

文章编号: 1001-0920(2013)10-1564-04

一种有效的增量顺序学习模糊神经网络

胡蓉^{1,2}, 徐蔚鸿^{1,3}

(1. 南京理工大学 计算机科学与技术学院, 南京 210094; 2. 长沙航空职业技术学院 航空
电子电器工程学院, 长沙 410014; 3. 长沙理工大学 计算机与通信工程学院, 长沙 410014)

摘要: 利用误差下降率定义输入数据对系统输出的敏感性, 并以此作为规则产生标准, 提出一种有效增量顺序学习模糊神经网络. 将修剪策略引入规则产生过程, 因此该算法产生的模糊神经网络不需要进行修剪. 通过仿真实验, 本算法在达到与其他算法相当性能的情况下, 能够获得更高的准确率和更简单的结构.

关键词: 增量学习; 模糊神经网络; 增长规则; 扩展卡尔曼算法

中图分类号: TP273

文献标志码: A

An efficient incremental sequential learning algorithm for fuzzy neural networks

HU Rong^{1,2}, XU Wei-hong^{1,3}

(1. School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China; 2. Aviation Institute of Electrical and Electronic Engineering, Changsha Aeronautical Vocational and Technical College, Changsha 410014, China; 3. College of Computer and Communications Engineering, Changsha University of Science & Technology, Changsha 410014, China. Correspondent: HU Rong, E-mail: 410907140@qq.com)

Abstract: The error reduction ratio is used to define the sensibility of input data to output, and the sensibility is used as growth criteria. An efficient incremental sequential learning algorithm for fuzzy neural networks (ISL-FNN) is presented, where the structure learning algorithm incorporating a pruning strategy into new growth criteria is developed. The performance of ISL-FNN is compared with several existing algorithms on some benchmark problems. Simulation results show that ISL-FNN has morosimple structure and better accuracy with less number of rules compared to other algorithms.

Key words: incrementl learning; fuzzy neural network; growing criteria; extended Kalman filter

0 引言

模糊系统 (FIS) 能够通过一系列模糊规则逼近任意非线性输入输出对, 将 RBF (radial basis function) 神经网络和 FIS 相结合^[1-2] 以实现上述任务的网络称为模糊神经网络 (FNN). 文献 [1] 提出的资源定位网络 (RAN) 基于新到数据的新颖性增加隐含层单元, 其改进算法称作扩展的卡尔曼 RAN 方法 (RANEKF)^[3], 用 EKF 替代 LLS 方法调整网络参数, 一旦规则建立将不再被删除. 为了解决此缺陷, 提出了最少资源定位网络 (MRAN)^[4-5], 在学习过程中, 检测不活跃的神经元并将其删除, 这样便可以产生更加精简的网络. 此后提出的增加和修剪 RBF (GAP-RBF) 方法通过简化高斯函数反射每个神经元的重要性, 文献 [6] 提出一种广义的 GAP-RBF 算法用来训练任意分布密度的样

本. 文献 [7] 提出了正交最小方差 (OLS) 学习算法, 其结构和参数识别同时进行. 文献 [8] 提出了一种基于 RBF 的动态模糊神经网络 (DFNN) 算法, 该算法可以自适应增加和删除规则. 文献 [9] 在此基础上提出了一种广义 DFNN (GDFNN) 算法, 基于椭圆基函数 (EBF) 提出在线参数定位机制, 利用 ε 完备性为每个输入变量的维数定位初始宽度, 但该方法是以牺牲计算速度为代价的. 类似的在线方法均存在一定缺陷, 复杂的增长和修剪策略降低了学习速度, 且使得网络难以理解.

文献 [10] 提出了一种快速准确自组织模糊神经网络 (FAOS-PFNN), 通过将修剪策略引入模糊规则的增加过程来提高学习效率. 在此基础上, 本文提出一种有效增量顺序学习模糊神经网络 (ISL-FNN). 初始

收稿日期: 2012-06-21; 修回日期: 2012-10-26.

基金项目: 国家自然科学基金项目(61163040); 湖南省教育厅科研项目(11C0009).

作者简介: 胡蓉(1974-), 女, 副教授, 博士生, 从事模式识别、机器学习的研究; 徐蔚鸿(1963-), 男, 教授, 博士生导师, 从事人工智能、模式识别等研究.

时刻没有隐含层神经元, 根据误差下降率定义输入数据对于系统的输出贡献, 以此作为增长标准. 在参数学习阶段, 使用赢者规则策略^[11], 规定 EKF 算法每一步只对赢了的规则的相关参数更新. 通过仿真实验表明了 ISL-FNN 的有效性.

1 ISL-FNN 结构

ISL-RBF 采用与 FAOS-PFNN 类似的 4 层网络结构, 为

$$\begin{aligned} R_j: & \text{ if } x_1 \text{ is } A_{1j}, \text{ and } \cdots, \text{ and } x_r \text{ is } A_{1j}, \\ & \text{ then } y \text{ is } w_j, \quad j = 1, 2, \cdots, u. \end{aligned} \quad (1)$$

假设 r 为输入变量的维数, 第 1 层的每个输入变量 $x_i (i = 1, 2, \cdots, r)$ 有 u 个模糊子集 $A_{ij} (j = 1, 2, \cdots, u)$. 第 2 层隶属度函数为

$$\begin{aligned} \mu_{ij}(x_i) &= \exp\{-(x_i - c_{ij})^2 / \sigma_{ij}^2\}, \\ i &= 1, 2, \cdots, r, \quad j = 1, 2, \cdots, u. \end{aligned} \quad (2)$$

其中: μ_{ij} 为 x_i 的第 j 个隶属度函数, c_{ij} 为高斯函数中心, σ_{ij} 为高斯函数宽度, r 为输入变量的维数, u 为每个输入变量的隶属函数的个数. 第 3 层的每个节点代表规则的 IF-PART 部分, 该层节点数反映了模糊规则数, 第 j 个规则的输出为

$$\begin{aligned} \varphi_j(x_1, x_2, \cdots, x_r) &= \exp\left(-\sum_{i=1}^r \frac{(x_i - c_{ij})^2}{\sigma_{ij}^2}\right), \\ j &= 1, 2, \cdots, u. \end{aligned} \quad (3)$$

第 4 层为输出层, 是多输入单输出 (MISO) 系统的单一输出节点, 有

$$y(x_1, x_2, \cdots, x_r) = \sum_{j=1}^u w_j \varphi_j, \quad (4)$$

其中 w_j 为第 j 个规则 then 部分的后件参数.

2 ISL-FNN 学习算法

对于每个观测数据对 (x^k, t^k) , 网络的实际输出 y^k 由式 (4) 获得. 其中: $k = 1, 2, \cdots, n$, n 为训练数据对数; $x^k \in R^r$ 和 $t^k \in R$ 为第 k 个输入和期望输出. 在学习过程中, 假设模糊神经网络已产生 u 个模糊规则, 为了得到更加精简和快速的模糊神经网络, 将修剪过程和增长过程同时考虑, 这样能够更加有效, 也能加速网络的学习过程.

2.1 规则增长标准

文献 [7] 采用误差下降率 (ERR) 作为模糊规则的增长标准. 假定有 n 个输入输出数据对 (x^k, d^k) , $k = 1, 2, \cdots, n$, 将式 (4) 改写为

$$w\Phi = Y. \quad (5)$$

作为线性回归模型的一种特殊情况, 有

$$D = \Phi W + E. \quad (6)$$

其中: $D = [d^1, d^2, \cdots, d^n]^T \in R^n$ 为期望输出向量; $W = [w_1, w_2, \cdots, w_u] \in R^u$ 为权值向量; $E = [e^1, e^2, \cdots, e^n] \in R^n$ 为误差向量, 假设与回归量不相关; $\Phi = [\psi_1, \psi_2, \cdots, \psi_u] \in R^{n \times u}$ 为第 3 层输出.

对于任意矩阵 Φ , 如果其行数大于列数, 则通过 QR 分解为

$$\Phi = PQ. \quad (7)$$

可以将 Φ 变成一组正交基向量 $P = [p_1, p_2, \cdots, p_u] \in R^{n \times u}$, 维数与 Φ 相同, $Q \in R^{u \times u}$ 为一个上三角矩阵, 具体推导过程见文献 [9]. 将式 (7) 代入 (6) 得到

$$D = PQW + E = PG + E. \quad (8)$$

G 的线性最小二乘解为

$$G = (P^T P)^{-1} P^T D. \quad (9)$$

因此定义误差下降率为

$$\text{err}_i = \frac{(p_i^T D)^2}{p_i^T p_i D^T D}, \quad i = 1, 2, \cdots, u. \quad (10)$$

ERR 提供了一个简单有效的方法寻找重要回归子集. 如果 ERR 值很大, 则表示 p_i 对于输出有显著影响. 将新得到的 $\text{err}_i (i = 1, 2, \cdots, (r+1)u)$ 重写为 $\Delta = (\delta_1, \delta_2, \cdots, \delta_u) \in R^{(r+1) \times u}$, Δ 的第 i 列 δ_i 即为与第 i 个规则相关的 $(r+1)$ 个误差下降率. 定义

$$\zeta_i = \sum_j \delta_i \quad (11)$$

为与第 i 个规则输入变量相关的总的误差下降率. 定义

$$\zeta_{ij} = \delta_{ij} / \zeta_i, \quad i = 1, 2, \cdots, u, \quad j = 1, 2, \cdots, r. \quad (12)$$

ζ_{ij} 为第 i 个规则中第 j 个输入变量的重要性, ζ_{ij} 值越大表示第 j 个输入对系统输出的敏感性越强. 如果

$$\zeta_{ij} < \text{ep}, \quad (13)$$

则网络需要调整参数或者增加更多的神经元以达到更好的泛化性能, 其中 ep 为预设的阈值.

2.2 参数调整

赢者策略是指修改胜出规则相关的参数, 即

$$\theta_{\text{EKF}} = [\theta_1, \theta_2, \cdots, \theta_u]^T =$$

$$[w_1, C_1^T, \sigma_1, \cdots, w_j, C_j^T, \sigma_j, \cdots, w_u, C_u^T, \sigma_u]^T.$$

赢者定义为与当前输入数据最近的规则, 采用欧氏距离定义当前输入数据与规则中心之间的距离. 假设当前数据 x^k 与规则 j 中心的距离最短, 有

$$d_j = \min(\|x^k - c_j\|),$$

$$k = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, u, \quad (14)$$

则 $H_j = [w_j, C_j^T, \sigma_k]$ 为最近的模糊规则, 其梯度推导如下:

$$\bar{w}_j = \partial y / \partial w = \varphi_j(x^k),$$

$$\begin{aligned} \bar{C}_j^T &= \partial y / \partial C = \varphi_j(x^k)(2w_j/\sigma_j^2)(x^k - C_j)^T, \\ \bar{\sigma}_j &= \partial y / \partial \sigma_j = \varphi_j(x^k)2w_j\|x^k - C_j\|^2/\sigma_j^3. \end{aligned} \quad (15)$$

获得了最近模糊规则的梯度向量 $H_j = [\bar{w}_j, \bar{C}_j^T, \bar{\sigma}_j]$ 后, 由 EKF 更新其参数如下:

$$\begin{aligned} K_n &= P_{n-1}H_j^T[R_n + H_j^T P_{n-1}H_n]^{-1}, \\ \theta_{\text{EKF}}(n) &= \theta_{\text{EKF}}(n-1) + K_n \times e^k, \\ P_n &= [I_{z \times z} - K_n H_j^T]P_{n-1} + qI_{z \times z}. \end{aligned} \quad (16)$$

其中: q 为决定梯度向量方向的任意步骤, z 为调整参数的维数.

2.3 完整的 ISL-FNN 算法

ISL-FNN 算法步骤如下.

Step 1: 初始化过程. 当第 1 个观测数据 (x_1, d_1) 到达时, ISL-FNN 还没有建立起来, 此时产生第 1 条规则, 参数如下:

$$c_1 = x_1, \sigma_1 = \sigma_0, w_1 = w_0, \quad (17)$$

其中 w_0, σ_0 为预先定义的参数.

Step 2: 第 k 个数据 (x^k, d^k) 到达时, 设 ISL-FNN 已有 u 个规则, 根据式 (11) 和 (14) 分别计算 $d_{k,\min}$ 和 ζ_{ij} , 并计算与准确率相关的增长标准, 有

$$\varepsilon_k = \max\{\varepsilon_{\max}\gamma^k, \varepsilon_{\min}\}, \quad 0 < \gamma < 1, \quad e^k = t^k - y^k.$$

若 $\|e^k\| > \varepsilon_k, d_{k,\min} > kd, \zeta_{ij} < \text{ep}, \varepsilon_k$ 和 kd 为阈值, 则增加一条规则, 新增加的规则参数设置为

$$c_{u+1} = x^k, \sigma_{u+1} = kd_{k,\min}, w_{u+1} = e^k; \quad (18)$$

否则, 若 $d_{k,\min} < kd, \|e^k\| > \varepsilon_k$, 则表明覆盖 x^k 的 RBF 单元的泛化能力不是很好, 调整最接近 x^k 的第 j 个 RBF 单元的宽度为

$$\sigma_j^i = k \times d_{k,\min}. \quad (19)$$

其他情况根据式 (16), 采用 EKF 算法调整最近规则的系统参数 $\bar{w}_j, \bar{C}_j^T, \bar{\sigma}_j$.

3 仿真实验

通过函数逼近验证 ISL 算法的有效性. 被逼近函数为如下 Hermite 多项式:

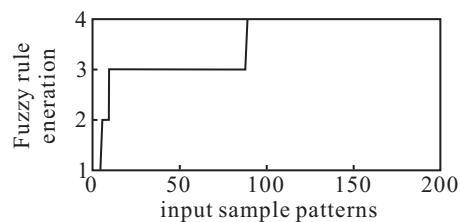
$$f(x) = 1.1 \times (1 - x + 2x^2) \exp(-x^2/2). \quad (20)$$

为了学习该函数, 利用区间 $(-4, 4)$ 内的随机样本函数产生 200 个输入输出数据作为训练集. 预先确定的参数为

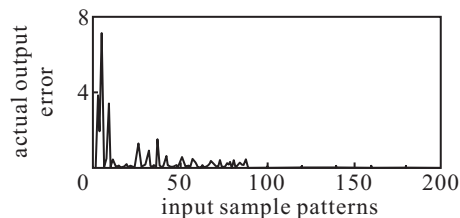
$$\begin{aligned} d_{\max} &= 2.0, d_{\min} = 0.1, \gamma = 0.99, \beta = 0.99, \\ e_{\max} &= 1.1, e_{\min} = 0.02, k = 1.1, k_{\text{err}} = 0.0015, \\ \sigma_0 &= 1.0, R_n = 1.0, \text{ep} = 1.0. \end{aligned}$$

仿真结果如图 1 所示. 由图 1 可见, 在 ISL-FNN 网络中只有 4 个模糊规则. 经过 100 个训练数据后, 规则数稳定下来, 且整个过程不需要修剪规则, 因此, 规

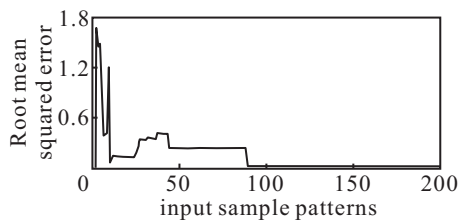
则数保持上升或稳定状态. RMSE 接近 0.02, 实际输出误差也在 0 附近.



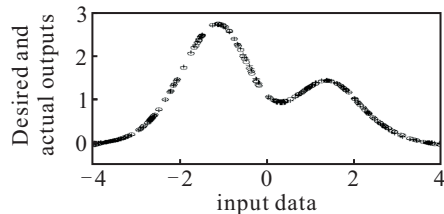
(a) 模糊规则的产生



(b) 训练时实际输出误差



(c) 均方根误差



(d) Hermite(+) 函数及其逼近 (o)

图 1 Hermite 函数逼近

ISL-FNN 算法与 ANFIS, GDFNN, SOFNN 和 FAOS-PFNN 算法的比较结果如表 1 所示.

表 1 ISL 算法与其他算法的比较

算 法	规则数	RMSE	训练时间
RANEKF	13	0.0262	-
MRAN	7	0.0090	-
OLS	7	0.0095	-
DFNN	6	0.0056	0.8703
GDFNN	6	0.0097	0.9347
FAOS-PFNN	5	0.0089	0.2639
ISL-FNN	4	0.0090	0.2544

由表 1 可见, ISL-FNN 达到了最简单的神经网络结构, 快速的学习和泛化能力与 FAOS-PFNN 几乎相同. 由于在参数学习过程中采用赢者策略, 只对胜出的神经元的相关参数进行调整, 提高了学习速度, 且由于规则的增长过程采用的是严格的检测网络的泛

化能力和误差等, 网络不会产生冗余的规则, 不需要等待网络形成后再进行检测和删减, 产生的网络结构更加紧凑。由此可见, ISL-FNN 不但能够实现与其他网络相似的逼近性能, 产生更加紧凑的网络, 而且减少了训练时间, 提高了计算速度。

4 结 论

本文在 FAOS-PFNN 的基础上提出了 ISL-FNN 算法。该算法不需要使用修剪策略, 因为修剪过程不可避免地要增加计算负担, 相应的学习速度也会减慢。另外, ERR 用来定义新输入数据的系统的敏感性, 并应用到规则产生过程, 这使得增长过程缓慢。仿真结果表明, ISL-FNN 能获得一个高性能、快速、自组织的学习过程, 与其他算法相比, 在准确性相当的情况下, 能够获得更加精简的网络结构。

参考文献(References)

- [1] Platt J. A resource-allocating network for function interpolation[J]. *Neural Computer*, 1991, 3: 213-225.
- [2] Hu Rong, Xu Wei-hong, Wang Jian-ping, et al. A novel way of face recognition to improve the quality of features extraction[C]. *IEEE Int Conf on Intelligent Computing and Intelligent Systems*. Piscataway, 2009, 4: 237-241.
- [3] Kadirkamanathan V, Niranjan M. A function estimation approach to sequential learning with neural networks[J]. *Neural Computer*, 1993, 5: 954-975.
- [4] Yingwei L, Sundararajan N, Saratchandran P. A sequential learning scheme for function approximation using minimal radial basis function neural networks[J]. *Neural Computer*, 1997, 9: 461-478.
- [5] Yingwei L, Sundararajan N, Saratchandran P. Performance evaluation of a sequential minimal radial basis function(RBF) neural network learning algorithm[J]. *IEEE Trans on Neural Networks*, 1998, 9: 308-318.
- [6] Huang G B, Saratchandran P, Sundararajan N. A generalized growing and pruning RBF(GGAP-RBF) neural network for function approximation[J]. *IEEE Trans on Neural Networks*, 2005, 16(1): 57-67.
- [7] Chen S, Cowan C F N, Grant P M. Orthogonal least squares algorithm for radial basis function network[J]. *IEEE Trans on Neural Networks*, 2008, 20(2): 302-310.
- [8] Wu S, Er M J. Dynamic fuzzy neural networks—A novel approach to function approximation[J]. *IEEE Trans on Systems Man Cybernet, Part B: Cybernet*, 2000, 30(2): 358-364.
- [9] Wu S Q, Er M J, Gao Y. A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks[J]. *IEEE Trans on Fuzzy System*, 2001, 9: 578-594.
- [10] Ning Wang, Meng Joo Er, Xianyao Meng. A fast and accurate online self-organizing scheme for parsimonious fuzzy neural networks[J]. *Neurocomputing*, 2009, 72: 3818-3829.
- [11] Rong Hai-jun, Sundararajan N, Huang Guang-bin, et al. Sequential adaptive fuzzy inference system(SAFIS) for nonlinear system identification and prediction[J]. *Fuzzy Sets and Systems*, 2006, 157: 1260-1275.
- [10] 张香燕, 张乃尧. 一般二叉树型分层模糊系统的通用逼近性[J]. *清华大学学报*, 2007, 47(1): 143-146.
(Zhang X Y, Zhang N Y. Universal approximation of general binary tree-type hierarchical fuzzy systems[J]. *J of Tsinghua University*, 2007, 47(1): 143-146.)
- [11] Liu Pu-yin, Li Hong-xin. Hierarchical TS fuzzy system and its universal approximation[J]. *Information Sciences*, 2005, 169(3/4): 279-303.
- [12] 王立新. 模糊系统: 挑战与机遇并存——十年研究之感悟[J]. *自动化学报*, 2001, 27(4): 585-590.
(Wang L X. Fuzzy systems: Challenges and chance — My experiences and perspectives[J]. *Acta Automatic Sinica*, 2001, 27(4): 585-590.)
- [13] Joo M G, Lee J S. A class of hierarchical fuzzy systems with constraints on the fuzzy rules[J]. *IEEE Trans on Fuzzy Systems*, 2005, 13(2): 194-203.
- [14] Raju G V S, Zhou Jun. Hierarchical fuzzy control[J]. *Int J Control*, 1991, 54(5): 1201-1216.
- [15] Chen Yue-hui, Yang Bo, Abraham Ajith, et al. Automatic design of hierarchical Takagi-Sugeno type fuzzy systems using evolutionary algorithms[J]. *IEEE Trans on Fuzzy Systems*, 2007, 15(3): 385-397.
- [16] Wang S. Cascaded centralized TSK fuzzy system: Universal approximator and high interpretation[J]. *Applied Soft Computing*, 2005, 5(2): 131-145.
- [17] 王立新. 模糊系统与模糊控制[M]. 北京: 清华大学出版社, 2003: 144-146.
(Wang L X. *Fuzzy systems and fuzzy control*[M]. Beijing: Tsinghua University Press, 2003: 144-146.)

(上接第1563页)