

一种连续型深度信念网的设计与应用

乔俊飞¹ 潘广源¹ 韩红桂¹

摘要 针对深度信念网 (Deep belief network, DBN) 学习连续数据时预测精度较差问题, 提出一种双隐层连续型深度信念网. 该网络首先对输入数据进行无监督训练, 利用连续型传递函数实现数据特征提取, 设计基于对比分歧算法的权值训练方法, 并通过误差反传对隐层权值进行局部寻优, 给出稳定性分析, 保证训练输出结果稳定在指定区域. 利用 Lorenz 混沌序列、CATS 序列和大气 CO₂ 预测实验对该网络进行测试, 结果表明, 连续型深度信念网具有结构精简、收敛速度快、预测精度高等优点.

关键词 深度学习, 神经网络, 结构设计, 稳定分析, 时序预测

引用格式 乔俊飞, 潘广源, 韩红桂. 一种连续型深度信念网的设计与应用. 自动化学报, 2015, 41(12): 2138–2146

DOI 10.16383/j.aas.2015.c150239

Design and Application of Continuous Deep Belief Network

QIAO Jun-Fei¹ PAN Guang-Yuan¹ HAN Hong-Gui¹

Abstract A continuous deep belief network (cDBN) with two hidden layers is proposed to solve the problem of low accuracy of traditional DBN in modeling continuous data. The whole process is to train the input data in an unsupervised way using continuous version of transfer function, to design the contrastive divergence in hidden-layer training process, and then to fine-tune the net by back propagation. Besides, hyper-parameters are analyzed according to stability analysis, as is given in the paper, to make sure the network finds the optimal. Experiments on Lorenz, CATS benchmark simulation and CO₂ forecasting show a simplified structure, fast convergence speed and accuracy of this cDBN.

Key words Deep learning, neural networks, structural design, stability, time series forecasting

Citation Qiao Jun-Fei, Pan Guang-Yuan, Han Hong-Gui. Design and application of continuous deep belief network. *Acta Automatica Sinica*, 2015, 41(12): 2138–2146

深度信念网 (Deep belief network, DBN) 是受人脑认知推理过程研究结果的启发而设计的一种人工神经网络, 其采用无监督训练方式, 具有较好的降维性能, 克服了传统人工神经网络需要大量有监督信号和易陷入局部极小等缺点^[1]. 自 2006 年该网络被提出以来, 成功解决了大规模数据计算耗时问题且精度较高, 至今已成功应用于多种人工智能问题的研究, 尤其在图像处理、声音辨识和智能网络分析

等问题中效果显著, 受到众多学者关注^[2–3]. DBN 通过将多个受限玻尔兹曼机 (Restricted Boltzmann machine, RBM) 逐层累加, 采用贪心无监督方式对数据进行预训练, 获取关于输入数据较好的分布式表示, 进而使用误差反传通过有监督训练方式对整个网络参数进行调整, 能够获得较优结果^[4]. DBN 的无监督训练阶段可以被视为有监督训练阶段网络权值初始化, 将其确定在一个较好的范围, 有利于提高无监督训练阶段的学习和收敛效率^[5]. DBN 通过这种方法克服了传统神经网络易受初始参数影响, 处理高维度数据速度较慢等缺点, 且与传统浅层神经网络相比, 更好地模拟人脑感知信息和处理信息的深层结构, 实现了对大量数据进行快速有效学习, 已成功应用于模式识别^[6]、特征检测^[7]、文档分类^[8]、信号采集和降噪^[9] 等领域.

虽然 DBN 在人工智能领域特别是模式识别任务中取得了较好的成果, 然而由于其在运算时使用数据离散化方法进行特征提取, 隐含层和可视层节点在采样过程中均为伯努利值, 限制了 DBN 在更多工业领域的研究和应用^[10–13], 例如在时序预测问题中难以学习并实现较高精度的预测^[14–15]. Taylor 等提出改进方法, 使输入层能够接受连续型信号, 但

收稿日期 2015-04-21 录用日期 2015-09-23
Manuscript received April 21, 2015; accepted September 23, 2015

国家自然科学基金 (61203099, 61225016, 61533002), 北京市科技计划课题 (Z141100001414005, Z141101004414058), 高等学校博士学科点专项科研基金资助课题 (20131103110016), 北京市科技新星计划 (Z131104000413007), 北京市教育委员会科研计划项目 (KZ201410005002, km201410005001) 资助

Supported by National Natural Science Foundation of China (61203099, 61225016, 61533002), Beijing Science and Technology Project (Z141100001414005, Z141101004414058), Program Foundation from Ministry of Education (20131103110016), Beijing Nova Program (Z131104000413007), Beijing Municipal Education Commission Science and Technology Development Program (KZ201410005002, km201410005001)

本文责任编辑 魏庆来

Recommended by Associate Editor WEI Qing-Lai

1. 北京工业大学电子信息与控制工程学院计算智能与智能系统北京市重点实验室 北京 100124

1. College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124

数据训练过程以及输出层数据仍为离散值^[16]. Chen 等对 RBM 的研究工作证明, RBM 理论上具有处理连续数据并进行函数建模能力^[17]. 最近的研究工作表明, RBM 能够很好应用于连续数据建模问题研究. Zhang 等将 RBM 改进成连续型 RBM, 用于预测股票数据, 通过尝试多种网络结构和参数, 获得较好的预测效果^[18]. Chen 等提出一种连续型 RBM, 预测淮河水水质参数并获得较好效果^[19], Song 等通过使用神经网络方法提出一种新的自适应动态规划 (Adaptive dynamic programming, ADP) 方法, 在非线性控制系统中通过误差逼近补偿实现稳定控制^[20], 此外 Song 也使用神经网络在连续数值的控制系统中实现有效控制^[21].

针对 DBN 处理连续数据的局限性, 文中提出一种双隐含层连续型深度信念网 (Continuous DBN, cDBN), 实现 DBN 在时序预测建模中的应用. 利用对比分歧 (Contrastive divergence, CD)^[22] 对每个隐含层进行逐层训练, 上一个 RBM 的输出作为下一个 RBM 的输入. 通过稳定性分析, 提出稳定性定理, 得到改进型网络参数选择方法. 将该改进网络应用于洛伦兹混沌序列、人工标准时间序列、大气二氧化碳变化预测问题的研究, 结果表明该连续型 DBN 能够更稳定和更精确地实现预测, 初步验证了该网络在工业应用中的可行性.

本文安排如下. 第 1 节介绍传统深度信念网及改进 cDBN 算法和结构; 第 2 节进行稳定性分析, 证明稳定性定理, 得到合理的参数选择方式; 第 3 节进行仿真实验, 展示实验结果和分析; 最后是总结.

1 连续型深度信念网

1.1 cDBN 结构

文中提出 cDBN 结构如图 1 所示. 图 1 中包含输入层 (或可视层) V , 接受时间序列 $t, t-1, \dots, t-n$ 时刻数据. 两个隐含层 $L1, L2$. V 和 $L1$ 构成第一个 RBM, $L1$ 和 $L2$ 构成第二个 RBM, 双向箭头表示相邻两层之间使用全连接, 同一层的内部神经元之间没有连接. 训练过程中每一层提取上一层的信息和特征, 获得对数据更好的表示. 输出层只有一个神经元, 其物理意义是学习到的输入数据特征, 数学意义是 $t+1$ 时刻预测值. 由于传统 DBN 使用双隐含层已经能够达到比较理想的效果^[23], 因此对连续型 DBN 也使用同样的结构, 它比单隐含层网络具有更好的特征提取能力, 同时又满足工业领域对运算速度的要求.

1.2 cDBN 算法

DBN 通常由数个 RBM 相互叠加, 对每个 RBM 使用无监督方式逐一训练, 最后将整个网络展开成一个前向型网络, 利用 Back propagation (BP)

对网络权值进行调整^[1]. 因此 DBN 属于递归网络, 其训练方式为无监督和有监督相结合. 式 (1) 和 (2) 分别为 RBM 中的学习过程和推理过程, 统称知识生成.

$$p(h_j = 1) = \frac{1}{1 + e^{-b_j - \sum_i v_i w_{ij}}} \quad (1)$$

$$p(v_i = 1) = \frac{1}{1 + e^{-c_i - \sum_j h_j w_{ji}}} \quad (2)$$

式中, v_i 是可视层 (输入层) 第 i 个节点值, h_j 是隐含层 (RBM 输出层) 第 j 个节点值, b 和 c 分别表示可视层和隐含层偏置值, w_{ij} 为可视节点 i 和隐藏节点 j 之间权值. 令 $\Phi = (W, b, c)$, 则式 (3) 表示了权值和偏置值更新方式.

$$\Phi^{(\tau+1)} = \Phi^{(\tau)} + \Delta\Phi = \langle h_j^0 v_i^0 \rangle - \langle h_j^1 v_i^1 \rangle \quad (3)$$

式中, $\langle \cdot \rangle$ 表示对状态采样获得的平均值, $h_j^0 v_i^0$ 为初始状态分布的乘积, $h_j^1 v_i^1$ 为运行马尔科夫一次迭代之后得到的状态, τ 表示无监督训练第 τ 次训练.

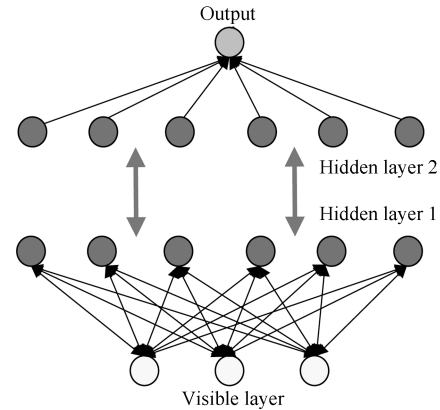


图 1 cDBN 结构

Fig. 1 cDBN's structure

在 cDBN 中, 令 s_j 和 s_i 代表隐含层神经元 j 和可视层神经元 i , 将式 (1) 和 (2) 的 Sigmoid 传递函数保留, 去掉之后将结果离散为二元制过程, 进一步添加噪声变量实现对连续化的转变, 公式描述如下.

$$s_j = \varphi_j \left(\sum_i w_{ij} s_i + \sigma \cdot N_j(0, 1) \right) \quad (4)$$

$$s_i = \varphi_i \left(\sum_j w_{ij} s_j + \sigma \cdot N_i(0, 1) \right) \quad (5)$$

同时

$$\varphi_j(x_j) = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + e^{-a_j x_j}} \quad (6)$$

$$\varphi_i(x_i) = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + e^{-a_i x_i}} \quad (7)$$

式 (4) 和 (5) 分别表示 cDBN 学习过程和推理过程, 式中, $N(0, 1)$ 代表均值为 0, 方差为 1 的高斯随机变量, σ 是常量, $\varphi(\cdot)$ 是渐近线为 θ_H 和 θ_L 的 Sigmoid 函数, a 是噪声控制变量, 代表对 Sigmoid 函数斜度的控制. 根据 CD 准则, 权值和偏置值更新公式为

$$\Delta w_{ij} = \eta_w (\langle s_i^0 s_j^0 \rangle - \langle s_i^1 s_j^1 \rangle) \quad (8)$$

$$\Delta b = \frac{\eta_b}{b^2} (\langle s_j^{0^2} \rangle - \langle s_j^{1^2} \rangle) \quad (9)$$

$$\Delta c = \frac{\eta_c}{c^2} (\langle s_i^{0^2} \rangle - \langle s_i^{1^2} \rangle) \quad (10)$$

式中, η_w, η_b, η_c 均为学习速率, $\langle \cdot \rangle$ 的定义同式 (3).

2 稳定性分析

cDBN 可以被视为一种包含多层隐单元的概率生成模型, 同时也是一种多层简单模型组合的复合模型. 实际上, cDBN 中的每一层都相当于连续型 RBM (即过去离散化的 RBM), 因此训练时可以使用高效的无监督训练方式. 连续型 RBM 作为一种无向的基于能量函数的生成模型, 在无监督训练阶段采用对比分歧 (CD) 算法^[22] 对网络权值进行进一步快速更新. 在传统方法中, 要把 Hidden layer 节点值置为二值化形式 (0 或 1), 而不是直接采用其概率值本身, 但在文中的连续型 DBN 中, 因为使用式 (6) 和 (7) 避免了运算过程中将概率值直接作为输出值的理论冲突, 仍可使用 CD 算法作为快速训练策略. 则有:

$$s_i^0 \in [0, 1] \quad (11)$$

$$s_j^0 = \varphi_j \left(\sum_i w_{ij} s_i^0 + \sigma \cdot N_j(0, 1) \right) \quad (12)$$

$$s_i^1 = \varphi_i \left(\sum_j w_{ij} s_j^0 + \sigma \cdot N_i(0, 1) \right) \quad (13)$$

$$s_j^1 = \varphi_j \left(\sum_i w_{ij} s_i^1 + \sigma \cdot N_j(0, 1) \right) \quad (14)$$

对比分歧算法是一种对最大似然的近似, 能够用于学习 cDBN 中每一个连续型 RBM 连接的权重. 当第一个 RBM 被训练完毕后, 第二个连续型 RBM 可被堆叠在已经训练完成的第一个连续型 RBM 上, 形成一个多层模型. 在堆叠过程中, 第一个连续型 RBM 的输入层被初始化为训练样本, 权重为先前训练得到的权重, 该网络的输出作为第二个连续型 RBM 的输入, 之后重复刚才的单层训练过程, 直到达到某个期望的终止条件. 例如在式 (11)~(14) 中, s_j^0 是第一个 RBM 输出, 同时也是第二个 RBM 的输入.

由于 CD 算法并不在任何函数的梯度方向上, 因此其对最大似然的近似十分粗略, 因此分析 cDBN 稳定性, 选取合理的参数, 保证神经网络稳定收敛于解空间, 是 CD 算法有效应用于 cDBN 的理论依据和必要条件. 通过分析, 在上式中, $s_j^1(RBM1) = s_i^0(RBM2)$, 根据式 (11), RBM 输出范围是 $[0, 1]$, 只有当训练网络的输出在规定范围之内, 不会出现发散, 才可以保证训练的可行性. 因此, 为了保证网络的稳定, 需保证 $s_j^1 \in [0, 1]$. 提出以下定理.

定理 1. 若系统输出稳定, 即 $s_j^1 \in [0, 1]$, 当且仅当中间状态满足 $s_j^0, s_i^1 \in [0, 1]$.

证明. 设 $s_j^0 \in [t_L^1, t_H^1]$, $s_i^1 \in [t_L^2, t_H^2]$, 其中 t_L^1, t_H^1 分别为 s_j^0 的假设下界和上界, t_L^2, t_H^2 分别为 s_i^1 的假设下界和上界.

1) 充分条件

当中间状态 $s_j^0, s_i^1 \in [0, 1]$, 在训练一个 RBM 时, 根据式 (12)~(14) 可知, 网络输出 s_j^1 满足 $s_j^1 \in [0, 1]$, 输出在规定范围之内, 网络稳定.

2) 必要条件

若网络稳定, 且,

a) 若 $t_L^1 > 0$, 因传递函数为单调函数, 通过 CD 算法得到中间状态 s_j^0, s_i^1 , 必有 $s_i^1 > s_j^0$, 因此 $t_L^2 > t_L^1$, 根据式 (14) 可知, $s_j^1 > s_i^1$, 因此网络输出必大于 0, 矛盾.

因此 $t_L^1 \leq 0$.

b) 同理可知 $t_L^1 \geq 0$.

因此 $t_L^1 = 0$.

c) 同样证法得知 $t_H^1 = 1$.

因此 $s_j^0 \in [0, 1]$.

同理可证, 当使用 CD 算法进行 Gibbs 采样的另外两个过程中, 输出也为 $[0, 1]$, 即 $s_i^1 \in [0, 1]$. $\square$

因此, 需保证

$$s_j^0 = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + e^{-a_j x_j}} = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + e^{-a_j (\sum_i w_{ij} s_i^0 + \sigma \cdot N_j(0, 1))}} \in [0, 1]$$

成立, 上式即为 cDBN 的参数选择条件. 若令高斯噪声为 0, 对于 $x > 0$, 有 $0 < e^{-x} < 1$, 所以有 $s_j^0 \in [(\theta_H + \theta_L)/2, \theta_H]$, 因此可选取 $\theta_L = -1, \theta_H = 1, a_j = 1$, 使网络输出为 $[0, 1]$. 实际上, 按照该定理, 不管使用多少 RBM 叠加, 当网络满足输入为区间 $[0, 1]$ 之间的连续值时, 运算中间状态和输出仍为 $[0, 1]$ 连续值, 网络输出收敛于规定范围之内, cDBN 的计算和知识推理具有可靠性. cDBN 的训练过程为, 首先, 采用无监督训练寻找较优解区域, 训练过程中采用对比分歧算法加快训练速度; 然后, 有监督训练过程中采用误差反向传播算法进行局部寻优和

权值调整.

3 实验及分析

时序预测是现实生活中的常见问题之一, 通过统计和分析历史数据, 获得其变化规律和趋势, 对未来数据进行预测和分析, 从而在生活、工业等领域对未来可能出现的情况进行预先调整^[24]. 因易受自身及外部环境影响的特点, 预测建模是一个较复杂的非线性问题, 重要且充满挑战. 传统的线性技术需要对数据进行特殊化处理, 且往往得到较差结果, 因此不适用于解决此问题. cDBN 采用无监督训练方式, 具有较好的特征学习和特征提取能力. 根据任意逼近定理, 神经网络能以任意精度逼近任意非线性函数^[25]. 此外, 作为一个映射工具, cDBN 可被视为黑箱模型而不需要知道内部参数运算机理, 而且 cDBN 的无监督训练阶段降低了有监督训练过程中随机初始权值导致训练结果失败的可能性. 将 cDBN 应用于时序预测建模一步预测及多步预测实验, 实验目的是测试 cDBN 的函数逼近能力. 网络输入为时间序列中的真实数据, 其中多步预测实验需将上一步预测输出作为下一步的输入. 对于深度信念网, 目前理论上尚无统一方法确定隐含层神经元数, 不同问题需要不同规模的神经元数, 实验中将凭经验选择. 网络输出为单神经元, 代表预测数据.

3.1 Lorenz 时序预测

为了验证 cDBN 能够以精简结构达到较好的预测精度. 混沌时间序列预测是检验神经网络结构和方法有效性的基准问题之一, Lorenz 效应因其相位图酷似蝴蝶又称蝴蝶效应, 是指气象对初始条件具有敏感依赖性, 因此其长期预测几乎是不可行的^[26]. 其方程表示如下,

$$\begin{cases} \frac{dx}{dt} = \sigma(y_t - x_t) \\ \frac{dy}{dt} = -x_t z_t + \gamma(x_t - y_t) \\ \frac{dz}{dt} = x_t y_t - \beta z_t \end{cases} \quad (15)$$

式中, x 是对流强度, y 表示上升与下沉气流之间水平温度差, z 为垂直温度差. 参数 $\sigma = 10, \gamma = 28, \beta = 8/3$.

实验中, 生成 x 维度上的时间序列, 去掉前 1000 组数据从而降低过渡阶段对时间序列的影响, 训练组为 1500 组数据, 测试组为 1000 组数据. 网络使用 3 输入进行一步预测, 即用 $t-2, t-1, t$ 时刻信息推测 $t+1$ 时刻数据, 双隐含层神经元数均为 3, 即网络结构为 3-3-3-1. 均方根误差 (Root mean square error, RMSE) 作为性能评价指标, 如下所

示:

$$\text{RMSE} = \sqrt{\frac{\sum_N (y(k) - \hat{y}(k))^2}{N}} \quad (16)$$

式中, $y(k), \hat{y}(k)$ 分别代表 k 时刻的真实值和网络预测值. 知识生成阶段 (无监督学习过程) 迭代周期 100, 调优阶段 (有监督学习过程) 训练速率为 2, 训练步数 5000, 在相同的计算机条件下比较, 实验结果如图 2.

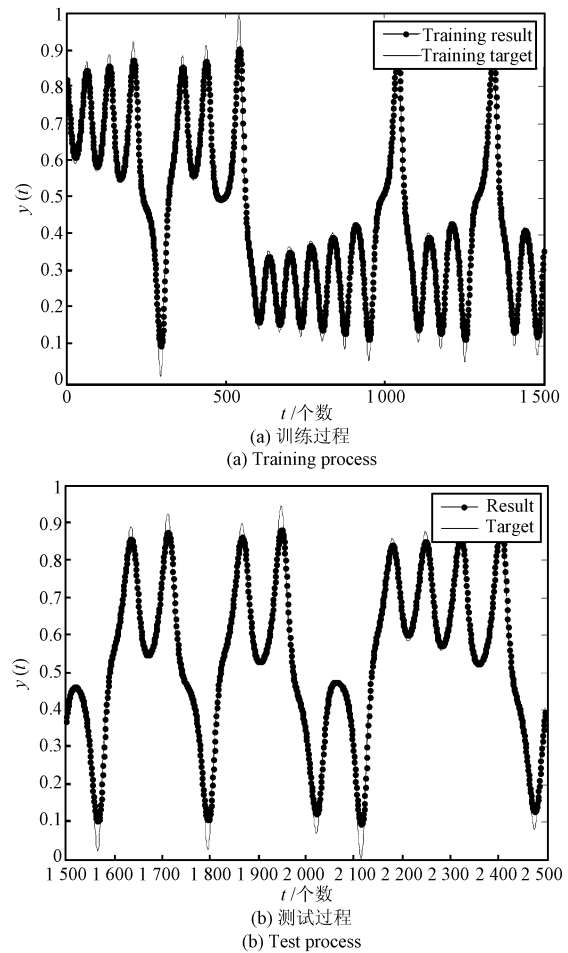


图 2 Lorenz 时序预测 RMSE
Fig. 2 RMSE of training and testing in Lorenz experiment

可以看出, 尽管在有些较陡的峰值部分学习困难, 网络仍能够在训练阶段较快地实现对函数的逼近和跟踪, 最终训练误差为 0.0208. 在测试阶段, 情况与训练阶段并无太大差异, 网络表现出较好的逼近和泛化能力, cDBN 较好地学习到了 Lorenz 状态转移规律并实现一步预测, 测试误差为 0.0223. 将 cDBN 与其他算法相比较, 如表 1 所示.

表 1 Lorenz 实验结果对比
Table 1 Results of Lorenz experiment

方法	神经 元数	训练 RMSE	测试 RMSE	运算 时间 (s)
cDBN	3-3	0.0208	0.0223	3.245
SRNN ^[27]	6	0.0232	0.0302	6.750
BPNN ^[28]	6	0.0700	0.0835	>10
HYBRID			0.0544	
ARIMA+MLP ^[29]				

在总神经元数相同的情况下, cDBN 的逼近效果更好, 因为 cDBN 使用第二层隐含层对第一层的信息进行进一步模拟和提取, 获得了更好的内在数据特征, 并且运算时间更短. 实际上, 经验表明若有规律地增加 cDBN 神经元数, 可以获得比表 1 中所示更好的逼近效果, 因此, Lorenz 实验证明 cDBN 具有较好的学习能力和较快的运算速度.

3.2 CATS benchmark 缺失时序预测

为了检验 cDBN 的长期预测效果. CATS benchmark 又称人工时间序列计算, 在 IJCNN04 被首次提出并应用于不同人工智能模型学习效果对比^[17]. 该时间序列包含 5 个数据块, 每个数据块有 1000 个数据, 前 980 个为已知数据, 后 20 为丢失数据, 如图 3 所示.

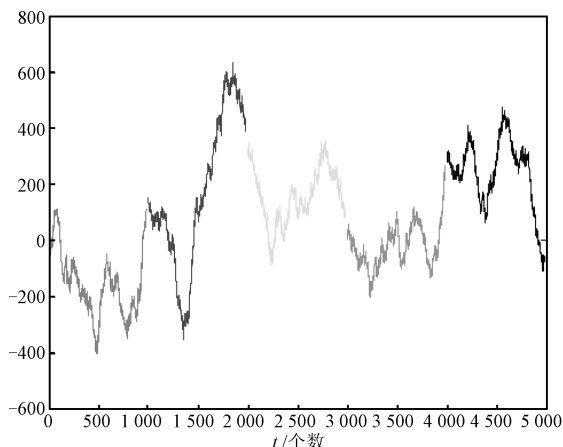


图 3 CATS benchmark 序列

Fig.3 CATS benchmark data

实验任务是预测丢失的数据共 100 个, 并通过 $E1$ 和 $E2$ 评价预测效果.

$$E1 = \frac{\sum_{981}^{1000} (x(t) - \hat{x}(t))^2}{100} + \frac{\sum_{1981}^{2000} (x(t) - \hat{x}(t))^2}{100} + \frac{\sum_{2981}^{3000} (x(t) - \hat{x}(t))^2}{100} + \frac{\sum_{3981}^{4000} (x(t) - \hat{x}(t))^2}{100} +$$

$$\frac{\sum_{4981}^{5000} (x(t) - \hat{x}(t))^2}{100} \quad (17)$$

$$E2 = \frac{\sum_{981}^{1000} (x(t) - \hat{x}(t))^2}{80} + \frac{\sum_{1981}^{2000} (x(t) - \hat{x}(t))^2}{80} + \frac{\sum_{2981}^{3000} (x(t) - \hat{x}(t))^2}{80} + \frac{\sum_{3981}^{4000} (x(t) - \hat{x}(t))^2}{80} \quad (18)$$

$E1$ 和 $E2$ 分别计算了 5 个数据块共 100 个数据的均方误差和前 4 个数据块共 80 个数据的均方误差.

因人工时间序列具有一定内部规律, 如果某一组丢失数据前后均有较多已知数据, 根据学习模型的不同, 利用前后已知数据对模型进行同时训练, 可以实现对丢失数据的较好预测. 而对于最后一组, 由于其之后没有更多已知数据, 因此, 预测前 4 组丢失数据难度要小于第 5 组. 因此, 任务 $E1$ 比任务 $E2$ 更难以实现较好的预测.

由于数据是缺失的, 实验中需要用 cDBN 进行长期预测, 网络使用 10 输入, 在测试阶段, 使用上一次预测值作为输入值 $x(t)$, 预测 $x(t+1)$. 网络结构采用 10-20-20-1, 将所有 4900 组数据输入网络进行训练, 然后分别对每块的缺失数据进行预测. 知识生成阶段迭代周期为 200, 调优训练速率为 2~3.5, 训练步数 5000~10000, 实验结果如图 4.

结果表明, 对每一个数据块, cDBN 均达到了较好的学习效果和预测效果. CATS 曲线在短期内的变化明显且迅速, 很多地方出现较陡坡度, 而 cDBN 学习曲线虽相对平滑, 但能达到较好的跟踪. 对于缺失时序预测阶段, 预测数据与真实数据之间差距逐渐增大, 且除了最后一块的预测方向与实际方向相反, 其他均预测到了相同的趋势, 真实值与预测值从基本吻合到最终相差 30~40, 其中最后一块预测值与真实值差距达到 50 左右.

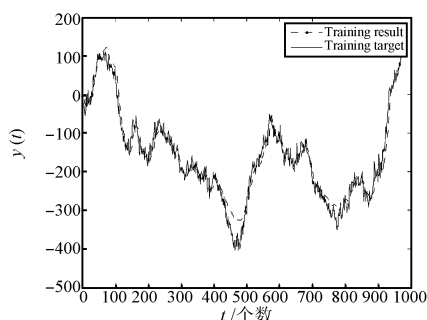
表 2 对比了多种模型对 CATS benchmark 的学习结果. 模型 Kalman smoother 和 Ensemble models 分别为 IJCNN04 中预测 $E1$ 和 $E2$ 的最优. 使用 cDBN 训练得到 $E1 = 771.455$, $E2 = 618.912$, 略低于前两种模型, 但优于 Hierarchical Bayesian learning 和 MLP.

由于 cDBN 尚缺少好的方法确定更优的网络结构和参数, 因此实验中只凭借经验选取, 仍存在改进空间, 实际上在实验过程中已经发现, 当适当增加第二隐含层单元数时, 将使 $E1$ 精确度得以提高, 当调优过程学习速率适当增加时, 前 4 个数据块的结果将得到优化从而使 $E2$ 精确度得以提高.

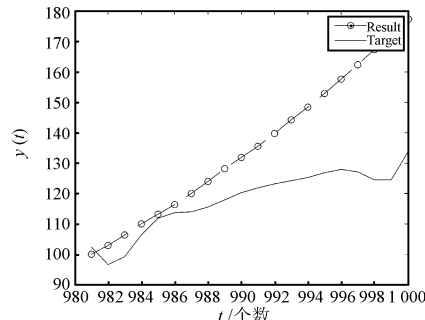
3.3 大气 CO₂ 变化趋势预测

cDBN 可以使用深度神经网络预训练的结果作为网络的初始权重, 再使用 Back propagation 算法作为调优手段. 这种方法在训练数据较为缺乏时很

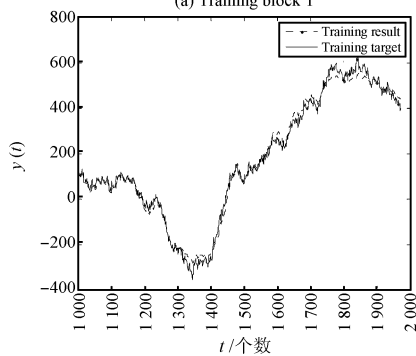
有价值, 因为初始化权重对最终模型的性能有显著影响, 而预训练获得的权重在权值空间中比随机权重更接近最优的权重. 为了验证 cDBN 在含有噪声的实际问题中实现较好的预测并稳定收敛于最优解. 环境问题一直是世界各国最关心的主题之一. CO₂



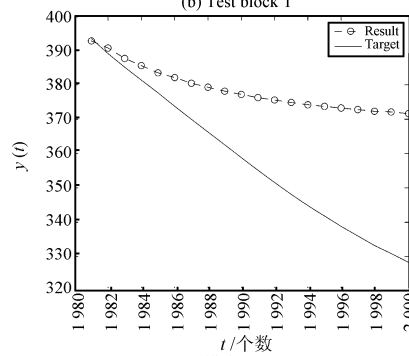
(a) 模块 1 训练图
(a) Training block 1



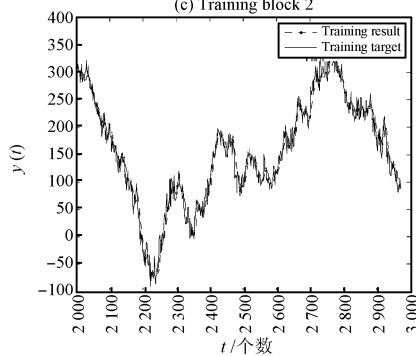
(b) 模块 1 测试图
(b) Test block 1



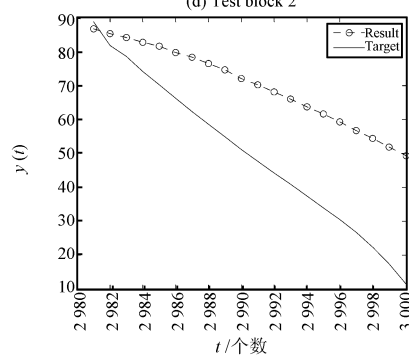
(c) 模块 2 训练图
(c) Training block 2



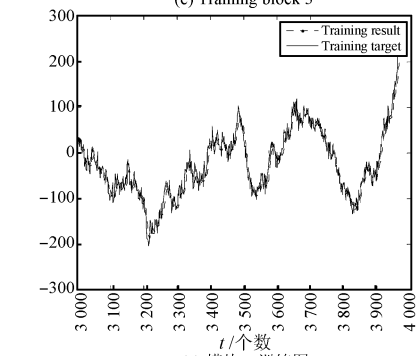
(d) 模块 2 测试图
(d) Test block 2



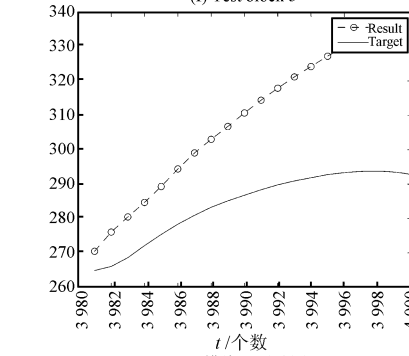
(e) 模块 3 训练图
(e) Training block 3



(f) 模块 3 测试图
(f) Test block 3



(g) 模块 4 训练图
(g) Training block 4



(h) 模块 4 测试图
(h) Test block 4

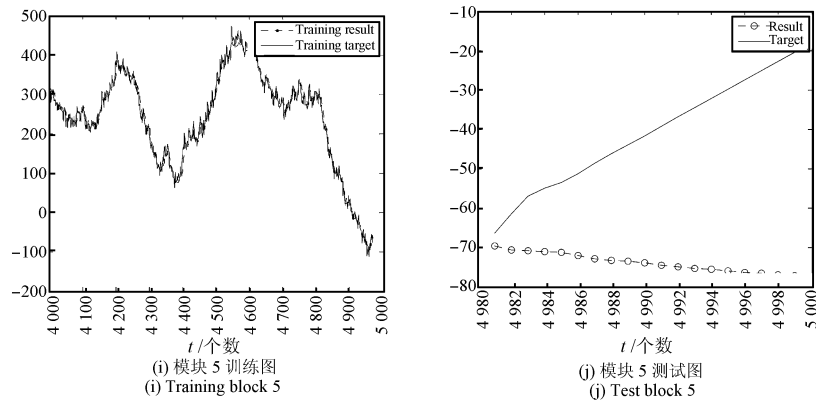


图4 CATS benchmark 缺失时序预测结果

Fig.4 Experimental results of using CATS benchmark

浓度变化由多种因素造成,例如地理环境因素、人类社会活动、海洋季风影响等。 CO_2 浓度变化会对人类社会造成影响,例如当 CO_2 浓度过高,出现温室效应,生物新陈代谢受到影响。因此预测某一地区 CO_2 浓度变化,除了具有较好的科学考察意义,也能够为制定地区和企业 CO_2 排放提供参考,具有一定社会研究价值。检测 CO_2 浓度时由于环境因素、气候因素和人为因素造成误差,短期内的 CO_2 浓度往往不能反映该地区的 CO_2 变化,因此应研究长期内 CO_2 浓度变化,应使用合适工具去除噪声影响。为了保证数据准确,预测全球气候变化的 CO_2 数据需要在不受城市排放和污染影响的地区采集,因此位于太平洋中心的夏威夷成为绝佳的科研考察地。实验使用夏威夷 Muana Loa 的 CO_2 浓度数据(1965 年~1980 年)共 190 组^[30],前 103 组数据为训练组共产生 100 组训练数据,后 87 组作为测试组。使用结构为 3-20-40-1 的 cDBN 对该地区 CO_2 浓度数据进行学习和一步预测,知识生成阶段迭代周期为 200,调优过程训练速率为 3,训练步数 10 000,得到结果如图 5。

图 5(a) 为训练效果,图 5(b) 是预测效果,5(c) 是训练过程中误差反传阶段的均方根误差变化曲线。图 5 中可以看出,该地区 CO_2 浓度在不同季节有规律的变化,并总体而言逐年上升。使用 cDBN 能够较好地去除噪声影响,实现对 CO_2 浓度变化的学习

和预测。通过误差变化曲线可知,训练过程中误差不断下降,最终找到最优点。

在本实验中为了防止 BP 网络出现过拟合,通过尝试,采用 BP 网络结构为 3-6-1,训练目标误差为 0.1,将 cDBN 和 BP 各运行 10 次,得到结果如表 3。BP 网络的测试 RMSE 分别为 3.7860, 2.8131, 1.3560, 5.8821, 6.6448, 1.7755, 1.2578, 2.2267, 5.9952, 3.3660, 测试 RMSE 均值为 3.4603, 方差为 4.3288。从结果可以看出, BP 网络测试误差大小不一,且测试均值和方差都较大,说明 BP 网络训练很容易落入局部极小点,相比之下 cDBN 均值和方差明显较小,且具有更好的稳定性能。而 cDBN 的测试 RMSE 基本相同,方差约等于 0,具有收敛性能稳定的特点,能够保证每次训练收敛于最优区域,具有较好的实际应用价值,见图 5(d),点代表十次 BP 实验得到的预测结果,线表示使用 cDBN 计算得到的结果。由于外来扰动对网络预测结果会有一定影响,而实验结果表明 cDBN 具有较好的降噪能力。

表 3 CO_2 预测结果

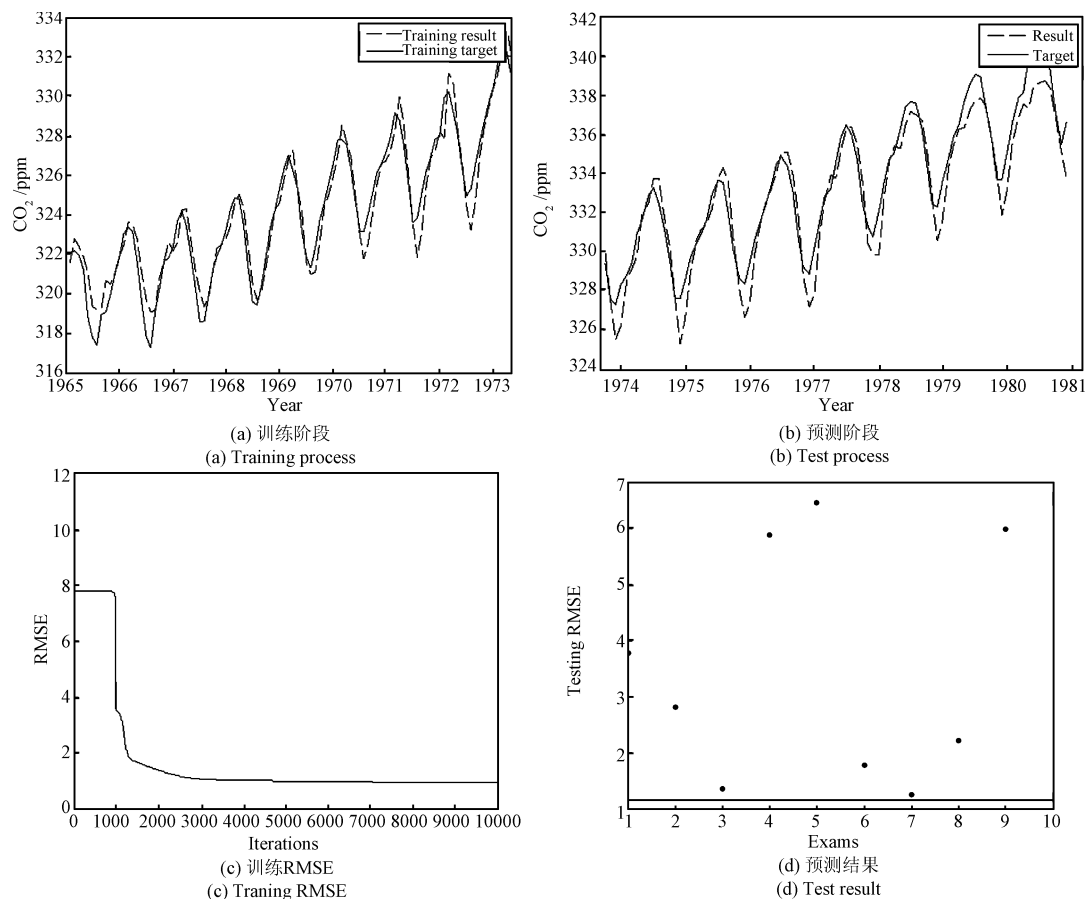
Table 3 Result data of CO_2 forecasting

训练 RMSE	测试 RMSE	均值	方差	方法
0.9133	1.1504	0.9133	<0.001	cDBN
>0.1	1.3~6.6	3.4603	4.3288	BP

表 2 CATS benchmark 模型结果对比

Table 2 Comparing different models in CATS benchmark

E1	Model	E2	Model
408	Kalman smoother ^[31]	222	Ensemble models ^[17]
771.455	cDBN	618.912	cDBN
1 247	Hierarchical Bayesian learning ^[30]	1 229	Hierarchical Bayesian learning
2 184	MLP ^[4]	2 337	MLP

图5 CO₂ 浓度趋势预测结果Fig.5 Experimental result of CO₂ forecasting

4 总结

cDBN 是将离散型深度信念网的输入层和输出层进行数据连续化的网络. 本文主要工作如下: 1) 提出双隐层结构的连续型 DBN, 其训练过程采用 CD 准则, 提出收敛性定理并证明, 通过收敛性分析给出了 cDBN 的传递函数中参数的选择依据; 2) 通过实验表明 cDBN 能够实现对连续数值的函数逼近和时序预测, 具有精度高、结构简、稳定性强的特点, 具有较好理论研究价值和实际应用价值. 目前, 深度信念网日益成为人工智能中重要的模式识别和辨识工具, 其在智能系统中的应用必将给世界带来深远影响, 而 cDBN 通过对 DBN 改进, 使其应用领域更加丰富, 较好地实现在时序预测中的应用, 未来可在一些工业问题中进行应用.

下一步工作将从以下两方面进行: 1) 理论方面, 改进 cDBN 算法, 获得更好的逼近效果, 研究网络自组织方法, 在训练中自我调整结构, 减少人为经验, 使网络结构针对某一问题达到最优状态; 2) 应用方面, 研究 cDBN 在其他领域中的研究如智能控制, 因为环境和工业中某些变量不易直接获得, 也可

以研究其软测量等方面的应用性能.

References

- 1 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504–507
- 2 Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527–1554
- 3 Deselaers T, Hasan S, Bender O, Ney H. A deep learning approach to machine transliteration. In: *Proceedings of the 4th EACL Workshop on Statistical Machine Translation*. Athens, Greece: Association for Computational Linguistics, 2009. 233–241
- 4 Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. *Nature*, 1986, **323**(6088): 533–536
- 5 Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. In: *Proceedings of the 20th Advances in Neural Information Processing Systems*. Vancouver, British Columbia, Canada: NIPS, 2007. 153–160
- 6 Arel I, Rose D C, Karnowski T P. Deep machine learning—a new frontier in Artificial intelligence research. *IEEE Computational Intelligence Magazine*, 2010, **5**(4): 13–18
- 7 Dahl G E, Dong Y, Deng L, Acero A. Large vocabulary continuous speech recognition with context-dependent DBN-HMMS. In: *Proceedings of the 2011 IEEE International*

- Conference on Acoustics, Speech and Signal Processing. Prague: IEEE, 2011. 4688–4691
- 8 Fasel I, Berry J. Deep belief networks for real-time extraction of tongue contours from ultrasound during speech. In: Proceedings of the 20th International Conference on Pattern Recognition. Istanbul: IEEE, 2010. 1493–1496
 - 9 Zhang S L, Bao Y B, Zhou P, Jiang H, Dai L R. Improving deep neural networks for LVCSR using dropout and shrinking structure. In: Proceedings of the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Florence: IEEE, 2014. 6849–6853
 - 10 Bengio Y. Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2009, **2**(1): 1–127
 - 11 Wei Q L, Wang F Y, Liu D R, Yang X. Finite-approximation-error-based discrete-time iterative adaptive dynamic programming. *IEEE Transactions on Cybernetics*, 2014, **44**(12): 2820–2833
 - 12 Liu D R, Wei Q L. Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(3): 621–34
 - 13 Zhang H G, Qin C B, Jiang B, Luo Y H. Online adaptive policy learning algorithm for H_∞ state feedback control of unknown affine nonlinear discrete-time systems. *IEEE Transactions on Cybernetics*, 2014, **44**(12): 2706–2718
 - 14 Schmidhuber J. Deep learning in neural networks: an overview. *Neural Network*, 2015, **61**: 85–117
 - 15 Längkvist M, Karlsson L, Loutfi A. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters*, 2014, **42**: 11–24
 - 16 Taylor G M, Hinton G E, Roweis S. Modeling human motion using binary latent variables. In: Proceedings of the 20th Advances in Neural Information Processing Systems. Vancouver, British Columbia, Canada: NIPS, 2007. 1345–1352
 - 17 Chen H, Murray A F. A continuous restricted Boltzmann machine with an implementable training algorithm. *IEEE Proceedings Vision, Image and Signal Processing*, 2003, **150**(3): 153–158
 - 18 Zhang R, Shen F R, Zhao J X. A model with fuzzy granulation and deep belief networks for exchange rate forecasting. In: Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN). Beijing, China: IEEE, 2014. 366–373
 - 19 Chen J F, Jin Q J, Chao J. Design of deep belief networks for short-term prediction of drought index using data in the Huaihe river basin. *Mathematical Problems in Engineering*, 2012, **2012**: 235929
 - 20 Song R Z, Xiao W D, Zhang H G, Sun C Y. Adaptive dynamic programming for a class of complex-valued nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **25**(9): 1733–1739
 - 21 Song R Z, Lewis F, Wei Q L, Zhang H G, Jiang Z P, Levine D. Multiple actor-critic structures for continuous-time optimal control using input-output data. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(4): 851–865
 - 22 Hinton G E. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 2002, **14**(8): 1771–1800
 - 23 Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 2012, **13**(1): 281–305
 - 24 Wu Zhi-Wei, Chai Tian-You, Wu Yong-Jian. A hybrid prediction model of energy consumption per ton for fused magnesia. *Acta Automatica Sinica*, 2013, **39**(12): 2002–2011 (吴志伟, 柴天佑, 吴永建. 电熔镁砂产品单吨能耗混合预报模型. 自动化学报, 2013, **39**(12): 2002–2011)
 - 25 Patan K. Stability analysis and the stabilization of a class of discrete-time dynamic neural networks. *IEEE Transactions on Neural Networks*, 2007, **18**(3): 660–673
 - 26 Crone S F, Nikolopoulos K. Results of the NN3 neural network forecasting competition. In: Proceedings of the 27th International Symposium on Forecasting Program. New York, USA: ISF, 2007. 1–129
 - 27 Chen Q L, Chai W, Qiao J F. A stable online self-constructing recurrent neural network. *Advances in Neural Networks — ISNN 2011*. Berlin Heidelberg: Springer, 2011, **6677**: 122–131
 - 28 Chang L C, Chen P A, Chang F J. Reinforced two-step-ahead weight adjustment technique for online training of recurrent neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, **23**(8): 1269–1278
 - 29 Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 2003, **50**: 159–175
 - 30 <http://CO2now.org>, September 23, 2015
 - 31 Lmaury A, Oja E, Simula O, Verleysen M. Time series prediction competition: the CATS benchmark. In: Proceedings of IJCNN'2004 — International Joint Conference on Neural Networks. Budapest: IJCNN, 2004. 1615–1620



乔俊飞 北京工业大学教授。主要研究方向为智能控制, 神经网络分析与设计。
E-mail: junfeiq@bjut.edu.cn
(QIAO Jun-Fei Ph.D., professor at Beijing University of Technology. His research interest covers intelligent control, analysis and design of neural networks.)



潘广源 北京工业大学博士研究生。主要研究方向为智能信息处理, 深度学习, 神经网络结构设计和优化。本文通信作者。E-mail: garrypan@outlook.com
(PAN Guang-Yuan Ph.D. candidate in Beijing University of Technology. His research interest covers intelligence information processing, deep learning, analysis and design of neural networks. Corresponding author of this paper.)



韩红桂 北京工业大学教授, 主要研究方向为污水处理工艺复杂建模与控制, 神经网络分析与设计。
E-mail: rechardhan@sina.com
(HAN Hong-Gui Ph.D., professor at Beijing University of Technology. His research interest covers modeling and control in water and waste water process, analysis and design of neural networks.)