文章编号:1000-8152(2000)04-0537-05

全连接回归神经网络的稳定性分析*

李远清　韦　岗　刘永清
(华南理工大学电子与信息学院·广州,510640)

摘要:讨论全连接回归神经网络的稳定性,具体来说,就是分析网络输入和网络权值的扰动对网络输出的影响,得到了在上述扰动下,网络输出误差不累积的条件,即整个网络稳定的条件.

关键词:回归神经网络;扰动;误差;稳定
文献标识码:A

Stability of Completely Connected Recursive Neural Networks

LI Yuanqing, WEI Gang and LIU Yongqing
(College of Electronic and Information Engineering, South China University of Technology · Guanzhou, 510640, P.R.China)

Abstract: This paper discusses the stability of completely connected recursive neural networks, that is, analyzes the affection of output of the neural networks under which the errors don't accumulate, brought by disturbances of inputs and weights. The corresponding conditions are obtained, which are also the conditions of structural stability of the neural networks.

Key words: recursive neural networks; disturbance; error; stability

1 引言(Introduction)

全连接回归神经网络是一种通用的神经网络模型,在这种网络中,所有的神经元是相互连接的.但是,不同的神经元,在功能上可能不同.如某些神经元可能接受外部输入,某些神经元的输出作为整个网络的输出.因此,全连接回归神经网络虽在连接关系上看是没有层次结构的,但在功能上,却有层次结构.图 1 为一全连接回归神经网络示意图.

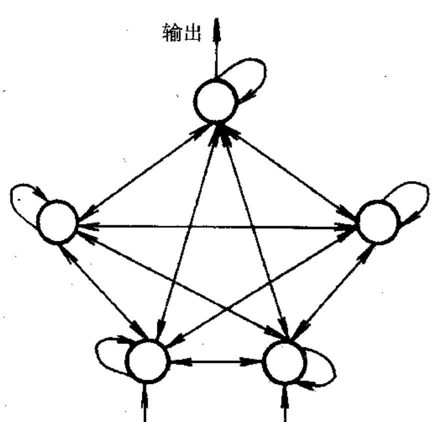

图 1　全连接回归神经网络模型

Fig.1 Completely connected recursive neural networks

直观上,可以想象全连接回归神经网络的学习是相当复杂的.全连接回归神经网络包含了几乎所有其它类型的神经网络.如,当网络的神经元间连接权值具有对称性时,即为 Hopfield 网络,去掉反馈支路即为多层前馈网络.模型越是通用,一般来说,其描述也越复杂.因此,在实际实用中,应尽量简化,以最小的复杂性实现要求的功能.但是,有一些问题,可能是只有用全连接回归神经网络才能解决.例如,对于一般的离散信息源,用 Markov 链就可以描述了,但有许多复杂信息源不能用 Markov 链描述.所以,研究全连接回归神经网络是必要的.全连接回归神经网络的结构已足够复杂,只要参数选择适当,是可以解决复杂问题的.

下面介绍全连接回归神经网络的学习算法.设网络具有 n 个神经元,m 个外部输入.$y(t)$ 表示网络在 t 时刻 n 个神经元的输出,$x(t)$ 为 t 时刻 m 个输入.我们将 $x(t)$,$y(t)$ 联合组成 $z(t)$,即

$$z(t)=[z_1(t),\cdots,z_{n+m}(t)]=[y_1(t),\cdots,y_n(t),x_1(t),\cdots,x_m(t)].$$

设 W 表示网络的权值矩阵,其分量表示神经元之间的连接权值,或者是外部输入到神经元的连接权值.根据上述假设,有:

* 基金项目:广东省自然科学基金(990584)资助项目.
收稿日期:1998-03-30;收修改稿日期:1999-09-24.

$$W=\begin{bmatrix} w_{11} & \cdots & w_{1(n+m)} \\ \square & \square & \square \\ w_{n1} & \cdots & w_{n(n+m)} \end{bmatrix} \tag{1.1}$$

为 $n\times(m+n)$ 矩阵. 设神经元是由内部求和器及非线性变换构成,非线性变换为 $f(\cdot)$.

神经元 k 在时刻 t 的内部求和器输出 $S_k(t)$ 可写为:

$$S_k(t)=\sum_{l=1}^{n+m} w_{kl}z_l(t),\quad k=1,\cdots,n, \tag{1.2}$$

则此神经元在下一时刻的输出为:

$$y_k(t+1)=f(S_k(t)). \tag{1.3}$$

设 $T(t)$ 表示 t 时刻,输出必须与要求的值相匹配的神经元的下标 k 构成的集合,$d_k(t)$ 为第 k 个神经元期望的输出,定义时变输出误差为:

$$e_k(t)=\begin{cases} d_k(t)-y_k(t), & k\in T(t), \\ 0, & k\notin T(t), \end{cases} \tag{1.4}$$

则 t 时刻网络总的输出误差可记为:

$$J(t)=\sum_{k=1}^{n}(e_k(t))^2. \tag{1.5}$$

假定我们考虑的时间区间为 $[t_0,t_1]$,则在此区间内,网络总输出误差为:

$$J_\Sigma(t_0,t_1)=\sum_{t=t_0+1}^{t_1} J(t), \tag{1.6}$$

网络学习的目的是 $J_\Sigma(t_0,t_1)$ 最小,可通过梯度算法来完成,即沿 $J_\Sigma(t_0,t_1)$ 关于 W 的负梯度方向,调整权值矢量 W. 由于我们考虑的输出误差是在某个时间区间上的误差,因此,具体计算时,可首先计算不同时刻的梯度 $\nabla J(t)$,然后求和,即权值调整为:

$$\Delta w_{ij}=\sum_{t=t_0+1}^{t_1}\Delta w_{ij}(t), \tag{1.7}$$

其中

$$\Delta w_{ij}(t)=-\mu\frac{\partial J(t)}{\partial w_{ij}}, \tag{1.8}$$

式中,μ 为正常数,称为学习律.

至此,我们给出了在时间间隔 $[t_0,t_1]$ 内有监督学习算法的过程. 这一算法是每隔一段时间,对权值进行学习. 在这个时间间隔内,权值是不变的.

上述算法有收敛性问题,一般来讲,如果目标函数是凸函数,则梯度算法是收敛的. 算法收敛性不是本文研究的内容. 本文要研究的则是有关网络的另一重要问题 稳定性问题.

定义 1.1 1) 设 W 为通过学习得到的最优权值,此时,如果输入为理想输入 $x(t)$,则输出为最优输出 $y(t)$. 当输入发生有界扰动 $\Delta x(t)$ 时,设其界为 ε_1,输出可设为 $y(t)+\Delta y(t)$,若存在一正常数 M_1,使得 $\|\Delta y(t)\|\leqslant M_1\varepsilon_1$,则称网络对输入扰动是稳定的.

2) 设权值 W 有扰动 ΔW,且 $\|\Delta W\|\leqslant\varepsilon_2$,网络输入为理想输入 $x(t)$,输出应记为 $y(t)+\Delta y(t)$,若存在一正常数 M_2,使得 $\|\Delta y(t)\|\leqslant M_2\varepsilon_2$,则称网络对权重扰动是稳定的.

上述定义中第一个问题的重要性是不言而喻的,第二个问题也是非常重要的,因为我们常常找到的往往不是最优权值,而是其近似值,可以看作是最优权值发生了扰动.

本文总假定 f 满足:

$$|\dot f|\leqslant 1, \tag{1.9}$$

如 $f(s)=\dfrac{1}{1+\exp^{-s}}$ 满足(1.9).

关于本文中用到的向量和矩阵范数,以 $y=(y_1,\cdots,y_n)^{\mathrm T}$,$B=(b_{ij})_{n\times m}$ 为例,意义如下:

$$\begin{cases} \|y\| = \max\{|y_i|,\ i=1,\cdots,n\}, \\ \|B\| = \max\{\sum_{j=1}^{m}|b_{ij}|,\ i=1,\cdots,n\}. \end{cases} \tag{1.10}$$

另外,本文将用到矩阵理论中的如下结果:

引理 1.1[4] 如果矩阵 A 的谱集(特征值集合)在幂级数 $\sum_{k=0}^{\infty}c_kx^k$ 的收敛圆内,则矩阵幂级数 $\sum_{k=0}^{\infty}c_kA^k$ 绝对收敛 如果矩阵 A 有一个特征值在幂级数 $\sum_{k=0}^{\infty}c_kx^k$ 的收敛圆外,则矩阵幂级数 $\sum_{k=0}^{\infty}c_kA^k$ 发散.

注意到 $\sum_{k=0}^{\infty}x^k$ 的收敛圆为单位圆,下面将会用到这一点.

2 输入存在扰动时网络的稳定性(Stability of the neural netwotks with disturbance of inputs)

设 W 为网络的权值矩阵,输入为 $x(t)+\Delta x(t)$,$\Delta x(t)$ 为扰动,输出为 $y(t)+\Delta y(t)$,$\Delta y(t)$ 为由 $\Delta x(t)$ 引起的误差.

设输入扰动满足:

$$\|\Delta x(t)\|\leqslant\varepsilon, \tag{2.1}$$

此处,$\varepsilon>0$ 为一常数.

我们现在寻求一种条件,使得存在一正常数

M_1，$\|\Delta y(t)\| \leqslant M_1\varepsilon$ 成立.

在上述情形下

$$S_k(t) = \sum_{l=1}^{n} w_{kl}(y_l(t) + \Delta y_l(t)) + \sum_{l=n+1}^{n+m} w_{kl}(x_l(t) + \Delta x_l(t)), \tag{2.2}$$

则此神经元在下一时刻的输出为：

$$\begin{aligned} y_k(t+1) + \Delta y_k(t+1) &= f(S_k(t)) = \\ & f[\sum_{l=1}^{n} w_{kl}(y_l(t) + \Delta y_l(t)) + \sum_{l=n+1}^{n+m} w_{kl}(x_l(t) + \Delta x_l(t))]. \end{aligned} \tag{2.3}$$

由 Tailor 公式或中值定理

$$\begin{aligned} \Delta y_k(t+1) = & f[\sum_{l=1}^{n} w_{kl}(y_l(t) + \Delta y_l(t)) + \sum_{l=n+1}^{n+m} w_{kl}(x_l(t) + \Delta x_l(t))] - y_k(t+1) = \\ & f[\sum_{l=1}^{n} w_{kl}(y_l(t) + \Delta y_l(t)) + \sum_{l=n+1}^{n+m} w_{kl}(x_l(t) + \Delta x_l(t))] - \\ & f[\sum_{l=1}^{n} w_{kl}y_l(t) + \sum_{l=n+1}^{n+m} w_{kl}x_l(t)] = \\ & \dot{f}(\theta_k(t))[\sum_{l=1}^{n} w_{kl}\Delta y_l(t) + \sum_{l=n+1}^{n+m} w_{kl}\Delta x_l(t)]. \end{aligned} \tag{2.4}$$

其中，$\theta_k(t)$ 介于 $\sum_{l=1}^{n} w_{kl}y_l(t) + \sum_{l=n+1}^{n+m} w_{kl}x_l(t)$ 与 $\sum_{l=1}^{n} w_{kl}y_l(t) + \sum_{l=n+1}^{n+m} w_{kl}x_l(t) + \sum_{l=1}^{n} w_{kl}\Delta y_l(t) + \sum_{l=n+1}^{n+m} w_{kl}\Delta x_l(t)$ 之间.

故

$$\Delta y(t+1) = \mathrm{diag}[\dot{f}(\theta_1(t)), \cdots, \dot{f}(\theta_n(t))] \cdot [W_1\Delta y(t) + W_2\Delta x(t)], \tag{2.5}$$

其中

$$W_1 = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \square & \square & \square \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}, \quad W_2 = \begin{bmatrix} w_{1(n+1)} & \cdots & w_{1(n+m)} \\ \square & \square & \square \\ w_{n(n+1)} & \cdots & w_{n(n+m)} \end{bmatrix}. \tag{2.6}$$

由(1.9)

$$\|\Delta y(t+1)\| \leqslant \|W_1\| \|\Delta y(t)\| + \|W_2\| \|\Delta x(t)\|. \tag{2.7}$$

由(2.7)及 $\Delta y(0) = 0$，有：

$$\|\Delta y(1)\| \leqslant \|W_2\|\varepsilon, \tag{2.8}$$

$$\begin{aligned} \|\Delta y(2)\| &\leqslant \|W_1\| \|\Delta y(1)\| + \|W_2\| \|\Delta x(1)\| \leqslant \\ & \|W_1\| \|W_2\|\varepsilon + \|W_2\|\varepsilon, \end{aligned} \tag{2.9}$$

$$\begin{aligned} \|\Delta y(3)\| &\leqslant \|W_1\| \|\Delta y(2)\| + \|W_2\| \|\Delta x(2)\| \leqslant \\ & \|W_1\|(\|W_1\| \|W_2\|\varepsilon + \|W_2\|\varepsilon) + \\ & \|W_2\|\varepsilon = (\|W_1\|^2 + \|W_1\| + 1)\|W_2\|\varepsilon. \end{aligned} \tag{2.10}$$

设

$$\|\Delta y(t)\| \leqslant (\|W_1\|^{t-1} + \|W_1\|^{t-2} + \cdots + \|W_1\| + 1)\|W_2\|\varepsilon, \tag{2.11}$$

则由(2.7)，不难有：

$$|\Delta y(t+1)| \leqslant (\|W_1\|^{t} + \|W_1\|^{t-1} + \cdots + \|W_1\| + 1)\|W_2\|\varepsilon, \tag{2.12}$$

由数学归纳法可知，$\forall t$，(2.11) 成立.

故若级数 $\sum_{i=0}^{\infty} \|W_1\|^i$ 收敛，即矩阵级数 $\sum_{i=0}^{\infty} W_1^i$ 绝对收敛，则有：

$$\|\Delta y(t)\| \leqslant M_1\varepsilon \tag{2.13}$$

对 $\forall t$ 成立，其中

$$M_1 = (\sum_{i=0}^{\infty} \|W_1\|^i)\|W_2\|,$$

从而使得输出误差不累积.

由引理 1.1，我们有：

定理 2.1　若矩阵 W_1 的特征值均在单位圆盘内，则网络对输入扰动是稳定的.

3　权值存在扰动时网络的稳定性(Stability of the neural networks with disturbance of weights)

设权值 W 有一扰动 ΔW，网络输入为理想输入 $x(t)$，输出为 $y(t) + \Delta y(t)$，$\Delta y(t)$ 为由 ΔW 引起的误差. 我们现在寻求条件，使得误差不累积.

在上述情形下，

$$S_k(t) = \sum_{l=1}^{n} (w_{kl} + \Delta w_{kl})(y_l(t) + \Delta y_l(t)) + \sum_{l=n+1}^{n+m} (w_{kl} + \Delta w_{kl}) x_l(t), \tag{3.1}$$

则此神经元在下一时刻的输出为：

$$y_k(t+1)+\Delta y_k(t+1)=$$
$$f[\sum_{l=1}^{n}(w_{kl}+\Delta w_{kl})(y_l(t)+\Delta y_l(t))+$$
$$\sum_{l=n+1}^{n+m}(w_{kl}+\Delta w_{kl})x_l(t)]. \tag{3.2}$$

由 Tailor 公式

$$\Delta y_k(t+1)=$$
$$f[\sum_{l=1}^{n}(w_{kl}+\Delta w_{kl})(y_l(t)+\Delta y_l(t))+$$
$$\sum_{l=n+1}^{n+m}(w_{kl}+\Delta w_{kl})x_l(t)]-y_t(t+1)=$$
$$f[\sum_{l=1}^{n}(w_{kl}+\Delta w_{kl})(y_l(t)+\Delta y_l(t))+$$
$$\sum_{l=n+1}^{n+m}(w_{kl}+\Delta w_{kl})x_l(t)]-$$
$$f[\sum_{l=1}^{n}w_{kl}y_l(t)+\sum_{l=n+1}^{n+m}w_{kl}x_l(t)]=$$
$$\dot f(\bar\theta_k(t))[\sum_{l=1}^{n}(\Delta w_{kl}y_l(t)+$$
$$w_{kl}\Delta y_l(t)+\Delta w_{kl}\Delta y_l(t)+$$
$$\sum_{l=n+1}^{n+m}\Delta w_{kl}x_l(t)], \tag{3.3}$$

其中，$\bar\theta_k(t)$ 介于 $\sum_{l=1}^{n}w_{kl}y_l(t)+\sum_{l=n+1}^{n+m}w_{kl}x_l(t)$ 与 $\sum_{l=1}^{n}w_{kl}y_l(t)+\sum_{l=n+1}^{n+m}w_{kl}x_l(t)+\sum_{l=1}^{n}(\Delta w_{kl}y_l(t)+w_{kl}\Delta y_l(t)+\Delta w_{kl}\Delta y_l(t)+\sum_{l=n+1}^{n+m}\Delta w_{kl}x_l(t))$ 之间.

故

$$\begin{aligned}\Delta y(t+1)=&\operatorname{diag}[\dot f(\bar\theta_1(t)),\cdots,\dot f(\bar\theta_n(t))][\Delta W_1y(t)+\\&W_1\Delta y(t)+\Delta W_1\Delta y(t)+\Delta W_2\Delta x(t)]=\\&\operatorname{diag}[\dot f(\bar\theta_1(t)),\cdots,\dot f(\bar\theta_n(t))][(W_1+\\&\Delta W_1)\Delta y(t)+\Delta W_1y(t)+\Delta W_2x(t)],\end{aligned} \tag{3.4}$$

其中，W_1,W_2 的意义同上节，而

$$\Delta W_1=\begin{bmatrix}\Delta w_{11} & \cdots & \Delta w_{1n}\\ \square & \square & \square\\ \Delta w_{n1} & \cdots & \Delta w_{nn}\end{bmatrix}, \tag{3.5}$$
$$\Delta W_2=\begin{bmatrix}\Delta w_{1(n+1)} & \cdots & \Delta w_{1(n+m)}\\ \square & \square & \square\\ \Delta w_{n(n+1)} & \cdots & \Delta w_{n(n+m)}\end{bmatrix}.$$

设理想输入 $x(t)$，理想输出 $y(t)$ 均有界，即满足

$$\|x(t)\|\leqslant a,\quad \|y(t)\|\leqslant a, \tag{3.6}$$

此处，$a>0$ 为一常数.

由(1.9)，(3.4)，(3.6)，有：

$$\begin{aligned}&\|\Delta y(t+1)\|\leqslant\\&\|W_1+\Delta W_1\|\,\|\Delta y(t)\|+\\&\|\Delta W_1\|\,\|y(t)\|+\|\Delta W_2\|\,\|x(t)\|\leqslant\\&\|W_1+\Delta W_1\|\,\|\Delta y(t)\|+\\&(\|\Delta W_1\|+\|\Delta W_2\|)a.\end{aligned} \tag{3.7}$$

不妨设

$$y(0)=0,\quad \Delta y(0)=0,$$

由(3.7)有：

$$\|\Delta y(1)\|\leqslant(\|\Delta W_1\|+\|\Delta W_2\|)a, \tag{3.8}$$

$$\begin{aligned}\|\Delta y(2)\|\leqslant&\|W_1+\Delta W_1\|\,\|\Delta y(1)\|+\\&(\|\Delta W_1\|+\|\Delta W_2\|)a\leqslant\\&\|W_1+\Delta W_1\|(\|\Delta W_1\|+\\&\|\Delta W_2\|)a+(\|\Delta W_1\|+\|\Delta W_2\|)a.\end{aligned} \tag{3.9}$$

设

$$\begin{aligned}&\|\Delta y(t)\|\leqslant\\&[(\|W_1+\Delta W_1\|)^{t-1}+\cdots+\\&\|W_1+\Delta W_1\|+1](\|\Delta W_1\|+\|\Delta W_2\|)a,\end{aligned} \tag{3.10}$$

则由(3.7)

$$\begin{aligned}&\|\Delta y(t+1)\|\leqslant\\&(\|W_1+\Delta W_1\|)[(\|W_1+\Delta W_1\|)^{t-1}+\\&\cdots+\|W_1+\Delta W_1\|+1]\cdot\\&(\|\Delta W_1\|+\|\Delta W_2\|)a+\\&(\|\Delta W_1\|+\|\Delta W_2\|)a\leqslant\\&[(\|W_1+\Delta W_1\|)^t+\cdots+\\&\|W_1+\Delta W_1\|+1](\|\Delta W_1\|+\|\Delta W_2\|)a.\end{aligned} \tag{3.11}$$

由数学归纳法可知，$\forall t$，(3.10) 成立. 故若级数 $\sum_{i=0}^{\infty}(\|W_1+\Delta W_1\|)^i$ 收敛，即矩阵级数 $\sum_{i=0}^{\infty}(W_1+\Delta W_1)^i$ 绝对收敛，则不难发现，存在一正常数

$$M_2=\sum_{i=0}^{\infty}(\|W_1+\Delta W_1\|)^ia,$$

使得 $\forall t$，

$$\|\Delta y(t)\|\leqslant M_2(\|\Delta W_1\|+\|\Delta W_2\|). \tag{3.12}$$

从而使得输出误差不累积，于是由引理 1.1，我们有：

定理 3.1　在理想输入和理想输出有界的前提下,若权值 W 的有界扰动 ΔW,使得矩阵($W_1+\Delta W_1$)的特征值均在单位圆盘内,则网络对权重扰动是稳定的.

将上节与本节的推导结合起来,不难有:

定理 3.2　在理想输入和理想输出有界的前提下,若权值 W 的有界扰动 ΔW,使得矩阵 $W_1+\Delta W_1$ 的特征值均在单位圆盘内,则网络对输入与权重的同时扰动是稳定的.

4　结束语(Concluding remarks)

若网络的权值矩阵 W 可使矩阵 W_1 的特征值均在单位圆盘内,则只要扰动 ΔW 充分小,仍可保证($W_1+\Delta W_1$)的特征值均在单位圆盘内,从而输入扰动与权值扰动使网络输出产生的误差不累积,网络对输入与权重的同时扰动是稳定的.可见,若网络的权值矩阵 W 使得矩阵 W_1 的特征值不全在单位圆盘内,则输入扰动与权值扰动使网络输出产生的误差会累积,此时网络是不稳定的.

本文只研究了网络的稳定性问题,还有很多问题有待进一步研究,如算法的收敛性问题;在什么条件下,网络存在权值矩阵 W,其子矩阵 W_1 的特征值均在单位圆盘内;以及如何通过有限的样本训练,得到该权值矩阵 W 等.

参考文献(References)

[1]　韦岗,贺前华.神经网络模型,学习及应用[M].北京:电子工业出版社,1994

[2]　徐秉铮,张百灵,韦岗.神经网络理论与应用[M].广州:华南理工大学出版社,1994

[3]　张立明.人工神经网络模型及应用[M].上海:复旦大学出版社,1992

[4]　王耕禄,史荣昌.矩阵理论[M].北京:国防工业出版社,1998

本文作者简介

李远清　1966 年生.1988 年毕业于武汉大学数学系,获学士学位,1994 年毕业于华南师范大学数学系,获硕士学位,1997 年毕业于华南理工大学自动控制系,获博士学位.现为华南理工大学电子与信息学院自动控制系副教授.主要研究方向:滞后广义系统,神经网络,信号盲处理.

韦　岗　1963 年生.华南理工大学电子与信息学院副院长,教授,博士生导师.主要研究方向:信号处理,模式识别,神经网络等.

刘永清　见本刊 2000 年第 1 期第 8 页.