}$$

其中下标 k 表示进化的代数. 对于固定的 k , 从 A_k 到 C_k 的状态转换构成了马尔可夫链, 虽然 A_{k+1} 的状态与前面各变量的状态有关, 但是随机过程 $\{A_k | k = 1, 2, \dots\}$ 仍是一个马尔可夫链. 这里定义若 $x \in \Omega$, 则 $f(x) = 0$. 我们将规模为 N 的群体认为是状态空间 $S = X^N$ 中的一个点, 其每个分量就是群体中的一个个体. 用 $s_i \in S$ 表示 s_i 是 S 中的第 i 个状态; 用 $s_i \subseteq s_j$ 表示将 s_i 和 s_j 分别看成是由其分量组成的 X 的子集时的包含关系. 用 V_k 表示随机变量 V_k 处于状态 s_i , 或者说在第 k 代时群体 V_k 处于状态 s_i . 设 f 是 X 上的适应度函数, 令

$$X^* = \{x \in X | f(x) = \max_{x_i \in X} f(x_i)\}$$

则定义算法的收敛性如下:

定义 1 如果对于任意的初始分布均有

$$\lim_{k \rightarrow \infty} \sum_{s_i \cap X^* \neq \emptyset} P\{A^k_i\} = 1$$

则称免疫进化算法收敛.

该定义表明: 算法收敛是指当算法迭代到足够多的次数以后, 群体中包含全局准最佳个体的概率接近于 1. 这种定义即为通常所说的概率 1 收敛.

定理 1 免疫进化算法是以概率 1 收敛的.

我们可以依据随机过程等相关理论证明上述定理是正确的, 此处鉴于篇幅有限, 定理 1 的证明过程参见文献[10].

3.3 子波函数的参数初始化

由前面的分析可知, 子波特征提取层的功能是提取可分性最大的一组信号子波特征. 在目标识别中, 应用子波变换的目的也在于提取信号有利于分类识别的信息, 即提取的子波特征在特征空间中可分性大于原始信号在信号空间中的可分性. 如果有一种对可分性的评价函数可以直接指导训练子波基函数参数的选择, 则会大大降低网络的训练时间. 有鉴于此, 我们可以结合原进化算法的特点, 采用如下所示的一种对具体信号进行子波基函数参数初始化的方法.

采用免疫进化算法解决优化问题的前提是确定合理的适应度函数, 针对目标分类问题, 我们即可采用可分性测度做为适应度函数. 具体而言, 在模式识别中, 度量同一类样本间的相似性和不属于同一类样本间的差异性时, 我们经常使用的测度之一是欧氏距离, 在 n 维空间中, 点 a 和点 b 之间的欧氏距离一般定义为如下两种形式:

$$D(a, b) = \|a - b\| = \sqrt{\sum_n (a_k - b_k)^2}$$

或

$$D^2(a, b) = (a - b)'(a - b) = \sum_{k=1}^n (a_k - b_k)^2$$

当考虑有两个以上的类别时, 类间距离的可分性可定义成两类集合 $\{a^{(i)}\}$ 和 $\{b^{(i)}\}$ 之间的类间距离, 即:

$$d_{ij} = \sqrt{\sum_{k=1}^l (m_{i, k} - m_{j, k})^2}$$

其中 $m_{i,k}$ 为第 i 类的样本集合的均值向量的第 k 维分量; $m_{j,k}$ 为第 j 类的样本集合的均值向量的第 k 维分量; L 表示样本的特征维数. 显然, d_j 越大, 则类 i 和类 j 之间的可分性越好.

另一方面, 同一类内部的距离也是衡量模式可分性的重要指标. 类内距离 d 可定义为:

$$\bar{d} = \sqrt{\frac{1}{L} \sum_{k=1}^L \delta_k^2}$$

其中, L 为样本的特征维数, δ_k^2 由下式决定:

$$\delta_k^2 = \frac{1}{K-1} \sum_{i=1}^K (a_k^i - \bar{a}_k)^2$$

其中

$$\bar{a}_k = \frac{1}{K-1} \sum_{i=1}^K a_k^i$$

为 $\{a^{(i)}\}$ 在第 k 分量方向的均值.

为使类内距离和类间距离统一起来, 我们重新定义可分性测度 $J_{i,j}$ 为如下形式:

$$J_{i,j} = \frac{d_{i,j}}{d_i + d_j}$$

显然, $J_{i,j}$ 越大, 表示 i 类和 j 类间的可分性越好; $J_{i,j}$ 越小, 表示 i 类和 j 类间的可分性越差. 有鉴于此, 我们以可分性测度 $J_{i,j}$ 做为适应度函数, 即适应度函数值越大, 表明提取的子波特征可分性越好.

在子波基函数参数的初始化方面, 我们首先确定一个母波函数以及对特定目标信号的伸缩、平移参数的取值范围; 其次, 利用免疫进化算法进行优化搜索; 最后, 获得一组有利于分类识别的信号子波特征.

3.4 免疫疫苗的提取

如前所述, 免疫算子是由接种疫苗和免疫选择两部分操作构成的. 其中, 疫苗是对最佳个体某些基因的一种估计, 它建立在人们对待求问题的或多或少的先验知识, 并从中提取出特征信息的基础上; 抗体是指根据疫苗而得出的一类个体. 前者也可以看作是对待求的最佳个体所能匹配的模式 (Schema) 的一种估计; 后者则是对这种模式进行匹配而形成的样本. 疫苗不是一个成熟或完整的个体, 它仅仅具备最佳个体某些局部基因位上的特征. 疫苗的正确选择会对群体的进化产生积极的推动作用, 进而对算法的运行效率具有十分重要的意义. 但是, 选取疫苗的优劣, 生成抗体的好坏, 只会影响到免疫算子中接种疫苗作用的发挥, 不至于涉及到算法的收敛性. 因为免疫算法的收敛性归根结底是由免疫算子中的免疫选择来保证的, 如下所示.

定理 2 在免疫选择作用下, 若疫苗使抗体适应度得到提高, 且高于当前群体的平均适应度, 则疫苗所对应的模式将在群体中呈指数级扩散; 否则, 它将被遏制或呈指数级衰减.

我们可以参照遗传算法中的模式定理证明上述结果的正确性, 此处不再赘述. 由此我们可以看出免疫选择在加强接种疫苗的积极作用, 消除其负面影响方面具有鲁棒性.

在前期的一些工作中, 我们已经就 TSP 问题讨论了疫苗提取的基本过程, 即分析问题, 搜集特征信息, 进而根据特征信息制作免疫疫苗, 并且给出了其间用到的基本方法以及理

论上的解释. 这里需要补充说明的是: 在某些情况下, 我们一方面会因为对待求问题一时难以形成较为成熟的先验知识, 而无法从分析问题的过程中提取出合适的特征信息, 故得不到有效的免疫疫苗; 另一方面, 为寻求用于全局求解的局部方案需要付出的代价超出其应占的比例, 使计算成本增加、效率降低, 从而使提取疫苗的工作失去了意义. 所以, 为了提高算法的通用性与应用上的便利性, 本文试图在群体进化过程中, 从其最佳个体的基因中提取有效信息, 进而开辟一条制作免疫疫苗的途径. 基于这种思想, 这里给出一种疫苗的自适应选取及其相应的接种方法.

这里为表述方便起见, 我们首先说明以下将要用到的一些特殊符号. 设 $a_{H,k}^i$ 为对第 k 代第 i 个个体 a_k^i 接种疫苗后所得到的抗体; P_V 为个体接种疫苗的概率; $Vaccine(a_k^i, h_j)$ 表示按模式 h_j 修改个体 a_k^i 上基因的接种疫苗操作; n 和 m 分别为群体规模和个体的基因长度; $random(m)$ 表示随机产生 1 到 m 中间的一个任意正整数的操作.

算法 2 疫苗的自适应提取与接种算法

Begin:

$k = 0$;

while (Conditions = True)

 统计父代群体, 确定最佳个体: $a_k^{\text{optimal}} = \text{Statistics}(a_k^i | i = 1, \dots, n)$;

 分解最佳个体, 抽取免疫基因: $H = \{h_j = a_{k,j}^{\text{optimal}} | j = 1, 2, \dots, m\}$;

 执行变异操作: $a_k^i = \text{Mutation}(a_k^i), i = 1, \dots, n$;

 for $i = 1$ to n

 if $\{P_V\} = \text{True}$

$J = \text{random}(m)$;

 接种疫苗: $a_{H,k}^i = \text{Vaccine}(a_k^i, h_j)$;

 免疫检验: if $a_{H,k}^i < a_{k-1}^i$, then $a_k^i = a_{k-1}^i$;

 else $a_k^i = a_{H,k}^i$;

 end

 end

 end

 选择: $A_{k+1} = S(A_k)$;

$k = k + 1$;

end

3.5 免疫进化子波网络的学习算法

子波网络的拓扑结构如图 1 所示, 我们在这里介绍采用免疫进化算法对该网络进行学习训练的过程.

算法 3 免疫进化子波网络学习算法

(1) 初始化. 将任意选取 n 组权值以及初始化后的子波基参数做为初始群体, 其中, 每一组权值与一组子波基参数所构成的结构体视为一个个体;

(2) 根据先验知识抽取疫苗 H . 由前面的介绍可知, 疫苗只是对最佳个体在某些基因位上所具有的基因信息的一种估计. 它对算法的搜索过程产生辅助的指导作用, 所以其准确性对算法的执行效率具有积极意义, 但它不是导致算法收敛的根本因素. 有鉴于此, 我们可以根据对问题的先验知识或其应

用背景方面的特征信息, 来确定个体在某些基因上的取值特征或基因之间的相互制约关系, 并以此做为待求问题的免疫疫苗, 经编码处理后即可视为 H ; 另一方面, 若以上条件尚不具备, 我们即可采用算法 2 来动态寻找 H , 并将该过程置于第 (4) 与第 (5) 步骤之间进行;

(3) 计算当前群体中所有个体的适应度, 并从中确定最佳个体, 然后判断停机条件是否满足: 若是, 则对最佳个体进行解码并输出相应结果; 否则, 继续;

(4) 对当前群体实施变异操作, 具体过程与要求参见算法 1 之步骤 (4);

(5) 对当前群体实施接种疫苗操作, 即按照疫苗 H 在一定概率的条件下修改目标个体的相应基因;

(6) 对接种了疫苗的个体进行检验, 并对所注射的疫苗做出评价;

(7) 计算当前群体中所有个体的适应度, 并以此为根据在一定的选择机制下(本文推荐使用的是模拟退火选择, 其选择概率的计算参见式(19)), 挑选出 n 个个体组成下一代进化的群体, 然后转至第(3)步。

这里需要指出的是, 免疫进化子波网络的学习训练过程包括两个方面, 即子波特征的提取和网络在传统意义上的学习训练。在实施阶段, 由于它们的具体任务或目标不尽相同, 如图 2 所示, 所以我们可以分两个过程进行, 但在算法方面没有区别。

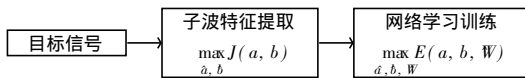


图 3 子波网络的学习过程

根据前面的分析可知, 在子波特征预提取时, 选用目标信号的子波特征可分性测度 J 做为适应度函数。因此, 通过免疫进化算法的学习, 最终提取的子波特征可分性最大。

4 仿真结果与分析

双螺旋线问题属于一个两类模式的非线性分类问题, 近十年来, 基于神经网络求解该问题方法研究已有不少结果^[15, 16], 在一定程度上也被间接地作为对各种非线性优化方法的评价准则。双螺旋线问题的表述很简单, 即要求将属于两个不同螺旋线的点加以正确分类即可, 如图 3 所示。其中, 螺旋线的极坐标方程一般表述为如下所示的形式:

$$\rho = \alpha\theta + \beta \quad (20)$$

其中, α 和 β 都是常数, 分别表示螺旋线的角速度和相对于极点的起始距离。另外, 螺旋线还可以采用参数方程的形式表示。设螺旋线上任意一点的直角坐标为 (x, y) , 则:

$$\begin{cases} x = (\alpha\theta + \beta) \cos\theta \\ y = (\alpha\theta + \beta) \sin\theta \end{cases} \quad (21)$$

采用免疫子波神经网络对双螺旋线进行识别与分类时, 我们首先可以根据该问题的特征信息来提取疫苗。具体而言, 当前情况下, 提取疫苗的任务就是根据双螺旋线的特性找出其上任意一点的坐标变量之间所应满足的约束关系。所以, 由螺旋线的参数方程式(21)易得:

$$\frac{y}{x} = \tan \frac{\sqrt{x^2 + y^2} - \alpha}{\beta} \quad (22)$$

由上式可以看出, 如果通过训练能够确定某一点的坐标与含 α, β 的螺旋线满足上式的要求, 则该点即属于此螺旋线, 否则属于另一螺旋线。因此, 上式可以作为基于当前问题所提取的特征信息。

接下来, 我们应根据双螺旋线的特点来选择子波基函数。由于子波的理论与应用研究已相当广泛, 就其对称性、连续性、变量数目以及维数等特性来说, 子波可以分成对称与非对称子波; 正交与非正交子波; 单元与多元子波; 一维与高维子波等等类型。对本问题而言, 由于平面上任意一点坐标 (x, y) 的两个分量的意义完全一样, 换句话说, 它们所能反映的信息特征是等价的, 所以我们可以从一维子波类型中进行选择。故鉴于本题的特性, 我们选择如下所示的函数做为子波基函数:

$$\psi(t) = \cos 1.75t \cdot \exp(-\frac{t^2}{2})$$

该子波函数的波形与 Mexico 草帽子波类似, 如图 4 所示。

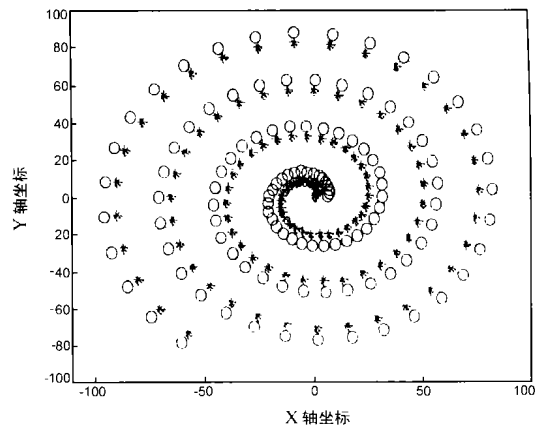


图 4 双螺旋线问题的平面示意图

在实验中, 我们以一个双螺旋线的分类问题为仿真对象来考查免疫进化子波神经网络的工作能力。具体的方法是: 首先, 产生带有随机干扰的样本点, 共 640 个, 且分别属于两类螺旋线 ρ_1 和 ρ_2 (各自具有 320 个样本点), 如图 5

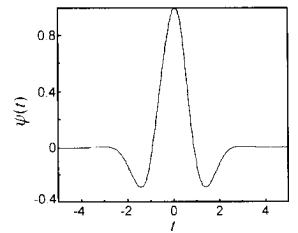


图 5 选定的子波基函数

所示。其中, 用于生成两类螺旋线的极坐标方程中的角速度相同, 为 4, 而起始距离分别为 1 和 7; 我们按照样本生成的先后次序间隔地取出一半的样本点用于网络的学习训练, 剩余的一半则用于网络的测试; 最后, 采用前面式(22)所示的特征信息提取疫苗向量并根据第 4.3 节所述方法来训练网络。

在具体的实现方面, 我们在 Pentium 233 的个人计算机上采用 MatLab 5.3 来编写实验程序。将训练好的网络用于对测试样本的分类, 其中错误识别的样本共有 15 个, 正确识别率为 95.3125%。如果去除样本点的随机干扰, 并采用相同的实验方法和实现手段, 在对可调参数进行优化的时间上大致相同, 但正确识别率为 100%。这与文献[15]和[16]中所报道的

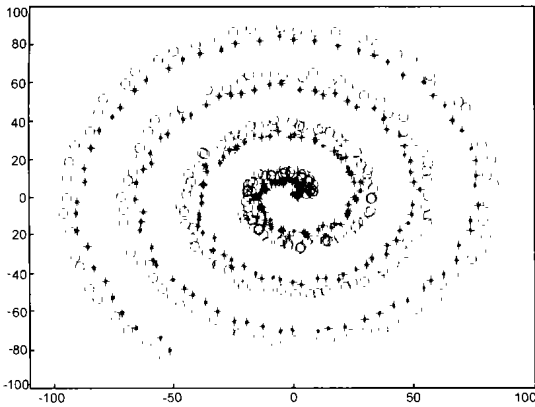


图 6 带有随机干扰的双螺旋线的平面分布图

情况相比有较大的提高。

5 结论与讨论

我们在分析子波神经网络内在机理的过程指出,对于任意多类识别问题的子波神经网络分类器,寻求最小均方差的过程也就是寻求可分性最大的子波特征以及对子波特征进行最大可分性划分的过程;对于特征提取层,就是使可分性测度最大;对于输出层,就是对于特征提取层的输出(即提取的信号子波特征)进行最大可分性的划分。在此基础上,我们给出了采用免疫进化算法对子波网络进行学习及训练的模型与方法。该训练算法是一种集免疫机制和进化机制于一体的一种新的全局并行算法,它可以通过对进化环境的自适应和自学习,有针对性地抑制由原进化算子操作的盲目性而引起的退化现象。理论分析和用于双螺旋线分类的仿真结果表明,免疫进化子波网络不仅是可行的,也是十分有效的。由于免疫进化算法本身所固有的并行运算规则、智能搜索方式和概率判断准则,从而有利于弱化子波网络的应用条件,加强其联想记忆和信息处理的能力。

参考文献:

- [1] Rogers S K, et al. Neural networks for automatic target recognition, neural networks [J]. 1996, 8(8): 1153- 1184.
- [2] Zurada J M. Introduction to Artificial Neural Systems, West Publishing Company [M], New York, 1992.
- [3] 焦李成, 保铮. 神经网络与 AI 发展现状、深思与综合 [J]. 系统工程与电子技术, 1992, 2.
- [4] Young R K. Wavelet Theory and Its Application [M]. Kluwer Academic Publishers, America, 1993.

- [5] 赵松年, 熊小芸. 子波变换与子波分析 [M]. 北京: 电子工业出版社, 1997.
- [6] Szu H H, Kadanebe S. Neural network adaptive wavelets for signal representation and classification [J]. Optical Engineering, 1992, 31(9): 1907- 1916.
- [7] Holland J H. Genetic algorithms and classifier systems: foundations and their applications [A]. Proceedings of the Second International Conference on Genetic Algorithms [C], 1987: 82- 89.
- [8] Rudolph G. Convergence analysis of canonical genetic algorithms [J]. IEEE Trans. on Neural Networks, 1994, 5(1): 96- 101.
- [9] 陈国良, 王煦法, 庄镇泉, 王东生. 遗传算法及其应用 [M]. 北京: 人民邮电出版社, 1996.
- [10] Jiao L C, Wang L. A novel genetic algorithm based on immunity [J]. IEEE Trans. on Systems, Man, And Cybernetics Part A: Systems and Humans, September 2000, 30(5): 552- 561.
- [11] 王磊, 潘进, 焦李成. 基于免疫策略的进化算法 [J]. 自然科学进展, 2000, 10(5): 451- 455.
- [12] 张艳宁. 智能目标识别方法研究 [R]. 西安电子科技大学博士后研究报告, 1999.
- [13] 陈仁. 免疫学基础 [M]. 北京: 人民卫生出版社, 1982.
- [14] 张讲社, 徐宗本, 梁怡. 整体退火遗传算法及其收敛充要条件 [J]. 中国科学(E), 1997, 27(2): 154- 164.
- [15] Fahlman S E, Leibiore C. The cascade correlation learning architecture [J]. In: Touretzky D. S. (Ed), Advances in Neural Information Processing System. San Mateo, CA: Morgan Kaufmann, 1990: 524- 532.
- [16] Chen Q C. Generating shrinking algorithm for learning arbitrary classification. Neural Networks, 1994, 7: 1477- 1489.

作者简介:

王 磊 男, 1972 年生于西安市。现为西安电子科技大学在读博士生, 主要研究领域包括: 人工神经网络、进化算法与数据挖掘等。

焦李成 男, 1959 年生。现为西安电子科技大学研究生院常务副院长, 教授, 博士生导师, 主要研究领域包括: 非线性理论、人工神经网络、子波理论与应用、进化算法、数据挖掘与多用户检测等。

刘 芳 女, 1963 年生于北京市。现为西安电子科技大学副教授, 主要研究领域包括: 网络智能信息处理与模式识别等。

张艳宁 女, 1968 年生于陕西兴平。现为西北工业大学副教授, 主要研究领域为网络智能信息处理等。