

最大绝对误差结合微遗传算法优化径向基概率神经网络

赵温波¹ 王立明¹ 黄德双²¹(解放军炮兵学院三系 合肥 230031)²(中国科学院合肥智能机械研究所 合肥 230031)

(wbzhao@iim.ac.cn)

Structure Optimization of Radial Basis Probabilistic Neural Networks by the Maximum Absolute Error Combined with the Micro-Genetic Algorithm

Zhao Wenbo¹, Wang Liming¹, and Huang Deshuang¹¹(Department Three, Artillery Academy of People Liberation Army, Hefei 230031)²(Institute of Hefei Intelligent Machines, Chinese Academy of Sciences, Hefei 230031)

Abstract The maximum absolute error algorithm (MAEA) is used to optimally selecting the hidden centers vectors of the radial basis probabilistic neural networks (RBPNN). The MAEA is combined with the micro-genetic algorithm (μ GA), which is used to optimize the controlling parameter of the kernel function of the RBPNN, i.e., MAE- μ GA, so as to carry out optimizing the overall structure of RBPNN. The experiments demonstrate that the RBPNN, optimized by the MAE- μ GA, has the best simple structure compared with results by the other optimization methods introduced. Furthermore, in the aspect of the generalization performance of the optimized networks, the RBPNN by the MAE- μ GA is a little better than ones by the other methods. In addition, the MAE- μ GA can also be used to optimize the radial basis function neural networks (RBFNN).

Key words RBPNN; structure optimization; μ GA; MAE- μ GA

摘要 使用最大绝对误差算法(MAEA)优选径向基概率神经网络(RBPNN)隐中心矢量,将MAEA与求解RBPNN最优核函数控制参数的微遗传算法(μ GA)相结合(MAE- μ GA)来共同实现RBPNN的全结构优化.实验结果显示,对比其他几种算法,MAE- μ GA优化后的RBPNN结构最简,而且在推广能力方面略好于其他几种优化方法.另外,MAE- μ GA对径向基函数网络也有很好的适用性.

关键词 径向基概率神经网络;结构优化;微遗传算法;最大绝对误差-微遗传算法

中图分类号 TP183

1 引言

径向基概率神经网络模型(RBPNN)^[1]是由径向基函数神经网络(RBFNN)^[2]和概率神经网络(PNN)^[3]发展而来的,因此,RBPNN具有RBFNN和PNN的优点,如快速的收敛性能;考虑了模式类

别间的交叉影响等.同时,RBPNN也减少了RBFNN和PNN的缺点,如计算复杂度的降低等.RBPNN的网络结构如图1所示,在拓扑结构上该网络有4层结构;即1个输入层、2个隐层和1个输出层.第1隐层是非线性处理层,由训练集中优选出的样本(又称隐中心矢量)组成.第2隐层根据隐中心矢量的类别对第1隐层的输出进行选择性地求和.输出层

最终完成 RBPNN 的非线性映射功能,如分类、函数逼近和预测等.从 RBPNN 结构的本质上来看,第 1 隐层是网络性能发挥的关键.

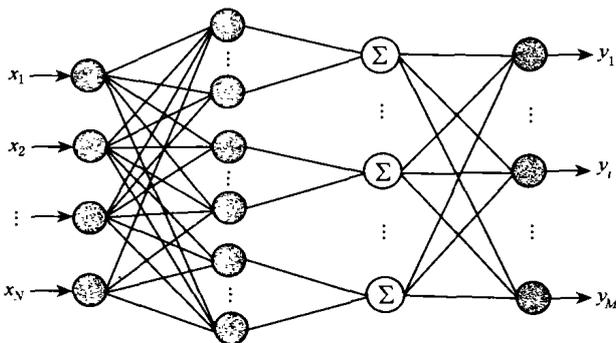


Fig. 1 The structure of the radial basis probabilistic neural networks.

图 1 径向基概率神经网络结构

推广能力是评价前馈神经网络性能的重要指标,决定了神经网络是否具有实用性.影响推广能力的因素有很多,如训练集的选取、学习算法构造,网络规模、以及过学习(overfitting)等.其中,过学习又称为过训练(overtraining),是决定网络推广能力至关重要的因素.有关过学习的研究有很多. Chi^[4]从训练集的角度来分析过学习产生的原因,将过学习划分为绝对过学习和相对过学习. Chi 认为相对过学习主要是由于训练集中模式类别相互交叉所致,并相应地给出 5 种解决方案. 而 Reed^[5]则从网络结构方面研究了过学习,指出网络结构冗余可能是导致过学习重要因素. 如果网络结构过于复杂,将“简单问题”“复杂化”,从而导致推广能力下降. 从这方面来看,神经网络结构并不是越复杂越好,网络结构过于复杂不仅造成训练与测试时间长,更重要的是有可能导致过学习造成推广能力下降;而小规模的神经网络不仅运算速度快,软硬件开销小,网络结构易于理解,更重要的是网络推广能力大大改观. 当然也不能过于简单,网络所蕴含的结构自由度不足以包容训练集的信息自由度,那么无论在训练的收敛性能、还是网络的推广能力方面都不可能满足求解实际问题的需要. 因此,适度规模的网络结构选择研究非常重要,是一个制约神经网络得以广泛应用的难题,也是当前神经网络领域研究的热点.

对于一般前馈神经网络来说,比较成熟的网络结构优化算法有减构法(pruning algorithms)^[5,6]和增构法(constructive algorithms)^[5,6]. 减构法的基本原理是从较大规模且具有较好推广能力的网络结构开始,删去结构中冗余的不重要的结点,从而使网络

的推广能力得以提高. 减构法的不足之处在于起始训练的神经网络规模比较大,使得减构之前的网络训练时间较长,从而增加了网络训练的时间耗费和存储开销. 增构法^[7,8]优化网络结构与减构法相反,它从一个相对简单的网络结构开始训练,通过不断地增加结点使得网络的推广能力得以提高. 增构法的不足之处在于有可能造成网络的输入至输出的传输延迟,而且新结点的随机增加可能破坏网络收敛性. 另外,遗传算法(GA)也可以用于前馈神经网络的结构优化^[9,10]. 遗传算法优化神经网络仍然需要较大的计算需求.

RBPNN 是一种特殊类型的前馈神经网络,隐中心矢量的数目、位置以及核函数控制参数(或称为感受野)是网络结构的重要组成部分. RBPNN 结构优化问题不仅涉及结点的增加或删除,还应包括核函数控制参数的优化. 特别地,当 RBPNN 用于模式识别问题时,网络结构优化仅限于第 1 隐层,因为第 2 隐层的结点数目由被求解问题的模式类别特性所决定,无须进行优化. 由于 RBPNN 第 1 隐层权重的特殊性,传统的结构优化算法并不适合的 RBPNN 的隐中心矢量的选取. 除此之外, RBPNN 的隐中心矢量的优化原理与其他前馈神经网络没有太大区别,即如果隐中心矢量数目过多,势必会导致网络训练及测试时间过长且出现过学习;而隐中心矢量数目太少,也有可能使得网络的收敛误差太大. 也就是说, RBPNN 用于解决特定的问题时第 1 隐层的隐中心矢量数目的合适选择对于网络性能发挥至关重要. 此外,核函数控制参数同样也是决定 RBPNN 性能的重要结构参数. 核函数控制参数的作用机理与人类视神经系统相似,如果视野宽度太大,物体的细微之处可能无从分辨;相反,宽度太小,物体的整体信息可能丢失. 因此,在解决实际问题时, RBPNN 的核函数控制参数必须适当选取和优化. 另外,从图 1 中我们还可以看出,隐中心矢量与核函数控制参数在结构上是紧密相关的,共同作用于 RBPNN 的第 1 隐层. 从几何意义上来看, RBPNN 第 1 隐层主要完成对样本空间的超划分与覆盖. 不同的数目、位置的隐中心矢量可能需要不同的核函数控制参数才能实现对本样本空间的全覆盖,反过来,不同取值的核函数控制参数如果满足全覆盖要求也需要不同位置、数目的隐中心矢量. 隐中心矢量和核函数控制参数在结构上的密切关联特性决定了我们在研究 RBPNN 结构优化时必须同时考虑这两种因素.

关于 RBPNN 隐中心矢量的优化选取方法,文

献[11]介绍了遗传算法(GA)和ROLSA优化RBPNN隐中心矢量的特点,即GA具有全局搜索能力,但计算速度慢,而ROLSA计算速度快,但容易陷入局部最小。另外,文献[11]也指出这两种算法只能优化隐中心矢量的数目和位置,并没有涉及核函数控制参数的优化求解。关于核函数控制参数的优化取值算法,文献[11]使用微遗传算法来实现是一个有意义的尝试。

为了克服GA与ROLSA的不足,本文使用最大绝对误差算法(MAEA)来优选隐中心矢量。该算法的特点在于继承了增构法的优点,能够从一个相对较为简单的网络结构开始搜索,直至找到一个结构最简单而且能够满足性能要求的网络结构为止,具有计算速度快等优点。同时为了求解与MAEA优选的隐中心矢量相匹配的核函数控制参数,本文使用微遗传算法(μ GA)来优化求解。

本文首先阐述了运用MAEA优化选取隐中心矢量数目及空间位置分布的原理及步骤;然后介绍了使用 μ GA来求解核函数控制参数;接着给出了MAEA与 μ GA结合(MEA- μ GA)优化RBPNN结构的算法描述;最后进行了实验模拟与总结。

2 运用MAEA优化选取RBPNN隐中心矢量

2.1 RBPNN的训练方法

假定训练样本集 φ 中包含 N 个样本且这些样本归属于 M 个模式类别($N \geq M$)。对于输入样本 $x_i (i=1, 2, \dots, N)$,则RBPNN输出层第 $j (j=1, 2, \dots, M)$ 个结点的输出为

$$y_j(x_i) = \sum_{k=1}^M w_{jk} h_k(x_i) = \mathbf{h}_i \mathbf{w}_j, \quad (1)$$

$$h_k(x_i) = \sum_{l=1}^{n_k} \phi(x_i, \mathbf{c}_k^{(l)}, \sigma) = \sum_{l=1}^{n_k} \phi(\|x_i - \mathbf{c}_k^{(l)}\|_F, \sigma), \quad (2)$$

其中, $\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jM}]^T$, $\mathbf{h}(x_i) = [h_{i1}, h_{i2}, \dots, h_{iM}]$, $y_j(x_i)$ 或者 y_{ij} 为对应于输入样本 x_i RBPNN的输出层第 j 个单元的输出; $h_k(x_i)$ 或 h_{ik} 为第2隐层第 k 个结点的输出; w_{jk} 是第2隐层第 k 个结点与输出层第 j 个结点的连接权重; n_k 为第1隐层第 k 个模式的隐中心矢量数目; $\varphi(\cdot)$ 是高斯核函数; $\mathbf{c}_k^{(l)}$ 是与第2隐层第 k 个结点相连的第 l 个隐中心矢量; σ 是核函数控制参数; $\|\cdot\|_F$ 是矩阵 F -范数。

关于RBPNN的权重训练有很多方法可供选择,如矩阵伪逆算法^[12],ROLSA^[13],RLSA^[14]等。其中前两种算法能够利用成熟的矩阵算法,权重学习无需大规模的迭代计算过程,具有较好的收敛精度等优点。但也存在不足,如在计算过程中需要求取矩阵的逆、计算量较大等。而RLSA利用成熟的自适应滤波(卡尔曼滤波器)原理来获得网络连接权值的递推迭代,该算法在每一时刻的迭代是由输入样本形成的协方差矩阵来递推求得权值的精确解(误差梯度为0时的解)。所以,RLSA具有迭代的方向性强、收敛速度快等优点。

类似于其他前馈神经网络学习训练方法,RLSA也是基于误差代价函数为优化目标函数的训练算法,与传统的误差代价函数的区别在于,RLSA每一步迭代的误差代价是基于前面迭代误差的加权累加,因此,又称为加权误差代价函数,即

$$J(k) = \frac{1}{2} \sum_{t=1}^k \lambda^{k-t} \sum_{j=1}^M \epsilon_j^2(x_t) = \frac{1}{2} \sum_{t=1}^k \lambda^{k-t} \sum_{j=1}^M (d_j(x_t) - y_j(x_t))^2, \quad (3)$$

其中, $d_j(x_t)$ 与 $\epsilon_j(x_t)$ 分别是对应于输入样本 x_t 输出层第 j 个结点的期望信号和误差; λ 是加权遗忘因子,其作用是淡化过去迭代的输入样本对当前估值参数的影响,尽量反映当前输入样本的特性,通常的取值范围为 $0 < \lambda < 1$,取值越大,反映出当前估值参数对过去样本的依赖性越大。

根据文献[14],基于误差代价函数 $J(k)$ 的RBPNN的权重更新过程为

$$\mathbf{w}_j(k) = \mathbf{w}_j(k-1) + \mathbf{g}(k)[d_j(k) - \mathbf{h}(k)\mathbf{w}_j(k-1)], \quad (4)$$

$$\mathbf{g}(k) = \frac{\mathbf{P}(k-1)\mathbf{h}^T(k)}{\lambda + \mathbf{h}(k)\mathbf{P}(k-1)\mathbf{h}^T(k)}, \quad (5)$$

$$\mathbf{P}(k) = \frac{1}{\lambda} [\mathbf{P}(k-1) - \mathbf{g}(k)\mathbf{h}(k)\mathbf{P}(k-1)], \quad (6)$$

其中, $\mathbf{h}(k) \in \mathcal{R}^{M \times 1}$ 为对应于第 k 个迭代时输入样本的第2隐层输出向量; $\mathbf{g}(k) \in \mathcal{R}^{M \times 1}$ 为 k 时刻的卡尔曼增益; $\mathbf{P}(k) \in \mathcal{R}^{M \times M}$ 为逆相关矩阵。

根据RLSA得到的RBPNN权重矩阵 $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_M]$,即可得到收敛误差:

$$e = \|\mathbf{D} - \mathbf{H}\mathbf{W}\|_F^2, \quad (7)$$

其中, \mathbf{D} 为期望信号矩阵, $\mathbf{H} = [\mathbf{h}(1), \mathbf{h}(2), \dots, \mathbf{h}(N)]^T$ 为第2层的输出矩阵; e 在本文中又称之为绝对误差(absolute error, AE)。

对于模式识别问题,式(7)进一步表示为

$$e' = \|\mathbf{D} - \text{round}(\mathbf{H}\mathbf{W})\|_F^2, \quad (8)$$

其中, $round(\cdot)$ 为四舍五入算子. 当期望信号矩阵是单位矩阵时, e 的取值必为 $0, 1, 2, 3, \dots, N$, e 在文中又称之为分类误差(classification error, CE).

2.2 MAEA 优选隐中心矢量

由于初选的网络结构相对简单, 造成许多关键的结构信息丢失, 当网络对所有训练样本收敛以后, 对应于网络输出端必然会有不同的误差体现, 导致其中最大的输出误差主要是由于在网络结构中, 即第 1 隐层中缺失该最大输出误差对应的输入样本信息, 因此, 有必要把这一输入样本增加为网络结构的新成员, 使得变化后的网络具有很好的非线性映射能力.

所以 MAEA 优选 RBPNN 的隐中心矢量的基本步骤可以总结为, 在当前没有被选中为隐中心矢量的训练样本中, 把引起网络输出误差最大的样本增加为新的隐中心矢量, 直至新构造的 RBPNN 能够满足给定的停止条件时才停止算法. 假定当前训练样本 S_1, S_2, \dots, S_p ($p < N - M$) 未被选中为初始隐中心矢量, 则新的隐中心矢量 S_q 需满足:

$$q^* = \arg \max_{j=1}^p \|e_j\|_F = \arg \max_{j=1}^p \|d_j - y_j\|_F, \quad (9)$$

式中, d_j 与 y_j 分别为对应于第 j 样本的 RBPNN 输出的期望信号向量及实际输出向量; e_j 为对应于第 j 个输入样本时网络输出的误差向量, 即 $e_j = [e_{j1}, \dots, e_{jp}]^T$.

MAEA 需要初始的网络结构, 一般来说, 这一初始结构要尽可能简单, 而且还要包含关键的隐中心矢量. 为此, 我们使用“最小-最大中心”(min-max center)^[15]方法来初选隐中心矢量, 初始网络结构中每类只选一个隐中心矢量. 所谓最大最小中心就是在训练样本集的每个类别中, 每个样本与其他样本距离最大值中的最小值所对应的那个样本. 假定训练集中的某一类别 ω 中有训练样本 S_1, S_2, \dots, S_j , 则最小-最大中心 S_{l^*} 可以表示为

$$\delta(S_{l^*}, S_l) \leq \min_{1 \leq m \leq j} \max_{1 \leq i \leq j} \delta(S_m, S_i), \quad (10)$$

其中, $\delta(\cdot)$ 表示距离算子, 在本文中使用 F -范数, 即 $\delta(\cdot) = \|\cdot\|_F$.

MAEA 是一个迭代的过程, 其停止条件为, 由当前优选的隐中心矢量构造的 RBPNN, 训练收敛后的分类误差满足给定的误差标准时算法即停止.

3 微遗传算法优化核函数控制参数

微遗传算法^[16](μ GA)是基于小种群和再初始

化操作的特殊遗传算法, 具有较好的收敛速度和精度^[17]. 文献[11]首次使用 μ GA 来优化求取 RBPNN 核函数控制参数, 实验结果良好. 由于其单独使用变异算子来进化纯实数编码, 变异概率要恰当选取才能取得很好的结果. 本文仍然使用 μ GA 来优化核函数控制参数, 并加以改进.

由于实数编码能够跨越整个正实数空间搜索最优解, 而且搜索精度高, 故本文仍将实数编码方式作为 μ GA 优化控制参数的惟一编码方案. 但是直接实数编码很难实现交叉操作, 没有交叉算子就不能充分发挥遗传算法的优势^[18], 因此, 我们使用固定长度的十进制数编码来实现核函数控制参数空间至遗传空间的映射, 编码方式如图 2 所示, 其中, 每一位代表一个十进制数 ($0 \sim 9$), L 为编码长度. 但固定长度的十进制数编码还是只能表示有限范围的实数, 对于没有先验知识可以利用的核函数参数取值, 我们还是不知道这 L 位编码如何取值才能正确表示核函数控制参数, 为此, 我们增加了弹性系数 k 来拉长或压缩该编码的取值范围. 假定这 L 位编码代表 0 与 1 之间的一个实数 R , 则在遗传进化过程中, 此编码所表示的核函数控制参数取值为 kR . 初始情况下, $k_0 = 1$. 下面我们来讨论遗传迭代中 k 值的选取. k 值的选取依赖于种群中最优个体取值, 假定当前种群中最优个体相应的核函数控制参数取值为 $\sigma_{opt} = k_{i-1}R$, 当 σ_{opt} 趋近于某一整数 P_i , 而且 $P_i \neq P_{i-1}$, 则下一轮的遗传操作中, 相应的 k 值就要改变, 即 $k_i = P_i + 1$. 也就是说, k 的求取可以通过四舍五入算子来实现, 即 $k_i = round(k_{i-1}R) + 1$.

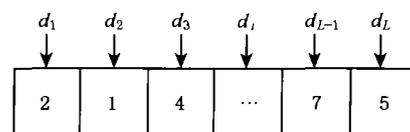


Fig. 2 Encoding scheme of the controlling parameter of kernel function with decimalization form.

图 2 核函数控制参数的十进制编码方式

本文 μ GA 的交叉算子使用两点随机交叉, 而不使用变异操作. 由于是小种群, 选择算子使用确定性锦标赛方法, 即当前种群中所有个体以概率 1 被选中为交叉操作的配对父母. 而个体适应度为绝对误差的函数, 即

$$f_i = \frac{1}{e_i + eps}, \quad (11)$$

其中, e_i 是个体的绝对误差; eps 是一个大于 0 的常数.

运用 μ GA 优化核函数控制参数的具体过程^[11]如下:

(1) 随机产生 5 个体组成初始种群,或者随机产生 4 个体,第 5 个体为前一代种群的最优个体,组成新一轮迭代的起始种群;

(2) 评估个体适应度,查找最优个体并将其标识为种群中的第 5 个体,然后将该个体保留为下一代种群成员;

(3) 选择种群中的所有个体(包括第 5 个体)进行交叉操作,由于种群规模小,因而交叉率为 1,即所有个体都参与交叉;

(4) 检查当前种群是否符合名义收敛条件,如果满足条件,则执行下一步,否则转向(2);

(5) 如果满足给定的误差条件,终止优化过程,否则转向(1).

上述算法包含两个收敛判据:种群收敛判据(局部收敛判据)和算法终止判据(全局收敛判据).前者是指小种群在没有“再初始化操作”的条件下,也就是在没有增加非进化新个体的情况下,完全依靠交叉算子而进化的小种群收敛性判据.后者是指退出算法执行时的判据,也就是算法的停止条件.不论是局部收敛判据,还是全收敛判据,一般来说,名义收敛条件被广泛采用.名义收敛一般包括基因型和显型收敛.所谓基因型收敛是指收敛后的种群为齐次种群或接近齐次种群,也就是种群中的个体具有全部相同或大部分相同的基因构成.而显型收敛是指算法满足给定的停止条件(如给定误差标准,最大进化代数等)时的一种间收敛形式.本文使用显型收敛,对种群收敛性判据,规定算法的最大进化代数;全局收敛判据使用给定的误差准则: μ GA 所优化的核函数控制参数结合 MAEA 选择的隐中心矢量构造的 RBPNN,对于所有的训练样本分类误差达到给定的误差标准则立即终止算法.

4 MAE- μ GA 全结构优化 RBPNN

为了实现 RBPNN 全结构优化,我们将 μ GA 引入 MAEA 中,使得优化选取隐中心矢量与核函数控制参数的优化取值同时进行. μ GA 的主要作用是优化选取与当前的隐中心矢量(或初始的隐中心矢量)相匹配能够实现对训练样本集全覆盖的核函数控制参数.而 MAEA 则是根据优化取值的核函数控制参数而进一步优选隐中心矢量并压缩其数目.从算法的执行特点来看, μ GA 贯穿于 MAEA 中,隐中心

矢量优选与核函数控制参数优化交替进行,因此,本算法又称之为 MAE- μ GA.算法描述如下:

(1) 使用最小-最大方法选取初始 RBPNN 的隐中心矢量.设定误差标准 ϵ ,令迭代次数 $i=1$;

(2) 利用 μ GA 优化与当前优选的隐中心矢量组合相匹配的核函数控制参数;

(3) 把每个未被选中为隐中心矢量的训练样本输入现有的 RBPNN 中,计算网络输出绝对误差 $e(i)$ ($e(i)=[e_1, e_2, \dots, e_p]^T$ ($1 \leq p \leq (N-M)$)),选取新的隐中心矢量;

(4) 计算当前优选的隐中心矢量组合为结构的 RBPNN 在 RLSA 训练收敛后,对于所有训练样本为输入时的分类误差 $e'(i)$.如果 $e'(i)$ 满足给定的误差标准,则终止算法,否则, $i=i+1$,转向执行(2).

5 实验模拟与讨论

5.1 双螺旋分类问题

我们来讨论使用 RBPNN 来分类双螺旋问题. RBPNN 全结构优化使用 MAE- μ GA.假定训练样本集的采样步长为 1,共有 200 个样本点参与训练,即每旋有 100 个采样点组成训练样本集(样本数据的产生见文献[12]).

利用 MAE- μ GA 优化后的 RBPNN 有 30 个隐中心矢量组成,相应的核函数控制参数为 0.0699(核函数控制参数变化如图 3 所示).分类误差曲线如图 4 所示.可以看出,算法在第 29 步迭代时分类误差为 0,即满足构建的 RBPNN 能够正确分类所有训练样本.利用同 MAE- μ GA 优选的隐中心矢量及核函数控制参数训练 RBPNN 收敛后,所形成的分界面如图 5 所示.

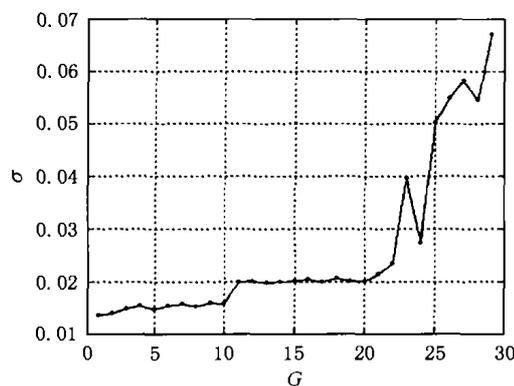


Fig. 3 The genetic curves of the controlling parameters of the kernel function for two spirals discrimination problem by RBPNN optimized with MAE- μ GA.

图 3 RBPNN 分类双螺旋问题时使用 MAE- μ GA 进行结构优化时核函数控制参数(σ)进化曲线

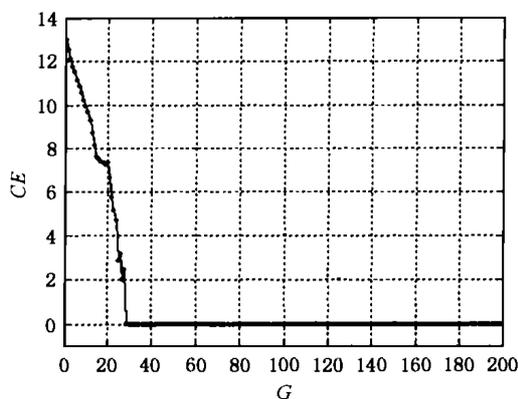


Fig. 4 The variety curve of the classification errors for two spirals discrimination problem by RBPNN optimized with the MAE-μGA.

图4 RBPNN 分类双螺问题使用 MAE-μGA 进行结构优化时的分类误差(CE)曲线

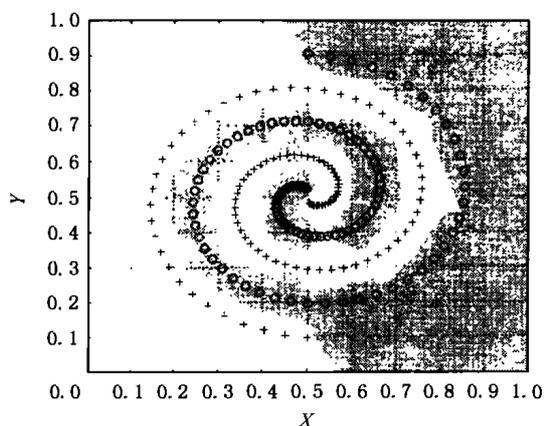


Fig. 5 The classification curves formed by the converged RBPNN optimized by MAE-μGA for two spirals discrimination problem.

图5 使用 MAE-μGA 优化分类双螺问题的 RBPNN 训练收敛后所形成的分界面

在同约束条件下,分别使用 ROLSA^[13]和改进 K-均值算法(modified k-means algorithms, MKA)^[15]来优化分类双螺旋问题的 RBPNN 结构. 由于这两种算法只能用来优选 RBPNN 的隐中心矢量,而不具有求解核函数控制参数的能力,因此,我们使用文献[19]提供的方法来求取控制参数. 即

$$\sigma = \frac{d^*}{\sqrt{N}}, \quad (12)$$

其中, N, d^* 分别为训练样数及样本间的最大欧拉距离.

对于 ROLSA,所有 200 个训练样本都作为初始隐中心矢量. ROLSA 的停止条件为如果所选择的隐中心矢量训练的 RBPNN 不能正确分类所有训练

样本时,算法即停止. 为了尽可能地压缩 RBPNN 的隐中心矢量数目, MKA 只选用每个聚类的中心为 RBPNN 的隐中心矢量. 当 MKA 开始时,每一类训练样本即为一个聚类,随着 MKA 新聚类的增加,由所选的隐中心矢量所构建的 RBPNN 能够正确分类所有训练样本时,立即停止算法,此时,所有聚类中心即是 MKA 的优选结果. 3 种算法的优化结果见表 1. 可以看出, MAE-μGA 的优化效果最好,优化后的 RBPNN 结构最简,而且能够同时提供与所选隐中心矢量相匹配的核函数控制参数,这是 ROLSA 与 MKA 所不具备的.

Table 1 The Performance Compare of the Structure Optimization of RBPNNs by the Three Algorithms for Two Spirals Discrimination Problem

表 1 三种算法对于分类双螺旋问题的 RBPNN 结构优化性能对比

算法	MAE-μGA	ROLSA	MKA
隐中心矢量数	30	34	38
结构压缩率(%)	85	83	81
控制参数值	0.0416	0.0566	0.0566

我们用测试样本集来检测该网络的识别能力. 测试样本集是这样构造的:采样步长由训练集的 1 缩减到 0.1,这样每旋就有 991 个采样点,两旋共有 1982 个采样点组成原始的测试样本集. 然后把 1982 个原始测试样本增加 20 组 0 均值不同方差的高斯白噪声(N_G),即共有 20 组,每组 1982 个样本构成一个测试样本集组合. 为了从统计意义上来检测 RBPNN

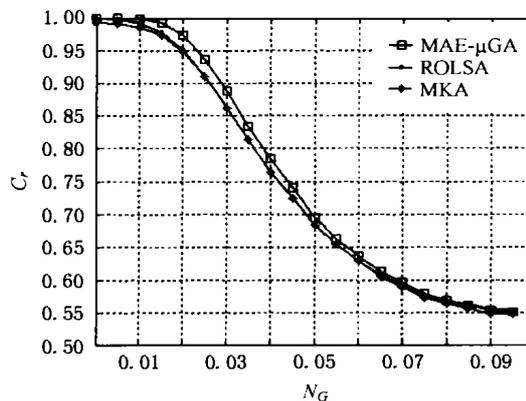


Fig. 6 The generalization performance compare of the RBPNNs optimized by MAE-μGA, ROLSA and MKA for two spirals discrimination.

图6 MAE-μGA, ROLSA 及 MKA 优化后的 RBPNN 对于双螺旋分类问题的推广能力(正确识别率 C_r)比较

的推广能力,克服随机噪声影响,我们按照上述的测试样本集组合的生成方法生成 50 个测试集组合,分别来测试 3 种算法优化后的 RBPNN,具体的测试结果见图 6. 尽管 MAE- μ GA 优化的 RBPNN 结构最简、优化效率最高,但推广能力并没有降低,反而略好于 ROLSA 及 MKA 优化的 RBPNN.

我们还把 MAE- μ GA 用于优化 RBFNN,算法将 RBFNN 的 200 个初始隐中心矢量压缩到 46,相应的核函数控制参数为 0.3313. 可以看出,MAE- μ GA 对 RBFNN 也有很好的压缩效率,尽管结构压缩效率不如 RBPNN 高.MAE- μ GA 优化求解的 RBPNN 及 RBFNN 的核函数控制参数在取值上有显著差异,RBFNN 要明显大于 RBPNN,这是由于两种网络不同的映射机理所致. 由此可见,MAE- μ GA 对于 RBPNN 与 RBFNN 结构优化具有很强的针对性. 另外,我们还使用上述的 50 个测试集组合来测试优化后的 RBFNN 的推广能力,并与 RBPNN 的测试结果进行比较(见图 7). 可以看出,虽然优化后的 RBPNN 要比 RBFNN 在结构上更加简单,但在推广能力上却没有降低反而明显好于优化后的 RBFNN. 由此可见,MAE- μ GA 更加适合 RBPNN 的结构优化.

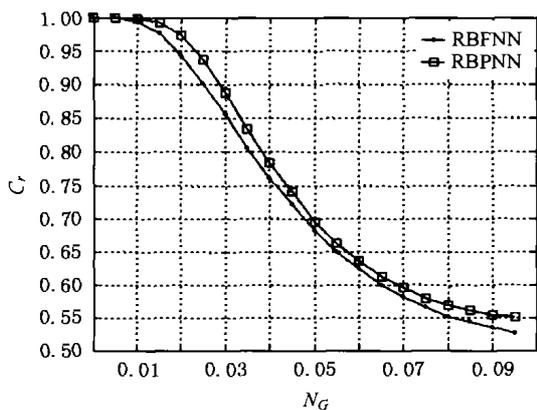


Fig. 7 The generalization performance compare of the RBPNN and RBFNN optimized by MAE- μ GA for two spirals discrimination.

图 7 使用 MAE- μ GA 分类双螺问题的 RBPNN 及 RBFNN 进行结构优化后的网络推广能力(正确识别率 C_r)比较

5.2 IRIS 分类问题

此外,IRIS 分类问题^[20]也被用来检验 MAE- μ GA 的有效性. 文献[20]提供 150 个样本,即每种植物有 50 个样本. 本文将 150 个样本平均划分为 75 个样本的训练集和 75 个样本的原始测试集,即

每种植物的 50 组数据平均划分为两部分. 使用 RBPNN 来完成分类任务,并运用 MAE- μ GA 优化 RBPNN 结构.

MAE- μ GA 优化后的 RBPNN 有 6 个隐中心矢量组成,相应的核函数控制参数为 0.6343(如图 8 所示). MAE- μ GA 的分类误差变化如图 9 所示. 可以看出,算法在第 4 步时分类误差达到 0. 与上例相似,为了对比 MAE- μ GA 的优化性能,同样也使用 ROLSA 和 MKA 来优化 RBPNN 结构,相应的核函数控制参数的求取方法同上例. 优化结果如表 2 所示. 在 3 种算法中,MAE- μ GA 具有最好的优化性能,而且还能够给出与所选的隐中心矢量相匹配的核函数控制参数.

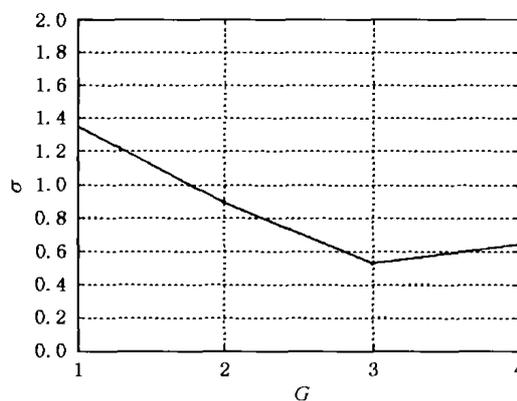


Fig. 8 The genetic curves of the controlling parameters of the kernel function for classifying IRIS problem by RBPNN optimized with MAE- μ GA.

图 8 RBPNN 分类 IRIS 问题时使用 MAE- μ GA 进行结构优化时的核函数控制参数(σ)的进化曲线

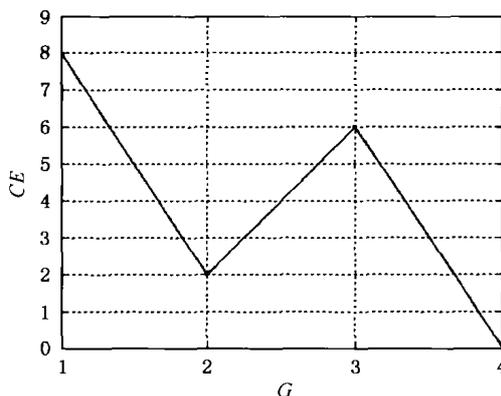


Fig. 9 The variety curve of the classification error for classifying IRIS problem by RBPNN optimized with the MAE- μ GA.

图 9 RBPNN 分类 IRIS 问题时使用 MAE- μ GA 进行结构优化时的分类误差(CE)曲线

Table 2 The Performance Compare of the Structure Optimization of RBPNNs by the Three Algorithms for Classifying IRIS Problem

表 2 MAE- μ GA, ROLSA 及 MKA 对于分类 IRIS 问题的 RBPNN 结构优化性能对比

算法	MAE- μ GA	ROLSA	MKA
隐中心矢量数	6	12	26
结构压缩率/%	92	84	65.3
控制参数值	0.3443	0.5998	0.5998

我们把原始的 75 个测试样本混合成 20 组不同方差、0 均值的高斯白噪声测试样本集。再按照这一方法,重复产生 50 个测试样本集组合,测试后的正识别率的均值如图 10 所示。从测试结果来看,尽管 MAE- μ GA 所优化的 RBPNN 的结构最简,但在推广能力上并没有降低,反而明显好于 ROLSA 及 MKA 的优化后的 RBPNN,这说明我们的算法对 RBPNN 的结构优化是成功的。

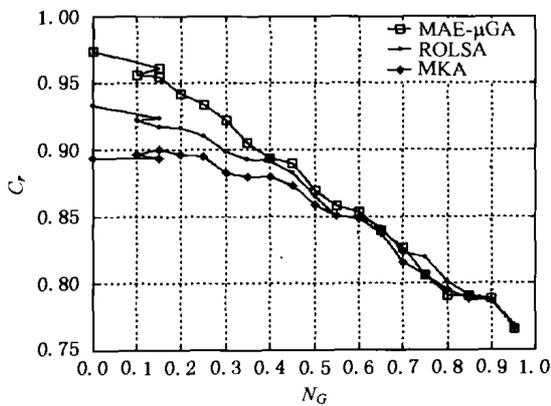


Fig. 10 The generalization performance compare of the RBPNNs optimized by MAE- μ GA, ROLSA and MKA for classifying IRIS problem.

图 10 MAE- μ GA, ROLSA 及 MKA 优化后的 RBPNN 对于 IRIS 分类问题的推广能力(正确识别率 C_r)比较

我们也把本文算法用于优化分类 IRIS 问题的 RBFNN。由 75 个初始隐中心矢量组成的 RBFNN 优化后被压缩到 17 个,优化求取的控制参数为 0.9089。MAE- μ GA 对分类 IRIS 问题的 RBFNN 的结构压缩率为 77.3%。MAE- μ GA 对 RBPNN 与 RBFNN 优化求解的核函数控制参数在数值上有很大的差异,更进一步证明了 MAE- μ GA 具有很强的针对性。我们还把上述的 50 个测试集组合来测试结构优化后的 RBFNN,测试结果与 RBPNN 的比较结果如图 11 所示。可以看出,优化后的 RBPNN 在

推广能力上明显好于优化后的 RBFNN,这说明 MAE- μ GA 更加适合 RBPNN 的结构优化。

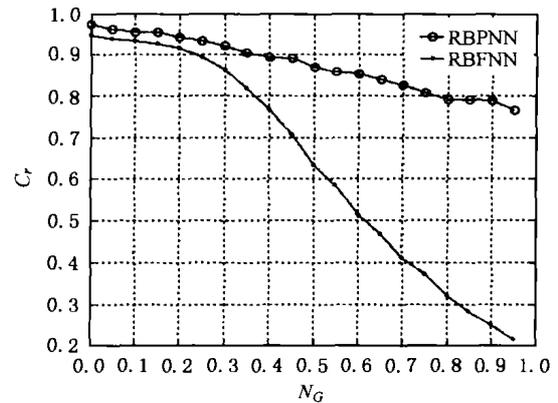


Fig. 11 The generalization performance compare of the RBPNN and RBFNN optimized by MAE- μ GA for classifying IRIS problem.

图 11 使用 MAE- μ GA 对分类 IRIS 问题的 RBPNN 及 RBFNN 进行结构优化后的网络推广能力(正确识别率 C_r)比较

6 小 结

本文提出了一种能够实现 RBPNN 的全结构优化算法,即 MAE- μ GA。该算法在优选隐中心矢量的同时能够实现核函数控制参数的优化取值,在优化效率上,明显好于 ROLSA 及 MKA。而且,由 MAE- μ GA 优化后的 RBPNN 并没有因结构压缩率高而导致推广能力下降。实验还证明了 MAE- μ GA 对 RBFNN 的结构优化也同样适用,但在优化效率及优化后的网络性能方面更加适合 RBPNN 的结构优化。

参 考 文 献

- 1 D. S. Huang. Radial basis probabilistic neural networks: Model and application. *Int'l Journal of Pattern Recognition and Artificial Intelligence*, 1999, 13(7): 1083~1101
- 2 D. Lowe. Adaptive radial basis function nonlinearities and the problem of generalization. *The 1st Int'l Conf. on Artificial Neural Networks*, London, 1989
- 3 D. F. Specht. Probabilistic neural networks. *Neural Networks*, 1990, 3(1): 109~118
- 4 Z. R. Chi. MLP classifiers: Overtraining and solutions. *Proceeding of Int'l Conf. on Neural Networks*, 1995, 5(27): 2821~2824
- 5 R. Reed. Pruning algorithms—A survey. *IEEE Trans. on Neural Networks*, 1993, 4(5): 740~747

- 6 B. George, G. Michael. Feed-forward neural networks. *IEEE Potentials*, 1994, 13(4): 27~31
- 7 M. Freat. The upstart algorithm; A method for constructing and training feed-forward networks. *Neural Computation*, 1990, 2(2): 198~209
- 8 S. Falhman, C. Lebiere. The cascade-correlation learning architecture. In: D Touretzky ed. *Advances in Neural Information Processing System II*. San Francisco: Morgan Kaufmann Publishers, 1990. 524~532
- 9 D. Whitely, C. Bogart. The evolutionary of connectivity: Pruning neural networks using genetic algorithms. *The Int'l Joint Conf. on Neural Networks*, Washington, DC, 1990
- 10 G. Miller, P. Todd, S. Hegde. Designing neural networks using genetic algorithms. *The 3rd Int'l Conf. on Genetic Algorithms*, San Francisco, 1989
- 11 W. B. Zhao, D. S. Huang, L. Guo. Optimizing radial basis probabilistic neural networks using recursive orthogonal least squares algorithms combined with micro-genetic algorithms. *The Int'l Joint Conf. on Neural Networks*, Portland, Oregon, 2003
- 12 W. B. Zhao, D. S. Huang. The structure optimization of radial basis probabilistic neural networks based on genetic algorithms. *The Int'l Joint Conf. on Neural Networks*, Hawaii, 2002
- 13 W. B. Zhao, D. S. Huang. Application of recursive orthogonal least squares algorithm to the structure optimization of radial basis probabilistic neural networks. *The Int'l Conf. on Signal Precessing'02*, Beijing, 2002
- 14 Huang Deshuang. *Systematic Theory of Neural Networks for Pattern Recognition*. Beijing: Publishing House of Electronic Industry. 1996. 49~53 (in Chinese)
(黄德双. *神经网络模式识别系统理论*. 北京: 电子工业出版社, 1996. 49~53)
- 15 J. G. Wilpon, L. R. Rabiner. A modified K-means clustering algorithm for use in isolated work recognition. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 1985, 33(3): 587~594
- 16 David E. Goldberg. Sizing population for serial and parallel genetic algorithms. *The 3rd Int'l Conf. on Genetic Algorithms*, San Mateo, California, 1989
- 17 K. Krishnakumar. Micro-genetic algorithms for stationary and non-stationary function optimization. *SPIE; Intelligence Control and Adaptive System*, Philadelphia, USA, 1989
- 18 Ren Qingsheng, Ye Zhongxing, Zeng Jin, *et al.*. The searching ability of the crossover operator. *Journal of Computer Research and Development*, 1999, 36(11): 1317~1321 (in Chinese)
(任庆生, 叶中行, 曾进, 等. 交叉算子的搜索能力. *计算机研究与发展*, 1999, 36(11): 1317~1321)
- 19 Ferdric M. Ham, Ivica Kostanic. *Principles of Neurocomputing for Science and Engineering*. New York: McGraw-Hill Higher Education, 2001. 140~152
- 20 R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936, 7(2): 179~188



Zhao Wenbo, born in 1972. Ph. D. His research interests include artificial neural networks, pattern recognition, data fusion, etc.

赵温波, 1972年生, 博士, 主要研究方向为人工神经网络、模式识别、数据融合等。



Wang Liming, born in 1972. Master. His research interests include pattern recognition, signal processing, etc.

王立明, 1974年生, 硕士, 主要研究方向为模式识别、信号处理等。



Huang Deshuang, born in 1964. Ph.D., research fellow and Ph. D. supervisor. He received the fund supporting by Hundred Talents Program of Chinese Academy of Sciences. His research interests include artificial neural networks, intelligence computation,

signal processing, pattern recognition, etc. In these research fields, he had published tens of papers.

黄德双, 1964年生, 博士, 研究员, 博士生导师, 中国科学院“百人计划”获得者, 主要研究方向为人工神经网络、智能计算、信号处理、模式识别等, 已经在该领域发表几十篇论文。

Research Background

The study of the radial basis probabilistic neural networks (RBPNN) was supported by Nature Science Fund (No.60173050). The research content of this paper is sub-item of the above project. The study goal of the paper was pursuing fast realization of the structure optimization of the RBPNN. From the viewpoint of the experimental results, we can draw the conclusion that MAE- μ GA, proposed in the paper, was preferred compared with other algorithms. We are affirmative that the RBPNN, optimized by MAE- μ GA, would be widely used to solve real-world problems in future.