

利用最大似然准则的双向联想网络研究

韦晓亮^{1,2}, 冯祖仁^{1,2}, 常宏^{1,2}

(1. 西安交通大学系统工程研究所, 710049, 西安; 2. 西安交通大学机械制造系统工程国家重点实验室, 710049, 西安)

摘要: 针对现有双向联想网络(BAM)存在的存储容量小、抗干扰能力弱的缺点,提出了一种利用最大似然准则的BAM网络(MLBAM)及其训练算法。MLBAM网络采用双向网络结构建立了神经元的发放以及抑制模型,充分利用似然函数的特性以及网络的双向联想特性,很好地完成了自联想和异联想功能,并且准确计算出关联样本对之间的关联度,使MLBAM网络在随机环境中具有很强的抗噪能力。利用最速下降算法,给出了MLBAM网络的训练算法,根据训练权重的Hessian矩阵负定,判定算法能够获得全局最优解,从而证明了算法的收敛性。该训练算法能够训练出最优的连接权重和神经元阈值。通过2个典型实验验证了MLBAM网络的抗噪能力和联想能力,在存在1位随机噪声的情况下,该网络的联想正确率达到了100%。

关键词: 双向联想;最大似然准则;自联想;异联想

中图分类号: TP183 **文献标志码:** A **文章编号:** 0253-987X(2008)08-0963-04

Bidirectional Associative Memory Utilizing Maximum Likelihood

WEI Xiaoliang^{1,2}, FENG Zuren^{1,2}, CHANG Hong^{1,2}

(1. Systems Engineering Institute of Xi'an Jiaotong University, Xi'an 710049, China; 2. State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A new learning rule and theoretical analysis of an extended bidirectional associative memory network (MLBAM) are presented by using the maximum likelihood criterion based on two well recognized and essential criteria, i. e., the convergence of the learning rule, and the noise tolerance of the network. Traditional methods fail to distinguish highly closing patterns. However, this disadvantage is improved by using the newly developed method, since the maximum likelihood method is used to seek the best possible closing mapping of two patterns. Experiments are made to verify the validity and efficiency of the proposed method. The method displays excellent anti-noise property and MLBAM network exhibits association at a 100% accuracy under one-bit inversion which implies that 100% of the one-bit errors is corrected.

Keywords: bidirectional associative memory; maximum likelihood criterion; auto-association; hetero-association

联想记忆是大脑的一个重要生物特征。与人脑类似,联想记忆神经网络把样本模式存储在神经元之间的权重中,经过大规模的并行计算学习,存储并记忆了确定的输入模式后,在输入带有噪声或无噪声的不完整模式时,网络仍能够通过计算给出与输入模式最接近的完整模式。联想记忆可分成自联想

记忆和异联想记忆2种^[1],其中异联想记忆也称为双向联想记忆(BAM),是指在成对的存储模式 $(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}), (\mathbf{X}^{(2)}, \mathbf{Y}^{(2)}), \dots, (\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})$ 情况下,如果输入为 $\mathbf{X}^{(i)}$,则输出为 $\mathbf{Y}^{(i)}$ 的联想形式。当存在噪声或缺损在一定范围时,联想功能仍可使样本对复原。Kosko^[2]在1987年提出的双向联想网络模

型就能够完成上述的联想功能.但是,这种传统的BAM网络的训练算法^[2-8]始终存在的缺点是记忆的训练模式个数少,尤其是对无法记忆强关联的近似样本.

本文利用最大似然估计准则提出了一种新的双向联想网络(MLBAM),能够很好地完成联想功能,尤其是对记忆强关联的近似样本,并且能够计算出极为近似的样本之间的关联度.

1 问题描述

联想网络按照处理误差的方式分为准确网络和差值网络.对于准确网络,它的输入输出关系可以表示为 $\psi(\mathbf{X}+\boldsymbol{\varepsilon})=\mathbf{Y}$,其中 $\mathbf{X}$ 是输入, $\mathbf{Y}$ 是输出, $\boldsymbol{\varepsilon}$ 是一个小干扰.对于差值网络,它的输入输出关系可以表示为 $\psi(\mathbf{X}+\boldsymbol{\varepsilon})=\mathbf{Y}+\boldsymbol{\delta}$,其中当 $|\boldsymbol{\varepsilon}|\rightarrow 0$ 时, $|\boldsymbol{\delta}|\rightarrow 0$.在很多情况下需要 $\mathbf{X}$ 最大程度地关联到 $\mathbf{Y}$,即完成差值网络的功能,我们称这种关联行为为联想.本文提出的MLBAM网络的基本思想就是对于需要关联的 K 对数据进行学习,如 $(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)})$, $(\mathbf{X}^{(2)}, \mathbf{Y}^{(2)})$, $\dots$, $(\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})$.假定 $\mathbf{X}^{(i)}+\boldsymbol{\varepsilon}$ 输入网络($\boldsymbol{\varepsilon}$ 为干扰信号),输出为 $\mathbf{Y}^{(i)}+\boldsymbol{\delta}$ ($\boldsymbol{\delta}$ 为偏差),当 $|\boldsymbol{\varepsilon}|\rightarrow 0$ 时, $|\boldsymbol{\delta}|\rightarrow 0$;反之亦然.

1.1 双向联想网络模型描述^[2-3]

图1所示的双向联想神经网络中,存储了 n 维神经元矢量 $\mathbf{X}=\{x_1, x_2, \dots, x_i, \dots, x_n\} \in \{-1, 1\}^n$ 与 m 维神经元矢量 $\mathbf{Y}=\{y_1, y_2, \dots, y_j, \dots, y_m\} \in \{-1, 1\}^m$,那么存储容量为 K 对关联数据,即 $(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)})$, $(\mathbf{X}^{(2)}, \mathbf{Y}^{(2)})$, $\dots$, $(\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})$,其中 $\mathbf{X}^{(i)}=\{x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)}\}$, $\mathbf{Y}^{(i)}=\{y_1^{(i)}, y_2^{(i)}, \dots, y_m^{(i)}\}$,与矢量 $\mathbf{X}$ 对应的一层有 n 个节点,与矢量 $\mathbf{Y}$ 对应的一层有 m 个节点,两层之间双向连接.假定 $\mathbf{Y}$ 到 $\mathbf{X}$ 的权重矩阵为 $\mathbf{W}$,则 $\mathbf{X}$ 到 $\mathbf{Y}$ 的权重矩阵为 $\mathbf{W}^T$, $\mathbf{W}=\mathbf{W}^T$.

1.2 神经元模型^[4]

向量 $\mathbf{S}=\{s_j, j=1, 2, \dots, m\}$ 作为神经元 x_i 的输入,神经元的输出为发放(+1)或抑制(-1),其中膜电位 $a_{x_i} \cdot \mathbf{S}$ 等于输入 $\mathbf{S}$ 和轴突权重 w_{ij} 的乘积加上阈值 θ_{x_i} .膜电位的数学表达式为

$$a_{x_i} \cdot \mathbf{S} = \theta_{x_i} + \sum_{j=1}^m w_{ij} s_j \quad (1)$$

具体神经元模型见图2.

定义1 输入 $\mathbf{S}$ 使得 x_i 神经元在输出 $x_i^{(k)}$ 为发放时的概率为

$$P(x_i^{(k)} = 1 | \mathbf{S}) = \sigma(a_{x_i} \cdot \mathbf{S}) \quad (2)$$

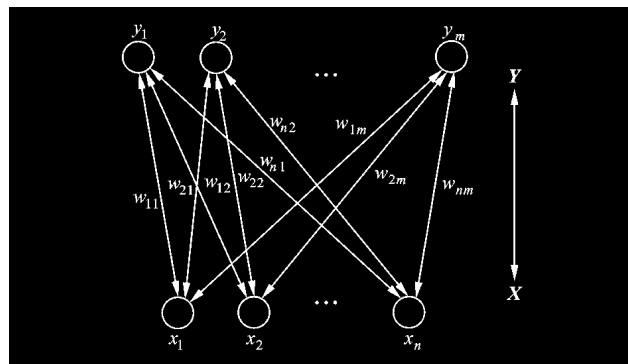


图1 BAM网络模型示意图

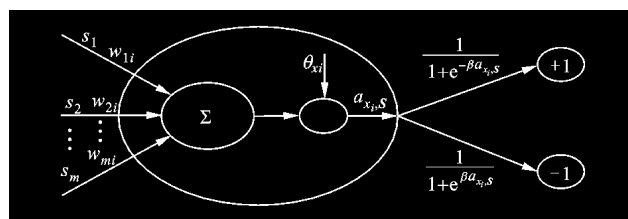


图2 神经元 x_i 的模型示意图

式中: $\sigma(x)=1/(1+e^{-\beta x})$, β 为控制系数, $\sigma(-x)=1-\sigma(x)$. 由此可以得到

$$P(x_i^{(k)} = -1 | \mathbf{S}) = 1 - P(x_i^{(k)} = 1 | \mathbf{S}) = 1 - \sigma(a_{x_i} \cdot \mathbf{S}) = \sigma(-a_{x_i} \cdot \mathbf{S}) \quad (3)$$

所以,神经元 x_i 的输出概率为

$$P(x_i^{(k)} | \mathbf{S}) = \sigma(x_i^{(k)} a_{x_i} \cdot \mathbf{S}) \quad (4)$$

2 基于最大似然估计的双向联想网络构建

2.1 似然函数表达式的建立

设 $\mathbf{A}$ 事件为 $\mathbf{X} \leftrightarrow \mathbf{Y}$,其中 $\mathbf{A}^{(i)}$ 事件为独立事件,表述为 $\mathbf{X}^{(i)} \leftrightarrow \mathbf{Y}^{(i)}$.

定义2 $\mathbf{A}^{(i)}$ 为 $\mathbf{X}^{(i)} \rightarrow \mathbf{Y}^{(i)}$ 与 $\mathbf{X}^{(i)} \leftarrow \mathbf{Y}^{(i)}$ 的联合事件,则事件 $\mathbf{A}^{(i)}$ 发生的可能性可以定义为

$$p(\mathbf{A}^{(i)}) = P(\mathbf{X}^{(i)} | \mathbf{Y}^{(i)})P(\mathbf{Y}^{(i)} | \mathbf{X}^{(i)}) \quad (5)$$

因而得到

$$p(\mathbf{A}) = P(\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \dots, \mathbf{A}^{(K)}) = \prod_{k=1}^K P(\mathbf{X}^{(k)} | \mathbf{Y}^{(k)})P(\mathbf{Y}^{(k)} | \mathbf{X}^{(k)}) \quad (6)$$

所以,似然函数表述为

$$L = \lg \prod_{k=1}^K P(\mathbf{X}^{(k)} | \mathbf{Y}^{(k)})P(\mathbf{Y}^{(k)} | \mathbf{X}^{(k)}) = \sum_{k=1}^K \sum_{i=1}^n \lg(\sigma(x_i^{(k)} a_{x_i} \cdot \mathbf{y}^{(k)})) + \sum_{k=1}^K \sum_{j=1}^m \lg(\sigma(y_j^{(k)} a_{y_j} \cdot \mathbf{x}^{(k)})) =$$

$$\sum_{k=1}^K \sum_{i=1}^N \lg(1/(1 + e^{-\beta x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}})) + \sum_{k=1}^K \sum_{j=1}^M \lg(1/(1 + e^{-\beta y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}})) \quad (7)$$

式中: $a_{x_i, \mathbf{Y}^{(k)}}$ 、 $a_{y_j, \mathbf{X}^{(k)}}$ 分别为神经元 x_i 、 y_j 在 $\mathbf{X}^{(k)}$ 、 $\mathbf{Y}^{(k)}$ 下的膜电位。

通过上面的问题描述即可获得似然函数,进而可以利用最大似然估计准则来训练网络,得到最大似然估计准则下最优的 $n \times m$ 个连接权重 w_{ij} 、 n 个神经元阈值 θ_{x_i} 和 m 个神经元阈值 θ_{y_j} 。

2.2 MLBAM 训练算法

2.2.1 递推公式 由最速下降算法,可以得到连接权重和神经元阈值

$$\left. \begin{aligned} w_{ij}^{\text{new}} &= w_{ij} + \eta \frac{dL}{dw_{ij}} = w_{ij} + \\ &\eta \sum_{k=1}^K \beta \sigma(x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}) e^{-\beta x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}} x_i^{(k)} y_j^{(k)} + \\ &\eta \sum_{k=1}^K \beta \sigma(y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}) e^{-\beta y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}} \\ \theta_{x_i}^{\text{new}} &= \theta_{x_i} + \eta \frac{dL}{d\theta_{x_i}} = \theta_{x_i} + \\ &\eta \sum_{k=1}^K \beta \sigma(x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}) e^{-\beta x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}} \\ \theta_{y_j}^{\text{new}} &= \theta_{y_j} + \eta \frac{dL}{d\theta_{y_j}} = \theta_{y_j} + \\ &\eta \sum_{k=1}^K \beta \sigma(y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}) e^{-\beta y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}} y_j^{(k)} \end{aligned} \right\} \quad (8)$$

2.2.2 全局最优判定 根据训练权重的 Hessian 矩阵可以判定是否能够获得最优解,训练权重的 Hessian 矩阵表述为

$$\frac{d^2 L}{dw_{ij} dw_{ql}} = -\beta^2 \sum_{k=1}^K [1 - \sigma(x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}})] \cdot \sigma(x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}) x_i^{(k)} y_j^{(k)} x_q^{(k)} y_l^{(k)} \delta_{ij} - \beta^2 \sum_{k=1}^K [1 - \sigma(y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}})] \sigma(y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}) y_j^{(k)} x_i^{(k)} x_q^{(k)} y_l^{(k)} \delta_{jl}$$

由于 $[1 - \sigma(x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}})] \sigma(x_i^{(k)} a_{x_i, \mathbf{Y}^{(k)}}) > 0$ 且 $[1 - \sigma(y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}})] \sigma(y_j^{(k)} a_{y_j, \mathbf{X}^{(k)}}) > 0$, 因此 Hessian 矩阵负定,因而证明了利用公式(8)可以训练出最优的连接权重和神经元阈值。

2.3 MLBAM 联想过程及关联度计算

在获得网络的结构描述参数,即连接权重和神经元阈值之后,已经得到了整个 MLBAM 网络的描

述,因而可以利用该网络完成自联想和异联想这 2 种功能,由此就能够得到下面的联想过程和关联度计算。

(1)联想过程. 假定输入为 $\mathbf{X}' = \mathbf{X}^{(i)} + \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}$ 为噪声,输出为 $\hat{\mathbf{Y}}$, 即 $\hat{\mathbf{Y}} = f(\mathbf{X}')$. 我们通过公式 $\hat{y}_i = \begin{cases} +1, \sigma(a_{x_i, \mathbf{Y}^{(k)}}) \geq 0.5 \\ -1, \sigma(a_{x_i, \mathbf{Y}^{(k)}}) < 0.5 \end{cases}$ 获得 $\hat{y}_i$, 再将 $\hat{y}_i$ 送入网络, 通过学习最后得到 $\hat{x}_i$, 从而得到了联想对 $(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$, 即 $\mathbf{X}' \rightarrow \hat{\mathbf{Y}} \rightarrow \hat{\mathbf{X}}$. 联想示意图如图 3 所示。

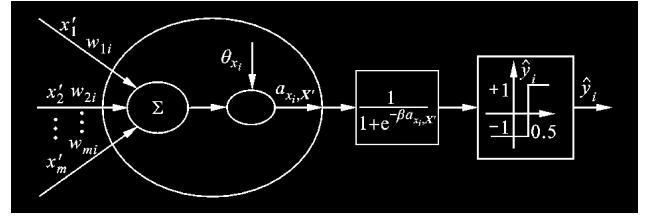


图 3 神经元 y_i 的联想示意图

(2)关联度计算. 利用 MLBAM 双向联想网络能够存储的记忆模式对,可以得到似然函数

$$L\{(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}), (\mathbf{X}^{(2)}, \mathbf{Y}^{(2)}), \dots, (\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})\} = \lg \prod_{k=1}^K P(\mathbf{X}^{(k)} | \mathbf{Y}^{(k)}) P(\mathbf{Y}^{(k)} | \mathbf{X}^{(k)}) \quad (9)$$

定义 3 向量 $\mathbf{c}$ 和 $\mathbf{d}$ 之间的关联度 $C_A(\mathbf{c}, \mathbf{d})$ 定义为

$$C_A(\mathbf{c}, \mathbf{d}) = P(\mathbf{c} | \mathbf{d}) P(\mathbf{d} | \mathbf{c}) \quad (10)$$

C_A 值越大,即代表 2 个向量的关联性越强。

定义 4 向量 $\mathbf{c}$ 和样本集 $\boldsymbol{\Psi} = \{\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(K)}\}$ 之间的关联度 $C_A(\mathbf{c}, \boldsymbol{\Psi})$ 定义为

$$C_A(\mathbf{c}, \boldsymbol{\Psi}) = \max\{C_A(\mathbf{c}, \mathbf{Y}^{(1)}), C_A(\mathbf{c}, \mathbf{Y}^{(2)}), \dots, C_A(\mathbf{c}, \mathbf{Y}^{(K)})\} \quad (11)$$

C_A 值越大,即代表输入向量与样本集的关联性越强。

假设需要计算某一个输入 $\mathbf{c}$ 与训练样本集 $\boldsymbol{\Psi} = \{\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \dots, \mathbf{Y}^{(K)}\}$ 之间的关联度 $C_A(\mathbf{c}, \boldsymbol{\Psi})$, 首先计算出一组似然函数值, 分别为 $L^{(1)} = L\{(\mathbf{c}, \mathbf{Y}^{(1)}), (\mathbf{X}^{(2)}, \mathbf{Y}^{(2)}), \dots, (\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})\}$, $L^{(2)} = L\{(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}), (\mathbf{c}, \mathbf{Y}^{(2)}), \dots, (\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})\}$, $\dots$, $L^{(P)} = L\{(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}), (\mathbf{X}^{(2)}, \mathbf{Y}^{(2)}), \dots, (\mathbf{c}, \mathbf{Y}^{(P)})\}$, 然后找出最大的似然值。

这个问题可以描述为

$$\max_{1 \leq i \leq K, i \in N} L\{(\mathbf{X}^{(1)}, \mathbf{Y}^{(1)}), (\mathbf{X}^{(2)}, \mathbf{Y}^{(2)}), \dots, (\mathbf{c}, \mathbf{Y}^{(i)}), \dots, (\mathbf{X}^{(K)}, \mathbf{Y}^{(K)})\} \quad (12)$$

在实际计算中, $P(\mathbf{X}^{(k)} | \mathbf{Y}^{(k)}) \approx 1$, $P(\mathbf{Y}^{(k)} | \mathbf{X}^{(k)})$

4 结束语

本文基于最大似然准则提出了一种新的 BAM 网络(MLBAM),并提出了该网络的学习算法,分析了该网络学习算法的收敛性以及网络的容错能力. MLBAM 网络能够实现相互关联的样本对的联想功能,同时能够准确计算关联样本对之间的关联度,确保了网络在随机环境中具有很强的抗噪能力,异联想和自联想实验验证了网络的抗噪能力和联想能力.

参考文献:

- [1] HECHT-NIELSEN R. Neurocomputing [M]. Reading, Berkshire, UK: Addison Wesley Publishing Company, 1991.
- [2] KOSKO B. Adaptive bidirectional associative memories [J]. Applied Optics, 1987, 26(23): 4947-4960.
- [3] KOSKO B. Bidirectional associative memories [J]. IEEE Trans System, Man, and Cybernetics, 1988, 18(1): 49-60.
- [4] BARBER D, AGAKOV F. Correlated sequence learning in a network of spiking neurons using maximum likelihood, EDI-INF-RR-0149 [R]. Edinburgh, Scotland, UK: University of Edinburgh, School of Informatics, 2002.
- [5] WANG Y F, CRUZ J B Jr, MULLIGAN J H Jr. Two coding strategies for bidirectional associative memory [J]. IEEE Transactions on Neural Networks, 1990, 1(1): 81-91.
- [6] WANG Zhengou. A bidirectional associative memory based on optimal linear associative memory [J]. IEEE Trans on Computers, 1996, 46(10): 1171-1179.
- [7] YUKIO K. Functional representation of recalling process and memory capacity in associative memory [C]// IEEE-INNS-ENNS International Joint Conference on Neural Networks, Piscataway, NJ, USA: IEEE, 2000: 625-632.
- [8] CHARTIER S, BOUKADOUM M. A bidirectional heteroassociative memory for binary and grey-level patterns [J]. IEEE Transactions on Neural Networks, 2006, 17(2): 385-396.

(编辑 刘杨)