

构建复杂回响状态网络的新方法

宋青松, 冯祖仁

(西安交通大学系统工程研究所, 710049, 西安)

摘要: 研究了人工神经网络混沌时间序列预测应用中的拓扑选择问题,受大脑皮层生长发育过程启发,提出一种类皮层网络——复杂回响状态网络(CESN)的构建方法.将影响皮层网络生长发育过程的依赖距离和依赖时间窗口的 2 个生长机制中的规模因子、距离敏感因子、密度敏感因子、种子神经元个数,以及时间窗宽度等调控因子,类比地定义为 CESN 构建过程中的构造参数.实验发现,CESN 的拓扑结构能够被这些构造参数唯一地确定,并且当种子神经元个数取值约为网络规模的 10% 时,构建出的 CESN 几乎具有最优的预测精度.根据这个发现,仿真证实了 CESN 能够被更省时地构建,表明所提方法是一种高效的网络构建方法.

关键词: 混沌时间序列预测;复杂度;回响状态网络

中图分类号: TP183 **文献标志码:** A **文章编号:** 0253-987X(2009)04-0001-04

A New Method to Construct Complex Echo State Networks

SONG Qingsong, FENG Zuren

(Institute of System Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: The issue of topology choosing in artificial neural networks for chaotic time series forecasting is studied. Motivated by the growth and development process of cerebral cortex, an algorithm for constructing cortex-like neural network, complex echo state network (CESN), is proposed. Some factors in both the distance-dependent growth mechanism and time-window-dependent growth mechanism, which have effects on the process of the growth and development of the cortex networks, are analogously defined as construction parameters that controls the process in creating a CESN, such as the size factor, the distance-sensitive factor, the density-sensitive factor, the seed-neuron number and the time-window width, etc. It is found from experiments that the topological characters of a CESN can be uniquely determined by these parameters, and that when the seed-neuron number is set to be about 10% of the whole network size, the created CESN can have almost the best forecasting performance. Simulation results show that CESNs can be created with less time-consumption and better forecasting performance. It can be concluded that the proposed algorithm is a more efficient network constructing method.

Keywords: chaotic time series prediction; complexity; echo state network

人工神经网络(ANN)可用于混沌时间序列预测,然而在 ANN 实施过程中存在两个难点:一是基于误差梯度的权重学习算法易陷入局部极小甚至不收敛;二是确定网络结构需采用耗时的“trial-and-error”或者进化方法.

回响状态网络(ESN)^[1]常用于 Mackey-Glass (MG)混沌时间序列预测,且性能获得了质的改善,其归一化均方根误差(NRMSE) R_{MSE} 刻画的预测结果约为 $10^{-4.2}$,远优于其他常见方法的预测结果($10^{-1.2} \sim 10^{-1.7}$)^[2].高维稀疏随机连接的动态库

(DR)是 ESN 核心之一,该库对输入激励具有丰富的表达能力.读取器通过线性或者非线性地组合该库中个体的活性状态,产生期望输出.简单的线性回归可胜任 ESN 训练任务,回避 ANN 方法遭遇的权重学习难题.如何精确地构建预测性能更好的 ESN,目前仍是亟待解决的问题,如具有相同谱半径设置的 ESN 会导致显著差异的预测结果等^[3].

针对 ESN 拓扑选择问题,受哺乳动物大脑皮层研究成果的启发,本文提出一种由 5 个构造参数调控的复杂回响状态网络(CESN)构建方法.

1 复杂回响状态网络构建方法

1.1 复杂网络生长算法

受具有多源信息融合、长、短时段记忆等强大信息处理能力的哺乳动物大脑皮层生长机制的启发^[4],本文设计了更高效的 ESN.为真实模拟部分大脑皮层的拓扑结构,Kaiser 等人提出一类复杂网络构造算法^[5].在该算法中采用 2 个机制来确定生物学似是而非的大规模网络的生长过程:依赖距离的生长机制,即 2 个神经元节点之间建立连接的概率随着二者间空间距离的增加指数级减小,连接“趋近避远”;依赖特效时间窗的生长机制,即时间窗性质由种子神经元个数和时间窗宽度确定,每个种子神经元内在具有一个特效时间窗,种子神经元个数等同于最终网络中“簇”的个数.由于特效时间窗的存在,归属同一个种子神经元的非种子神经元被优先地与该种子神经元建立连接,所以 2 个神经元节点之间建立连接的概率同时依赖于它们之间的空间距离和各自对应种子神经元的特效时间窗.

本文提出了面向混沌时间序列预测的复杂网络生长算法,其中依赖距离的生长机制由距离敏感指数和网络密度敏感指数来确定,依赖特效时间窗的生长机制由种子神经元的个数和特效时间窗的宽度来确定.将网络的最终规模数、距离敏感指数、密度敏感指数、种子神经元个数、时间窗宽度(假定所有时间窗等宽)作为网络构造参数,依次记为 N, α, β, n 和 w .复杂网络生长算法描述如下.

(1)初始化.设置 N, α, β, n 和 w ;选择依赖距离的概率 $P_{\text{dist}}^{u,v}(d(u, v))$ 和依赖时间窗口的概率 $P_{\text{time}}^{n_j}(t; n, w)$,其中 u 和 v 为任意 2 个神经元, n_j 为种子神经元, $j=1, \dots, n$. $P_{\text{dist}}^{u,v}(d(u, v)) = \beta e^{-\alpha d(u, v)}$,其中 $d(u, v)$ 是 u 和 v 之间的欧氏距离,而 $P_{\text{time}}^{n_j}(t; n, w)$ 为文献[6]中定义的函数形式.将 n 个种子神经元以时间步长 $j/(n+1)$ 在长、宽均为 $[0, 1]$ 的平

面上任意布置,则 $m=N-n$ 个非种子神经元将以时间步长 $1/m$ 被逐个释放.记步骤 $s=1$.

(2)计算建立连接的概率.随机释放一个非种子神经元 u ,并将距离 u 最近的种子神经元选为 u 的种子神经元 n_u .若多于一个的种子神经元具有最短距离,则从中任意抽取一个,再根据式(1)计算 u 与其他任意神经元 v 建立连接的概率

$$P(u, v) = P_{\text{dist}}^{u,v}(d(u, v)) P_{\text{time}}^{n_u}(s/m) \cdot P_{\text{time}}^{n_v}(s/m) \quad (1)$$

为了加速生长,如果在 u 上没有新的连接,即 $P(u, v)$ 总是小于某个固定常数,则将 u 随机与一个种子神经元连接.此时,记步骤 $s=s+1$.

(3)如果 $s=m$,则生长过程结束,由此得到的网络用其邻接矩阵 $\mathbf{W}_{\text{adj}}$ 表示;否则,转(2).

此网络构建算法与 Kaiser 的算法比较,生长过程更快,随机化程度更高.

1.2 复杂回响状态网络

由 K 个个体神经元构成一个输入单元,由 L 个个体神经元构成一个输出单元,由 N 个个体神经元构成一个 DR.将输入单元中每个个体到 DR 中的每个个体的连接权重、DR 中的每个个体到输出单元中每个个体的连接权重以及 DR 中个体之间的连接权重,分别记作一个大小为 $N \times K$ 的矩阵 $\mathbf{W}^{\text{in}}$ 、一个大小为 $L \times N$ 的矩阵 $\mathbf{W}^{\text{out}}$ 以及一个大小为 $N \times N$ 的矩阵 $\mathbf{W}$,其中 $\mathbf{W}$ 为邻接矩阵 $\mathbf{W}_{\text{adj}}$ 与一个随机方阵的点积;将 $\mathbf{W}^{\text{back}}$ 记作反馈连接权重矩阵,输入、输出和 DR 单元中的个体在 t 时刻的状态分别记为 $\mathbf{I}(t)$ 、 $\mathbf{Y}(t)$ 和 $\mathbf{X}(t)$.CESN 体系如图 1 所示,可以看到生成的网络取代了常规的 ESN(JESN)中的 DR.

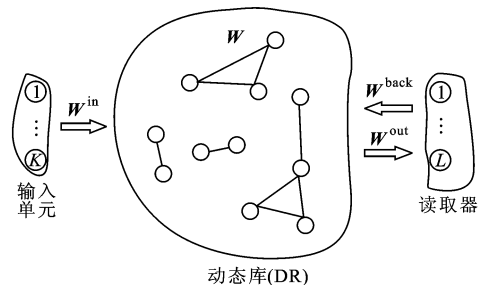


图1 复杂回响状态网络体系示意图

CESN 的状态方程和输出方程为

$$\mathbf{X}(t+1) = f^{\text{PE}}(\mathbf{W}^{\text{in}} \mathbf{I}(t) + \mathbf{W} \mathbf{X}(t) + \mathbf{W}^{\text{back}} \mathbf{Y}(t)) \quad (2)$$

$$\mathbf{Y}(t+1) = f^{\text{O}}(\mathbf{W}^{\text{out}} \mathbf{X}(t+1)) \quad (3)$$

式中: f^{PE} 和 f^{O} 分别是 DR 单元和输出单元中个体

的活性函数,二者均为双曲正切函数。

2 实验及结果

2.1 实验设计

实验的目的是研究网络构造参数对 CESN 预测性能的影响,比较 CESN 与 JESN 的预测精度。MG 时间序列预测被认为是一类标准问题,其系统差分方程为

$$\frac{dz}{dt} = \frac{0.2z(t-\tau)}{1+z(t-\tau)^{10}} - 0.1z(t) \quad (4)$$

当延时常数值大于或等于 17 时,混沌发生。取延时常数值等于 17,使用 Matlab 中的函数 dde23 求解方程(4),得到一个长度为 10 000 时间步的混沌时间序列(序列最大李雅普诺夫指数约为 0.006),本文将将其作为预测数据集。

对 MG 混沌时间序列进行 84 步前向预测,选取 NRMSE 评价预测结果。由于 MG 混沌时间序列预测本质上是一类系统辨识问题,所以必须首先建立 MG 动力系统的“黑箱”模型,即学习 CESN 的输出权重矩阵 $\mathbf{W}^{\text{out}}$,训练方法见文献[1]。

首先将 MG 时间预测的真实值作为教师信号 $\mathbf{Y}_T$,通过 $\mathbf{W}^{\text{back}}$ 送入 DR,然后从第 1 001 个时间步开始到第 3 000 个时间步结束,逐步将 DR 状态 $\mathbf{X}(t)$ 汇聚成一个状态收集矩阵 $\mathbf{M}$ 。当状态收集过程结束时,对 $\mathbf{Y}_T$ 和 $\mathbf{M}$ 用线性回归方法计算输出连接权重 $\mathbf{W}^{\text{out}} = \mathbf{M}^{-1} \mathbf{Y}_T$,学习过程结束。之后,利用学习获得的 CESN,对该序列的其他片断进行预测。

每生成一个 CESN,对 MG 混沌时间序列中 50 个不同片断依次预测,记录下每个片段的第 84 步预测值 $y_{\text{out}}^i(84)$ 和真实值 $y_{\text{true}}^i(84)$, $i = 1, \dots, 50$,再根据

$$R_{\text{MSE}} = \left(\frac{\sum_{i=1}^{50} (y_{\text{true}}^i(84) - y_{\text{out}}^i(84))^2}{50\sigma^2} \right)^{1/2} \quad (5)$$

计算 NRMSE,其中 σ 是 MG 时间序列的标准差。

在构造参数组 (N, α, β, n, w) 的每一次赋值上, CESN 可独立实现 20 次。

$$\bar{R} = \frac{1}{20} \sum_{i=1}^{20} (R_{\text{MSE}})_i \quad (6)$$

$$R_{\text{std}} = \left[\frac{1}{20-1} \sum_{i=1}^{20} [(R_{\text{MSE}})_i - \bar{R}]^2 \right]^{\frac{1}{2}} \quad (7)$$

根据式(6)和式(7),可求得 CESN 的 20 次独立产生的 20 个 NRMSE 的均值和标准差,其可作为构造参数赋值对应的预测结果。

2.2 实验结果

固定 α, β 和 w ,研究 N 和 n 对 NRMSE 的影响。 N 的赋值分别为 100、200、300 和 500,对应 N 的每一次赋值, n 依次赋值为 1、 $N \times 10\%$ 、 $N \times 30\%$ 、 $N \times 50\%$ 和 $N \times 80\%$ 。定义 $r = n/N$,则 r 分别取 0(约等于 0)、0.1、0.3、0.5、0.8。 $\bar{R}$ 随 r 的变化如图 2 所示。 R_{std} 约为对应 $\bar{R}$ 的 10%。

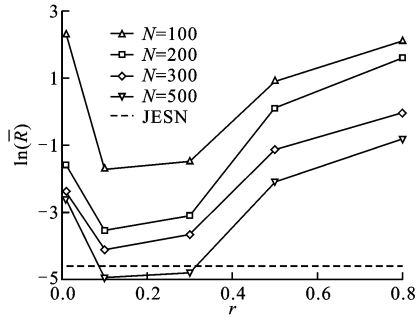


图 2 不同构造参数赋值对应 CESN 的预测误差

从图 2 看到了 3 种现象:第 1,当 n 固定,例如 $n = 1$,即在 $r \approx 0$ 的情形下, $\bar{R}$ 随 N 的增加而减小,也就是规模越大的 DR,预测精度越高;第 2,当 N 固定, $\bar{R}$ 和 r 与 n 之间突现出非单调函数关系,即取得较小 $\bar{R}$ 值的 CESN,其对应的 n 取值既不得太大,也不得太小,应取中间值,特别是当 $r \approx 0.1$,也就是 $n \approx N \times 10\%$,相应地 CESN 获得最小的 $\bar{R}$,亦为最好的预测精度;第 3,当 N 赋值为 500, r 取值为 0.1~0.3,即 n 取值为 50~150 时, CESN 比等规模、等连通密度和等谱半径的 JESN 能够取得更好的预测精度。这 3 种现象表明, CESN 用于 MG 混沌时间序列预测不仅是有效的,而且在适当构造参数赋值的情形下,能够获得比 JESN 更好的预测性能,即 CESN 比 JESN 效率更高。

用同样的方法,研究 α, β 和 w 对 NRMSE 的影响。实验中,最优预测精度 CESN 的 DR 邻接矩阵如图 3 所示,其中神经元 $i, j = 1, \dots, 500$,连通密度

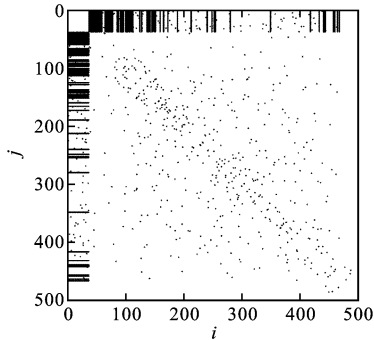


图 3 最优预测精度 CESN 的 DR 邻接矩阵示意图

为 0.027 8, 谱半径为 0.85, N, n, α, β, w 取值分别为 500、50、40、0.5 和 0.5.

3 动态库对激励的表达能力分析

CESN 预测精度的差异本质上是其 DR 对输入激励的表达能力的存在差异, 该能力越强, CESN 的预测精度越高^[3]. 用 Omega 复杂性 Ω 可以度量 DR 对输入激励的表达能力的. Ω 常用于刻画多通道脑电信号空间的复杂度, 以判断癫痫是否发作或熟睡的程度^[7].

可以通过下式在 DR 上定义 Ω , 即

$$C = \frac{1}{2\ 000} \sum_{t=1\ 001}^{3\ 000} \mathbf{X}^T(t) \mathbf{X}(t) \quad (8)$$

$$\ln \Omega = - \sum_i \lambda'_i \ln \lambda'_i \quad (9)$$

式中: $\mathbf{X}(t)$ 是训练过程中 DR 的状态向量; $\lambda'_i = \lambda_i / \sum_i \lambda_i$, λ_i ($i = 1, \dots, N$) 是状态协方差矩阵 C 的特征值. 当所有个体的活动过程彼此完全同步时, 则 C 的特征值仅一个不为 0 且 Ω 取最小的值, 这表明此情形下 DR 具有最弱的激励表达能力; 当所有个体的活动过程彼此非同步时, 则 C 具有 N 个均匀分布、非 0 特征值的对角阵且 Ω 取最大值, 这表明此情形下 DR 具有最强的激励表达能力.

$N=500$ 时不同种子神经元个数对应 CESN 的复杂性度量如图 4 所示, 图 4 中 n 取值与图 2 中 r 取值一一对应. 结合图 4 和图 2 发现, Ω 和 $\bar{R}$ 之间呈现较强的负相关性, 即 Ω 越大, $\bar{R}$ 越小, 反之亦然. 此外, 在 $n \approx N \times 10\%$ 处同时突现最大 Ω 值和最小 $\bar{R}$ 值, 这表明在不同的构造参数下, DR 对输入激励的表达能力的差异能够被 Ω 有效地刻画. 另一方面, JESN 的 Ω 值小于 $n = N \times 10\%$ 情形下对应的 CESN 的 Ω 值, 这说明在此构造参数赋值情形下, CESN 的预测精度优于 JESN.

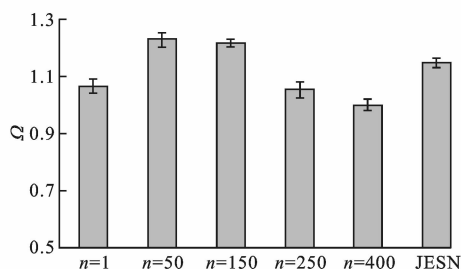


图 4 $N=500$ 时不同种子神经元个数对应 CESN 的复杂性度量

4 结束语

本文提出一类由 5 个构造参数决定的 CESN 构建算法. 实验表明, CESN 不仅继承了 JESN 的所有内在优点, 诸如能有效避免回归神经网络实施中常见的训练算法难点和耗时的拓扑搜索过程, 而且在构造参数 $n \approx N \times 10\%$ 赋值情形下, 能够获取比 JESN 更好的预测精度.

参考文献:

- [1] JAEGER H, HAASS H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication [J]. Science, 2004, 304(5667):78-80.
- [2] 王军平, 陈全世, 田光宇. 随机模糊神经网络及在随机混沌时间序列预测中的应用 [J]. 西安交通大学学报, 2003, 37(10):991-994.
WANG Junping, CHEN Quanshi, TIAN Guangyu. Stochastic fuzzy neural network and its application to prediction of stochastic chaotic time series [J]. Journal of Xi'an Jiaotong University, 2003, 37(10):991-994.
- [3] OZTURK M C, XU D, PRINCIPE J C. Analysis and design of echo state networks [J]. Neural Computation, 2007, 19(1):111-138.
- [4] EDELMAN G M, TONONI G. A universe of consciousness: how matter becomes imagination [M]. New York, USA: Basic Books, 2000.
- [5] KAISER M, HILGETAG C C. Development of multi-cluster cortical networks by time windows for spatial growth [J]. Neurocomputing, 2007, 70(10/12):1829-1832.
- [6] NISBACH F, KAISER M. Developmental time windows for spatial growth generate multiple-cluster small-world networks [J]. The European Physical Journal: B, 2007, 58(2): 185-191.
- [7] 裴晓梅, 郑崇勋, 宾光宇. 基于多通道脑电特征运动意识任务的分类 [J]. 西安交通大学学报, 2005, 39(8): 904-907.
PEI Xiaomei, ZHENG Chongxun, BIN Guangyu. Classification of the mental tasks based on multichannel electroencephalogram features [J]. Journal of Xi'an Jiaotong University, 2005, 39(8):904-907.

(编辑 苗凌)