文章编号:1000-582x(2000)02-0102-04

(26)
102-104,119

线性方程神经网络——Minimax准则的光滑法方法*

0241.6

叶仲泉
(重庆大学 理学院,重庆 400044)

TP18

摘　要:将神经网络用于线性系统方程的求解问题,但所用的准则函数不是通常的最小二乘函数,而是 l_∞(或 Chebyshev) 范数准则函数,即 $E_\infty(x)=\max\limits_{x\in R^n}\{|r_i(x)|\}$ 或 $E_\infty(x)=\max\limits_{x\in R^n}\left\{\frac{1}{2}r_i^2(x)\right\}$,先将 $E_\infty(x)$ 光滑化,再利用神经网络来求解无约束光滑优化问题。并讨论了网络的收敛性和稳定性。

关键词:线性方程;收敛性 / 神经网络;l_∞范数准则函数
中图分类号:TP 202.1　　　　**文献标识码**:A

Minimax准则.
光滑法

求解线性方程是科学技术中广泛遇到的基本问题之一。信号处理、机器人、自动控制、系统理论和统计学中都要经常遇到线性方程组 $Ax\approx b$ 的估计问题。在许多应用中在线或实时求解线性方程是所希望的甚至是必须的。数字计算机常不能满足希望的计算时间,或代价太高。解决这个问题的有效方法是应用人工神经网络(可被认为依赖于大大简化了的神经元模型的模拟计算机)。近年来的神经网络热潮已导致了 VLSI 技术方面新的理论结果和进步,并且使制造具有高度复杂的微电子网络成为可能。

1　基本公式和问题

考察线性参数估计模型

$$Ax=b+r=b_{true} \tag{1}$$

其中 $A=[a_{ij}]\in R^{m\times n}(m\geqslant n)$ 为模型矩阵,$b\in R^m$ 为观察或观测向量,$r\in R^m$ 为未知测量误差向量,b_{true} 为真值向量。我们希望能实时求得系统的精确解或近似解。

利用人工神经网络解决上述问题的关键步骤是构造适当的计算能量函数(Lyapunov)函数 $E(x)$ 使最低能量状态对应于期望的解 x^*,能量函数的导数使我们能将极小化问题转化为一组微分方程或差分方程。

与式(1)相对应的优化问题可叙述为:求向量 $x^*\in R^n$,使如下能量函数达到极小:

$$E(x)=\sum_{i=1}^{m}\sigma_i(r_i(x)) \tag{2}$$

其中
$$r_i(x)=a_i^Tx-b_i=\sum_{j=1}^{n}a_{ij}x_j-b_i \tag{3}$$

$\sigma_i(r_i)$ 表示适当的凸函数。令:

$$r(x)=[r_1(x)\cdots r_m(x)]^T=Ax-b \tag{4}$$

在实践中,$\sigma_i(r_i)$ 的如下选择特别重要:

1) 取 $\sigma_i(r_i)=r_i^2/2$,即通常的最小二乘问题

$$E_2(x)=\frac{1}{2}\sum_{i=1}^{m}r_i^2(x)=\frac{1}{2}r^T(x)r(x)=\frac{1}{2}(Ax-b)^T(Ax-b)=\frac{1}{2}\|Ax-b\|_2^2 \tag{5}$$

2) $\sigma_i(r_i)=|r_i|$,这时

$$E_1(x)=\sum_{i=1}^{m}|r_i(x)| \tag{6}$$

除了上述准则函数外,l_∞(或 Chebyshev) 范数准则也经常用到,这时问题就成为如下的 minimax 问题了:

$$\min_{x\in R^n}\max_{1\leqslant i\leqslant m}\{|r_i(x)|\} \tag{7}$$

或修正的 minimax 问题:

$$\min_{x\in R^n}\max_{1\leqslant i\leqslant m}\left\{\frac{1}{2}r_i^2(x)\right\} \tag{8}$$

准则的适当选择不仅依赖于具体的应用而且强

* 收稿日期:1999-05-17
作者简介:叶仲泉(1961-),男,四川自贡人,重庆大学副教授,博士,主要从事控制论、非线性分析、人工神经网络领域研究。

烈依赖于测量 b 的误差分布[1-4]。标准的最小二乘准则对噪声为 Gauss 分布时是最优的。但由于误差的不同来源，如仪器误差、模型误差、抽样误差、人为误差等，假定观测或观察误差具有 Gauss 分布往往是不现实的。

在许多应用中，minimax（Chebyshev 范数，l_∞ 范数）准则和最小绝对值（l_1— 范数）准则往往较最小二乘（l_2— 范数）准则更实用。从统计学的观点看，l_∞ 和 l_1 范数准则对某些误差分布是最优的，例如，如误差为一致均匀分布，则从极大似然法的观点看，用 l_∞ 范数准则将是最合适的。

但是，比起最小二乘准则，minimax 和最小绝对值准则仍很少用于参数估计问题，这主要是因为目标函数的不可微性，因而导致分析和数值问题。且在实际实现时也会产生一些问题。首先，$E_1(x)$ 和 $E_\infty(x) = \max\limits_{1\le i\le m}\{|r_i(x)|\}$ 有不连续的一阶偏导数，导数的不连续性常导致一些异常的结果（如错误输出，锯齿现象）；其次，符号函数的精确实现也相当困难；类似的，直接实现激活函数：

$$g[r(x)_i] = \begin{cases} \mathrm{sign}[r_i(x)], & |r_i(x)| = \max\limits_{1\le k\le m}\{|r_k(x)|\} \\ 0, & \text{otherwise} \end{cases}$$

也有困难且不现实。

文献[2] 讨论了线性方程神经网络的 minimax 和最小绝对值准则问题，将 minimax 问题：$\min\limits_{x\in R^n}\max\limits_{1\le i\le m}\{|r_i(x)|\}$ 转化为等价的约束优化问题：

$$\begin{aligned} &\textit{minimize}\ \varepsilon \\ &\text{subject to } |r_i(x)| \le \varepsilon, \varepsilon \ge 0 \end{aligned} \tag{9}$$

然后利用罚函数方法建立了一个目标函数。尽管得到了一个可微的优化问题，但却以解一个带约束的优化问题为代价。

本文利用文献[5] 提出的方法，先将 minimax 问题(8) 转化为无约束可微优化问题，然后再利用神经网络来研究该无约束优化问题。

2　minimax 问题的光滑法方法

minimax 问题(7) 或(8) 是不可微优化问题，因而不能利用标准的无约束优化算法来求解，在用神经网络求解时，也会产生一些问题。

考察如下的无约束 minimax 问题(P)：

$$\min_{x\in R^n} f(x) \equiv \max_{1\le i\le m}|f_i(x)| \tag{10}$$

其中

$$f(x) \equiv \max_{1\le i\le m}\{f_i(x)\}$$

显然，若取 $f_i(x) = \frac{1}{2}r_i^2(x)$ 则 minimax 问题(10) 即为(8)，也就是说(8) 只是(10) 的一种特殊情形。

minimax 问题(10) 是一个特别典型的不可微优化问题，文献[5] 提出了一个光滑法方法，其思路是力图保持问题的无约束特点，解决问题的途径是找一个光滑函数，以便计算时可以用它代替不可微的目标函数 $f(x) \equiv \max\limits_{1\le i\le m}\{f_i(x)\}$，minimax 问题(10) 的 Lagrange 函数为：

$$L(x,\lambda) = \sum_{i=1}^{m}\lambda f_i(x) \tag{11}$$

其中 Lagrange 乘子 λ 应满足

$$\lambda \in W \equiv \left\{\lambda \,\middle|\, \lambda_i \ge 0, \sum_{i=1}^{m}\lambda_i = 1\right\} \tag{12}$$

Lagrange 乘子 λ_i 可以看作分量函数 $f_i(x)$ 等于极大值函数 $f(x)$ 的概率，这样，Lagrange 函数 $L(x,\lambda)$ 可视为各分量函数的“均值”。在完全没有任何信息的情况下、这一概率分配应满足下面“最大熵”问题：

$$\max_{\lambda\in W} H(\lambda) \equiv -\sum_{i=1}^{m}\lambda_i\ln\lambda_i \tag{13}$$

$H(\lambda)$ 代表 Shannon 熵。

考虑如下优化问题：

$$\max_{\lambda\in W} L_p(x,\lambda) \equiv L(x,\lambda) + H(\lambda)/p \tag{14}$$

其中 p 为一个加权系数。$\forall\lambda\in W$，$H(\lambda)$ 总取一个非负的有限值，故当参数 $p\to+\infty$ 时，目标函数 $L_p(x,\lambda)$ 中的第二项将消失，这时问题(14) 退化为问题：

$$\max_{\lambda\in W} L(x,\lambda) \equiv \sum_{i=1}^{m}\lambda f_i(x) \tag{15}$$

当 p 取有限值时，问题(14) 仅有唯一解：

$$\lambda_i^*(x,p) = \exp[pf_i(x)]/\sum_{i=1}^{m}\exp[pf_i(x)],\quad 1\le i\le m \tag{16}$$

当 p 取有限值时，所有的 $\lambda_i^*(x,p) > 0, 1\le i\le m$，这意味着 λ^* 位于单纯形 W 的内部。所以(14) 中的熵函数 $H(\lambda)$ 实际起着对数障碍函数的作用、$H(\lambda)$ 迫使解点远离边界。

将(16) 代入 $L_p(x,\lambda)$，消去乘子 λ，则可得到一个仅依赖于变量 x 与 p 的函数。

$$F_p(x) = L_p(x,\lambda^*(x,p)) = \frac{1}{p}\ln\left\{\sum_{i=1}^{m}\exp[pf_i(x)]\right\} \tag{17}$$

这就是要求的“凝聚”函数。$F_p(x)$ 具有如下重要性质：

定理 1([5])　对任何 $\lambda\in W$，$F_p(x)$ 满足下列不

等式关系：

$$F_p(\boldsymbol{x}) \geq L_p(\boldsymbol{x}, \boldsymbol{\lambda}) \tag{18}$$

当且仅当 λ 取(16) 式所给的值时，上述不等式才成为等式。

定理 2([5]) 对任何 $\boldsymbol{x} \in R^n$，函数 $F_p(\boldsymbol{x})$ 都随着参数 p 的增加而单调减少，且当 $p \to +\infty$ 时，以 $f(\boldsymbol{x})$ 为极限，即

$$F_s(\boldsymbol{x}) \leqslant F_r(\boldsymbol{x}), \quad r \leqslant s \tag{19}$$

$$\lim_{p \to \infty} F_p(\boldsymbol{x}) = f(\boldsymbol{x}) \tag{20}$$

定理 3([5])]对任何 $\boldsymbol{x} \in R^n$，$F_p(\boldsymbol{x})$ 与 $f(\boldsymbol{x})$ 间均有下列不等式关系：

$$0 \leqslant F_p(\boldsymbol{x}) - f(\boldsymbol{x}) \leqslant \ln(m)/p$$

式中 m 代表分量函数的个数。

上述几个定理充分显示了 $F_p(\boldsymbol{x})$ 的优良性态，因此可以用它代替目标函数 $f(\boldsymbol{x}) = \max\limits_{1 \leqslant i \leqslant m} f_i(\boldsymbol{x})$，这样就将问题(13) 转化成可微的无约束优化问题：

$$\min_{x \in R^n} F_p(\boldsymbol{x}) \equiv \frac{1}{p} \ln\Big\{\sum_{i=1}^{m} \exp[p f_i(\boldsymbol{x})]\Big\} \tag{21}$$

的求解，当 p 取得足够大，一次无约束优化问题即能得到很高精度的解。

3 神经网络模型及收敛性分析

根据上节的讨论，我们可以利用光滑函数：

$$F_p(\boldsymbol{x}) \equiv \frac{1}{p} \ln\Big\{\sum_{i=1}^{m} \exp[p f_i(\boldsymbol{x})]\Big\}$$

其中 $f_i(\boldsymbol{x} = \frac{1}{2} r_i^2(\boldsymbol{x})$

来代替非光滑函数

$$f(x) = \max_{1 \leqslant i \leqslant m} \frac{1}{2} r_i^2(\boldsymbol{x})$$

当 p 足够大时，$F_p(x) - f(x)$ 充分小，于是 minimax 优化问题：

$$\min_{x \in R^n} \max_{1 \leqslant i \leqslant m} \frac{1}{2} r_i^2(x) \tag{22}$$

可以近似化为极小化问题

$$\min F_p(x) \equiv \frac{1}{p} \ln\Big\{\sum_{i=1}^{m} \exp[p f_i(x)]\Big\} = \frac{1}{p} \ln\Big\{\sum_{i=1}^{m} \exp\Big[\frac{1}{2} p r_i^2(x)\Big]\Big\} \tag{23}$$

我们建立如下的微分方程控制的神经网络来研究极小化问题：

$$\left.\begin{aligned} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} &= -\nabla F_p(\boldsymbol{x}) \\ x(0) &= x^{(0)} \end{aligned}\right\} \tag{24}$$

其中

$$\nabla F_p(\boldsymbol{x}) = \frac{\partial F_p(\boldsymbol{x})}{\partial x} = \frac{\sum\limits_{i=1}^{m} \exp\Big[\frac{1}{2} p r_i^2(\boldsymbol{x})\Big] r_i(\boldsymbol{x}) a_i}{\sum\limits_{i=1}^{m} \exp\Big[\frac{1}{2} p r_i^2(\boldsymbol{x})\Big]} \tag{25}$$

a_i 为矩阵 A_{mn} 的第 i 行向量的转置。将(24) 写成分量的形式即为

$$\left.\begin{aligned} \frac{\mathrm{d}x_j}{\mathrm{d}t} &= \frac{-\sum\limits_{i=1}^{m} \exp\Big[\frac{1}{2} p r_i^2(\boldsymbol{x})\Big] r_i(\boldsymbol{x}) \alpha_{ij}}{\sum\limits_{i=1}^{m} \exp\Big[\frac{1}{2} p r_i^2(\boldsymbol{x})\Big]} \\ x_j(0) &= x_j^{(0)} \qquad j = 1, 2, \cdots, n \end{aligned}\right\} \tag{26}$$

动力系统(25) 或(26) 即为解无约束优化问题(23) 的(连续) 最速下降法，其基本思想是计算从初始点 $\boldsymbol{x}^{(0)}$ 出发的轨道 $\boldsymbol{x}(t)$，使其极限点为 $F_p(\boldsymbol{x})$ 的极小值点。网络的稳定性可由下式说明：

$$\frac{\mathrm{d}F_p}{\mathrm{d}t} = (\nabla F_p(\boldsymbol{x}))^{\mathrm{T}} \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -(\nabla F_p(\boldsymbol{x}))^{\mathrm{T}} F_p(\boldsymbol{x}) = -\|\nabla F_p(\boldsymbol{x})\|^2 \leqslant 0 \tag{27}$$

即 $F_p(\boldsymbol{x})$ 沿着动力系统(24) 或(26) 的轨道 $\boldsymbol{x}(t)$ 是单调下降的。又 $F_p(\boldsymbol{x}(t))$ 有下界，故网络必是稳定的，且稳定点对应着能量函数 $F_p(\boldsymbol{x})$ 的极小点。

关于网络(24) 或(26) 的电路实现问题，可用与文献[1]、[2] 类似的方法来实现。

参 考 文 献

[1] CICHOCKI A, UNBEBAUEN R. Neural networks for solving systems of linear equations and related problems[J]. IEEE Trans Circuits Syst, 1992, 39: 124 ~ 138.

[2] CICHOCKI A, UNBEBAUEN R. Neural networks for solving systems of linear equations—Part Ⅱ: Minimax and least absolute value problems[J]. IEEE Trans Circuits Syst, 1992, 39: 619 ~ 633.

[3] HUBER P J. Robust Statistie[M]. New York: Wieley, 1981.

[4] HAMPEL F R, RONCHETTI E M, ROUSSEW P, et al. Robust Statics – The Approach Based on Influence Functions[M]. New York: Wiely, 1987.

[5] 李兴斯. 一类不微优化问题的有效解法[J]. 中国科学(A辑), 1994, 24(4): 371 ~ 377.

(下转 119 页)

A Survey on Data Warehouse and Data Mining

YIN Yong[1], *CAO Chang-xiu*[2], *LIN Jing-dong*[2], *ZHANG Bang-li*[2]
(1. College of Communication and Measurement Control Engineering, Chongqing University, Chongqing 400044, China;
2. College of Automation, Chongqing University, Chongqing 400044, China)

ABSTRACT: The circumstances of rich data and lack of knowledge lead to the emergence of technology of data warehouse and data mining, and arouse the interests of people in various fields. This paper gives a comprehensive introduction about the basic ideas of data warehouse and data mining, key techniques, and the content of the main researches. This paper have also discussed the characteristics as well as developing potential of the combination of data warehouse and data mining.

KEYWORDS: data warehouse; data mining; knowledge discovery in databases

（责任编辑　李胜春）

※ ※

（上接104页）

Neural Networks for Linear Equation
—Smoothing Method of the Minimax's Critria

YE Zhong-quan
(College of Science, Chongqing University, Chongqing 400044, China)

ABSTRACT: In the paper, neural networks are applied to explore the solving problem of linear equations. The critria is l_∞—norm(or Chebyshev' snorm), i.e. $E_\infty(x) = \max\limits_{x \in R^n} \{|r_i(x)|\}$ 或 $E_\infty(x) = \max\limits_{x \in R^n} \left\{\frac{1}{2} r_i^2(x)\right\}$ instead of usual last – squares criteria. We first smooth the objective function $E_\infty(x)$. Then we explore the unconstrained optimization problem by using neural networks. The convergence and stability of the network are discussed.

KEYWORDS: linear equations; neural networks; the criteria fuction of l_∞—norm; convergence

（责任编辑　张小强）