

文章编号:1001-9081(2008)12-3052-03

基于小生境技术的神经网络进化集成

於时才,陈涓

(兰州理工大学 计算机与通信学院, 兰州 730050)

(yzzdej@163.com)

摘要:针对目前神经网络集成方法中生成个体网络差异度小、集成泛化能力较差等缺点,提出一种基于小生境技术的神经网络进化集成方法。利用小生境技术在增加进化群体的多样性、提高进化局部搜索能力方面的良好性能,通过个体间相似程度的共享函数来调整神经网络集成中个体网络的适应度,再依据调整后的新适应度进行选择,以维护群体的多样性,得到多样性的个体网络。理论分析和实验结果表明,该方法能有效生成差异度较大的个体网络,提高神经网络集成系统的泛化能力与计算精度。

关键词:进化神经网络;小生境;进化集成;聚类

中图分类号: TP183 **文献标志码:**A

Evolutionary ensemble of neural network based on niche technology

YU Shi-cai, CHEN Juan

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou Gansu 730050, China)

Abstract: In view of the current problems that neural network ensemble generates individual networks with low difference degree and its poor generalization ability, a method of evolutionary ensemble of neural network based on niche technique was proposed. Niche technique's good performance was used in increasing population diversity and improving local search capability of evolution, and the similarity degree's sharing function among individuals was adopted to adjust individual network's fitness. Then the individuals were selected according to the new adjusted fitness to get individual network with diversity. Theoretical analysis and experimental results show that this method can generate individual network with great difference degree and can improve the generalization ability and calculation accuracy of neural network ensemble system.

Key words: evolutionary neural network; niche; evolutionary Ensemble; clustering

0 引言

随着神经网络集成的深入研究和发展,多神经网络集成的性能远远高于单个神经网络已经得到理论和实践的证明。神经网络集成将不同的个体网络组合成一个计算模型,利用多个个体间的差异来提高模型的泛化性能和计算精度。文献[1]在1990年第一次提出了神经网络集成的概念,认为可以简单地通过训练多个神经网络并将其结果进行合成,显著地提高神经网络系统的泛化能力。之后文献[2]给出了神经网络集成泛化误差的分解公式,并表明只要个体神经网络泛化误差均值保持不变,增加差异性可以提高网络的泛化性能。因此在理论上认为构造差异性较大的个体网络,是集成方法最重要的任务。目前针对网络差异性为目的的神经网络集成方法有:遗传算法选择方法^[3]、基于神经网络的分类器集成方法^[4]、应用遗传规划的分类器集成方法^[5]等。

近几年来,进化神经网络被用来作为一种选择最优神经网络结构的方法,在众多领域得到应用。但是由于进化算法是基于多种群的并行搜索方法,进化完全是在多种群的基础上随机进行的,虽然这种随机化杂交形式寻优的初级阶段保持了解的多样性,但在进化的后期,大量个体集中在适应度较高的个体上,使得神经网络越来越趋于相似。基于文献[2]的理论研究,研究人员可以通过使用不同的训练集、不同的网络拓扑结构、不同的网络类型、不同的初始条件及不同的训练

算法来增大网络间的差异。

本文通过研究小生境技术,提出了一种新的集成方法,将可定向小生境技术引入到神经网络进化集成中,让种群中的个体不是聚集在一种环境中,而在不同特定的生存环境中进化。这样可以使算法在整个解空间中搜索,以找到更多的最优个体,生成具有差异性的个体网络,从而提高神经网络集成系统的泛化能力与计算精度。

1 可定向的小生境技术原理

生物学上,小生境(niche)是指特定环境下的一种生存环境。生物在其进化过程中,一般总是与自己相同的物种生活在一起,共同繁衍后代,它们也都是在某一特定的地理区域中生存。小生境技术是将每一代个体划分为若干个类,每个类中选出若干个适应度较大的个体作为一个类的优秀代表组成一个种群,再在该种群以及不同种群之间通过杂交、变异产生新一代个体群。应用于进化算法的小生境机制主要有预选择机制、排挤机制与共享机制。本文建立的小生境模型是基于共享机制^[6]的选择策略,同时可标记进化方向。其基本做法是通过个体之间的相似程度的共享函数来调整群体中各个个体的适应度,从而在群体的进化过程中,算法能够依据这个调整后的新适应度来进行选择,创造出小生境的进化环境,以维护群体的多样性。共享函数是表示群体中两个个体间密切关系程度的函数,其目的就是将搜索空间的多个不同峰值

收稿日期:2008-06-23;修回日期:2008-08-25。 基金项目:甘肃省自然科学基金资助项目(ZS031-A25-015-G)。

作者简介:於时才(1950-),男,甘肃岷县人,教授,主要研究方向:神经计算、模式识别; 陈涓(1981-),女,江苏扬州人,硕士研究生,主要研究方向:神经计算、智能网络、模式识别。

在地理上区分开,每一个峰值处接受一定比例数目的个体,比例大小与峰值高度有关。为了实现这样的分布,共享法将个体的目标适应度降低,即用适应度除以 niche 计数 C_i 得到共享函数,如式(1) :

$$f_{S_i} = f_i / C_i \quad (1)$$

$$Sh(d) = \begin{cases} 1 - \frac{d}{\sigma_{share}}, & d \leq \sigma_{share} \\ 0, & d > \sigma_{share} \end{cases} \quad (2)$$

$$\sigma = \frac{1}{2p^2} \sum_{i=1}^p \sum_{j=1}^p d_{ij} \quad (3)$$

其中 $C_i = \sum_{j \in P_{op}} Sh[d[i, j]]$ 是 niche 计数, f_i 是适应度。 $d[i, j]$ 是个体 i 和 j 的距离, $Sh(d)$ 是共享函数,如式(2)所示, σ_{share} 是 niche 半径,共享半径可由式(3)获得。在距离为 σ_{share} 的范围内的个体彼此消减适应度。因为这些个体 niche 大小相同,因此收敛在一个 niche 内,避免了整个种群的收敛,当一个 niche 填满时,其 niche 计数增大,使共享函数低于其他的 niche。

进化方向的选定可在上面的小生境实现中同步完成,设 x_{ik} 为个体 p_i 的第 k 个决策变量, x_{jk} 为个体 p_j 的第 k 个决策变量。当 $\|p_i - p_j\| < \varepsilon$ 时, $\|p_i - p_j\|$ 是个体 p_i 和 p_j 之间的欧氏距离, ε 为一较小的正数,可以通过比较 $f(p_i)$ 与 $f(p_j)$ ($f(\cdot)$ 为适应度函数), x_{ik} 与 x_{jk} 的大小关系确定 p_i 与 p_j 中更优者的进化方向分量 ε 。若目标函数为 $F = \max f(\cdot)$, 当 $f(p_i) > f(p_j)$ 时, p_i 的进化方向为:

$$\varepsilon = \begin{cases} -1, & x_{ik} < x_{jk} \\ 0, & x_{ik} = x_{jk} \\ 1, & x_{ik} > x_{jk} \end{cases} \quad (4)$$

当 $f(p_i) < f(p_j)$ 时, p_j 的进化方向为:

$$\varepsilon = \begin{cases} -1, & x_{ik} > x_{jk} \\ 0, & x_{ik} = x_{jk} \\ 1, & x_{ik} < x_{jk} \end{cases} \quad (5)$$

2 神经网络进化集成设计

基于小生境技术的神经网络进化集成就是通过在进化的过程中产生差异度较大的个体网络,以此来得到集成精度高、泛化能力强的集成系统。

首先随机初始化神经网络的权值,选择部分数据作为训练数据对神经网络进行训练,以调整神经网络权值,再计算种群的适应度,通过本文提出的基于共享策略的小生境技术,调整种群的适应度。然后依一定比例选择最优的多个个体,通过遗传操作的交叉与变异产生新一代的种群,反复操作直到迭代数超过设定的最大迭代值或者适应度是 1。最后使用训练数据对得到的个体网络进行全面的训练,使用聚类的方法从种群中选择代表性的个体网络,通过绝对多数投票法得到集成结果。具体流程如图 1 所示。

2.1 编码设计

每个神经网络结构都需要表示为一个个体染色体编码,然后才能对其用遗传算法进行优化。在进化算法中,对神经网络进行编码可以使用以下几种方案,如二进制、树形、链表和矩阵表示。

本文使用矩阵表示对神经网络进行编码。设神经网络有 N 个神经元,序号是从 1 到 N 排列的输入层、隐层、输出层节

点。设计一个 $N \times N$ 矩阵表示其结构,如表 1 所示,矩阵中 (i, j) 的元素表示第 i 个神经元到第 j 个神经元的连接关系,在矩阵的右上三角中,“0”表示没有连接,“1”表示有连接。在矩阵的左下三角中, (i, j) 表示相应连接权值,其相对应的网络拓扑结构如图 2 所示,有两个输入节点($I_1 I_2$),三个隐层节点($H_1 H_2 H_3$),一个输出节点(O_1)。

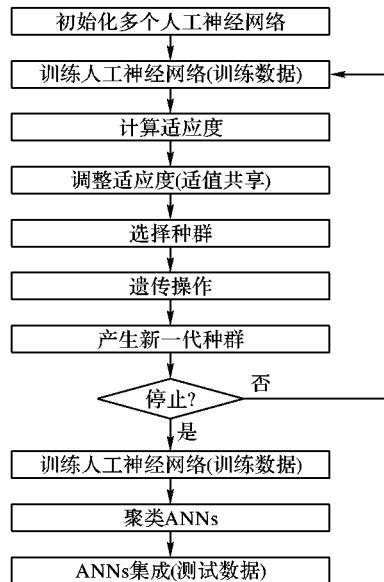


图 1 神经网络进化集成的流程

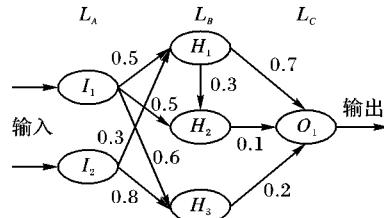


图 2 神经网络结构示意图

表 1 图 2 所示网络中神经元连接关系

from	I_1	I_2	H_1	H_2	H_3	O_1
I_1	0	0	1	1	1	0
I_2	0.0	0	1	0	1	0
H_1	0.5	0.3	0	1	0	1
H_2	0.5	0.0	0.3	0	0	1
H_3	0.6	0.8	0.0	0.0	0	1
O_1	0.0	0.0	0.7	0.1	0.2	0

2.2 适值函数设计

在进化策略中,将所求优化问题的目标函数值作为每个个体的适应值,设 $L = \{a_1, a_2, \dots, a_k\}$ 是一个有 k 个个体网络的神经网络集,训练样本集为 $D = \{(x_1, d_1), (x_2, d_2), \dots, (x_n, d_n)\}$, $x_j \in X^p$, d 是目标期望输出, n 表示样本集的大小,输出为一个 m 维的向量 $[a_{i1}(x), a_{i2}(x), \dots, a_{im}(x)]$ 。则第 i 个神经网络的目标函数定义为:

$$E_i = \frac{1}{N} \sum_{n=1}^N E_i(x) = \frac{1}{N} \sum_{n=1}^N |a_i(x_j) - d_j| \quad (6)$$

BP 神经网络一般以系统总能量函数作为网络训练的评判标准,当网络的总能量 E 小于某一设定值时,网络训练结束。对于每个染色体 X_i , 定义其适应度函数为 $f_i = 1/E_i$ 。

2.3 遗传算子设计

1) 交叉操作。交叉操作通过交换种群中两个父代来获得一个新的个体。首先确定进行遗传操作的父代个体,再以

概率 pc 做交叉操作, 则种群中每次有 $POP_Num \times pc$ 个染色体进行交叉操作。交叉操作的父代采用随机产生的办法。交叉算子定义为:

$$x_i^{n+1} = (1 - \alpha) \times x_i^n + \alpha \times x_j^n \quad (7)$$

$$x_j^{n+1} = (1 - \alpha) \times x_j^n + \alpha \times x_i^n \quad (8)$$

其中 α 为 0 到 1 之间的一个随机数。

2) 变异操作。以概率 pm 做变异操作, 种群中每次都有 $POP_Num \times pm$ 个染色体进行变异操作。变异操作以下面三种方式进行:

- a) 增加隐含层节点数, 随机生成相应的权值;
- b) 删除某些连接, 也就是将权值设为 0;
- c) 删除隐含层中的某些节点和相应的连接。

进行多次变异, 直到变异后新个体适应值优于父代个体或者达到指定的变异次数为止。然后对每次变异结果还原出神经网络并进行性能评估, 如果后代优于父代, 则停止变异; 否则对该父代进行下一次变异, 直到找出优于父代的后代。

2.4 结论集成

进化结束后获得的多样性的种群, 首先通过分层次聚类的方法将种群分类, 再分别选取各个类别中的一个个体, 通过使用绝对多数投票法(某分类成为最终结果当且仅当有超过半数的神经网络输出结果为该分类)来得到最终的输出。

分层次聚类就是将样本中每个个体看成一类, 若有 C 个神经网络, 则分层聚类将种群样本初始化分为 C 类, 然后寻找“距离”最近的两个神经网络结合, 样本被划分为 $C - 1$ 类, 同样的方法应用于 $C - 1$ 类的样本中, 依次计算直到全部样本被划分为一类。两类间的距离表达式如下: $d(C_i, C_j) = \min_{i \in C_i, j \in C_j} d_{ij}, d_{ij} = \|p_i - p_j\|$ 是个体 p_i 和 p_j 之间的欧氏距离。

图 3 即为一个分层聚类的树状图示例, 横轴代表 8 个个体网络, 竖轴代表个体网络的相似度。

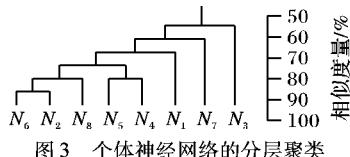


图 3 个体神经网络的分层聚类

神经网络的集成是通过多个个体神经网络对同一个问题进行学习而实现的, 对于某个输入, 它在集成上的输出是由集成系统中多个网络共同决定的^[7]。当神经网络集成用于分类器时, 集成的输出由个体网络的输出投票产生, 通常采用绝对多数投票法, 即当超过半数的个体神经网络的分类是同一分类, 则该分类将作为最后的输出。

3 实验分析

为了测试本文提出的集成方法的可行性和有效性, 采用了 UCI 机器学习知识库下面的两个标准数据集。首先使用 Iris 数据集介绍本文方法的工作流程; 然后再使用 pima-indians-diabetes 数据集做进一步探讨, 验证基于小生境的进化集成能够得到差异性更大的个体网络, 同时将本文的方法与以往的进化神经网络进行比较。在每个数据集中, 使用 60% 的数据作为 NN 的训练数据, 40% 的数据用于对得到的集成系统进行测试。

Iris 数据集由 150 组数据组成, 分成三类 (Setosa, Versicolour, Virginica), 每一类 50 组。每组数据有四个属性,

分别是萼片与花瓣的长度和宽度。Pima-indians-diabetes 数据集由 768 组数据组成, 分成 2 类, 其中 positive 为 268 组, negative 为 500 组, 每组数据有 8 个属性。遗传算法的参数设置如下: 种群大小为 20, 最大进化代数是 150, 交叉率与变异率分别是 0.6 和 0.1, 选择率为 0.5。神经网络设置如下: 个体网络是前馈神经网络, 使用 BP 算法进行训练, 输入节点数即为数据集的属性值, 隐层节点最大值是 7, 输出节点数是数据集的类数。通过两个数据集的实验仿真, 分别得到个体网络差异性较大的集成系统, 其对训练数据的识别率较高, 如表 2、3 所示。

表 2 Iris 数据的分类效果

数据集	训练数据	错误识别	准确率/%
Setosa	30	0	100.00
Virginica	30	2	93.30
Versicolor	30	1	96.70
Overall	90	3	96.67

表 3 Diabetes 数据的分类效果

数据集	训练数据	错误识别	准确率/%
positive	161	6	96.3
negative	300	5	98.3
Overall	461	11	97.6

进一步对 40% 的测试数据进行分类识别, 并与未引入小生境的进化集成系统比较, 图 4 展示了两种进化集成的多样性分析。多样性的测量是通过种群中所有个体的平均欧氏距离来获得的。

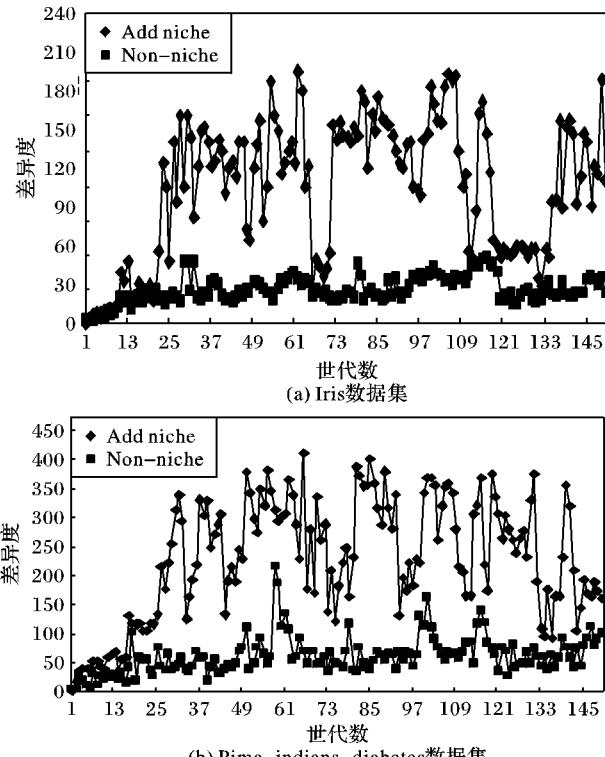


图 4 引入小生境的进化集成多样性分析

在图 4 中, 与一般的进化集成比较, 本文基于小生境技术的进化集成保持着较高的多样性。表 4 是两种进化集成的对比实验结果, 对于未被神经网络使用的 40% 的测试数据, 本文方法的分类准确率高于一般的进化集成方法。

(下转第 3083 页)

零,由以上证明和引理不难得出 $p_N^w \leq p_N$ 这一结论。

证毕。

定理 1 表明,就最优分类器而言,如果用正确分类概率或错误概率作为文本分类器的性能评估指标,特征权向量起不到改善分类效果的作用,反而会增加计算工作量。正确的选择是不使用特征权向量或者说采用均权,即令 $w = (1, 1, \dots, 1)$ 。

3 特征权对 BEP 值的影响

下面讨论两类分类问题,记号不变。先考虑特征权向量对应的对角矩阵为非奇异矩阵时的情况。

当特征权为均权,且特征向量经过规范化变换后,贝叶斯分类器犯第一类的概率为:

$$e_1 = P_1 p(\Theta_2 \mid \omega_1) \quad (25)$$

而用特征权向量 $w = (w_1, w_2, \dots, w_d)$ 作用于特征向量 x 并进行规范化处理之后,贝叶斯分类器犯第一类的概率为:

$$e_1^w = P_1 p(\Theta_2^w \mid \omega_1) \quad (26)$$

当对角矩阵 W 为非奇异矩阵时, $\Theta_i = \Theta_i^w, i = 1, 2$, 显然有:

$$e_1 = e_1^w \quad (27)$$

由式(9)、(11)知,特征权向量 w 对微平均 BEP 值、宏平均 BEP 值均不可能产生有效影响。

当特征权向量对应的对角矩阵为奇异矩阵时,由推论知:

$$P_1 e_1 + P_2 e_{\text{II}} \leq P_1 e_1^w + P_2 e_{\text{II}}^w \quad (28)$$

考虑到 BEP 值的计算应满足式(6),即:

$$e_1 / e_{\text{II}} = P_2 / P_1 \quad (29)$$

和:

$$e_1^w / e_{\text{II}}^w = P_2 / P_1 \quad (30)$$

不难推出:

$$e_1 \leq e_1^w \quad (31)$$

由式(9)、(11)知,特征权向量 w 只可能降低微平均 BEP 值和宏平均 BEP 值。

(上接第 3054 页)

表 4 引入小生境与未引入的进化集成方法比较

数据集	测试数据	错误识别		准确率/%	
		Add niche	Non-niche	Add niche	Non-niche
Setosa	20	2	2	90.0	90.0
Virginica	20	3	5	85.0	75.0
Versicolor	20	1	4	95.0	80.0
Positive	107	8	14	92.5	86.9
Negative	200	6	11	97.0	94.5

4 结语

神经网络集成因为能显著提高神经网络系统的泛化能力和精度,已经成功应用于很多领域,然而进化算法的应用都是建立在使用进化的最优一代的理念上,虽然取得了较小的训练误差,但却失去了种群的多样性,得不到较好的泛化能力。本文提出的基于可定向小生境技术的神经网络进化集成,通过个体之间的相似程度来调整群体中各个个体的适应度,以维护群体的多样性,使得集成中的个体网络具有较大的差异,从而获得泛化能力较强集成系统。通过对 UCI 下两个标准数据集的实验仿真,证明了本方法的可行性和有效性。通过与没有引入小生境技术的进化集成进行比较,验证了本方

综合以上讨论有以下定理。

定理 2 在文本分类中,特征权向量不能有效提高最优分类器——贝叶斯分类器的微平均 BEP 值以及宏平均 BEP 值。

4 结语

由定理 1、定理 2 可以看出,在类先验概率和类条件概率完全已知,并采用贝叶斯最优分类器进行分类的前提下,特征权并不能有效提高文本分类效果。在实际应用中,需要依据训练样本和各种分类算法对上述诸概率进行直接或间接的估计,以至于特征权对文本分类器的分类性能有着或多或少的影响。

参考文献:

- [1] SEBASTIANI F. Machine learning in automated text categorization [J]. ACM Computing Surveys, 2002, 34(1): 1–47.
- [2] SALTON G, MCGILL M J. An introduction to modern information retrieval [M]. [S. l.]: McGraw-Hill, 1983.
- [3] SHANKAR S, KARYPIS G. A feature weight adjustment algorithm for document categorization [C]// 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2000.
- [4] DEBOLE F, SEBASTIANI F. Supervised term weighting for automated text categorization [C]// SAC 03: 18th ACM Symposium on Applied Computing. New York: ACM, 2003: 784–788.
- [5] 宋枫溪,陈才扣,刘树海,等. 文本表示方式对线性支持向量机分类性能的影响[J]. 模式识别与人工智能,2004,17(2):161–166.
- [6] 宋枫溪,高林. 文本分类器性能评估指标[J]. 计算机工程,2004,30(13):107–109.
- [7] YANG YI-MING. An evaluation of statistical approaches to text categorization [J]. Information Retrieval, 1999, 1(1): 69–90.
- [8] 边肇祺,张学工. 模式识别 [M]. 北京: 清华大学出版社,2000.

法在解决个体网络差异性问题的有效性。

参考文献:

- [1] HANSEN L K, SALAMON P. Neural network ensembles [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(10): 993–1001.
- [2] KROGH A, VEDELBY J. Neural network ensembles, cross validation, and active learning [C]// Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 1995: 231–238.
- [3] ZHOU Z H, WU J X, TANG W. Ensembling neural networks: Many could be better than all [J]. Artificial Intelligence, 2002, 137(1/2): 239–263.
- [4] 李凯,黄厚宽. 一种提高神经网络集成差异性的学习方法[J]. 电子学报,2005,33(8): 1387–1390.
- [5] IMAMURA K, SOULE T, HECKENDORN R B, et al. Behavioral diversity and a probabilistically optimal GP ensemble [J]. Genetic Programming and Evolvable Machines, 2003, 4(3): 235–253.
- [6] 崔明义. 浮点数编码小生境遗传算法的研究[J]. 计算机科学,2007,34(14): 225–228.
- [7] 周志华,陈世福. 神经网络集成[J]. 计算机学报,2002,25(1): 1–8.