

基于卡尔曼滤波算法和双径向转移函数的RBF型神经网络

宋绍云

(玉溪师范学院 信息技术工程学院, 云南 玉溪 653100)

摘要: RBF径向基函数神经网络具有训练简洁、学习效率快、不易陷入局部极小等优点, 广泛应用于信号处理与模式识别。虽然常用的RBF网络比较容易构建, 但因其结构通常固定或者复杂度较高, 从而导致学习时间过长或网络资源的浪费。针对上述原因, 提出利用扩展卡尔曼滤波器作为RBF的学习算法, 并在隐层中使用双径向函数。通过对逼近基准的结果分析, 清楚地表明该算法比其他分类网络模型具有更强的泛化性。

关键词: RBF神经网络, 卡尔曼滤波, 转移函数, 径向基函数

中图分类号: TP183

文献标识码: A

The RBF neural network based on the Kalman filter algorithm and biradial transfer function

SONG Shao Yun

(College of Information Technology Engineering, Yuxi Normal University, Yuxi 653100, China)

Abstract: RBF neural network, with simple training, high efficiency of study, and be not easy to lost in local minimum, etc., is widely used in signal processing and pattern recognition. Although the commonly used RBF network is easier to built, their fixed structural or high complexity, usually causes too long learning time or a waste of network resources. For the above reasons, the extended Kalman filter is proposed as RBF learning algorithm, and the biradial function is used in hidden layer. By approaching the results of the benchmark analysis, we make it clear that this alorithm is more generalized nature than any other network model.

Keywords: RBF neural network; Kalman filter; transfer function; radial basis function

人工神经网络(ANN)可用于解决问题的各个方面。例如, 分类、逼近、模式识别、信号处理、预测、特征提取等。大多数可以通过ANN在输入与输出空间的映射学习解决, 设输入与输出数据对为:

$$S = \{ \langle x_1, y_1 \rangle, \dots, \langle x_n, y_n \rangle \} (x_n \in R^n, y_n \in R^n),$$

$F(\bullet)$ 为:

$$F(x_i) = y_i + \eta \quad i = 1, \dots, n \quad (1)$$

其中, η 是具有方差为 σ_{ns}^2 的零均值白噪声。

由于每个神经元具有局部特性, 最终整个RBF网也呈现“局部映射”特性, 即RBF网是一种局部网络, 如果网络有较大的输出, 必定激活一个或多个隐节点。隐层单元仅对“中心”附近的输入敏感, 随着与“中心”距离的加大, 隐层单元的输出很快减小到很小值, 表现为一种“局

部敏感性”。而减小的快慢由宽度决定。

RBF网络用于解决多维空间中近似问题。典型的RBF网络形式为:

$$f(x; w, p) = \sum_{i=1}^M w_i G_i(\|x\|_i, p_i) \quad (2)$$

其中, M 是隐层的神经元数量。 $G_i(\|x\|_i, p_i)$ 是第 i 个径向基函数, 通常是高斯函数($e^{-\|x-t\|^2/b^2}$)、二次多项式或薄板样条函数; p_i 是调整参数, 例如中心、偏值等。

资源分配网络(RAN)是一种满足下列条件的RBF网络:

$$y_n - f(x_n) = e_n > e_{\min} \quad \|x_n - t_c\| > e_{\min} \quad (3)$$

其中, e_n 是当前错误, t_c 是输入矢量 x_n 的基函数的最近中心, e_{\min} 是实验选择的较小常数。若转移函数在输入空间

有限的区域满足公式(3), 则称为局部的, 否则称为非局部的。

1 扩展卡尔曼滤波学习算法

卡尔曼滤波技术是一种最优估计技术, 即递推线性最小方差估计, 状态估值是量测值的线性函数, 其估计准则是所估计的均方误差最小。扩展卡尔曼滤波器(EKF)用来作为一种学习算法, 因为它具有收敛速度快, 隐层使用较少神经元的特点。该算法的计算量如下:

$$\begin{aligned} e_n &= y_n - f(x_n; p_{n-1}) \\ d_n &= \partial f(x_n; p_{n-1}) / \partial p_{n-1} \\ R_y &= R_n + d_n^T P_{n-1} d_n \\ k_n &= P_{n-1} d_n / R_n \end{aligned} \quad (4)$$

$$p_n = p_{n-1} + e_n k_n$$

$$P_n = [I - k_n d_n^T] P_{n-1} + Q_o(n) I$$

其中, P_n 由所有自适应参数组成, 包括权重、中心、偏值等。

快速扩展卡尔曼滤波^[7]: 协方差矩阵 P_n 的计算可以用一些简化方法进行。假设不同神经元之间参数的相互关系并不非常重要, 可以简化矩阵 P_n 为块对角线结构的 \tilde{P}_n , 它由 \tilde{P}_n^i , ($i=1 \cdots M$)组成。这些对角线元素代表第 i 个神经元适应性参数之间的关系。对于给定问题 p 的矩阵 P_n 复杂性是 $o(M^2)$, 并且矩阵 \tilde{P}_n 仅为 $o(M)$ (M 是神经元数量)。使用这种快速扩展卡尔曼滤波逼近的定义是:

$$\begin{aligned} e_n &= y_n - f(x_n, p_{n-1}) \\ d_n^i &= \partial f(x_n, p_{n-1}) / \partial p_{n-1}^i \\ R_y &= R_n + d_n^{1T} \tilde{P}_{n-1}^1 d_n^1 + \cdots + d_n^{MT} \tilde{P}_{n-1}^M d_n^M \end{aligned} \quad (5)$$

$$k_n^i = \tilde{P}_{n-1}^i d_n^i / R_y$$

$$p_n^i = p_{n-1}^i + e_n k_n^i$$

$$\tilde{P}_n^i = [I - k_n^i d_n^{iT}] \tilde{P}_{n-1}^i + Q_o(n) I \quad i=1, \dots, M$$

新标准: 设在学习期间每一个参数的不确定性使用方差进行估计, 且网络输出的不确定性可确定。以下是新标准的使用:

$$H_o: \frac{e_n^2}{R_y} = \frac{e^2}{\text{Var}[f(x; p) + \eta]} < x_{n,\theta}^2 \quad (6)$$

其中, $x_{n,\theta}^2$ 是以 $\theta\%$ 信任 η 自由度的 x^2 分布。 $e = y - f(x, p)$ 是误差。如果目前的模式不满足这一假设, 则网络应增长。

剪枝标准: 检查(7)式中不相等的 ρ , 以决定是否对网络进行剪枝。也可以选择具有最小 L 值的神经元进行剪枝。

$$\begin{aligned} \rho &: L / R_y < \chi_{1,\theta}^2 \\ L &= \min_i w^2 / [P_w]_{ii} \end{aligned} \quad (7)$$

其中 $x_{n,\theta}^2$ 是 $\theta\%$ 关于 x^2 分布的一个自由度。如果 L 太小或者网络输出 R_y 太大, 那么相应的神经元将被剪枝。

2 径向转移函数

S型函数可与几种方式合并为一个局部化窗口类型的函数, 例如, 对每一个变量采用不同的两个S型函数的差 $\sigma(x) - \sigma(x - \theta)$ 或一对S型函数的乘积 $\sigma(x)(1 - \sigma(x))$ 。这些转移函数是非常灵活的, 并能生产适合分类的凸形决策区域。 N 对S型函数 $\sigma(x) = 1/(1 + e^{-x})$ 具有以下一般形式:

$$B_i(x; t, b, s) = \prod_{j=1}^N \sigma(e^{s_j}(x_i - t_j + e^{b_j})) (1 - \sigma(e^{s_j}(x_i - t_j - e^{b_j}))) \quad (8)$$

乘积中的第一个S形因子随着输入向量 x_i 的增长而增长, 同时第二个因子减少, 并围绕 t_j 局部化。密度 $B_i(x; t, b, s)$ 的形状自适应由移动中心 t 、改变 b 和 s 来实现。在这种情况下每个处理单元可调参数的数量(不计权重 w_j)为 $3N$ 。在高斯条中可进行降维, 获得更加灵活的密度形状, 这样在网络中可以减少自适应单元的数量。 e^{s_j} 和 e^{b_j} 用来代替 s_j 和 b_j 防止学习过程中的震荡(学习变得更加稳定)。

双径向函数的旋转: 上述提到的双径向函数的每一个节点包含 $3N$ 个参数, 可以灵活地表示各种概率密度。下一步需要由每一个单元提供密度的单独旋转以便获得更大的灵活性。当然可以采用旋转矩阵操作输入向量 R_x , 但在实践中把这个 $N \times N$ 矩阵分解为 $N-1$ 个独立对角矩阵(例如, 欧拉对角矩阵), 计算反向传播训练过程中的导数是很困难的(见图1)。

$$C_p(x; t, R) = \prod_i^N (\sigma(R_i x + t_i) - \sigma(R_i x + t_i')) \quad (9)$$

其中: R_i 是下列旋转矩阵 R 的第 i 行。

$$R = \begin{bmatrix} s_1 & \alpha_1 & & 0 \\ & \ddots & \ddots & \\ & & s_{N-1} & \alpha_{N-1} \\ 0 & & & s_N \end{bmatrix} \quad (10)$$

对其他双径向转移函数的扩展参看参考文献。

网络与通信

Communication and Network

应用IncNet分类：对每个给出问题的类建立独立的IncNet网络。每一个网络接收输入向量 x ，如果第 i 层子网等于所需类的数量，则索引值为1，否则为0。第 i 层网络的输出定义一种给出情况有多少属于第 i 层的类。赢者通吃的策略是用来决定一种情况的最终类。图2中给出了IncNet网络分类的结构。请注意，每个子网单独学习(这有助于实现并行算法)，其最终结构尝试匹配复杂的第 i 类，而不是所有类(每个子网的结构通常是不同的)。

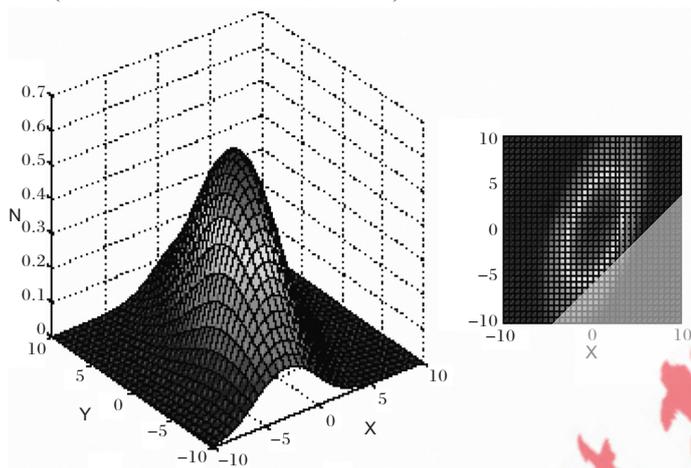


图1 双径向函数的旋转

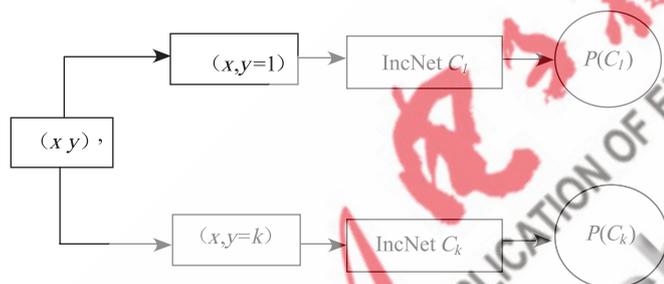


图2 IncNet分类网络

3 举例

Sugeno 函数定义为：

$$f(x, y, z) = (1 + x^{0.5} + y^{-1} + z^{-1.5})^2$$

使用双径向、旋转双径向和转移函数IncNet模型的结果与其他结果比较由Sugeno、Kosinski和Horikawa给出，如表1所示。虽然这一函数经常用于测试自适应系统的逼近能力，但没有标准程序来选择训练点，因此，结果很难进行比较。从[1, 6]区间选择216个训练点，并且从[1.5, 5.5]之间随机选择125个测试点。所有测试使用相同的(如果可能的话)或类似的初始参数。平均百分比误差(APE)用来衡量逼近误差：

$$APE = 1/N \sum_{i=1}^N |(f(x_i) - y_i) / y_i| \times 100\%$$

最终网络的隐层中最多有11个神经元。

上述结果表明，在增量网络中使用双径向转移函数非常有效。在学习过程中应用快速扩展卡尔曼滤波EKF

比常规RBF神经网络有更好的效果，学习速度比梯度下降法明显加快，减少了计算负担。双径向旋转函数改进了复杂密度的估计，每神经元使用 $4N-1$ 参数(其中 N 为输入空间的维数)。这种网络可以成功地用来解决现实世界的分类、逼近、模式识别、信号处理、预测、特征提取等问题。

表1 Sugeno函数的逼近

模型	APE TRS	APE TES
GMDS model Kongo	4.7	5.7
Fuzzy model 1 Sugeno	1.5	2.1
Fuzzy model 2 Sugeno	0.59	3.4
FNN Type 1 Horikawa	0.84	1.22
FNN Type 2 Horikawa	0.73	1.28
FNN Type 3 Horikawa	0.63	1.25
M-Delta model	0.72	0.74
Fuzzy INET	0.18	0.24
Fuzzy VINET	0.076	0.18
IncNet	0.119	0.122
IncNet Rot	0.053	0.061

参考文献

- [1] ADAMCZAK R, DUCH W, JANKOWSKI N. New developments in the feature space mapping mode I. [A] In Third Conference on Neural Networks and Their Applications [C], 1997, 65-70.
- [2] BISHOP C. M. Improving the generalization properties of radial basis function neural networks [J]. Neural Computation, 1991, 3(4):579-588.
- [3] CANDY J. V. Signal processing: The model based approach. [J]. McGraw-Hill, New York, 1986.
- [4] DUCH W, JANKOWSKI N. Survey of neural transfer functions[J]. Neural Computing Surveys, 1999(7).
- [5] GIROSI F. An equivalence between sparse approximation and support vector machines[J]. Neural Computation, 1998. 10(6).
- [6] JANKOWSKI N. Ontogenic neural networks and their applications to classification of medical data[J]. PhD thesis, Department of Computer Methods, Nicholas Copernicus University, Torun, Poland, 1999.
- [7] JANKOWSKI N, KADIRKAMANATHAN V. Statistical control of RBF-like networks for classification[A]. In 7th International Conference on Artificial Neural Networks[C]. 1997. 385-390.
- [8] KADIRKAMANATHAN V, NIRANJAN M. A function estimation approach to sequential learning with neural networks[J]. Neural Computation, 1993, 5(6):954-975.

(收稿日期：2009-01-12)