

动态自适应模块化神经网络结构设计

张昭昭

(辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘要: 针对全连接前馈神经网络不能有效应对时变系统的问题, 提出一种动态自适应模块化神经网络结构. 该网络采用减法聚类算法在线辨识工况数据的空间分布, 利用 RBF 神经元实现对数据样本空间的划分, 并结合模糊策略将不同子样本空间的数据动态分配给不同的子网络, 最后对各子网络的输出进行集成. 该模块化网络中子网络数量和子网络规模都能根据所学时变任务动态自适应调整. 通过对不同时变系统的预测表明了该网络能够有效跟踪时变系统.

关键词: 模块化神经网络; 数据驱动; 时变系统; 在线学习

中图分类号: TP183

文献标志码: A

Dynamic adaptive modular neural network architecture design

ZHANG Zhao-zhao

(Institute of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China.
E-mail: zzzhao123@126.com)

Abstract: Due to the fact that the fully coupled feedforward neural network can not effectively deal with the problem of time-varying systems, a dynamic adaptive modular neural network model is proposed. In this model, the subtractive cluster algorithm is applied to online identification of the spatial distribution of the condition data. RBF neurons are used to decompose the learning sample space and combined with fuzzy strategy to dynamically allocate different sub-sample space learning data to different sub-networks. Finally, the output of the modular neural network can be achieved by integrating the output of the sub-networks. The number of the sub-networks and the architecture of the sub-networks can be adaptively adjusted based on the current learning time-varying task. Experiment results on different time-varying systems show that proposed model can effectively track the time-varying system.

Key words: modular neural network; data driven; time-varying system; online learning

0 引言

前馈神经网络(FNN)以其简单的结构和强大的非线性逼近能力已广泛地应用于非线性系统建模和控制等方面^[1-3]. 但对于时变系统而言, FNN 具有全连接的特点, 新学习样本将激活所有隐节点参与当前任务的学习, 因此 FNN 对当前样本的学习将导致对过去已学得的旧样本知识的破坏, 其所建立的模型往往只是系统的近似模型^[4]. 此外, 随着 FNN 结构复杂度的增大, FNN 全连接的特性也会导致网络收敛速度缓慢或者难以收敛于全局最小点^[5].

具有局部连接特性的模块化神经网络是解决 FNN 应用于非线性时变系统的有效方法. 文献 [6] 和文献 [7] 分别提出了适用于时变系统的模块化神经网络

SAMNN (Structure adaptive modular neural network) 和 SGMN (Self-regulating growing multi-experts network), 二者都能随着时变系统的变化自适应调整网络结构. 然而, SAMNN 子网络的增加或删除采取固定步长策略, 不能及时反映时变系统的时变特性. SGMN 采用梯度下降法修正网络参数, 导致其难以跟踪快速时变系统. 同时, SAMNN 和 SGMN 中子网络增加或删除的条件由子网络输出的局部误差决定, 使得网络的生长速度缓慢, 而且子网络的结构在整个学习过程中都固定不变, 使得它们对时变系统的自适应能力较差.

针对上述问题, 本文提出一种动态自适应模块化神经网络 (DAMNN). DAMNN 中子网络的增加和删

收稿日期: 2012-09-23; 修回日期: 2013-01-30.

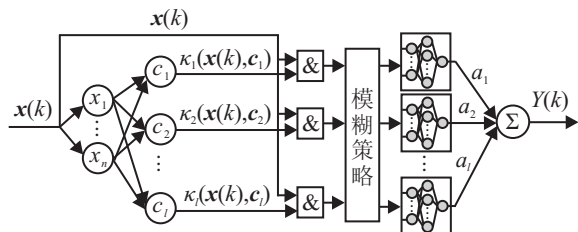
基金项目: 国家自然科学基金项目(60971048); 辽宁省教育厅基金项目(L2013129).

作者简介: 张昭昭(1973-), 男, 讲师, 博士, 从事智能信息处理、神经网络结构与优化的研究.

除以学习对象的实时数据流为驱动, 根据在线工况数据的空间分布动态调整网络系统中子网络的个数, 同时, 每个子网络也能根据所学习任务的状况自适应调整自身的规模. 最后通过仿真实验验证了 DAMNN 的性能.

1 DAMNN 结构

DAMNN 是基于在线工作机制的一种网络结构, 其结构如图 1 所示, 共有 5 层, 分别为: 输入层、任务分解层、任务分配层、子网络层和子网络集成输出层.



输入层 任务分解层 任务分配层 子网络层 集成输出层

图 1 DAMNN 神经网络结构框图

各层的数学描述和具体功能如下.

第 1 层: 输入层.

该层中有 n 个输入节点, n 为 k 时刻输入向量 $\mathbf{x}(k)$ 的维数, $\mathbf{x}(k) = [x_1(k), x_2(k), \dots, x_n(k)]^T$.

第 2 层: 任务分解层.

该层中包括 l 个 RBF 神经元, 每个 RBF 神经元的激活函数为

$$\kappa_i(\mathbf{x}(k), \mathbf{c}_i) = e^{(-\|\mathbf{x}(k) - \mathbf{c}_i\|^2 / \delta_i^2)}. \quad (1)$$

其中: \mathbf{c}_i 为第 i 个 RBF 神经元的数据中心, δ_i 为核函数的扩展宽度. 在 DAMNN 中, $\kappa_i(\mathbf{x}(k), \mathbf{c}_i)$ 被定义为硬限幅函数, 即

$$\kappa_i(\mathbf{x}(k), \mathbf{c}_i) = \begin{cases} 1, & \|\mathbf{x}(k) - \mathbf{c}_i\| \leq \delta_i; \\ 0, & \|\mathbf{x}(k) - \mathbf{c}_i\| > \delta_i. \end{cases} \quad (2)$$

DAMNN 中 RBF 神经元的物理意义是在输入样本空间中, 以 \mathbf{c}_i 为球心, 以 δ_i 为半径作一个超球面. 当输入样本在此超球面内时 RBF 神经元的输出为 1, 在超球面以外时 RBF 神经元的输出为 0. 因此, 该层中 RBF 神经元的功能是实现输入样本空间的分解.

第 3 层: 任务分配层.

该层中共包括 l 个“与门”和一个模糊策略. 每个“与门”的输入为 $\kappa_i(\mathbf{x}(k), \mathbf{c}_i)$ 和 $\mathbf{x}(k)$. 显然, 若 $\kappa_i(\mathbf{x}(k), \mathbf{c}_i) = 1$, 则该“与门”后面的子网络将被激活而有可能参与对 k 时刻输入样本 $\mathbf{x}(k)$ 的学习.

在 DAMNN 中, 对任务的分解属于软分类, 因此, 当 k 时刻的输入样本 $\mathbf{x}(k)$ 落在多个 RBF 神经元的“重叠”区域时, 该样本将会激活多个子网络参与学习. 从集成学习和提高 DAMNN 学习速度的角度出

发, 本文采取一种模糊的策略, 从被 $\mathbf{x}(k)$ 激活的多个子网络中选择部分最适合的子网络对 $\mathbf{x}(k)$ 进行学习.

第 4 层: 子网络层.

该层中共包括 l 个结构相对简单的子网络, 每个子网络均为前馈神经网络, 其主要任务是学习任务分配层分配来的学习样本.

设 k 时刻 DAMNN 中第 h 个子网络的网络结构为 $n - m - 1$ (n 个输入节点, m 个隐节点, 1 个输出节点), 则该子网络的输出为

$$y_h(k) = \sum_{j=1}^m w_j(k) f\left(\sum_{i=1}^n v_{ji}(k) x_i(k)\right). \quad (3)$$

其中: w_j 为子网络隐含层中第 j 个隐节点与输出节点之间的连接权值; v_{ji} 为子网络中输入层第 i 个输入节点与隐含层第 j 个隐节点之间的连接权值; 函数 $f(x) = 1/(1 + e^{-x})$, 为隐节点激活函数.

第 5 层: 集成输出层.

该层的主要功能是对 k 时刻参与学习的子网络的输出进行集成, 集成输出为

$$Y(k) = \sum_{i=1}^p \alpha_i y_i(k). \quad (4)$$

其中: p 为 k 时刻参与学习的子网络个数, 满足 $1 \leq p \leq l$; α_i 为各子网络的集成权值; $y_i(k)$ 为 k 时刻参与学习的第 i 个子网络的输出.

2 DAMNN 结构设计及学习

2.1 RBF 神经元参数在线辨识

DAMNN 针对时变系统的动态特性主要体现在任务分配层 RBF 神经元数据中心 \mathbf{c}_i 上, 即 \mathbf{c}_i 能够随着时变数据流的变化而动态调整, 以达到实时分解和跟踪时变任务的目的. 减法聚类以当前时刻的数据密度和聚类中心密度为聚类准则, 它们在一定程度上能够反映出时变系统的未来变化趋势. 因此, RBF 神经元数据中心 \mathbf{c}_i 可以通过在线减法聚类算法获得.

减法聚类的数据密度是以该数据到其他所有数据的距离为度量标准, 其形式为

$$P(\mathbf{x}(i)) = \frac{1}{1 + \frac{1}{N-1} \sum_{j=1, j \neq i}^N (\mathbf{x}(i) - \mathbf{x}(j))^T (\mathbf{x}(i) - \mathbf{x}(j))}. \quad (5)$$

其中: $X = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)]$, N 为数据 X 中的数据样本量.

在线减法聚类时, 首先以第 1 个数据 $\mathbf{x}(1)$ 作为初始聚类中心, 并赋值 $P_1(\mathbf{x}(1)) = 1$ (P_k 表示第 k 时刻计算的密度). 在 k 时刻, 如果已经存在 l 个聚类中心 \mathbf{c}_i ($i = 1, 2, \dots, l$), 则需计算当前数据 $\mathbf{x}(k)$ 与原聚类

中心 \mathbf{c}_i 的密度并进行比较. $\mathbf{x}(k)$ 密度计算公式为

$$P_k(\mathbf{x}(k)) = \frac{1}{1 + \frac{1}{k-1} \sum_{j=1}^{k-1} (\mathbf{x}(k) - \mathbf{x}(j))^T (\mathbf{x}(k) - \mathbf{x}(j))} = \frac{k-1}{(k-1)(\gamma(k)+1) - 2\boldsymbol{\eta}(k)\mathbf{x}(k) + \sigma(k)}. \quad (6)$$

其中

$$\boldsymbol{\eta}(k) = \sum_{j=1}^{k-1} \mathbf{x}^T(j), \quad \sigma(k) = \sum_{j=1}^{k-1} \mathbf{x}^T(j)\mathbf{x}(j),$$

$$\gamma(k) = \mathbf{x}^T(k)\mathbf{x}(k).$$

式(6)中, $\gamma(k)$ 可通过当前数据 $\mathbf{x}(k)$ 计算而得, 其他参数可通过如下递推关系计算得到:

$$\boldsymbol{\eta}(k) = \boldsymbol{\eta}(k-1) + \mathbf{x}^T(k-1), \quad (7)$$

$$\sigma(k) = \sigma(k-1) + \mathbf{x}^T(k-1)\mathbf{x}(k-1). \quad (8)$$

由式(6)~(8)可知, 只需保存 $\boldsymbol{\eta}(k-1)$ 、 $\sigma(k-1)$ 、 $\mathbf{x}(k-1)$, 即可由当前数据 $\mathbf{x}(k)$ 递推求得 $P_k(\mathbf{x}(k))$.

对于聚类中心 \mathbf{c}_i , 由式(5)可计算出 $k-1$ 时刻的密度为

$$P_{k-1}(\mathbf{c}_i) = \frac{1}{1 + \frac{1}{k-2} \sum_{j=1, j \neq i}^{k-1} (\mathbf{c}_i - \mathbf{x}(j))^T (\mathbf{c}_i - \mathbf{x}(j))}. \quad (9)$$

k 时刻聚类中心 \mathbf{c}_i 密度递推公式为

$$P_k(\mathbf{c}_i) = \frac{1}{1 + \frac{1}{k-1} \sum_{j=1, j \neq i}^k (\mathbf{c}_i - \mathbf{x}(j))^T (\mathbf{c}_i - \mathbf{x}(j))} = \frac{(k-1)P_{k-1}(\mathbf{c}_i)}{k-2 + P_{k-1}(\mathbf{c}_i) + \zeta(k)P_{k-1}(\mathbf{c}_i)}, \quad (10)$$

其中 $\zeta(k) = (\mathbf{c}_i - \mathbf{x}(k))^T (\mathbf{c}_i - \mathbf{x}(k))$ 可由当前数据 $\mathbf{x}(k)$ 计算得到.

由式(5)知, 数据的密度反映了该数据在所有数据中的描述能力, 数据周围的数据越多, 密度值越大. 因此在聚类过程中, 通过对 $P_k(\mathbf{x}(k))$ 与 $P_k(\mathbf{c}_i)$ ($i = 1, 2, \dots, l$) 进行比较, 并考虑以下几种情况实现对聚类中心 \mathbf{c}_i 的动态调整^[10-11].

情况 1

$$\text{If } \min_{i=1,2,\dots,l} \|\mathbf{x}(k) - \mathbf{c}_i\|_2 > r_1 \text{ and}$$

$$P_k(\mathbf{x}(k)) > \bar{\varepsilon},$$

$$\text{Then } \mathbf{c}_{l+1} = \mathbf{x}(k), \quad l = l + 1.$$

其中: r_1 和 $\bar{\varepsilon}$ 为预先设定的阈值, $r_1 = (0.3 \sim 0.5) \times$

$$\max_{i \neq j, i,j=1,\dots,N} \|\mathbf{x}(i) - \mathbf{x}(j)\|_2, \quad \bar{\varepsilon} = 0.5 \max_{i=1,\dots,l} P_{k-1}(\mathbf{c}_i).$$

情况 1 表明新数据与最近的聚类中心的距离足够大, 而且其密度也大于给定的阈值, 说明该数据比周围其他数据的描述能力更强, 意味着一个新工况的出现, 因此, 需要增加一个新的聚类中心以跟踪该新工况. DAMNN 结构设计时, 每增加一个新的聚类, 则对应增加一个新的子网络.

情况 2

$$\text{If } \min_{i=1,2,\dots,l} \|\mathbf{x}(k) - \mathbf{c}_i\|_2 \leq r_1 \text{ and}$$

$$P_k(\mathbf{x}(k)) > P_k(\mathbf{c}_t),$$

$$\text{Then } \mathbf{c}_t = \mathbf{x}(k),$$

其中 $t = \arg \min_{i=1,2,\dots,l} \|\mathbf{x}(k) - \mathbf{c}_i\|_2$.

情况 2 表明新数据与某个聚类中心距离很近, 而且该新数据的密度大于该聚类中心的密度值, 说明系统的工况中心发生了转移, 因此必须由新的数据代替与其距离最近的聚类中心.

情况 3

$$\text{If } \min_{i \neq j, i,j=1,2,\dots,l} \|\mathbf{c}_i - \mathbf{c}_j\| \leq r_2,$$

$$\text{Then } \mathbf{c}_i = \mathbf{c}_d, \quad l = l - 1.$$

其中: 若 $P_k(\mathbf{c}_j) > P_k(\mathbf{c}_i)$, 则 $d = j$, 否则 $d = i$; $r_2 \approx (0.5 \sim 0.7)r_1$.

当两个聚类中心距离很近时, 这两个聚类可以合并成一个聚类, 以使聚类更加紧凑. 由图 1 可知, 此时应将每个聚类中心后面的两个子网络合并, 本文采取保留隐节点多而删除隐节点少的子网络.

通过在线辨识前层 RBF 神经元的参数 \mathbf{c}_i , 便可根据实时工况数据流的空间分布自适应确定 DAMNN 网络系统中子网络的个数.

2.2 子网络输出动态集成

子网络输出动态集成是指从被学习样本 $\mathbf{x}(k)$ 激活的所有子网络中选择部分最适合学习该样本的子网络参与学习, 并为被选中的子网络确定一组最优的集成权值^[12].

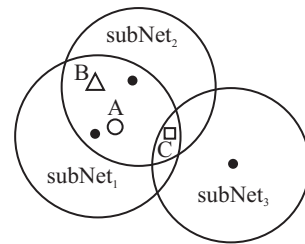


图 2 样本分布示意图

由于 DAMNN 利用 RBF 神经元实现对输入样本空间的划分, 子网络选择与学习样本在样本划分空间中的位置有关. 如图 2 所示, 当学习样本落在 A 点时, 虽然该样本激活 subNet₁ 和 subNet₂ 两个子网络, 但

由 subNet₁ 单独进行学习便能取得较好的效果; 而当样本落在 B 点或 C 点时(分类边界区域), 任何一个子网络对该样本的学习能力均有限. 因此选择 subNet₁ 和 subNet₂ 对该样本进行学习, 并将两个不同子网络的学习结果进行集成处理才能取得较好的效果.

设 $\mathbf{c} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_a\}$ 为被 $\mathbf{x}(k)$ 激活的 RBF 神经元数据中心, 令

$$dr_i = \|\mathbf{x}(k) - \mathbf{c}_i\|/da_i, \quad (11)$$

$$da_i = \frac{1}{a} \sum_{i=1}^a \|\mathbf{x}(k) - \mathbf{c}_i\|, \quad (12)$$

其中 da_i 和 dr_i 分别为 $\mathbf{x}(k)$ 到被激活 RBF 神经元数据中心的平均距离和相对距离. 显然, 对于 $\mathbf{x}(k)$ 应优先选择相对距离小的子网络.

对 dr_i 进行归一化处理, 有

$$u_i = dr_i / \sum_{i=1}^a dr_i. \quad (13)$$

其中: $u_i \in [0, 1]$, $\sum u_i = 1$.

设 u_i 的模糊集 $F = \{\text{很小}(V), \text{小}(S), \text{中等}(M), \text{大}(B)\}$, 模糊隶属度函数为

$$g(x) = e^{-(x-\beta)^2/0.02}, \quad (14)$$

其中 β 为模糊子集的中心.

设置隶属度函数中模糊子集 $\{V, S, M, B\}$ 的中心分别为: $r_2/4r_1, r_2/3r_1, r_2/2r_1, r_2/r_1$, 且模糊隶属度函数满足:

- 1) 当 $u_i \leq r_2/4r_1$ 时, $u_i \in \{V\}$ 的隶属度为 1;
- 2) 当 $u_i \geq r_2/r_1$ 时, $u_i \in \{S\}$ 的隶属度为 1.

u_i 对 F 的模糊隶属度曲线如图 3 所示.

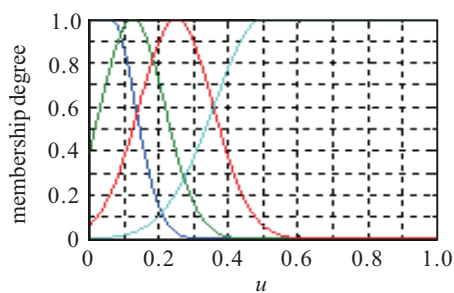


图 3 u_i 对 F 的隶属度曲线

基于模糊策略的子网络选择方法如下:

- 1) 计算 $\mathbf{x}(k)$ 到被激活 RBF 神经元数据中心的 u_i 值;
- 2) 由式 (14) 计算 u_i 对模糊集 F 中各模糊子集的隶属度;
- 3) 用最大隶属度法判断 u_i 属于哪个模糊子集, 并按照从 V 到 B 的顺序选择属于同一子集的子网络进行集成. 即首先选择 $u_i \in \{V\}$ 的子网络对 $\mathbf{x}(k)$ 进行学习, 其余的子网络均被舍弃; 若 $u_i \in \{V\}$ 的子网

络不存在, 则选择 $u_i \in \{S\}$ 的子网络; 依此类推.

设被选中进行集成的子网络对应的 RBF 神经元的数据中心为 $\mathbf{c} = \{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_p\}$, p 为被选择的子网络个数, 则 k 时刻 DAMNN 的集成输出如式 (4) 所示.

式 (4) 中 α_i 为

$$\alpha_i = (1/d_i) / \sum_{i=1}^p (1/d_i). \quad (15)$$

其中: $d_i = \|\mathbf{x}(k) - \mathbf{c}_i\|$, $i = 1, 2, \dots, p$; 显然, $\alpha_i \in (0, 1]$, $\sum_{i=1}^p \alpha_i = 1$.

2.3 DAMNN 学习算法

各子网络采用增量学习方法, 则 DAMNN 学习算法和子网络结构在线调整方法如下.

Step 1 采集 k 时刻的学习样本 $(\mathbf{x}(k), \mathbf{y}(k))$, 在线递推计算 $P_k(\mathbf{x}(k))$ 和 $P_k(\mathbf{c}_i)$.

Step 2 在线更新任务分解层 RBF 神经元数据中心 \mathbf{c}_i , 包括以下几种情况:

情况 1

If $\min_{i=1,2,\dots,l} \|\mathbf{x}(k) - \mathbf{c}_i\|_2 > r_1$ and

$$P_k(\mathbf{x}(k)) > \bar{\epsilon},$$

Then $\mathbf{c}_{l+1} = \mathbf{x}(k)$, $l = l + 1$.

增加一个聚类中心, 则对应增加一个隐节点数为 1 的子网络.

情况 2

If $\min_{i=1,2,\dots,l} \|\mathbf{x}(k) - \mathbf{c}_i\|_2 \leq r_1$ and

$$P_k(\mathbf{x}(k)) > P_k(\mathbf{c}_t),$$

Then $\mathbf{c}_t = \mathbf{x}(k)$.

当前工况中心发生转移, 故在线修正相应的 RBF 神经元数据中心 \mathbf{c}_i .

情况 3

If $\min_{i \neq j, i,j=1,2,\dots,l} \|\mathbf{c}_i - \mathbf{c}_j\| \leq r_2$,

Then $\mathbf{c}_i = \mathbf{c}_d$, $l = l - 1$.

合并数据中心很近的 RBF 神经元, 同时删除与 RBF 神经元对应的隐节点数少的子网络.

Step 3 计算 $\mathbf{x}(k)$ 到各激活 RBF 神经元数据中心的 u_i 值, 并计算 u_i 对模糊集 F 中各模糊子集的隶属度.

Step 4 以最大隶属度法选择属于同一集合的子网络学习 $\mathbf{x}(k)$ (每个子网络独立学习 $\mathbf{x}(k)$), 并对各子网络的输出集成.

Step 5 若集成输出没有达到精度要求, 则对 k

时刻所有参与集成的子网络进行如下操作:

1) 计算子网络 subNet_i 独立学习 $\boldsymbol{x}(k)$ 时的输出误差 $E_i(k)$ 和误差下降趋势 $\text{PE}_i(k)$, 其中 $\text{PE}_i(k)$ 为

$$\text{PE}_i(k) = (E_i(k) - E_i(k-1))/E_i(k) \times 100\%.$$

2) 若 $E_i(k)$ 未达到精度要求且 subNet_i 累计激活次数 > 5 , 则转 4); 若 subNet_i 累计激活次数 ≤ 5 , 则转 3).

3) 若 $\text{PE}_i(k) > 0.2$, 则转 Step 6; 若 $\text{PE}_i(k) \leq 0.2$, 表明学习有可能陷入局部极小点, 此时应跳出极小点, 则 $k+1$ 时刻子网络 subNet_i 输入层-隐层间的权连接矩阵为

$$W_{\text{subNet}_i}^{ih}(k+1) = -W_{\text{subNet}_i}^{ih}(k).$$

4) 在 subNet_i 隐层增加一个隐节点, 即选择 subNet_i 中输出贡献最大的隐节点 h_a 将其分裂为两个隐节点 h_b 和 h_c , 新隐节点 h_b 和 h_c 的连接权为

$$w_b = (1 + \xi)w_a, w_c = -\xi w_a.$$

其中 ξ 是很小的一个随机数.

Step 6 $k = k + 1$, 返回 Step 1.

3 仿真实验

本文选取动态非线性系统预测的一个基准问题 Mackey-Glass 进行仿真实验. 实验时子网络采用增量 LM (Levenberg-Marquardt) 算法^[13].

混沌现象是一种由确定的非线性动力系统生成的复杂行为, 由如下时延微分方程产生:

$$\frac{dx}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t). \quad (16)$$

本文以嵌入矢量 $[x(t) \ x(t-6) \ x(t-12) \ x(t-18)]$ 预测 $x(t+20)$ 的值.

为检验 DAMNN 网络的性能共做了 2 组不同的实验. 第 1 组实验针对静态 Mackey-Glass 混沌时间序列 (时不变系统) 展开, 实验数据在 $\tau = 17$ 的情况下, 采用四阶-龙库塔方法产生 6 000 组数据, 其中 $201 \leq t \leq 3 200$ 的数据作为训练数据, $5 001 \leq t \leq 5 500$ 的数据作为测试数据.

图 4 所示为学习过程中 DAMNN 的训练 RMSE 和子网络数变化情况, 从该图可以看出, 在 DAMNN 学习的初始阶段, 网络结构便能快速构建起来. 除了在 DAMNN 网络结构构建初始阶段 ($300 \leq t$) 训练 RMSE 出现震荡外, 当网络结构构建完成后, 训练 RMSE 平滑下降, 表明在整个在线学习期间网络稳定收敛. 学习过程结束后, 整个 DAMNN 系统中共有 7 个子网络, 共 30 个隐节点.

图 5 所示为 500 组测试样本的测试效果, 从该图可以看出, 测试样本与实际输出拟合较好, 表明训练完的 DAMNN 具有较强的泛化能力.

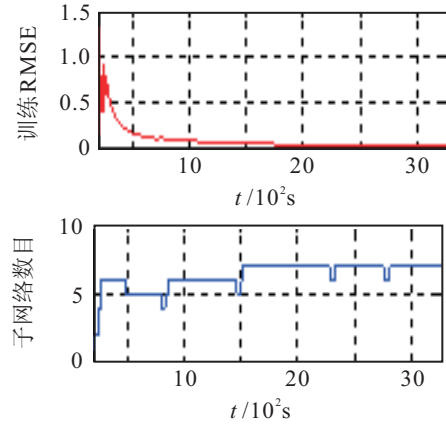


图 4 静态系统训练 RMSE 及子网络数动态

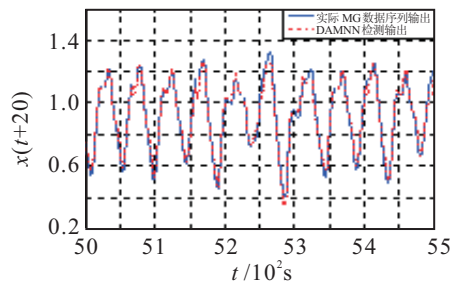


图 5 静态系统测试效果

第 2 组实验针对慢时变和快时变系统展开, 以检验 DAMNN 跟踪不同时变系统的能力. 式 (16) 中, 不同的时延系数 τ 对应于系统不同的混沌特性, 如图 6 所示, 当 τ 分别等于 17、30、50、100 时, $x(t+20)$ 相对于 $x(t)$ 的混沌特性随着 τ 的增大而增大.

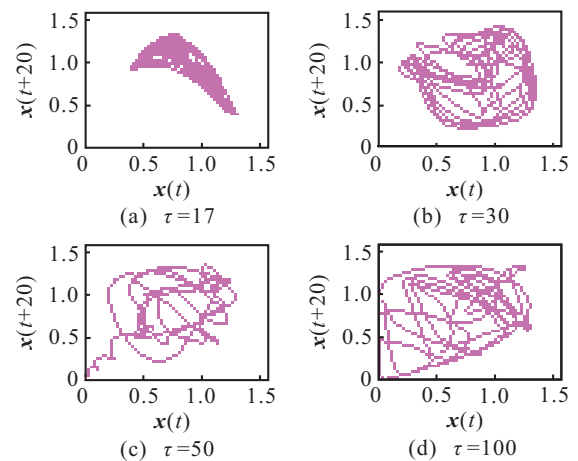


图 6 $x(t+20)$ 和 $x(t)$ 的混沌行为

慢时变和快时变系统模型可通过混合时延系数 τ 不同的静态 Mackey-Glass 混沌时间序列而获得. 混合方法如下^[6]:

$$f(x_t) = \alpha(t)f_1(x_t) + (1 - \alpha(t))f_2(x_t), \quad (17)$$

$$\alpha(t) = \exp(-5t/T), t = 1, 2, \dots, T. \quad (18)$$

慢时变系统模型通过式 (17) 和 (18) 混合 $\tau = 17, 30$ 的两个静态 Mackey-Glass 时间序列获得, 即前 300 个

数据由 $\tau = 17 \rightarrow \tau = 30$ 变化, 后 300 个数据由 $\tau = 30 \rightarrow \tau = 17$ 变化, 共 600 个数据. 快时变系统模型通过混合 4 个静态 Mackey-Glass 时间序列获得, 即由 $\tau = 17 \rightarrow \tau = 30, \tau = 30 \rightarrow \tau = 50, \tau = 50 \rightarrow \tau = 100, \tau = 100 \rightarrow \tau = 50, \tau = 50 \rightarrow \tau = 30, \tau = 30 \rightarrow \tau = 17$ 变化的数据各 100 个, 共 600 个数据.

图 7 和图 8 为 DAMNN 对不同时变系统的预测效果, 以及与典型全连接前馈神经网络 BP 和 RBF 神经网络的预测效果对比图(隐节点数与 DAMNN 的隐节点总数相同). 从这两个图可以看出, DAMNN 不仅能够有效预测不同的时变系统, 而且其预测效果明显优于典型前馈神经网络. 仔细观察图 7 和图 8 可以看出, RBF 神经网络对于时变系统也具有一定的预测能力, 原因是 RBF 神经元具有局部激活特性, 因此其对于时变系统也具有一定的知识积累特性. 但与 DAMNN 相比, RBF 神经网络中每个 RBF 神经元后面只有一个权连接, 而 DAMNN 的每个 RBF 神经元后面有一个子网络, 因此 DAMNN 不仅具有局部激活特性, 而且学习能力强于 RBF 神经网络.

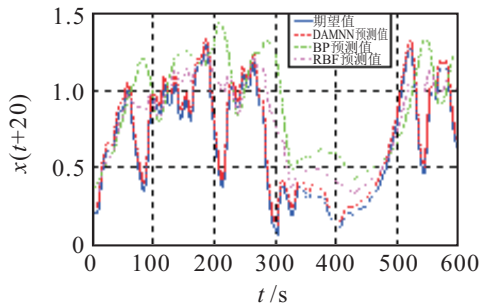


图 7 慢时变系统预测效果对比

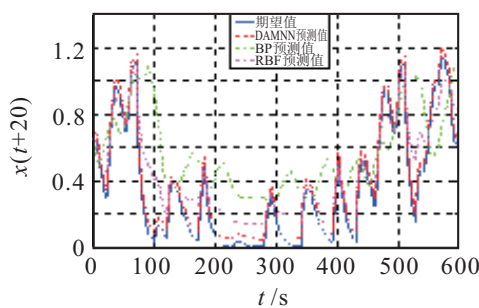


图 8 快时变系统预测效果对比

图 9 和图 10 为 DAMNN 预测不同时变系统时 RMSE 及网络系统中子网络数量的变化动态. 从这两个图可以看出, 无论是跟踪慢时变系统还是快时变系统, DAMNN 都能在学习的初始阶段很快构建自身的网络结构, 而且在整个学习过程中, 都能够根据当前所学习时变对象实时工况数据的分布动态调整自身的规模. 学习结束后, 跟踪慢时变系统的 DAMNN 中共有 8 个子网络, 共 39 个隐节点; 跟踪快时变系统的 DAMNN 中共有 12 个子网络, 共 60 个隐节点.

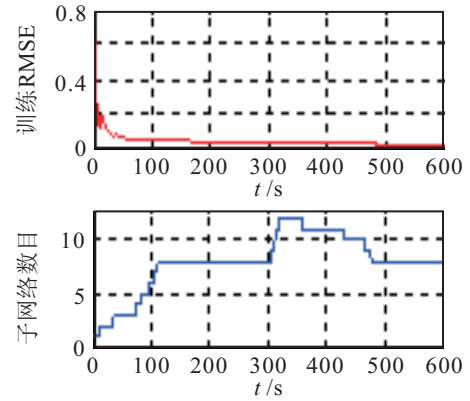


图 9 慢时变系统训练 RMSE 及子网络数动态

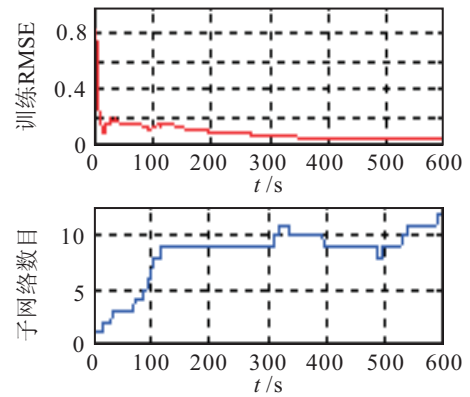


图 10 快时变系统训练 RMSE 及子网络数动态

对比图 4 与图 9 和图 10 还可看出, DAMNN 在跟踪静态 Mackey-Glass 时间序列(时不变系统)时, 网络结构一旦建立, 在整个学习期间网络结构基本保持不变(变化幅度较小); 而在跟踪时变系统时, 网络结构在整个学习期间一直都在调整, 这也表明 DAMNN 确实能够随着时变系统的变化而自组织调整自身网络规模.

表 1 所示为 DAMNN 在相同训练数据集上, 分别对静态 M-G 时间序列、慢时变系统和快时变系统的仿真结果与 BP 和 RBF 神经网络, 以及 SAMNN 和 SGMN 的对比结果. 实验时, 静态 Mackey-Glass 时间序列的数据集与文献 [7] 相同; 慢时变系统数据集与

表 1 不同神经网络模型性能比较

学习对象	神经网络类型	模块数	隐层节点总数	训练 RMSE	训练时间 s
静态 M-G 时间序列 ($\tau = 17$)	DAMNN	7	30	0.008 6	11.2
	SGMN	18	*	0.044 0	*
	BP	1	30	0.195 8	19.1
	RBF	1	30	0.083 0	7.8
慢时变系统	DAMNN	8	39	0.017 6	15.7
	SAMNN	21	*	0.021 5	*
	BP	1	41	0.367 3	29.5
	RBF	1	41	0.153 0	12.4
快时变系统	DAMNN	12	60	0.039 2	26.8
	BP	1	60	0.627 2	47.2
	RBF	1	60	0.336 4	21.7

注: * 结果在原文中没有给出.

文献[6]相同.表1中,SAMNN和SGMN的实验数据来自文献[6]和文献[7].

4 结 论

针对全连接前馈神经网络不能有效应对时变系统的问题,本文提出了一种动态自适应模块化神经网络,该网络具有“局部激活”、“自组织”和“动态集成”特性.利用该网络对不同时变系统进行预测,并详细分析了DAMNN的性能,得到以下结论:

1) DAMNN能够以实时工况数据流为驱动,根据所学习对象工况数据的空间分布自组织建立网络结构模型;

2) DAMNN的局部激活特性和系统动态集成特性使得DAMNN不仅具有知识积累特性,而且还具有较高的学习速度和精度.

参考文献(References)

- [1] Weisheng Chen, Licheng Jiao. Adaptive tracking for periodically time-varying and nonlinearly parameterized systems using multilayer neural networks[J]. IEEE Trans on Neural Networks, 2010, 21(2): 345-351.
- [2] Min Han, Jianchao Fan, Jun Wang. A dynamic feedforward neural network based on Gaussian particle swarm optimization and its application for predictive control[J]. IEEE Trans on Neural Networks, 2011, 22(9): 1458-1468.
- [3] Yuntong Wen, Xuemei Ren. Neural networks-based adaptive control for nonlinear time-varying systems with unknown control direction[J]. IEEE Trans on Neural Networks, 2011, 22(10): 1599-1612.
- [4] Hunt H, Sbarbaro D, Zbikow Ski R. Neural networks for control system—A survey[J]. Automatica, 1992, 28(6): 1083-1112.
- [5] Molina-Viaplana J, Feliu-Battle J, Lopez-Coronado J. A modular neural network architecture for step-wise learning of grasping tasks[J]. Neural Networks, 2007, 20(5): 631-645.
- [6] Viswanath Ramamurti, Joydeep Ghosh. Structurally adaptive modular networks for nonstationary environments[J]. IEEE Trans on Neural Networks, 1999, 10(1): 152-160.
- [7] Chu Kiong Loo, Mandava Rajeswari, Rao M V C. Novel direct and self-regulating approaches to determine optimum growing multi-experts network structure[J]. IEEE Trans on Neural Networks, 2004, 15(6): 1378-1395.
- [8] Jianye Sun. Local coupled feedforward neural network[J]. Neural Networks, 2010, 23(1): 108-113.
- [9] Angelov P P, Filev D P. An approach to online identification of Takagi-Sugeno fuzzy models[J]. IEEE Trans on Systems, Man, and Cybernetics, 2004, 34(1): 484-498.
- [10] 张昭昭, 乔俊飞. 基于在线减法聚类的RBF神经网络结构设计[J]. 控制与决策, 2012, 27(7): 997-1002. (Zhang Z Z, Qiao J F. Design RBF neural network architecture based on online subtractive clustering[J]. Control and Decision, 2012, 27(7): 997-1002.)
- [11] Plamen Angelov, Arthur Kordon. Adaptive inferential sensor based on evolving fuzzy models[J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2010, 40(2): 529-539.
- [12] Chris Tseng H, Bassam Almogahed. Modular neural networks with applications to pattern profiling problems[J]. Neurocomputing, 2009, 72(10): 2093-2100.
- [13] Hagan M T, Menhaj M B. Training feedforward networks with the Marquart algorithm[J]. IEEE Trans on Neural Networks, 1994, 5(6): 989-993.

(责任编辑: 李君玲)