1000-9825/2001/12(11)1693-06 ©2001 Journal of Software 软件学报 Vol.12, No.11

前向神经网络的容错性问题*

张 铃

(安徽大学 人工智能研究所,安徽 合肥 230039);
(清华大学 智能技术与系统国家重点实验室,北京 100084)
E-mail: zling@ahu.edu.cn
http://www.ahu.edu.cn

摘要: 在讨论神经网络的容错性的文献中,主要涉及的一直是关于输入噪音的容错问题.在这些文献中通常把该问题转换为某种优化问题,并用现成的优化方法进行求解,但很少涉及由网络故障所引起的容错问题,即结构容错问题.利用覆盖算法分析结构容错问题,给出一个神经网络容纳所有单节点故障的充要条件和构造这种网络的算法.这些结果揭示了神经网络结构容错能力的本质,并提供了一种分析神经网络容错的新方法.

关键词: 神经网络;容错性;结构容错性;单点故障
中图法分类号: TP18 **文献标识码**: A

早在20世纪70年代前后,人们曾对逻辑网络的故障测试、诊断以及网络的容错性能进行过较为深入的研究.但后来由于技术的进步,这种测试和诊断的工作成为不必要的了(因为硬件很便宜,发现有故障换一个新板即可,无须花大力气进行诊断).当然,逻辑网络的容错能力的设计仍然有研究的价值.

自20世纪80年代开始,随着神经网络研究的重新兴起,神经网络的容错问题和故障测试问题也就接踵而来.但是,因为神经网络的可读性较差,其研究的难度要比对逻辑网络的研究困难得多,到目前为止,对神经网络的容错性的研究尚未有很深刻的研究成果.

人工神经网络是对人脑神经的一个简单模拟,由于它具有分布式存储、并行处理能力、泛化能力以及处理非线性的能力,故得以广泛应用.这也算是对人脑能力的一个(简单的)模拟,但人脑具有非常强的容错能力,在这一点上,目前的前向神经网络似乎不大具备.比如,人脑每天约有10^4个细胞死亡,仍能照常工作;而前向神经网络只要有一个神经元发生故障,整个网络就不能正常进行工作了.Hammadi[1]认为,这是因为目前的人工神经网络从本质上来说不具有容错能力所致.下面,我们将分析并指出Hammadi的看法未必正确.

20世纪90年代以来,已有不少研究者在进行这方面的研究,并取得了一些成果,下面简单介绍目前国内外对这个问题的研究现状.

对神经网络的容错性的研究大体上可分为两类.其中一类是研究当输入发生偏差时,或者说输入有噪音时,网络的输出将如何发生变化,这也可归于"泛化问题"或"容噪问题".例如,Minnix[2]发现用噪音数据对网络进行训练可以提高网络的容错能力,Murray和Edwards[3]研究对权加入模拟噪音后对容错能力的影响,但这些研究从本质上来说还不能算是网络容错性的研究,因为它们只是对网络的容噪音能力的研究,即网络泛化能力的研究.真正意义下的容错能力应理解为,当网络的

* **收稿日期**: 2000-03-03; **修改日期**: 2000-06-26
基金项目: 国家自然科学基金资助项目(69675011);国家重点基础研究发展规划973资助项目(G1998030509)
作者简介: 张铃(1937-),男,福建福清人,教授,博士生导师,主要研究领域为人工智能理论,人工神经网络理论及应用.

硬件出现故障时，网络保持正确的能力，即网络的结构性故障容错问题.

对此，有人通过对网络增加冗余的方法来改善网络的容错性.例如，Emmerson[4]和 Phatak[5]通过复制隐结点来建立容错神经网络. Phatak[5]还给出为了容纳对所有单点故障(single fault)所需要的最小冗余. Chiru[6]发现大的连接权对故障的敏感性. Tan 和 Nanya[7]提出一种最小化故障和非故障网络之间差别的学习算法，并讨论泛化与容错之间的关系. Ito 和 Yagi[8]指出，通过使用纠错码来修正输出元错误的方法.另有一些学者将求解容错能力问题转化为某种优化问题，然后通过求解优化问题来改善容错性.例如，Neti[9]等人提出，当多层网络的单个结点失效时，针对每个输入，用某种方法最小化实际输出与期望输出之间的最大偏差，以改善网络的容错性. Deodhare[10]等人提出利用优化特定的误差函数的方法来改善网络的容错能力，并设计出两种不同的求解方法，一种是将原来的最小最大优化问题转化为一个无约束的最小平方优化问题，另一种是将原来的优化问题转化为一个无约束的最小化问题，且其解能一致逼近原问题的解. Nijhuis 等人[11]研究了在存在故障的情况下，学习率、训练时间、噪音训练数据对网络性能的影响.总之，已有的研究大多是将求解容错问题转化成某种优化问题，然后用通常的求优化方法进行求解.

本文讨论前向神经网络的单点故障问题，即讨论三层前向神经网络中的某个隐层元及其所有相关的连接权失效情况下的容错问题.

首先利用我们早先给出的 M-P 神经元的几何意义及相关的覆盖学习算法[12,13]，给出一个能容纳所有单点故障的网络的特征性质，即其充要条件，然后给出求解具有容纳“单点故障”的网络的算法.

1 M-P 神经元的几何意义

本文将利用我们在文献[12]中给出的 M-P 神经元的几何意义的方法以及文献[13]中给出的神经网络覆盖学习算法来证明“容错网络”的充要条件.下面简单介绍一下本文用到的文献[12]中的一些概念和性质.

M-P 神经元的模型.将神经元看成是一个有 n 个输入和 1 个输出的元件，此元件的功能函数可表示为 $y=\sigma(\langle w,x\rangle-\theta)$.其中 σ 是符号函数，

$$\sigma(x)=\begin{cases}1, & x>0\\ -1, & x\leqslant 0\end{cases}.$$

现在让我们来分析一下这个模型的几何意义.

(1) 超平面表示

由神经元的功能函数的定义可知，它是由两个函数复合而成，其第 1 个函数为 $\langle w,x\rangle-\theta$，若令其等于 0，得 $\langle w,x\rangle-\theta=0$.这个方程在 n 维空间中表示一个超平面 P，当 $(\langle w,x\rangle-\theta)>0$ 时，表示点 x 落在超平面的正半空间之内，此时 $\sigma(\langle w,x\rangle-\theta)=1$；当 $(\langle w,x\rangle-\theta)<0$ 时，表示点 x 落在 P 的负半空间内，此时 $\sigma(\langle w,x\rangle-\theta)=-1$.于是，一个 M-P 神经元从功能上可看成是一个由超平面划分的空间的位置的识别器.这是对神经元的一种相当直观的理解.人们也曾利用这种理解[14,15]来进行神经网络的学习方面的研究.当只有两三个平面时，尚可直观理解；而当 n,m 较大时，n 维空间中 m 个超平面的相交情况就变得非常复杂，很不直观.总之，很难用超平面的模型来帮助我们直观地理解神经网络的内在性质.因此，这以后就很少有人再借助神经元的几何意义(指用超平面的表示)来进行神经网络的学习和研究了.

(2) 球面上的“领域”表示

几何的直观往往是研究的好向导.以上使用超平面的方法虽然遭到失败,但是只要我们对此稍加改变,将会给我们提供新的直观帮助.

若我们限定输入向量的长度相等,即输入向量是限定在 n 维空间的某个球面上,那么,这时 $(\langle w,x\rangle-\theta)>0$,就表示球面上落在 P 的正半空间的部分,这个部分恰好是球面上的某个“球形领域”.若取 w 与 x 等长,则这个“球形领域”的中心恰好是 w,其半径为 $r(\theta)$,是 θ 的单调下降的函数.

若我们取
$$\sigma'(x)=\begin{matrix}1, & x>0\\ 0, & x\leqslant 0\end{matrix},$$
且取神经元的激励函数为 $\sigma'(\langle w,x\rangle-\theta)$,则一个神经元的功能函数正好是它所代表的球面上“球形领域”的特征函数.若取 $|w|=|x|$,则这个球形领域的中心就是 w,其半径由 θ 决定.这样,我们就能非常直观地进行神经网络的各种研究了.

2 覆盖算法

利用神经元的几何意义,下面给出一个简单的网络设计方法——领域覆盖算法.

设样本集 K 为 $K=r^t=(x^t,y^t),t=0,1,2,\ldots,p-1$. K 中不同的 y^t 只有 k 个,不妨设为 $y^0,y^1,\ldots,y^{k-1}$.令样本的输出为 y^t 的样本标号的集合为 $I(t)$,$I(t)=\{i\mid y^i=y^t\}$,其对应的输入集合记为 $P(t)$,$t=0,1,2,\ldots,k-1$,并将输入样本的上标按 $I(0),I(1),\ldots,I(k-1)$的顺序进行排序.

于是,如果我们能够取一批“球形领域”C_j^t,$t=0,1,\ldots,k-1$;$j=1,2,\ldots,k_t$,令 $C^t=\bigcup C_j^t$,使得 C^t 只覆盖 j 属于 $I(t)$的 x^j,而不覆盖 j 不属于 $I(t)$的 x^j,且 C^t 互不相交,这样的球形领域就能完成分类的要求.

对每个属于 $P(t)$的样本输入 x^I,均可按下面的方法构造覆盖 C_i^t,

固定 t,令 $P(t)=\{x^i\mid i\in I(t)\}$,对每个样本输入 $x^i\in P(t)$.

令
$$d^1(i)=\max_{j\notin I(t)}\{\langle x^i,x^j\rangle\},$$
$$d^2(i)=\min_{j\in I(t)}\{\langle x^i,x^j\rangle\mid\langle x^i,x^j\rangle d^1(i)\}.$$
$$d(i)=\frac{d^2(i)+d^1(i)}{2},$$
$$\theta_i=d(i),$$
$$w^i=x^i,\theta=(\theta_i),\quad i=1,2,\ldots,k_t,\tag{1}$$
得到覆盖 C_i^t 为 (x^i,θ_i),$i=1,2,\ldots,k_t$.其中 $\langle x,y\rangle$,表示 x,y 的内积.

定义 1. 令 $D_i(t)=\{x\mid\langle x,x^i\rangle>\theta_i\}$,称为以 x^i 为中心的邻域.以式(1)定义权和阈值的神经元记为 A_i^t 称为对应于 x^i 的神经元,其中 $i\in I(t)$.

定义 2. 令 $D(t)=\bigcup\limits_{i\in I(t)}D_i(t)$,称 $D(t)$为对应于 y^t 的领域组.

由 FP 算法[13]可知,当样本输入 $x^j\notin P(t)$时,$x^j\notin D(t)$,$x^j\in P(t)$,则 $x^j\in D(t)$.

覆盖算法.

(1) 对每个 y^t,求对应的领域组 $D(t)$,$t=0,1,\ldots,k-1$.

(2) 从 $D(t)$中选出 $P(t)$的子覆盖 $D'(t)$(可用贪心法求子覆盖),记 $D'(t)$中领域标号的集合为 $I'(t)$.

(3) 对 $D'(t)$中的每个邻域,作对应的神经元 A_i^t,$t=0,\ldots,k-1$;$i\in I'(t)$.

取网络隐层的神经元为 A_i^t,$i\in I'(t)$,$t=0,1,2,\ldots,k-1$.

(4) 网络的输出层:取 k 个元件(或门)B_t,$t=0,1,\ldots,k-1$.

(5) 网络连接:凡是 $i\in T'(t)$的神经元 A_i^t 的输出均输入 B_t.

所得到的网络 N 即为所求.

显然,一个覆盖中的球形领域对应于隐层中的一个隐层元.所以,如果我们每次能从 $D(t)$中选出最小覆盖 $D'(t)$,则在一定意义下,我们就给出了隐层元最少的网络结构.

由覆盖算法可知,对于任意一个样本(x^i,y^i),其输入 $x^i(i\in C_t)$必能被一个神经元 A_i^t 所覆盖,而不被其他神经元 A_j^s(s 不等于 t)覆盖.

3 基本命题的证明

有了上面的知识,就可以证明基本命题了.

Funabashi[16]和 Arai[17]证明了三层前向神经网络的万有性.我们还可以对网络作进一步要求:当输出是实数值时,第 3 层只用线性元件;当输出为{0,1}时,第 3 层只用或门.这样的网络仍然具有万有性.我们称第 3 层只用或门的网络为 O 型网络.

定义 3. 一个前向神经网络,如果它具有容纳所有"单点故障"的能力,则称此网络为"容错网络".

基本命题. 三层前向 O 型神经网络 N 为容错网络的充分必要条件是,其样本集中的任意一个样本输入必被隐层中两个或两个以上神经元所覆盖.

证明:(1) 充分性.若每个样本至少被两个隐层神经元所覆盖,现设有 1 个隐层神经元失效,则每个样本至少还被 1 个隐层神经元所覆盖.故其对应的隐层神经元 A_i^t 中至少还有 1 个神经元覆盖该样本输入,于是其对应的隐层神经元的输出为 1.这个输出连接到输出层的对应神经元 B_t,即神经元 B_t 的输入中至少有 1 个取值为 1.由于 B_t 是或门,故其对应的输出为 1.即对应于每个样本输入,其输出都是正确的.

(2) 必要性.设网络具有容纳所有单点故障的能力.用反证法,假设样本中有一个样本输入只被一个隐层神经元所覆盖,若这个神经元失效,则对这个样本进行输入.这时,它未被任何隐层元所覆盖,于是其对应的输出层的神经元 B_t 的各输入值均为 0,故其输出值为 0,即发生错误,矛盾.故每个样本输入至少被两个以上的隐层神经元所覆盖. □

有了基本命题,要验证一个三层前向神经网络是否为"容错网络",只要逐一检查所有样本是否均被两个隐层元所覆盖即可.

4 容错网络的构成

上面证明了"容错网络"的充要条件,下面我们讨论如何去构成这样的网络.

从上面的介绍可知,用覆盖算法可以很快解出满足所有样本的三层前向神经网络.在这个解的基础上我们再进行加工,使之满足"容错网络"的充要条件,这样,我们就可以得到"容错网络".下面给出一个简单的算法.

求"容错网络"的算法.

(1) 对给定的样本集,用覆盖算法求出一个对应的解,记为 N.

(2) 对 N 进行检查,将已被两个(或两个以上)隐层神经元覆盖的样本输入删去,所得的样本输入集记为 K'.

(3) 再用覆盖算法,对 K'求解,得到解 N',则 $N+N'$网络,就是一个"容错网络".在对 K'的样本输入求对应的覆盖时,其对应的公式(1)中的参数作如下的修改:

$$d^2(i)=\min_{j\in I(t)\cap K'}\{\langle x^i,x^j\rangle \mid \langle x,x^j\rangle d^1(i)\},$$

其中 $I(t)\cap K'=\{i\mid y^i=y^t,x^i\in K'\}$,其余不变.

例:给定样本集 $K=\{(0,0;0),(1,0;1),(0,1;1),(-1,0;1),(0,-1;1),(0,-1;1),(1,1;1),(1,-1;0),(0,1;10),(0,-1;1)\}$,求对应的三层前向容错网络.

解:先将各样本输入投影到半径为 $\sqrt{3}$ 的球面上,再求各样本输入的对应覆盖可得:

$A_1^0=((0,0,\sqrt{3}),(3+\sqrt{6})/2)$, $A_2^1=((1,0,\sqrt{2}),(3+\sqrt{6})/2)$,

$A_3^1=((0,1,\sqrt{2}),(3+\sqrt{6})/2)$, $A_4^1=((-1,0,\sqrt{2}),(3+\sqrt{6})/2)$,

$A_5^1=((0,-1,\sqrt{2}),(3+\sqrt{6})/2)$, $A_6^1=((1,1,1),(1+\sqrt{2}+\sqrt{3})/2)$,

$A_7^1=((1,-1,1),(1+\sqrt{2}+\sqrt{3})/2)$, $A_8^1=((-1,1,1),(1+\sqrt{2}+\sqrt{3})/2)$,

$A_9^1=((-1,-1,1),(1+\sqrt{2}+\sqrt{3})/2)$.

最后取 $A_1^0,A_6^1,A_7^1,A_8^1,A_9^1$ 这 5 个神经元(子覆盖)为隐层的神经元,输出层取 B_0,B_1(或门),得到网络 N.直接验证 N 不是容错网络,其中只有 4 个样本输入被两个隐层元覆盖,故再作 N',得到 $A_2^0=((0,0,\sqrt{3}),(3+\sqrt{6})/2)$,$A_{10}^1=((1,0,\sqrt{2}),(3+\sqrt{2})/2)$,$A_{11}^1=((0,1,\sqrt{2}),(3+\sqrt{2})/2)$,$A_{12}^1=((-1,0,\sqrt{2}),(3+\sqrt{2})/2)$,$A_{13}^1=((0,-1,\sqrt{2}),(3+\sqrt{2})/2)$.

最后得到,隐层共有 10 个神经元,将隐层元 A_*^0 (A_*^1) 的输出连到 $B_0(B_1)$ 的输入,记最后得到的网络为 N'',则 N''是容错网络.

5 结 论

本文利用文献[12,13]给出的 M-P 神经元的几何意义以及对应的覆盖算法,证明了"容错网络"的充要条件,并给出构成"容错网络"的具体算法,对"容错网络"的本质有了较彻底的了解,为分析神经网络的结构容错问题提供了一种新方法.

References:

[1] Hammadi, N.C., Ito, H. A learning algorithm for fault tolerant feedforward neural networks. IEICE Transactions on Information and Systems, 1997,E80-D(1):21~27.

[2] Minnix, J.I. Fault tolerance of backpropagation neural network trained on noise inputs. In: Proceedings of the International Joint Conference on Neural Networks. New York: IEEE Press, 1992. 847~852.

[3] Murray, A.F., Edwards, P.J. Enhanced MLP performance and fault tolerance resulting from synaptic weight noise during training. IEEE Transactions on Neural Networks, 1994,5(5):792~802.

[4] Emmerson, M.D., Damper, R.I. Determining and improving the fault-tolerance of multilayer perceptions in a pattern-recognition application. IEEE Transactions on Neural Networks, 1993,4(5):788~793.

[5] Phatak, D.S., Koren, I. Complete and partial fault tolerance of feedforward neural nets. IEEE Transactions on Neural Networks, 1995,6(2):446~456.

[6] Chiru, C.T., Mehrotra, K., Mohan, C.K., *et al.* Training techniques to obtain fault-tolerant neural networks. In: Proceedings of the International Symposium on Fault-Tolerant Computing. New York: IEEE Press, 1994. 360~369.

[7] Rumelhart, D., Hinton, G., Williams, R. Learning representations by backpropagating errors. Nature, 1986,323(9):318~362.

[8] Tan, Y., Nanya, T. Fault-Tolerant back-propagation model and its generation ability. In: Proceedings of the International Joint Conference on Neural Networks. New York: IEEE Press, 1993. 2516~2519.

[9] Neti, C., Schneider, M.H., Young, E.D. Maximally fault-tolerant neural networks. IEEE Transactions on Neural Networks, 1992,3(1):14~23.

[10] Deodhare, D., Vidyasagar, M., Keerthi, S.S. Synthesis of fault-tolerant feedforward neural networks using minimax optimization. IEEE Transactions on Neural Networks, 1998,9(5):891~900.

[11] Nijhuis, J., Hofflinger, B., Schaik, A., *et al*. Limits to fault-tolerance of a feedforward neural network with learning. In: Proceedings of the International Symposium on Fault-Tolerant Computing. New York: IEEE Press, 1990. 228~235.

[12] Zhang, L., Zhang, B. A geometrical representation of McCulloch-Pitts neural model and its applications. IEEE Transactions on Neural Networks, 1999,10(4):925~929.

[13] Zhang, L., Zhang, B. A Learning and synthesis algorithm of multilayered feedforward neural networks. Journal of Software, 1995,6(7):440~448 (in Chinese).

[14] Rujan, P., Marchand, M. A geometric approach to learning in neural networks. In: Proceedings of the International Joint Conference on Neural Networks'89. Washington, DC: IEEE Press, 1989. 105~110.

[15] Ramacher, U., Wesseling, M. A geometrical approach to neural network design. In: Proceedings of the International Joint Conference on Neural Networks'89. Washington, DC: IEEE Press, 1989. 147~154.

[16] Funahashi, K. On the approximate realization of continuous mappings by neural networks. Neural Networks, 1989,2(1):183~192.

[17] Arai, M. Mapping abilities of three-layer neural networks. In: Proceedings of the International Joint Conference on Neural Networks'89. Washington, DC: IEEE Press, 1989. 419~423.

附中文参考文献：

[13] 张铃,张钹.多层前馈网络的学习和综合算法.软件学报,1995,6(7):440~448.

On the Fault Tolerance Problem of the Feedforward Neural Networks*

ZHANG Ling

(*Institute of Artifical Intelligence, Anhui University, Hefei* 230039, *China*);
(*State Key Laboratory of Intelligent Technology and Systems, Tsinghua University, Beijing* 100084, *China*)
E-mail: zling@ahu.edu.cn
http://www.ahu.edu.cn

Abstract: Until recently, in the literatures of discussing the fault tolerance problem of neural networks, the tolerance about input noise is mainly concerned. The problem is generally transformed into that of optimization and solved by some well-known optimization approach. But only a few dealt with the tolerance due to the network structural failure, i.e., the structural fault tolerance. In the paper, the structural fault tolerance is analyzed by using the covering algorithms. The necessary and sufficient conditions of allowing of single node failure in a neural network and the algorithm for constructing such a network are given. These results reveal the essence of the structural fault tolerance capacity and show a new way for the analysis of the fault tolerance of neural networks.

Key words: neural network; fault tolerance; structural fault tolerance; single fault

* Received March 3, 2000; accepted June 26, 2000
Supported by the National Natural Science Foundation of China under Grant No. 69675011; the National Grand Fundamental Research 973 Program of China under Grant No. G1998030509