

光学神经网络中硬件误差对性能影响的分析*

杨建文 高胜泉 张延焯 母国光

(南开大学现代光学研究所, 天津 300071)

廉玉忠

(南开大学数学系, 天津 300071)

摘 要 利用统计方法, 分析了光学系统硬件误差对用 Hopfield 神经网络作联想记忆时出错率的影响, 导出了系统的误差与出错率关系的近似公式, 并给出了模拟结果。得到了一定限度的误差对出错率的影响并不显著的结论, 特别对探测器阵列动态范围的要求, 可以远远小于其探测到的最大值。这对光学神经网络系统的设计及硬件的选取, 具有指导意义。

关键词 光学神经网络, 硬件误差, 出错率。

1 引 言

当前国际上对人工神经网络的研究, 集中于它的联想、分类、复原以及控制等功能, 这些功能可以归结为将其作为一个联想记忆系统来进行研究, 人工神经网络作联想记忆, 研究得较多的是 Hopfield 模型^[1]。

目前在用 Hopfield 网络作联想记忆中, 人们已经用统计方法导出了它的收敛半径或记忆容量极限^[2]。在具体用光学方法实现时, 由于系统各个环节都不可避免地存在误差, 这必将影响收敛半径或存储容量。本文利用统计方法, 研究了硬件的误差与联想记忆出错率的关系, 导出了具体表达式并给出了模拟结果, 从而为光学神经网络系统硬件的设计和选取提供了依据。

2 Hopfield 网络作联想记忆的工作原理

作者以 Farhat^[3] 提出的, 用光学器件实现人工神经网络的结构为例, 来说明 Hopfield 网络联想记忆的工作原理。

设网络由 N 个神经元组成, 有 M 个独立随机记忆样本, 每个样本可表示为一个 N 维矢量 $X^{(m)}$, $X^{(m)} = [x_1^{(m)}, x_2^{(m)}, \dots, x_N^{(m)}]$, $m = 1, \dots, M$ 。系统的互连权重矩阵 W 由这 M 个样本

* 国家高科技计划和攀登计划认知科学(神经网络)重大关键课题的资助项目。

收稿日期: 1994 年 10 月 21 日; 收到修改稿日期: 1995 年 1 月 9 日

矢量的自外积和求得, 构成一个 $N \times N$ 维矩阵, 用标量形成表示为:

$$W_{ij} = \begin{cases} \sum_{m=1}^M X_i^{(m)} X_j^{(m)}, & i \neq j \\ 0, & i = j \end{cases} \quad i, j = 1, 2, \dots, N \quad (1)$$

对一个输入样本 $X^{(0)}$, 设它与某存贮样本 $X^{(m)}$ 的汉明距离最短, 其值为 ρ , 即 $X^{(0)}$ 与 $X^{(m)}$ 之间有 ρN 个分量反号, 神经元状态的演变规则为:

$$X_j(t+1) = \text{sgn} \left[\sum_{k=1, k \neq j}^N W_{jk} X_k(t) \right] \quad (2)$$

网络正确运算的结果是, 经若干次迭代后, 输出状态尽量接近或等于 $X^{(m)}$ 。否则就称网络不能收敛。这里定义一个变量:

$$S_j^{(m)} = \sum_{k=1, k \neq j}^N W_{jk} X_k^{(0)} X_j^{(m)} \quad (3)$$

由(3)式可知, 只要 $S_j^{(m)} \geq 0$, 对于 $j = 1, \dots, N$ 皆成立, 网络就能一次迭代收敛。(1)式代入(3)式并分解得:

$$S_j^{(m)} = \sum_{r=1, r \neq m}^M \sum_{k=1, k \neq j}^N X_j^{(r)} X_k^{(r)} X_k^{(0)} X_j^{(m)} + \sum_{k=1, k \neq j}^N X_j^{(m)} X_k^{(m)} X_k^{(0)} X_j^{(m)} \quad (4)$$

令 $Z_j^{(m)} = \sum_{r=1, r \neq m}^M \sum_{k=1, k \neq j}^N X_j^{(r)} X_k^{(r)} X_k^{(0)} X_j^{(m)}$, (4)式中, 第二项为:

$$\begin{aligned} \sum_{k=1, k \neq j}^N X_j^{(m)} X_k^{(m)} X_k^{(0)} X_j^{(m)} &= (1 - 2\rho)N - 1, \\ \therefore S_j^{(m)} &= Z_j^{(m)} + N(1 - 2\rho) - 1 \end{aligned} \quad (5)$$

$Z_j^{(m)}$ 是 $(N-1)(M-1)$ 个独立随机变量之和, 每个变量取 $+1$ 或 -1 的几率都是 $1/2$, 故经过一次迭代后, $X^{(m)}$ 与 $X^{(0)}$ 某个分量反号的几率为:

$$P_{E_1} = P_{(S_j^{(m)} < 0)} = P_{|Z_j^{(m)}| < [N(1-2\rho)-1]} \quad (6)$$

由中心极限定理, 当 M, N 较大时, $Z_j^{(m)}$ 取值服从正态分布, 它取值 $Z - Z + dZ$ 的几率为

$$\begin{aligned} P_{(Z, Z+dZ)} &= \frac{1}{\sqrt{2\pi} \sqrt{(M-1)(N-1)}} \exp \left[\frac{-Z^2}{2(M-1)(N-1)} \right] dz \\ &\doteq \frac{1}{\sqrt{2\pi} \sqrt{MN}} \exp \left[\frac{-Z^2}{2MN} \right] dz, \quad (M, N \gg 1) \end{aligned} \quad (7)$$

$$\therefore P_{E_1} = P_{(S_j^{(m)} < 0)} = P_{|Z_j^{(m)}| < [N(1-2\rho)-1]} \doteq \frac{1}{\sqrt{2\pi MN}} \int_{-\infty}^{-N(1-2\rho)+1} \exp \left[-\frac{Z^2}{2MN} \right] dz \quad (8)$$

其几率密度曲线如图 1 所示, 其中阴影部分的面积即为 P_{E_1} 的值。

从文献[2]可知, 当 $N \rightarrow \infty$ 时, 如果网络一次迭代收敛, 其容量的上限值为:

$$M = (1 - 2\rho)^2 N / 2 \ln N \quad (9)$$

光学实现的原理如图 2 所示。图中, 神经元状态矢量由发光二极管阵列(LED)的光强分布来表示。 W 矩阵由透明的照相掩膜版构成, 透射率的分布与加权矩阵 W 各元素之分布一一对立。它的另一侧是探测器阵列(PDA), 加权求和信号经光电转换之后送到放大器和取阈电路, 然后决定是否输出或进行下一次迭代。

由于 W 和 $X^{(0)}$ 都有正、负值元素, 光学不便直接表示负数, 为解决这个问题, 已提出了一些方案^[3~7]。而本文假设正、负值同时参与了运算, 不难理解, 这并不影响本文的讨论。

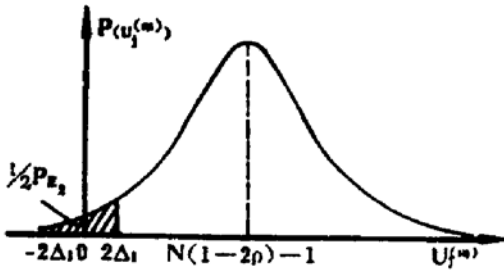


Fig. 1 The probability density curve of $S_j^{(m)}$ (The shadow area equals to P_{E_1})

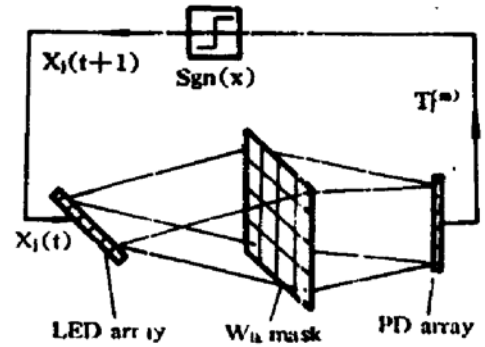


Fig. 2 Concept of optical neural network (Matrix-vector multiplier incorporating nonlinear electronic feedback)

3 误差对联想记忆出错率的影响

图 2 中, 系统的误差主要包括下列几部分(由于系统误差可以修正, 本文不予考虑):

- 1) 发光二极管发光的不均匀性(光强起伏), $X_i^{(0)'} = X_i^{(0)}(1 \pm \sigma_1)$;
- 2) 透明片制作时透光率的误差, 即透射率之分布并不正好等于 W 各元素值的大小,

$$W_{jk} = W_{jk}(1 \pm \sigma_2);$$

- 3) 探测器的噪声(包括放大器的噪声) Δ_1 , 即探测器存在分辨极限, 光强为 $I \pm \Delta_1$, 都判为光强值为 I .

另外还有杂散光、环境条件的变化等因素, 这些可以控制得很好, 与上述误差相比很小, 因而可以忽略。

首先看在没有误差时, 加权求和值的分布情况, 设:

$$T_j^{(m)} = \sum_{k=1, k \neq j}^N W_{jk} X_k^{(0)} \quad (10)$$

由于 $X_j^{(m)}$ 为 +1 或 -1, 比较(3)式和(10)式, 显然有:

$$\begin{aligned} T_j^{(m)} &= \frac{S_j^{(m)}}{X_j^{(m)}} = S_j^{(m)} X_j^{(m)} = X_j^{(m)} [Z_j^{(m)} + (1 - 2\rho)N - 1] \\ &= \begin{cases} Z_j^{(m)} + (1 - 2\rho)N - 1 & (X_j^{(m)} = +1) \\ - [Z_j^{(m)} + (1 - 2\rho)N - 1] & (X_j^{(m)} = -1) \end{cases} \end{aligned} \quad (11)$$

当考虑(1)、(2)两项误差时, 第 j 个探测器接收到的光强为:

$$U_j^{(m)} = \sum_{k=1}^N W_{jk} X_k^{(0)'} \quad (12)$$

根据误差理论, $U_j^{(m)}$ 期望值为: $E(U_j^{(m)}) = \sum_{k=1, k \neq j}^N W_{jk} X_k^{(0)} = T_j^{(m)}$, 标准差约为: $\sigma_{U_j^{(m)}} = \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$, 由于 $U_j^{(m)}$ 服从正态分布, 可将其极限误差值取为: $3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$, 具体推导过程见附录。

$$\begin{aligned} \therefore U_j^{(m)} &= T_j^{(m)} \pm 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2} \\ &= \begin{cases} + [Z_j^{(m)} + (1 - 2\rho)N - 1 \pm 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}] & (X_j^{(m)} = +1) \\ - [Z_j^{(m)} + (1 - 2\rho)N - 1 \pm 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}] & (X_j^{(m)} = -1) \end{cases} \end{aligned} \quad (13)$$

实际上, $U_j^{(m)}$ 中包括正值和负值两部分, 而在非相干光中, 光强无法直接相减, 因而在实际光学系统中, 通过适当编码, 把加权求和的正值与负值分开, 用两个探测器分别探测或用

一个探测器依次探测,然后把得到的这两个正值通过计算机或电子线路实现减法运算,即

$$U_j^{(m)} = U_j^{(m)}(+)-U_j^{(m)}(-) \quad (14)$$

又由于各个探测器的噪声为 $\pm \Delta_1$,则在探测 $U_j^{(m)}$ 时,得到的实际值为;

$$(U_j^{(m)}(+)\pm\Delta_1)-(U_j^{(m)}(-)\pm\Delta_1)=U_j^{(m)}\pm 2\Delta_1=I_j^{(m)} \quad (15)$$

输出 $X_j^{(m)}$ 是对 $I_j^{(m)}$ 硬取阈的结果,显然有下列关系:

$$X_j^{(m)} = \begin{cases} +1 & (U_j^{(m)} \geq 2\Delta_1) \\ -1 & (U_j^{(m)} \leq -2\Delta_1) \\ \text{不确定} & (|U_j^{(m)}| < 2\Delta_1) \end{cases} \quad (16)$$

即当 $|U_j^{(m)}| < 2\Delta_1$ 时,由于探测器带来的误差也为 $2\Delta_1$,这时 $X_j^{(m)}$ 就有可能由于探测器的误差而与 $U_j^{(m)}$ 反号,反号的几率为 $1/2$ 。

综上所述,只要算出 $U_j^{(m)}$ 落在区间 $[-2\Delta_1, +2\Delta_1]$ 中的几率,就可知由于系统的误差带来的误判几率 P_{E_2} ,其值为 $(1/2) P_{|U_j^{(m)}| \in (-2\Delta_1, +2\Delta_1)}$ 。(不管 $X_j^{(m)} = 1$ 或 $X_j^{(m)} = -1$,情况都一样,故只须考虑 $X_j^{(m)} = 1$ 就行。)

由(13)式可知:

$$\begin{aligned} P_{E_2} &= (1/2) P_{|U_j^{(m)}| \in (-2\Delta_1, +2\Delta_1)} \\ &= (1/2) P_{|(Z_j^{(m)} + (1-2\rho)N - 1) \pm 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}| \in (-2\Delta_1, 2\Delta_1)} \\ &= (1/2) P_{|Z_j^{(m)}| \in (-2\Delta_1 - (1-2\rho)N + 1 - 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}, 2\Delta_1 - (1-2\rho)N + 1 + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2})} \\ &= \frac{1}{2} \frac{1}{\sqrt{2\pi MN}} \int_{-2\Delta_1 - (1-2\rho)N + 1 - 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}}^{2\Delta_1 - (1-2\rho)N + 1 + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}} \exp\left(-\frac{Z^2}{2MN}\right) dZ \end{aligned} \quad (17)$$

(17)式的值等于图 3 中的阴影部分的面积的 $1/2$ 。

综合(8)式和(13)式可知系统的总的误判率(由软件和硬件共同造成的)为:

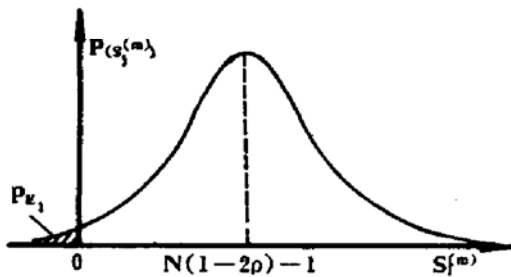


Fig. 3 The error probability caused by hardware error (It equals to half the shadow area)

$$\begin{aligned} P_E &= P_{E_1} + P_{E_2} = P_{|S_j^{(m)}| < -2\Delta_1 - 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}} \\ &\quad + (1/2) P_{|U_j^{(m)}| \in (-2\Delta_1 - 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}, 2\Delta_1 + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2})} \\ &= (1/2) P_{|S_j^{(m)}| < -2\Delta_1 - 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}} \\ &\quad + (1/2) P_{|U_j^{(m)}| < 2\Delta_1 + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2}} \end{aligned} \quad (18)$$

式中 $S_j^{(m)}$ 与 $U_j^{(m)}$ 的分布相同,均为中心处于 $(1-2\rho)N-1$ 的正态分布。

4 举例说明系统的噪声对误判率的影响

由于 $U_j^{(m)} = U_j^{(m)}(+)-U_j^{(m)}(-)$

$$\text{而: } U_j^{(m)}(+)+U_j^{(m)}(-) = MN + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2} = \text{constant} + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2} \quad (19)$$

所以,考虑噪声后,有:

$$U_j^{(m)} = 2U_j^{(m)}(+)- (MN + 3\sqrt{MN}\sqrt{\sigma_1^2 + \sigma_2^2})$$

$$= MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2} - 2U_j^{(m)}(-) \quad (20)$$

(16)式变为:

$$X_j^{(m)} = \begin{cases} +1 & (U_j^{(m)}(+)) > \frac{1}{2}(MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}) + \Delta_1 \\ -1 & (|U_j^{(m)}(-)|) > \frac{1}{2}(MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}) + \Delta_1 \\ \text{不确定} & \text{其他} \end{cases} \quad (21)$$

设探测器阵列为 CCD, 其动态范围为 G 个灰度阶, 根据(21)式, 由于只对 CCD 探测的光强作硬取阈, 故对于 $U_j^{(m)}(+)$ 或 $|U_j^{(m)}(-)| > \frac{1}{2}(MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}) + \Delta_1$ 的光强, 并不需要精确测出, 而只需让它使 CCD 达到饱和就行。这样, 适当调节光源强度, 使 $[U_j^{(m)}(+)]_{\max}$ 或 $[U_j^{(m)}(-)]_{\max} = (1/2) / (MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}) + \Delta_1$ 为 CCD 的饱和值 G , 这样, 更加有效地利用了 CCD 的动态范围, 这时, 探测器的误差为:

$$\Delta_1 = \frac{(1/2)(MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}) + \Delta_1}{G}, \text{ 即 } \Delta_1 = \frac{MN + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}}{2(G-1)}. \quad (22)$$

作者正在研制的二维光学神经网络系统, 神经元数为 $N = 1024$, 它是一个通用系统, 网络的工作原理为: 对有 $N \times N$ 个神经元的二维网络, 其神经元的动力学演变规则为:

$$V_a(t+1) = f(V_a(t)), \quad V_a(t) = \sum_{i=1}^N \sum_{j=1}^N W_{aij} V_{ij}(t), \quad (23)$$

其中, V_a 和 V_{ij} 是二维矢量(神经元状态), W_{aij} 是互连权重矩阵(IWM)。

如果神经元状态 V_{ij} 和互连权重 W_{aij} 值都是双极的, 则 V_{ij} 和 W_{aij} 各可分为正负两组:

$$V_{ij} = V_{ij}^{(+)} - V_{ij}^{(-)}, \quad T_{aij} = W_{aij}^{(+)} - W_{aij}^{(-)} \quad (24)$$

于是: $V_a(t) = \sum_{i=1}^N \sum_{j=1}^N W_{aij} V_{ij}(t) = \sum_{i=1}^N \sum_{j=1}^N [W_{aij}^{(+)} - W_{aij}^{(-)}] [V_{ij}^{(+)}(t) - V_{ij}^{(-)}(t)]$

$$\begin{aligned} &= \left[\sum_{i=1}^N \sum_{j=1}^N [W_{aij}^{(+)} V_{ij}^{(+)}(t)] + \sum_{i=1}^N \sum_{j=1}^N [W_{aij}^{(-)} V_{ij}^{(-)}(t)] \right] \\ &\quad - \left[\sum_{i=1}^N \sum_{j=1}^N [W_{aij}^{(-)} V_{ij}^{(+)}(t)] + \sum_{i=1}^N \sum_{j=1}^N [W_{aij}^{(+)} V_{ij}^{(-)}(t)] \right]. \end{aligned} \quad (25)$$

上式中, 前两项是正值部分, 后两项为负值部分。

IWM 可分解为 $N \times N$ 个子矩阵, 每个子矩阵有 $N \times N$ 个元素, 如图 4 所示。因此, 四维互连权重矩阵 IWM 可展开成二维子矩阵阵列。二维光学神经网络的光路如图 5 所示, 本结构与 Yu 等^[8]提出的结构基本相同, 只是用平行光照明互连权重矩阵的透明片来代替视频显示器, 并由透镜阵列和透镜 L_1 组成共轴结构以减少系统离轴象差。其优点是, 由于离轴象差的减小, 而使网络的规模极限提高, 并且这种结构还易于加工和调整。其工作原理为: P_1 是透镜阵列的前焦面。通过每个子矩阵的光束, 再通过相应的微透镜及成像透镜 L_1 , 而到达液晶开关阵列 LCSA 上。LCSA 是 IWM 的共轭面, 用来显示输入神经元的状态。由几何光学成像特性, $N \times N$ 个子矩阵的 $N \times N$ 个像将重迭在 LCSA 上。由于探测器阵列是透镜阵列后焦面的共轭面, 所以在探测器阵列面上, 形成 $N \times N$ 个表示(23)式所给出的加权求和关系的亮点。如果作全双极运算, 最简单的方法是把权重和状态的正负部分分别输入, 分四次探测并组合。本文作者提出了一种编码方法^[7], 全双极信号可一次给出, (25)式中前两项和后两项的求和关系分别用一个亮点表示。取阈和反馈运算由计算机完成。

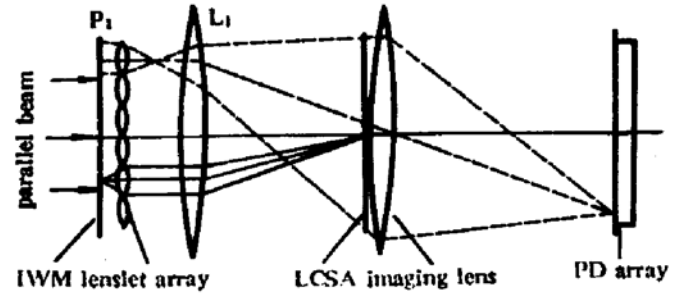
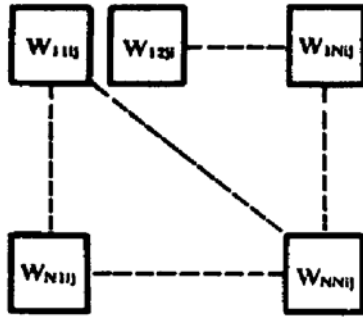


Fig. 4 Reconstructing the 4-D tensor into an array of 2-D submatrices

Fig. 5 Optical arrangement of the neural network

本系统中, 误差来源有以下几方面:

1) 权重模板 IWM 制作误差。IWM 上各元素值的大小, 是由处于该位置的透光孔开孔面积的大小表示的, 用精缩机或电子束曝光机图形合成的方法做成掩膜片, 其面积的误差是随机的, 服从正态分布, 相当于文中的 σ_2 。

2) 用透射式开关阵列(LCSA)表示神经元状态, 开关阵列制作时各单元面积及透光率的非均匀性, 服从正态分布。另外, 光源强度的涨落也服从正态分布。这几部分之综合对应于文中的 σ_1 。

3) 探测器阵列 PDA、A/D 转换、放大器、取阈电路等都有电子噪声, 它们是随机平稳的, 属随机误差, 其综合对应于 Δ_1 。

这里, 假设用它来作联想记忆。在方程(18)中, 改变 σ_1 和 σ_2 , 来考察 P_E 的变化情况。网络的神经元数 $N = 1024$, 其存储的样本数 M 由方程(9) 决定。所用 CCD 动态范围为 256 个灰阶。令光源的起伏 σ_1 和透明片的不均匀性 σ_2 均在 $0 \sim 0.2$ 之间变化, 对 $\rho = 0$ 和 $\rho = 0.3$ 两种情况, 通过计算机模拟, 由图 6(a)、图 6(b) 给出了 P_E 随 σ_1 和 σ_2 变化的等高线曲线和外形曲线。

从上述两组模拟曲线可知, 当 σ_1 或 $\sigma_2 = 0.2$ 时, 对 $\rho = 0$, 误判几率小于 1.8%, 而对 $\rho = 0.3$, 误判几率小于 0.4%, 这是可以允许的。

为了使系统有较好的性能, 同时又符合经济和工艺原则, 系统中各误差分量需合理分配, 一般人为引进条件为“等误差原则”, 按此原则考虑得到的误差分配值是初步的, 在此基础上还需进一步按技术条件、经济条件等实际情况作适当调整。

本文中, 系统硬件的总误差为: $\Delta = 2\Delta_1 + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$, (17) 式变为:

$$P_{E_2} = \frac{1}{2} \frac{1}{\sqrt{2\pi MN}} \int_{-(1-2\rho)N+1-\Delta}^{-(1-2\rho)N+1+\Delta} \exp\left(-\frac{Z^2}{2MN}\right) dZ \quad (26)$$

在进行系统设计时, 在限定的 P_{E_2} 条件下 (M, N 确定), 由(26) 式可得到总误差限 Δ , 然后由“等误差原则”, 初步确定误差分配为: $\sigma_1 = \sigma_2$, $2\Delta_1 = 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$, 且 $\Delta = 2\Delta_1 + 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$, 这三个方程可解出 σ_1, σ_2 , 及 Δ_1 。然后根据工艺性和经济性最好原则, 适当调整 σ_1, σ_2 和 Δ_1 , 使总合成误差不超过 Δ 。 $\sigma_1, \sigma_2, \Delta_1$, 各自是多个因素之综合, 它们内部的误差分配也可用上述原则。

结 论 在神经网络的光学实现中, 系统硬件存在的误差对网络的正确运行, 究竟有多大影响, 这在进行网络设计时一定要有所估计的, 以便为器件的选取提供依据。本文的分析说明, 硬件的误差在一定范围内并不显著影响运行结果, 比直觉上想象的误差限度要宽松的多。事

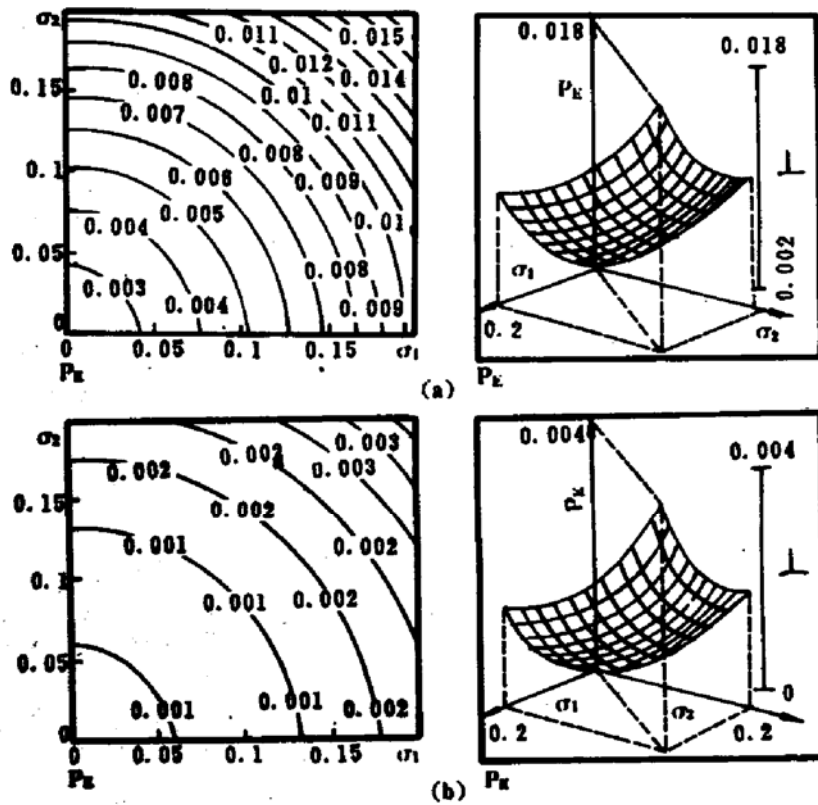


Fig. 6 The surface and contour of P_k vs. σ_1 and σ_2

(a) $\rho=0$, (b) $\rho=0.3$

实上，神经网络系统本身就具有相当大的容错性。因而，在设计光学神经网络时，并不需要过分追求器件的高精度。

附 录

推导 $U_j^{(n)}$ 的标准差 $\sigma_{U_j^{(n)}}^{[9, 10]}$

定义：对随机变量 x ，用 $E(x)$ 表示其期望值， $D(x)$ 表示其方差， $\sigma(X) = \sqrt{D(x)}$ 表示其标准差。对随机性误差函数 $y = f(x_1, x_2, \dots, x_n)$ ，各自变量 x_i 的标准差为 σ_i ， x_i 之间是相互独立的。则：

$$D(y) = \sum_i \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 \tag{1}$$

由于 $U_j^{(n)} = \sum_{k=1}^N W_{jk} x_k^{(n)}$ ，由(1)可得：

$$D(U_j^{(n)}) = \sum_{k=1}^N (W_{jk}^2 D(x_k^{(n)}) + (x_k^{(n)})^2 D(W_{jk})) \tag{2}$$

(2)中， $D(x_k^{(n)}) = \sigma_k^2$ ， $D(W_{jk}) = W_{jk}^2 \sigma_k^2$ ， $(x_k^{(n)})^2 = 1$ ，所以，

$$D(U_j^{(n)}) = (\sigma_k^2 + \sigma_k^2) \sum_{k=1}^N W_{jk}^2, \tag{3}$$

又由于 $W_{jk} = \sum_{m=1}^M x_j^{(m)} x_k^{(m)}$ ($j \neq k$)， $x_j^{(m)}$ ， $x_k^{(m)}$ 为相互独立的取值为 ± 1 ，几率均为 $1/2$ 的变量，所以 W_{jk} 相当于 M 个相互独立的取值为 ± 1 ，几率均为 $1/2$ 的变量之和，当 M 较大时 ($M > 15$)， W_{jk} 近似服从正态分布，且 $E(W_{jk}) = 0$ ， $D(W_{jk}) = M$ 。在(3)中，令 $y = \sum_{k=1}^N W_{jk}^2$ ，则 y 服从 χ^2 分布。由 χ^2 分布可得： $E(y) = NM$ ， $D(y) = 2NM^2$ 。当 N 充分大时 (本文 $N = 1024$)， y 趋于正态分布，故可以取 y 的极限偏差为 $3M \sqrt{2N}$ 。

所以 y 的极限值可估计为：

$$y_{\max} = MN + 3M \sqrt{2N} = MN(1 + 3 \sqrt{2} / \sqrt{N}), \quad \sqrt{y_{\max}} = \sqrt{MN} \sqrt{1 + 3 \sqrt{2} / \sqrt{N}}.$$

由于 N 很大(本文 $N = 1024$), 由误差舍弃原则, $\sqrt{1 + 3 \sqrt{2} / \sqrt{N}} \doteq 1$, $\sqrt{y_{\max}} \doteq \sqrt{MN}$ 。最后可得:
 $\sigma(U_i^{(n)}) \doteq \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$ 。

由于 $U_i^{(n)}$ 是大量(N 个)基本上独立且同分布的随机变量 $\zeta_k = W_{jk} X_k^{(0)}$ ($k = 1, \dots, N$) 的和, 由中心极限定理, $U_i^{(n)}$ 一定服从正态分布, 因而可以把其极限误差取为: $3\sigma(U_i^{(n)}) = 3 \sqrt{MN} \sqrt{\sigma_1^2 + \sigma_2^2}$ 。

参 考 文 献

- [1] J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, April, 1982, **79**: 2554~2558
- [2] Anthony Kuh, B. W. Dickinson, Information capacity of associative memories. *IEEE Trans. Inform. Theory*, 1989, **35**(1): 59~68
- [3] N. Farhat, D. Psaltis, A. Prata *et al.*, Optical implementation of the hopfield model. *Appl. Opt.*, 1985, **24**(10): 1469~1475
- [4] X. Wang, G. G. Mu, Optical neural network with bipolar neural states. *Appl. Opt.*, 1992, **31**(23): 4712~4719
- [5] J. S. Jang, S. W. Jung, S. Y. Lee *et al.*, Optical implementation of Hopfield model for two-dimensional associative memory. *Opt. Lett.*, 1988, **13**(3): 248~250
- [6] A. P. Ittycheriah, J. F. Walkup, T. F. Krile *et al.*, Outer processor using polarization encoding. *Appl. Opt.*, 1990, **29**(2): 275~283
- [7] J. Yang, S. Gao, Y. Zhang *et al.*, Optical fully-bipolar neural network using color encoding. *IEE Proc-Optoelectron*, 1995, **142**(2): 121~124
- [8] T. Lu, S. Wu, X. Xu *et al.*, Two-dimensional programmable optical neural network. *Appl. Opt.*, 1989, **28**(22): 4908~4913
- [9] 肖明耀编著, 实验误差估计与数据处理. 北京, 科学出版社, 1984,
- [10] 王尔祺, 宋德慧编著, 光学仪器精度分析(第一章). 北京, 测绘出版社, 1988,

Influence of Hardware Error on Optical Neural Network

Yang Jianwen Gao Shengquan Zhang Yanxin Mu Guoguang

(*Institute of Modern Optics, Nankai University, Tianjin 300071*)

Lian Yuzhong

(*Department of Mathematics, Nankai University, Tianjin 300071*)

(Received 21 October 1994; revised 9 January 1995)

Abstract The error of optical hardware, which influences the error probability of associative memory in Hopfield neural network, is analyzed by means of statistical method. The approximate formula and modulation result of error probability versus hardware error are proposed. It is concluded that the error within some extent does not affect the error probability sharply. Especially, the required gray level of detector array can be much less than the maximum value detected. This is very significant for designing the optical neural network.

Key words optical neural network, hardware error, error probability.