

径向基函数神经网络的软竞争学习算法

张志华, 郑南宁, 史 罡

(西安交通大学人工智能与机器人研究所, 陕西西安 710049)

摘要: 本文构造了径向基函数(RBF)神经网络的一类软竞争学习算法(SCLA). 该算法的主要思想是首先在高斯基函数中心向量的训练过程中引入了隶属度函数, 对每个输入样本, 所有中心向量根据该样本属于其代表的类的隶属度值的大小进行自适应地调整; 第二, 把隶属度函数的模糊因子的倒数与模拟退火算法中的温度等同起来, 在迭代过程中采用递增的方式来调整它. SCLA是RBF网络基于 k -均值方法训练中心向量的学习算法的软竞争格式, 它可以克服后者对初始值敏感和死节点的问题. 仿真实验论证了SCLA是有效的.

关键词: 径向基函数; 软竞争学习; 模糊因子; 模拟退火

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112(2002)01-0132-04

Soft-Competition Learning Algorithms of the Radial Basis Function Neural Networks

ZHANG Zhi-hua, ZHENG Nan-ning, SHI Gang

(Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China)

Abstract: In this paper, the soft-competition learning algorithms (SCLA) of RBF neural networks are designed. The main ideas of the algorithms are: firstly, membership functions are introduced into the training procedure of the center vectors of the Gaussian basis functions, and for each input sample, all center vectors are self-adaptively adjusted according to the values of the membership functions, in what degree the sample belongs to the classes that the center vectors represent; secondly, the reciprocal of the fuzzy factor of membership function are considered as the temperature of the simulated annealing algorithm, and increasingly adjusting method is used to the fuzzy factor during the learning procedure. SCLA are soft-competition schemes of the learning algorithms, in which the center vectors are trained based on the k -means algorithm, and can remedy the problems of the dead-node and the sensitivity to initial weight vectors that the latter algorithms have. The simulation results show that SCLA are efficient.

Key words: radial basis function; soft-competition learning; fuzzy factor; simulated annealing

1 引言

径向基函数(Radial Basis Function, RBF)神经网络是近几年来继多层感知机(Multi-Layer Perceptrons, MLP)之后又一种十分有效的多层前向网络. RBF网络输出节点计算隐节点给出的基函数的线性组合, 隐层的输出函数(即基函数)通常由高斯函数定义. 对于一个基函数的类型及个数选定好的网络, 为了能使其正确执行期望的数据处理, 有两个参数需要确定: 一是基函数的中心在输入样本数据空间中的位置; 二是隐含层到输出层的权向量集合. 因此一个RBF网络的训练可以分为两个阶段: 基函数中心的选取, 输出层权向量的确定.

目前, 第一阶段的训练方法有基于 k -均值的聚类法^[1], Kohonen聚类法^[2], 正交最小平方二乘算法^[3], 梯度下降法^[4]以及对每类样本协方差的Gram-Schmidt正交化法^[5]等. 第二阶段的训练方法主要有求目标函数最小的梯度下降法^[1-4], 最近Kaminski等提出了基函数的标准Gram-Schmidt正交化训

练法^[6]等.

本文的工作主要集中于第一阶段. k -均值和Kohonen法都属于Winner-Take-All, 即每次迭代只对获胜者进行调整. 它们计算简便, 但同时暴露出对初始值的选取具有大的敏感性等缺点. 误差回传算法虽然学习精度和泛化能力都较强, 但中心向量以及它相应宽度的梯度的计算是十分复杂的, 这使得学习速度较慢. 事实上, RBF网络的高斯隐含层同Kohonen自组织网络的竞争层相比, 它的优点在于样本对不同的节点都有响应, 只不过获胜节点响应程度要高于其它节点. 本文把这一思想利用去训练中心, 讨论了中心学习的一种软竞争策略. 文章其余部分是如下安排: 第二节给出RBF网络的一般结构; 第三节研究了RBF网络的软竞争学习算法及其有关性质; 在第四节把本文提出的算法用于一个著名时间序列仿真试验中; 最后, 总结了全文的工作以及今后的进一步工作的方向.

收稿日期: 1999-01-22; 修回日期: 2001-01-04

基金项目: 国家自然科学基金重点基金(No. 69735010); 博士点专项基金(No. 98069825)

2 径向基函数(RBF)网络

设训练样本集 $X = \{x_1, x_2, \dots, x_p\}$, p 为样本个数, 其中第 i 个输入样本 $x_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}^T \in R^n$; $y_i = \{y_{i1}, y_{i2}, \dots, y_{im}\} \in R^m$ 和 $o_i = \{o_{i1}, o_{i2}, \dots, o_{im}\} \in R^m$ 分别是网络对应 x_i 的实际输出和期望输出. 设网络隐含层的节点个数有 c 个, w_{jl} 表示第 l 个隐节点到第 j 个输出节点的连接权值, v_{lk} 表示第 k 个输入节点到第 l 个隐节点连接权值.

RBF 网络隐层的输出函数采用高斯基函数, 它对输入激励产生一个局部化的响应, 这个特点使高斯隐层对输入样本有一个聚类的作用, 隐层节点数代表聚类的类数, 隐层的中心为这一类的凝聚中心点. 基函数的形式通常为

$$R_l(x) = \frac{1}{2} \exp\left[-\frac{\|x - v_l\|^2}{2\sigma_l^2}\right] \quad (1)$$

其中 $v_l = (v_{l1}, v_{l2}, \dots, v_{ln})$, $\sigma_l^2 (l=1, 2, \dots, c)$, 分别是第 l 个隐单元响应函数的中心向量和宽度. 当输入样本为 x_i 时, 网络的实际输出为

$$y_{ij} = f(x_i) = \sum_{l=1}^c w_{jl} R_l(x_i), j=1, 2, \dots, m \quad (2)$$

或是采用正则化的形式计算:

$$y_{ij} = f(x_i) = \frac{\sum_{l=1}^c w_{jl} R_l(x_i)}{\sum_{l=1}^c R_l(x_i)} = \sum_{l=1}^c w_{jl} \rho_l(x_i), j=1, 2, \dots, m. \quad (3)$$

$$\text{其中, } \rho_l(x_i) = \frac{R_l(x_i)}{\sum_{l=1}^c R_l(x_i)} \quad (4)$$

RBF 网络的基本学习算法由三部分组成: 初始化网络结构和学习参数; 训练权向量 v_l , 计算 ρ_l 的值; 调整权向量 w_j . 通常采用 k -均值方法来训练中心向量 v_l :

$$v_l = \frac{1}{n_l} \sum_{x \in X_l} x, l=1, 2, \dots, c \quad (5)$$

其中 $X = X_1 \dots X_c$ 是由最近邻规则定义的 X 的一个硬 k -分割, $n_l = |X_l|$ 表示第 l 类的容量, 下文中的 n_l 的意义相同. 一般情况下, 宽度 σ_l 表示第 l 类的分散度

$$\sigma_l^2 = \frac{1}{n_l} \sum_{x \in X_l} (x - v_l)^T (x - v_l) \quad (6)$$

另外, 网络的权向量 $w_j = (w_{j1}, w_{j2}, \dots, w_{jm}) (j=1, 2, \dots, m)$ 的学习是基于求由下式定义的误差函数 E_i 的最小值的梯度下降法导出的, 它的第 $t+1$ 次学习规则由式(8)定义.

$$E_i = \frac{1}{2} \sum_{k=1}^m (y_{ik} - o_{ik})^2, i=1, 2, \dots, p \quad (7)$$

$$w_{jl}(t+1) = w_{jl}(t) + \eta(t) \frac{\partial E_i}{\partial w_{jl}} \quad (8)$$

$$\text{其中, } \frac{\partial E_i}{\partial w_{jl}} = (y_{ij} - o_{ij}) \rho_l(x_i), \eta(t) \text{ 为学习步长.} \quad (9)$$

3 RBF 网络的软竞争学习算法

k -均值聚类算法不仅计算简单, 而且性能良好, 但是用它来训练 RBF 网络的中心向量时一方面对初始值的选取敏感, 同时存在死节点(即无任一样本与之距离最近)问题, 此时, 中心向量 v_l 和宽度 σ_l 无法计算(因为式(5)和(6)的分母为零).

针对这种情况, 本文提出了 RBF 网络的一种软竞争学习算法(Soft-competition learning algorithm, SCLA). 该算法采用两阶段分别训练中心向量和输出层权向量 w_j , 权向量 w_j 仍然用梯度下降法进行训练, 而中心向量的调整则是利用一种软竞争方法.

3.1 软竞争学习算法(SCLA)

SCLA 的具体学习步骤如下:

SCLA1. 确定基函数的个数 c 和最大迭代次数 T ; 初始化中心 $v_l(0) = (v_{l1}(0), \dots, v_{ln}(0))$.

SCLA2. For $t=1$ to T

(a) 修正 (t) ,

(b) For $i=1, 2, \dots, p$

for $l=1, 2, \dots, c$

计算 $P_{li}(t)$

$$P_{li}(t) = \frac{\exp\left(-\frac{\|x_i - v_l(t-1)\|^2}{2\sigma_l^2}\right)}{\sum_{l=1}^c \exp\left(-\frac{\|x_i - v_l(t-1)\|^2}{2\sigma_l^2}\right)} \quad (10)$$

按下式调整 $v_l(t)$

$$v_l(t) = v_l(t-1) + \eta(t) P_{li}(t) (x_i - v_l(t-1)) \quad (11)$$

SCLA3. 计算 $E1(t+1) = \sum_{j=1}^m \sum_{i=1}^p |v_{ji}(t+1) - v_{ji}(t)|$.

SCLA4. 如果 $E1(t+1) \geq \epsilon$ 且 $t < T$, 令 $t = t+1$, 转入 SCLA2; 否则按照下式(12)计算 σ_l^2 , 进入网络的第二阶段学习, 即根据式(8)训练输出层权向量 w_j .

$$\sigma_l^2 = \frac{\sum_{i=1}^p P_{li}(t) (x_i - v_l)^T (x_i - v_l)}{\sum_{i=1}^p P_{li}(t)} \quad (12)$$

其中, 步长 $\eta(t)$ 一般根据 Kohonen 网络的收敛性^[2]来选取. 上述算法与由 Yair 等提出改进学习矢量量化的软竞争格式^[7]相似. 由式(10)可知, $0 < P_{li}(t) < 1$ 且 $\sum_{l=1}^c P_{li}(t) = 1$, 因而可以称 $P_{li}(t)$ 为第 i 个样本 x_i 属于第 l 类的模糊隶属度函数, $\eta(t)$ 是表征其模糊程度的一种因子, 这里称之为模糊因子. 所以 v_l 的调整方法是一种模糊自适应方法. 在每次迭代中, 所有隐节点的中心都同时进行调整, 这可以克服死节点问题.

SCLA 的关键是如何选取 $\eta(t)$, 我们发现上述训练中心向量的算法是 Gibbs 抽样的确定性的等价类, 类似于模拟退火算法^[8], 其中 $\frac{1}{\eta(t)}$ 可看作温度. 因而可以基于模拟退火算法的思想来选取 $\eta(t)$: 开始设置高的温度, 随着时间的变化, 温度逐渐降低. 也即 $\eta(t)$ 是迭代次数 t 的递增函数, 且满足 $\lim_{t \rightarrow +\infty} \eta(t) = 0$. 在本文下面的仿真实验中, 我们定义 $\eta(t)$ 为

$$\eta(t) = \eta_0 + \frac{T - t}{T} \eta_0 \quad (13)$$

上述的算法是一种序列训练方法, 它的每次调整与当时输入样本有关. 因而样本的输入顺序对其是有直接影响. k -均值方法的优点在于它是一种批处理方式, 同样我们可以定义上述算法相应的批处理方式. 首先考虑 v_l , 它的批处理形式如下定义:

$$v_l(t) = v_l(t-1) + \eta(t) \frac{\sum_{i=1}^p P_{li}(t) (x_i - v_l(t-1))}{\sum_{i=1}^p P_{li}(t)} \quad (14)$$

权向量 w_j 的批处理学习是基于网络整个误差的平均函数 $E = \frac{1}{P} \sum_{i=1}^P E_i$ 的最小化而导出的

$$w_{jl}(t+1) = w_{jl}(t) + \frac{(t)}{P} \sum_{i=1}^P (y_{ij} - o_{ij}) \cdot i(x_i) \quad (15)$$

3.2 软竞争学习算法的性质

由于在中心向量的学习和宽度的计算中引入了模糊隶属度函数,它极大地丰富了 SCLA 的性能,这节我们从理论上讨论 SCLA 的性质. 首先有下面的基本定理:

定理 1

$$(1) \lim_{(t) \rightarrow 0} P_{li}(t) = \begin{cases} 1, & l = \arg \min_r x_i - v_r(t-1); \quad i = 1, \dots, c; \\ 0, & \text{otherwise} \end{cases}$$

$$(2) \lim_{(t) \rightarrow +0} P_{li}(t) = \frac{1}{c}, l = 1, 2, \dots, c.$$

定理 1 表明 $(t) \rightarrow 0$ 当模糊程度最大,而 (t) 就是通常的硬格式,这正是我们称 (t) 为模糊因子的原因. 由定理 1, 可以立即得到 SCLA 有下面的性质:

定理 2 若中心向量 $v_l(l = 1, 2, \dots, c)$ 的学习规则由 (11) 定义, 则

$$(1) \lim_{(t) \rightarrow +1} v_l(t) = \begin{cases} v_l(t-1) + \frac{(t)}{P} \sum_{i=1}^P (x_i - v_l(t-1)), & l = \arg \min_r x_i - v_r(t-1); \\ v_l(t-1), & \text{其它} \end{cases}$$

$$(2) \lim_{(t) \rightarrow +0} v_l(t) = v_l(t-1) + \frac{(t)}{c} \sum_{i=1}^P (x_i - v_l(t-1)), l = 1, 2, \dots, c$$

定理 3 若按式 (12) 计算宽度, 则

$$(1) \lim_{(t) \rightarrow 0} \frac{2}{l} = \frac{1}{n_{l \times x}} \sum_{i=1}^P (x_i - v_l)^T (x_i - v_l), l = 1, 2, \dots, c;$$

$$(2) \lim_{(t) \rightarrow +0} \frac{2}{l} = \frac{1}{P} \sum_{i=1}^P (x_i - v_l)^T (x_i - v_l), l = 1, 2, \dots, c;$$

定理 4 若采用批处理 (14) 去训练中心向量, 则

$$(1) \lim_{(t) \rightarrow 0} v_l(t) = v_l(t-1) + \frac{(t)}{n_{l \times x}} \sum_{i=1}^P (x_i - v_l(t-1));$$

$$(2) \lim_{(t) \rightarrow +0} v_l(t) = v_l(t-1) + \frac{(t)}{P} \sum_{i=1}^P (x_i - v_l(t-1)), i = 1, 2, \dots, c$$

当学习率 $(t) = 1$ 时, 根据式 (14)

$$v_l(t) = v_l(t-1) + \frac{\sum_{i=1}^P P_{li}(t) (x_i - v_l(t-1))}{\sum_{i=1}^P P_{li}(t)} = \frac{\sum_{i=1}^P P_{li}(t) \cdot x_i}{\sum_{i=1}^P P_{li}(t)} \quad (16)$$

于是立即有下面的结论

推论 若采用批处理 (14) 去训练中心向量, 且令学习率 $(t) = 1$, 则

$$(1) \lim_{(t) \rightarrow 0} v_l(t) = \frac{1}{n_{l \times x}} \sum_{i=1}^P x_i;$$

$$(2) \lim_{(t) \rightarrow +0} v_l(t) = \frac{1}{P} \sum_{i=1}^P x_i, l = 1, 2, \dots, c$$

定理 3 和推论表明基于 (5)、(6) 训练中心向量和计算宽度的 k -均值方法是本文提出的算法 (t) 当的极限, 因而是本文算法的硬格式, 本文的算法是它的软竞争格式.

上节我们基于模拟退火的思想来选取 (t) , 事实上, 这节的几个定理表明当随着迭代次数 t 增加从零递增到无穷, 由于模糊隶属度函数 $P_{li}(t)$ 从模糊度最大变到最小 (即清晰), 相应的 SCLA 从模糊变清晰, 这符合人类认识世界的过程.

4 仿真实验与结果分析

这里我们用著名的 Mackey-Glass 混沌时间序列去检验本文提出的算法的性能. 混沌 Mackey-Glass 微分延迟方程是如下定义:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t)$$

当 $\tau \geq 17$ 时, 上述方程会产生混沌现象. 这里我们取 $\tau = 17$, 用欧拉格式

$$x(t+1) - x(t) = \frac{0.2x(t-17)}{1+x^{10}(t-17)} - 0.1x(t)$$

离散上述方程, 产生 600 个数据 (如图 1 所示). 产生 596 对输入-输出数据 $[x(k-1), x(k), x(k+1), x(k+2); x(k+3)] (k = 1, 2, \dots, 596)$, 前 298 个数据对被用来去训练网络, 而其余 298 个数据对用来去检验网络的预测性能. 在实验中, 一些参数的取值是: $c = 20, (0) = 0.6, (0) = 0.04$. 在这个例子中, (t) 由式 (13) 定义, 其中最大迭代次数 $T = 1000, \theta_0 = 2.5, \tau = 1000$. 图 2 是用基于 k -均值算法的 RBF 网络进行实验的结果, 在实验中我们把输入-输出训练样本集中前 c 对输入-输出作为算法的初始权值, 而当随机初始化权值时, 算法溢

表 1 两种算法的学习和预测结果

学习算法	迭代次数	总的学习误差	绝对预测误差和	平均预测误差
k -均值	$t_1 = 69, t_2 = 10000$	0.2276	8.2436	0.0278
SCLA1	$t_1 = 326, t_2 = 10000$	0.1437	7.3439	0.0248
SCLA2	$t_1 = 209, t_2 = 10000$	0.0193	2.8875	0.0097

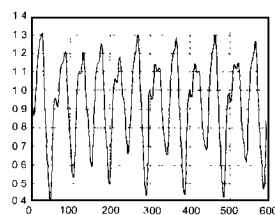


图 1 时间序列的 600 个样本

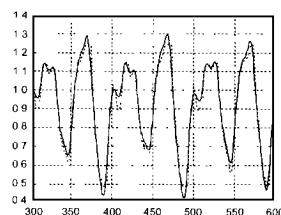


图 2 基于 k -均值的学习算法的实验结果

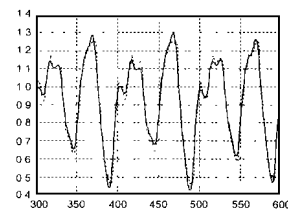


图 3 SCLA1 的实验结果

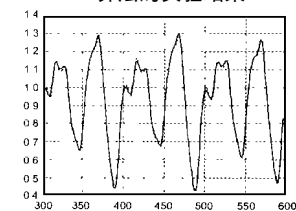


图 4 SCLA2 的实验结果

出,这是由于算法存在死节点.图3是用SCLA实验的结果,其中权值初始化方法同上(表1中的SCLA1);而图4是SCLA随机赋权值(表1中的SCLA2)的实验结果.图中实线表示期望输出,虚线代表网络的预测值.实验中的一些主要指标见表1.从图和表中可看出基于软竞争学习的RBF网络其性能要优于基于 k -均值的RBF网络.SCLA可以克服死结点问题,并且随机赋权值的结果相对而言还要好.表中的 t_1 、 t_2 分别表示学习算法中的第一和第二阶段的迭代次数.

5 结论

本文研究了RBF网络的软竞争学习算法.该算法主要的优点体现在两个方面:输入层到隐含层的连接权向量的学习规则不是基于 k -均值法的胜者为王的策略,而是基于一种软竞争格式,即引入了模糊度函数到中心向量的训练过程中,对每个输入样本,所有中心向量根据其隶属度值的大小都进行自适应地调整,从而可以克服 k -均值法中的对初始值敏感和死节点的问题,而且比基于误差回传的梯度下降法计算量要小;在软竞争学习算法中引入模拟退火机制,算法在一定程度上可以避免收敛到局部最小.

参考文献:

- [1] Bezdek J C. Pattern recognition with fuzzy objective function algorithm [M]. New York:Plenum,1981.
- [2] Kohonen T. Self-organization and associative memory [M]. Berlin,Germany:Springer-Verlag,1989,3rd. ed.
- [3] Chen S,et al. Orthogonal least squares learning algorithm for radial basis function networks [J]. IEEE Trans. on Neural Networks,1991,2(2):302-309.
- [4] Tarassenko B,et al. Supervised and unsupervised learning in radial basis function classifiers [J]. Vision Image Signal Processing,1994,141(1):210-216.
- [5] Musavi M F,et al. On the training of radial basis function neural networks [J]. Neural Networks,1992,5(3):595-603.
- [6] Kaminski W,et al. Kernel orthonormalization in radial basis function neural networks [J]. IEEE Trans. on Neural Networks,1997,8(5):1177-1183.
- [7] Yair E. Competitive learning and soft competition for vector quantizer design [J]. IEEE Trans. on Signal Processing,1992,40(2):294-308.
- [8] Kirkpatrick S. Optimization by simulated annealing [J]. Science,1983,220(3):671-680.

作者简介:



张志华 男,1970年5月出生于湖北.博士生.主要从事神经计算、统计模式识别及Bayesian统计学等领域的研究.现已在国内外各种重要期刊和会议上发表论文20余篇. e-mail:zhhaa@aiar.xjtu.edu.cn

郑南宁 男,博士,教授,中国工程院院士.主要从事计算机视觉、图像处理和模式识别等领域的理论及应用研究.现已在国内外各种重要期刊和会议上发表论文100余篇.