

一种新型随机二进制神经网络¹⁾

赵杰煜¹ 张泉方² 汪加才¹

¹(宁波大学信息科学与技术研究所 宁波 315211)

²(浙江大学计算机系 杭州 310027)

(E-mail: zhao-jieyu@sina.com)

摘要 二进制数据表示具有简洁高效的特点,随机噪声有助于系统摆脱局部极小.新型的随机神经网络模型采用随机加权联接,内部数据表示为随机二进制序列形式,实现十分高效.文中分别就前馈型网络和反馈型网络进行了深入的讨论,给出了前馈型网络的梯度下降学习算法,为反馈型网络设计了快速有效的模拟退火算法和渐进式 Boltzmann 学习算法.通过对 PARITY 问题的测试,发现了新模型的一些有趣特征,实验结果表明梯度下降学习效果显著.利用渐进式 Boltzmann 学习,反馈型网络被成功地用于带噪声人脸识别.

关键词 随机计算,随机二元神经网络,平稳分布,模拟退火,渐进式 Boltzmann 学习

中图分类号 TP183,TP391.41

A NOVEL STOCHASTIC BINARY NEURAL NETWORK

ZHAO Jie-Yu¹ ZHANG Quan-Fang² WANG Jia-Cai¹

¹(The IST Research Institute, Ningbo University, Ningbo 315211)

²(Department of Computer Science, Zhejiang University, Hangzhou 310027)

(E-mail: zhao-jieyu@sina.com)

Abstract The highly efficient binary representation scheme has dominated the world of computation for a long time. However, the deterministic binary representation seems to be challenged in stochastic neural computation which makes use of the random noise to escape from local minima. This paper presents a novel stochastic binary neural network which uses stochastic weights and the 'stochastic binary sequence' data representation. It is very attractive with its great potential to be implemented efficiently. Both the feedforward and the recurrent stochastic binary networks have been discussed in depth. The gradient descent learning techniques are described for the feedforward network. A novel simulated annealing method and an incremental Boltzmann learning algorithm have been proposed. Simulation results on the PARITY problem and a face recognition task show the excellent performance of the model.

Key words Stochastic computing, stochastic binary neural network, stationary distribution, simulated annealing, incremental Boltzmann learning

1) 国家自然科学基金(69805002)、教育部优秀青年教师资助项目和浙江省自然科学基金青年人才培养专项资助

收稿日期 2000-09-22 收修改稿日期 2001-11-21

1 引言

随机噪声在计算过程中通常被认为是非常有害的,然而在机器学习、模式识别、智能控制等众多领域,往往要求计算系统具有强容错性、能克服局部极小、造价低廉等特性,每个计算单元也许十分简单而且可靠性也不高,但通过适当的冗余设计和并行连接,使整个系统达到非常高的可靠性和计算精度. von Neumann 在一篇重要的论文中阐述了如何运用不可靠计算单元去构造一个可靠的系统^[1]. Rosenblatt 敏锐地洞察到神经系统中存在的噪声和随机性绝非是由于人脑结构的不合理或某种缺陷带来的不良后果,而恰恰是人脑所进行的计算所必备的^[2].

随机神经网络正是通过大规模地并行连接一些简单而且不太可靠的神经元来达到计算目的,通过采用随机离散序列表示模拟量^[3,4],使系统有能力利用随机噪声摆脱局部极小点,搜索整个系统状态空间,并收敛于全局极小,因此解决全局优化问题. 在随机神经网络内部,信息的表达往往是随机的、带噪声的,单个神经元的输入无需达到很高的精度,整个网络的输出由系统的全局状态决定而不是取决于单个神经元. 这种完全分布式的信息表达在噪声环境中具有极强的鲁棒性、抗干扰能力和很好的泛化特性,少量神经元或联接的损坏不会引起大的不良后果. 一个典型的随机神经网络模型是 Boltzmann 机^[5],与 Hopfield 网络相比, Boltzmann 机中的隐单元使系统有能力处理输入样本的高价相关性,从而提高网络的计算能力,然而,由于其学习算法和模拟退火过程十分耗时,阻碍了 Boltzmann 机的广泛应用.

本文提出一种新的随机神经网络模型——随机二元神经网络,利用概率型时间域信息表示法,将模拟量转换成随机二进制序列,使系统具有很高的实现效率. 类似的时间域信息表示法如“脉冲流表示法”^[6]已在传统的 Sigmoid 型神经网络中获得了应用,近来“spiking neuron”^[7~9]研究也利用类似的技术. 本文将要深入讨论的并不是将时间域的数据表示法用于传统 Sigmoid 型神经元,而是设计和实现一种全新的网络模型. 通过对不同结构网络的深入分析,提出梯度下降学习方法、模拟退火法和渐进式 Boltzmann 学习法,并实现模型的仿真和学习算法,将其用于解决实际问题.

随机神经网络是一个并行分布式信息处理系统,其结构呈有向图形式,可分为前馈型和反馈型二种. 前馈型网络是单向多层结构,计算比较直观,从输入层开始,逐层向前计算直至输出. 反馈型网络则是一个结点互连的动力学系统,需要一个稳定过程以达到“平衡态”,输出往往是在稳定以后给出.

随机二元神经网络^[10]是一种具有随机加权联接的二进制神经网络,它是一个符合下列条件的有向图:

- 1) 有向图的结点为随机二进制神经元,神经元输出为双态;
- 2) 有向图的弧为随机联接,其上定义的权值可调,权值大小决定联接强度,即信息的通过率;
- 3) 所有权值、输入和输出均采用随机二进制序列表示.

前馈型和反馈型随机二元神经网络除了结构迥异外,在系统运行方式、计算能力、学习算法等方面亦具有明显的区别. 前馈型网络适合于带噪声的模式识别、非线性分类等;而反馈型网络更接近于真正的生物神经网络,其计算能力更为强大,不仅可用于模式识别和分

类,还可运用模拟退火进行组合优化,实现联想记忆,进行吸引子计算等.下面首先讨论前馈型随机二元神经网络.

2 前馈型随机二元神经网络

前馈型随机二元神经网络中采用的神经元是一个十分简单的随机二进制序列处理单元,它将所有输入与相应的加权相乘,然后相加,如果总和大于神经元阈值则输出 1,否则输出 0.以下是用于前馈型网络的随机二进制神经元 AND-SBN 的确切定义.

定义 1. 一个带 n 输入的 AND 型随机二进制神经元(AND-SBN)具有 n 个加权 $w_i \in [-1, 1]$ 和 n 个输入 $v_i \in [0, 1]$ ($i=1, 2, \dots, n$),所有 w_i 和 v_i 均采用单相随机型 Bernoulli 序列表示,每个加权 Bernoulli 序列还带有一个附加符号位.神经元的阈值 θ 是一个取值于 $[-n, n]$ 之间的整型随机变量,其值根据概率密度函数 $\phi(\theta)$ 动态产生.每个时钟周期神经元进行如下操作:

1) n 个输入 Bernoulli 序列的当前位分别与 n 个加权 Bernoulli 序列的当前位相“与(AND)”(即乘法运算);

2) 将相与输出位与对应的加权符号相结合,然后把 n 个带符号的加权输入位相加,相加结果与当前动态产生的阈值相比较,如果总和大于阈值则输出 1,否则输出 0.

设 $P_0(\mathbf{w}, \mathbf{v}, \phi(\theta))$ 是 AND-SBN 的激活函数,其中 $\mathbf{w} = (w_1, \dots, w_n)$ 为神经元的权向量, $\mathbf{v} = (v_1, \dots, v_n)$ 为相应输入向量, $\phi(\theta)$ 为阈值概率密度函数,激活函数的值即神经元输出 1 的概率.采用 MGF (Moment Generating Function) 描述后,激活函数 $P_0(\mathbf{w}, \mathbf{v}, \phi(\theta))$ 可以表示如下^[11]:

$$P_0(\mathbf{w}, \mathbf{v}, \phi(\theta)) = \sum_{\theta=-n}^n \left(\phi(\theta) \sum_{i=\theta}^n g_i(\mathbf{w}, \mathbf{v}) \right),$$

其中 $g_i(\mathbf{w}, \mathbf{v})$ 是总加权输入的 MGF 展开项对应于各 e^{θ} 的系数.

可见,AND-SBN 激活函数与传统 BP 网络中采用的 Sigmoid 型激活函数十分不同,此激活函数并不是神经元加权输入总和的一个简单函数.

这给设计梯度下降学习算法带来了困难,实现梯度下降学习的关键是计算激活函数相对于其每一个输入权的偏导数.为了适当简化学习算法,假定所有 AND-SBN 的阈值均为确定型,即 θ 是确定型变量.为了方便起见,我们将 AND-SBN 激活函数重写为 $P_0(p_1, p_2, \dots, p_n, \theta)$, 其中 p_i 为第 i 个加权输入值 $w_i v_i$, θ 是神经元的阈值,所有权值在 $[-1, 1]$ 之间,输入值在 $[0, 1]$ 之间, $|p_i|$ 为加权输入二进制序列 X_i 取值 1 的概率.经过推导后^[11],可得

$$\frac{\partial P_0(p_1, p_2, \dots, p_n, \theta)}{\partial p_j} = \begin{cases} \text{Prob} \left(\sum_{i=1, i \neq j}^n X_i = \theta - 1 \right), & \text{如果 } p_j > 0, \\ \text{Prob} \left(\sum_{i=1, i \neq j}^n X_i = \theta \right), & \text{如果 } p_j < 0. \end{cases}$$

所以,对于 AND-SBN 神经元,如果第 j 个加权输入为正,神经元的输出相对于该加权输入的偏导数为除去第 j 个加权输入外的 $(n-1)$ 个加权输入的总和等于 $\theta-1$ 的概率;如果第 j 个加权输入为负,神经元的输出相对于该加权输入的偏导数为除去第 j 个加权输入外

的 $(n-1)$ 个加权输入的总和等于 θ 的概率.

现在讨论梯度下降学习,根据广义 delta 学习规则,每一个权值的改变与它对总误差 E 的贡献率成反比: $\Delta w_{ji} = -\alpha \partial E / \partial w_{ji}$,式中 α 为学习速率,它决定学习步伐的大小.根据链式公式,我们可以得到

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial O_j} \frac{\partial O_j}{\partial p_{ji}} \frac{\partial p_{ji}}{\partial w_{ji}},$$

其中 O_j 是第 j 个神经元的输出, $p_{ji} = w_{ji} O_i$ 是来自第 i 个神经元的加权输入.

令 $\delta_j = -\partial E / \partial O_j$,则

$$\Delta w_{ji} = \alpha \delta_j \frac{\partial O_j}{\partial p_{ji}} \frac{\partial p_{ji}}{\partial w_{ji}} = \alpha \delta_j O_i \frac{\partial O_j}{\partial p_{ji}},$$

偏导数 $\partial O_j / \partial p_{ji}$ 可以按照上述方法计算. δ_j 是用于反向传递^[12]的误差值.如果神经元 j 是网络的输出单元,假定 E 定义为平方误差 $E = (T_j - O_j)^2 / 2$,其中 T_j 是对应的目标值,那么

$$\delta_j = -\frac{\partial E}{\partial O_j} = T_j - O_j;$$

如果神经元 j 是网络的隐单元,则

$$\delta_j = -\sum_k \frac{\partial E}{\partial O_k} \frac{\partial O_k}{\partial p_{kj}} w_{kj} = \sum_k \delta_k w_{kj} \frac{\partial O_k}{\partial p_{kj}},$$

其中 k 是指所有直接连至神经元 j 的单元,偏导数 $\partial O_k / \partial p_{kj}$ 同样可以按照上面介绍的方法计算.

对于大型的随机二元神经网络,随着神经元输入数的增加,神经元激活函数的计算就变得十分繁琐,为了有效地计算多输入 SBN 的激活函数,可以采用近似算法.因为 SBN 的总加权输入是单个独立随机加权输入的和,根据大数定律,当输入较多时总加权输入和将近似服从高斯分布.因此,可以用积分型高斯函数来近似替代 SBN 激活函数

$$P_0(p_1, p_2, \dots, p_n, \theta) \approx P'_0(p_1, p_2, \dots, p_n, \theta) = \int_{\theta}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx,$$

其中期望值 $\mu = \sum_{i=1}^n p_i$, 均方差 $\sigma = \sqrt{\sum_{i=1}^n (|p_i| - p_i^2)}$.

采用上述近似后,同样可以求得梯度下降学习中所需的偏导数

$$\frac{\partial P'_0}{\partial p_j} = \frac{\partial P'_0}{\partial \sigma} \frac{\partial \sigma}{\partial p_j} + \frac{\partial P'_0}{\partial \mu} \frac{\partial \mu}{\partial p_j} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}} \left(\frac{\theta-\mu}{2\sigma^2} \times \frac{\partial |p_j|}{\partial p_j} - \frac{\theta-\mu}{\sigma^2} \times p_j + 1 \right).$$

由此可见,大型随机二元神经网络的梯度计算并不复杂.在大型随机二元神经网络中运用高斯逼近及相应的导数计算,可以大大提高学习速度,使大型随机二元神经网络的训练不仅可行而且非常高效.实验表明,通常只要随机二元神经网络中的神经元个数不小于 15,采用高斯逼近学习后所得的结果可以直接应用于随机二元神经网络,只有极个别的样本需要重新学习以达到所要求的精度,这种重新学习往往只需很少的学习周期.

3 反馈型随机二元神经网络

反馈型随机二元神经网络由随机二进制神经元任意互连而成,其典型结构是一个带有

环的有向图. 为了使系统具备对称性, 在神经网络实现时采用了双相随机 Bernoulli 序列表示的 XNOR 型随机二进制神经元^[10], XNOR-SBN 与 AND-SBN 十分相近, 均是将实数表示为随机二进制序列形式以简化乘法运算, 主要区别是 XNOR-SBN 神经元状态取值 1 或 -1. 由于本文不讨论具体实现, 为了方便起见, 直接给出反馈型随机二元神经网络的定义.

定义 2. 反馈型随机二元神经网络(RSBN)是一个具备以下特征的有向图 $N(V, W, F)$, 该图具有大小为 n 的顶点集 V , 状态集 $S \in \{-1, +1\}^n$ 和加权联接 W , W 是一个大小为 $n \times n$ 的随机矩阵, 其元素 W_{ij} 是一个具有以下 pdf 的随机变量:

$$f_{ij}(x) = \begin{cases} (1 + \beta w_{ij})/2, & x = 1, \\ (1 - \beta w_{ij})/2, & x = -1, \\ 0, & \text{其它,} \end{cases}$$

其中 $-1 \leq w_{ij} \leq +1$ 是随机联接 W_{ij} 的加权, $0 \leq \beta \leq 1$ 是一个控制参数. F 是神经网络的动态更新机制, 它随机地选择顶点并按下列规则更新其状态

$$S_i = \text{sgn}\left(\sum_{j \neq i} W_{ij} S_j\right), \quad \text{if } \sum_{j \neq i} W_{ij} S_j \neq 0.$$

假如加权和恰好为 0, 则 S_i 以 1/2 的等概率随机地取 1 或 -1.

如果顶点是串行地随机选择并更新状态的, 称之为“串行 RSBN”; 如果多个顶点被同时选中并更新其状态的, 则称之为“并行 RSBN”. 本文主要讨论串行 RSBN, 而且假定网络是对称的, 即所有 $W_{ij} = W_{ji}$.

根据统计物理和热力学定律, 当一个物理系统达到热平衡时, 系统的某一特定状态发生的概率服从 Boltzmann-Gibbs 分布. 现在来分析一下串行 RSBN 达到平稳态时的概率分布. 首先给出 RSBN 能量函数的定义.

定义 3. 反馈型随机二进制网络的能量函数 $H = -\frac{1}{2} \beta \sum_{ij} w_{ij} S_i S_j$, 其中 β 是控制参数, w_{ij} 是神经元 i 和神经元 j 之间随机联接的加权, S_i 和 S_j 分别是神经元 i 和神经元 j 的状态.

定理 1. 当控制参数 $\beta \in [0, 1]$ 时, 串行反馈型随机二进制网络 RSBN 存在唯一的平稳概率分布. 如果串行 RSBN 是对称的, 并且神经元个数较多时, RSBN 的平稳概率分布近似为 Boltzmann-Gibbs 分布.

定理的详细证明参见文献[10].

利用上述结论, 可以通过控制变量 β 来有效地实现模拟退火. 在反馈型随机二元网络中, 系统的随机程度由均方差 σ 决定, σ 的作用与传统模拟退火中的温度控制变量相同. 均方差 σ 越大, 网络的随机程度越大; 随着 σ 的逐渐下降, 网络的随机程度也逐渐下降, 当 σ 下降到 0 时, 随机型网络就退化为确定型网络. 均方差 σ 与控制参数 β 具有关系^[10]

$$\sigma = \sqrt{\sum_{l=1, l \neq k}^n (1 - \beta^2 w_{kl}^2)}.$$

开始进行 RSBN 模拟退火时, 首先设定 $\beta = 0$, 此时均方差 σ 达到其最大值, 网络变成一个完全随机系统. 然后, 按事先确定的冷却程序逐步将 β 从 0 增加到 1. 在此过程中, 网络的随机性渐渐下降, 能量逐渐降到其最低点.

利用模拟退火可以有效地解决组合优化问题, 只要将所要解决的问题的代价函数映射到网络的能量函数上, 模拟退火过程使能量降到最小, 相应的代价就降到最低, 由此得到组

合优化问题的最优解.

模拟退火是反馈型随机神经网络随着系统内部随机性的降低逐渐趋向全局稳定点的过程,是典型的吸引子计算,整个退火过程中网络的权值并不改变,显而易见,为了更好地发挥反馈型网络强大的计算优势,引入学习能力是十分必要的.

反馈型随机神经网络的学习就是通过调节网络的加权,使输入输出单元的状态服从特定的概率分布,也就是使网络内部的状态概率分布与外部环境的状态概率分布相吻合.当一个随机神经网络的状态服从 Boltzmann-Gibbs 分布时,可以采用 Boltzmann 学习规则:

$$\Delta w_{ij} = \epsilon(\langle \overline{S_i S_j} \rangle_c - \langle S_i S_j \rangle_f)$$

式中 $\langle \overline{S_i S_j} \rangle_c$ 是输入输出单元钳定时的 Hebb 学习项, $\langle S_i S_j \rangle_f$ 是系统自由运行时的 Hebb 学习项(加上负号也称反 Hebb 学习项), ϵ 是学习速率.

上述学习需要在平衡态进行,这就要求进行模拟退火过程,所以学习过程相当耗时.对于反馈型随机二元神经网络,采用一种改进的“渐进式 Boltzmann 学习法”,学习过程简要如下.

1) 当有学习样本出现时,系统处于钳定态,即所有输入神经元的状态钳定于输入样本,输出神经元钳定于输出样本;当没有学习样本时,系统处于自由态,神经元的状态可以自由更新;当所有输入神经元的状态钳定于输入样本,输出神经元状态可以自由更新时,系统处于测试态.

2) 系统无论是处于钳定态、自由态还是测试态,都进行 RSBN 模拟退火:先将控制参数 β 设为 0,此时网络变成一个完全随机系统,然后按事先确定的冷却程序逐步将 β 从 0 增加到 1.在此过程中,网络渐渐达到其平衡态,能量降到最低点.

3) 在处于钳定态时,如果两个神经元的状态相同,则将对应用于此两神经元的随机联接 W_{ij} 的加权加上一个小常数 ϵ ;如果两个神经元的状态不同,则将对应的随机联接 W_{ij} 的加权减去 ϵ .

4) 在处于自由态时,如果两个神经元的状态相同,则将对应用于此两神经元的随机联接 W_{ij} 的加权减去一个小常数 ϵ ;如果两个神经元的状态不同,则将对应的随机联接 W_{ij} 的加权加上 ϵ .

5) 在处于测试态时,如果网络输出错误,则执行自由态时网络加权变化规则.

6) 将系统轮流设置于钳定态、自由态和测试态,重复 1)~5) 直至输出单元的状态服从所要求的概率分布为止.

上述学习算法的工作原理可以通过状态空间的能量分布来解释.在处于钳定态时,单步学习使对应于学习样本的网络状态的能量下降,而此状态出现的概率由 Boltzmann-Gibbs 分布决定,即与能量负值的指数成正比,所以单步学习后该状态出现概率会增加;自由态和测试态时,单步学习是一个去耦合过程,它使网络各状态间的耦合降低.由于随机联接的加权值被限定在 $[-1, 1]$,信息在 RSBN 中的存储是完全并行和分布式的,这大大增强了系统的容错性和学习能力,渐进式 Boltzmann 学习又是区域性的,由此,RSBN 克服了 Boltzmann 机中常见的“过度学习”问题,“过度学习”就是随着学习的进行,网络权值变得过大而使学习过程难以继续.

4 实验测试

4.1 PARITY 问题

首先,用前馈型随机二元神经网络去解决小型 PARITY 问题,在训练过程中发现小型前馈型随机二元神经网络并不擅长于解决这些问题,与 BP 网络相比,随机二元神经网络需要更长的学习时间,主要原因是随机二进制神经元的离散型阈值造成学习过程不够稳定,对于小型随机二元神经网络,阈值的微小变化就能对网络的输出产生很大影响,从而使学习过程出现大幅振荡.出人意料的是随机二元神经网络获得的结果揭示了一些非常有趣的现象.对于那些具有逻辑目标函数的问题,随机二元神经网络中的权值似乎总是收敛于 $-1, 1$ 或 0 ,在学习过程中,所有处于 $(0, 1)$ 或 $(-1, 0)$ 的权值都自动地受到不同程度的惩罚.图 1 是前馈型随机二元神经网络对 4-PARITY 问题求得的解,可见所有权值均收敛于 -1 或 1 ,输入与输出间的逻辑关系一目了然.

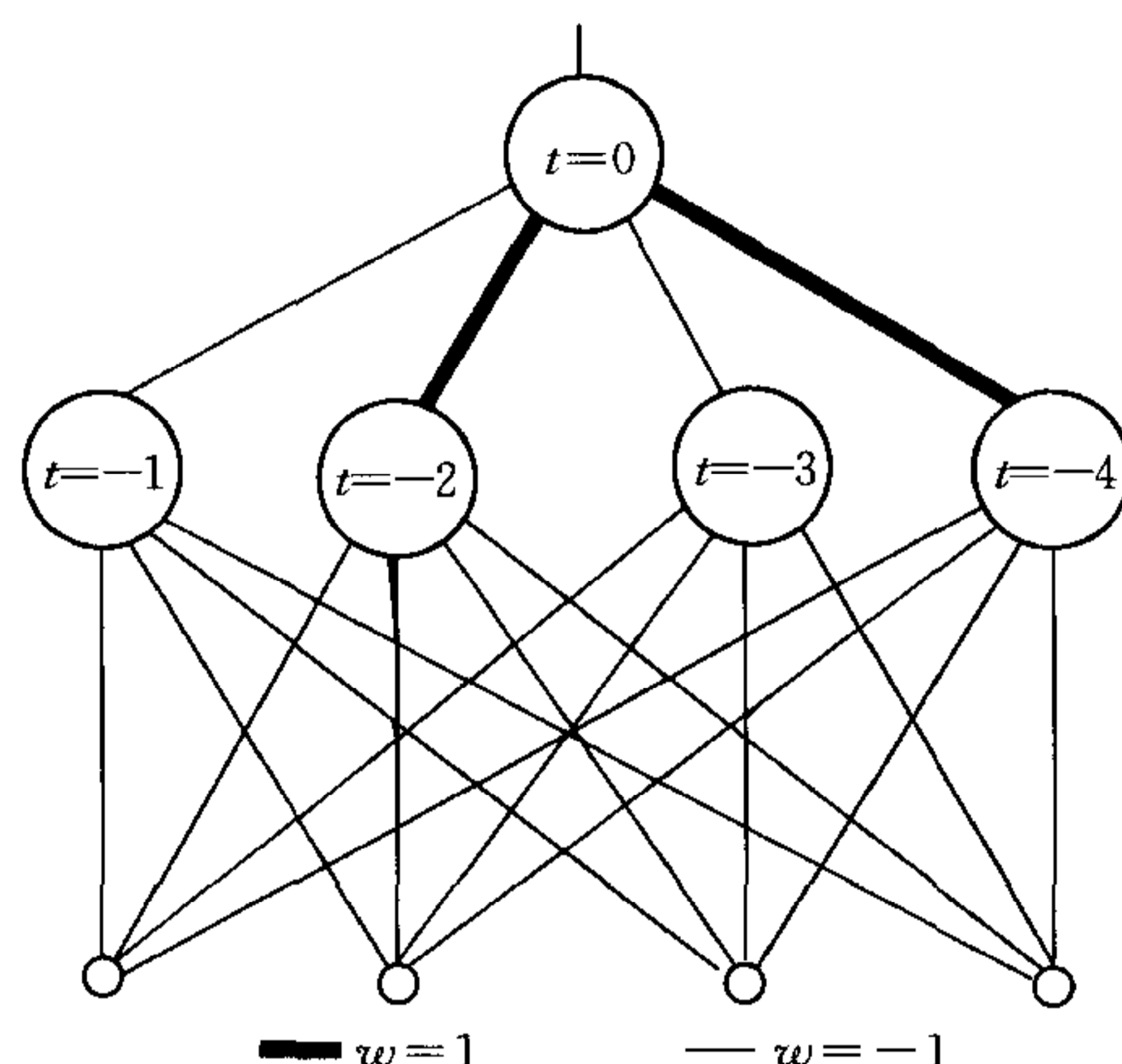


图 1 小型前馈随机二元神经网络对 4-PARITY 问题的解

然后,又分别对大小为 $7\sim 9$ 的 PARITY 问题作了测试,测试集是整个样本集合(样本数分别是 $128, 256, 512$),训练集是整个样本集合中任意抽取的 70% ,最后结果如表 1.

一般来说,小型随机二元神经网络的功能比不上相同大小的采用 Sigmoid 神经元的网络,但随着网络中神经元个数的增加,二者在能力上会越来越接近.当然,对中等规模或者大型的随机二元神经网络更感兴趣.

表 1 测试结果

	7-PARITY	8-PARITY	9-PARITY
训练正确率(%)	100	98.9	97.2
测试正确率(%)	97.6	96.9	96.5

4.2 带噪声人脸识别

反馈型随机二元神经网络具有广阔的应用前景,限于篇幅,在此只讨论其在模式识别方面的应用,有关采用模拟退火进行组合优化问题求解的应用参阅文献[10].为了有效地检验

RSBN 的学习性能和模式识别能力,选择了难度较大的人脸识别作为测试平台.虽然前馈型随机二元网络也可以胜任静态人脸的识别,但复杂的人物表情是动态变化的结果,只有反馈型网络才具备一定的动态记忆能力,尤其是以后进一步的动态人脸识别,RSBN 显然更为合适.传统的 Boltzmann 机由于其学习算法过于费时,无法处理人脸识别这类模式识别问题.

人脸图片取自于 MIT Media Lab 的 FERET 图片库,照片的大小为 128×192 像素、256 灰度,人脸图片库中共有 200 幅照片,其中的 100 幅作为学习样本,另外 100 幅作为测试样本,测试样本是学习样本中人物的某些变化,如微笑、眨眼、位移、旋转、发型变化、光线变化、带眼镜等,见图 2.所有测试样本被人为地加入了 10% 的随机噪声.

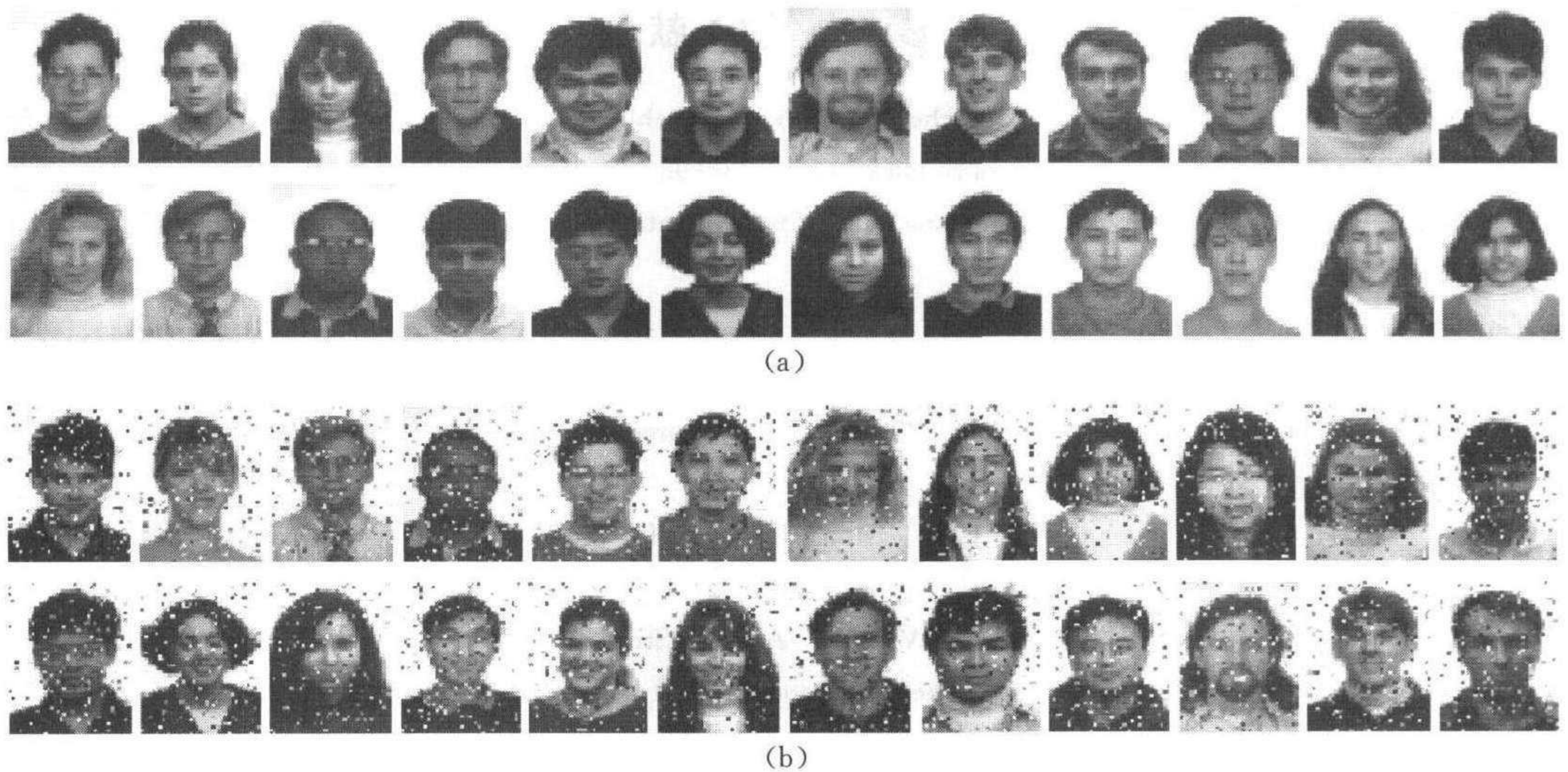


图 2 用于人脸识别的一些典型学习样本(a)和带噪声的测试样本(b)

用于人脸识别的 RSBN 网络是全相联反馈型结构,共包含 3910 个神经元,其中的 $72 \times 54 = 3888$ 个神经元作为输入神经元,7 个作为输出神经元,15 个隐单元.因为在学习过程中所有输入单元总是被钳定于特定的输入样本,输入单元之间的联接不起作用,所以事先就去掉了所有输入单元之间的联接以增加学习速度.采用的学习算法是前面介绍的“渐进式 Boltzmann 学习法”,模拟退火过程由参数 β 控制.RSBN 的每个学习周期包括输入样本钳定学习,自由运行和测试三种状态.

人脸图片样本在输入神经网络之前都经过预处理,具体包括照片尺寸的规范化和灰度的规范化,尺寸的规范化是将原照片按比例变换到 72×54 像素,灰度的规范化是将照片中最暗的点变换到灰度 0,最亮的点变换到灰度 255,中间灰度的点按比例变换到 $0 \sim 255$ 之间.在将图片输入到 RSBN 网络时,再把灰度 $0 \sim 255$ 的点线性地映射至 $[-1, 1]$ 区间内.

经过 637 个学习周期后,RSBN 顺利地学完了全部 100 个学习样本,对 100 个测试样本的识别率达 90%.

5 讨论

传统的 Sigmoid 型神经网络中权值的取值范围一般不受限制,因此在学习过程中,有

时会发生权值增长过快的情况,最终导致学习无法继续.如果简单地对权值范围进行限定,则会导致神经元输出无法接近 0 或 1,从而影响系统的正常运行.权衰减(Weight Decay)是避免权值过大的有效技术,但其代价是占用系统时间,减缓学习过程.

随机二元神经网络的所有权值均被限定在-1 和 1 之间,这种权值范围限制对系统的正常运行没有影响,因为所有单元的输出可以是 $[0,1]$ 之间的任意值.权值范围限制的一个重要结果是导致网络中知识的完全分布式表达,因此系统的容错性和鲁棒性大大提高,同时在一定程度上改善系统泛化性能.当然在一些需要集中式知识表达的特定场合,随机二元神经网络可能不太适合.

参 考 文 献

- 1 von Neumann J. Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components. In: Shannon C E, McCarthy J ed. Automata Studies, 1956. 43~98
- 2 Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958, **65**(6):386~408
- 3 Gaines B R. Stochastic Computing System. In: Tou J T ed. Advances in Information Systems Science. New York: Plenum, 1969. 37~172
- 4 Poppelbaum W J, Dollas A, Glickman J B, O'Toole C. Unary processing. *Advances in Computers*, 1987, **26**(1):47~92
- 5 Ackley D H, Hinton G E, Sejnowski T J. A learning algorithm for Boltzmann machines. *Cognitive Science*, 1985, **9**(1):147~169
- 6 Murray A, Tarassenko L. Analogue Neural VLSI—A Pulse Stream Approach. London: Chapman & Hall, 1994
- 7 Maass W. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 1997, **10**(9):1659~1671
- 8 Banerjee A. On the phase-space dynamics of systems of spiking neurons. I: model and experiments. *Neural Computation*, 2001, **13**(1):161~193
- 9 Banerjee A. On the phase-space dynamics of systems of spiking neurons. II: formal analysis. *Neural Computation*, 2001, **13**(1):195~225
- 10 Zhao J. A recurrent stochastic binary network. *Science in China(F)*, 2001, **44**(5):376~388
- 11 Zhao J, Shawe-Taylor J. Learning in stochastic bit stream neural networks. *Neural Networks, Elsevier Science*, 1996, **9**(6):991~998
- 12 Werbos P J. The Roots of Backpropagation. New York: John Wiley & Sons, 1994

赵杰煜 1988 年获得浙江大学计算机应用硕士学位,1995 年获英国伦敦大学 Royal Holloway 学院计算机科学博士学位,同年赴瑞士 IDSIA 人工智能研究所从事博士后研究,1998 年回宁波大学工作,现任该校教授.主要研究领域为机器学习、计算智能、图模型、随机神经网络、统计学习理论、模式识别等.

张泉方 1988 年毕业于浙江大学计算机系,获工学硕士学位.现任职于浙江大学计算机科学与技术学院讲师.主要研究方向为计算机应用、智能技术、计算机网络及通信技术.

汪加才 1985 年毕业于中国矿业大学自动化系计算机应用专业,1994 年调入宁波大学,2002 年获浙江大学计算机应用博士学位.研究方向为人工智能、神经网络、知识工程、数据挖掘等.