

文章编号: 1001-0920(2011)05-0757-04

基于信赖域 Newton 算法的 ELM 网络

韩 敏, 王新迎

(大连理工大学 电子信息与电气工程学部, 辽宁 大连 116024)

摘 要: 针对极端学习机 (ELM) 网络伪逆输出权值计算方法的运算复杂度制约其训练速度问题, 提出一种基于信赖域 Newton 算法的新型 ELM 网络 (TRON-ELM), 并采用信赖域 Newton 算法求解 ELM 网络的输出权值. 该算法首先构造一个 ELM 网络代价函数的 Newton 方程, 并将其作为一个无约束优化问题, 采用共轭梯度法求解, 避免了求代价函数 Hessian 矩阵逆的运算, 提高了训练速度, 信赖域条件的存在保证了算法的整体收敛性. 仿真实验结果验证了所提出方法的有效性.

关键词: 极端学习机; 信赖域 Newton 法; 共轭梯度法; 回归

中图分类号: TP273

文献标识码: A

ELM based on trust region Newton method

HAN Min, WANG Xin-ying

(Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China.

Correspondent: HAN Min, E-mail: minhan@dlut.edu.cn)

Abstract: Considering the problems that the complexity of generalized inverse limits the learning speed of extreme machine learning (ELM), a novel ELM, called TRON-ELM, is proposed based on the trust region Newton method in which the trust region Newton method is used to derive the output weights. The proposed method takes the Newton equation of the cost function of ELM as an unconstrained optimization, and a conjugate gradient method is used to solve the equation, which avoids solving the inverse of the Hessian matrix, thus the operation speed is improved. Meanwhile, the existence of trust region guarantees the global convergence. The experimental results show the effectiveness of the proposed method.

Key words: extreme machine learning; trust region Newton method; conjugate gradient method; regression

1 引 言

前馈神经网络是应用最为广泛的一类人工神经网络, 具有较强的非线性映射能力, 而其学习问题一直是人工神经网络研究领域的重要课题. Rumelhart 等人^[1]发现的 BP 算法, 是前馈神经网络的主要学习算法. 然而, BP 算法存在着收敛速度慢, 容易陷入局部最优等不足^[2]. 它的众多改进型算法, 虽然对提高网络学习的收敛速度和克服局部最优具有进步意义, 但仍具有一定的局限性^[3-4].

极端学习机 (ELM)^[5]是最近提出的一种新型前馈神经网络, 网络的输入权值和隐含层 (此处称之为静态储备池, 隐含层输出矩阵称之为储备池系数矩阵) 内部权值是随机生成的, 并且在生成后保持不变, 唯一需要确定的是网络的输出权值, 简单的线性回归

即可胜任 ELM 网络的训练任务. 因此, 与 BP 算法相比, 在训练速度和训练精度上都有了极大提高.

从目前的研究来看, 应用最多的 ELM 网络输出权值训练算法是伪逆算法^[6-7], 实际上是对储备池系数矩阵进行 SVD 分解, 然后用不为零的奇异值计算输出权值. 当训练样本较大时, SVD 分解的计算复杂度严重制约了 ELM 网络的训练速度^[8]. 信赖域 Newton 法是一种有效的大规模无约束优化算法, 它克服了传统 Newton 法需估计目标函数二阶 Hessian 阵、选取初始点困难及求解 Newton 方程费时等不足, 已被成功地应用于逻辑回归中^[9]. 基于此, 本文将信赖域 Newton 法引入 ELM 网络中, 快速有效地计算网络输出权值, 并通过 SinC 函数逼近及基准数据回归的仿真实验, 验证了本文方法的快速有效性.

收稿日期: 2010-02-22; 修回日期: 2010-04-06.

基金项目: 国家自然科学基金项目(60674073); 国家科技支撑计划项目(2006BAB14B05); 国家 973 计划项目(2006CB403405).

作者简介: 韩敏(1959—), 女, 教授, 博士生导师, 从事神经网络、专家系统等研究; 王新迎(1987—), 男, 硕士生, 从事神经网络、时间序列预测的研究.

2 ELM网络结构

ELM网络具有简单的3层结构: 输入层、输出层, 以及一个包含大量非线性节点的隐含层. 该隐含层被称之为储备池, 输入层和储备池之间采用的是随机连接. 相对应于前馈神经网络的隐含层, 输出矩阵在ELM网络中被称为储备池系数矩阵.

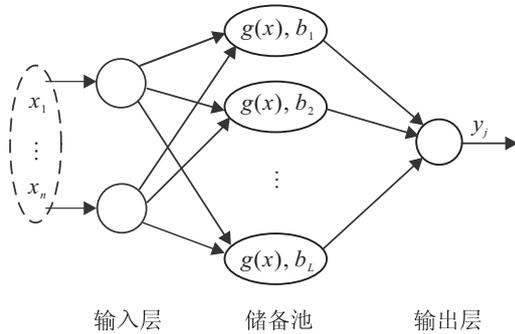


图1 ELM结构

ELM的网络方程为

$$\sum_{i=1}^L \beta_i g(\mathbf{x}_j) = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = y_j, \quad j = 1, 2, \dots, N. \quad (1)$$

其中: $\mathbf{x}_i \in \mathbf{R}^n$ 为输入向量, $y_i \in \mathbf{R}$ 为网络输出, \mathbf{w}_i 为网络的输入权值向量, b_i 为偏置值, β_i 为网络输出权值, N 为样本个数, $g(\cdot)$ 为激活函数(一般取 Sigmoid 函数), L 为储备池规模. 在算法的开始, \mathbf{w}_i 和 b_i 的值随机给定, 网络训练的目标是确定输出权值 $\beta = [\beta_1, \beta_2, \dots, \beta_L]$, 最小化如下的代价函数:

$$f(\beta) = \sum_{j=1}^N \left(\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) - t_j \right)^2. \quad (2)$$

在ELM网络中输出权值为 $\beta = \mathbf{H}^\dagger \mathbf{T}$, 其中

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1, b_1, \mathbf{x}_1) & \dots & g(\mathbf{w}_1, b_1, \mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_L, b_L, \mathbf{x}_1) & \dots & g(\mathbf{w}_L, b_L, \mathbf{x}_N) \end{bmatrix},$$

\mathbf{H}^\dagger 是 \mathbf{H} 的伪逆, $\mathbf{T} = [t_1, \dots, t_N]$ 为目标变量.

3 基于信赖域 Newton 算法的 ELM 网络

3.1 基于信赖域 Newton 算法的 ELM

ELM网络的训练过程, 即求解输出权值 β 使得代价函数 $f(\beta)$ 最小的过程. 信赖域 Newton 法用一个二次型模型

$$q_k(s) = \nabla f(\beta_k)^\top s + \frac{1}{2} s^\top \nabla^2 f(\beta_k) s \quad (3)$$

近似代价函数 $f(\beta_k + s_k) - f(\beta_k)$ 的差值. 其中: β_k 为输出权值的迭代变量, s_k 为 Newton 方向. 总可以找到一个 s^k 满足于信赖域约束条件 $\|s^k\| \leq \Delta_k$ 并使得 $q_k(s)$ 最小. 通过比较代价函数的实际减少值和二次型模型的预测减少值之间的比率 ρ_k 来更新 β_k 和 Δ_k , 即

$$\rho_k = \frac{f(\beta_k + s_k) - f(\beta_k)}{q_k(s^k)}. \quad (4)$$

如果 ρ_k 足够大, 说明计算出的 s^k 是代价函数值较快下降的方向, 因此 s^k 被接受, 有

$$\beta_{k+1} = \begin{cases} \beta_k + s_k, & \rho_k > \eta_0; \\ \beta_k, & \rho_k \leq \eta_0. \end{cases} \quad (5)$$

其中 η_0 是一个给定的值.

信赖域约束 Δ_k 按如下规则更新:

$$\begin{aligned} \Delta_{k+1} &\in [\sigma_1 \min\{\|s^k\|, \Delta_k\}, \sigma_2 \Delta_k], & \rho_k \leq \eta_1; \\ \Delta_{k+1} &\in [\sigma_1 \Delta_k, \sigma_3 \Delta_k], & \rho_k \in (\eta_1, \eta_2); \\ \Delta_{k+1} &\in [\Delta_k, \sigma_3 \Delta_k], & \rho_k \geq \eta_2. \end{aligned} \quad (6)$$

其中: $0 < \eta_1 < \eta_2 < 1$, $0 < \Delta_1 < \Delta_2 < 1 < \Delta_3$ 均为常数.

ELM网络的代价函数可写为

$$f(\beta) = E = (\beta^\top \mathbf{H} - \mathbf{T})(\beta^\top \mathbf{H} - \mathbf{T})^\top,$$

其梯度为

$$\nabla f(\beta) = 2\mathbf{H}\mathbf{H}^\top \beta - 2\mathbf{H}\mathbf{T}^\top,$$

其 Hessian 阵为

$$\nabla^2 f(\beta) = 2\mathbf{H}\mathbf{H}^\top.$$

因此, 用信赖域 Newton 法计算 ELM 网络输出权值的具体步骤如下:

Step 1: 初始化 $k = 0$, 任意给定 β^0 .

Step 2: 判断 $\nabla f(\beta^k) = 0$ 是否满足. 若满足, 则求得输出权值最终解 $\beta = \beta^k$, 算法结束; 否则 $k := k + 1$, 转 Step 3.

Step 3: 求得信赖域子问题的近似解 s^k , 即

$$\begin{aligned} \min_s q_k(s) &= \nabla f(\beta_k)^\top s + \frac{1}{2} s^\top \nabla^2 f(\beta_k) s, \\ \text{s.t. } \|s^k\| &\leq \Delta_k. \end{aligned} \quad (7)$$

Step 4: 根据式 (4) 计算 ρ_k .

Step 5: 根据式 (5) 更新 β^k 到 β^{k+1} .

Step 6: 根据式 (6) 更新 Δ_k 到 Δ_{k+1} , 转 Step 2.

信赖域 Newton 算法在内层采用共轭梯度法求解信赖域子问题 (7), 其算法具体步骤如下:

Step 1: 赋初值, $i = 0$, 对给定的 $\xi_k < 1$, $\Delta_k > 0$, 令

$$\bar{s}^0 = 0, \quad r^0 = -\nabla f(\beta^k), \quad d^0 = r^0.$$

Step 2: 判断

$$\|r^i\| \leq \xi_k \|\nabla f(\beta^k)\| \quad (8)$$

是否满足. 若满足, 则返回 Newton 方向的近似解 $s^k = \bar{s}^i$, 算法结束; 否则, $i := i + 1$, 转 Step 3.

Step 3: 计算 $\alpha_i = \frac{\|r^i\|^2}{((d^i)^\top \nabla^2 f(\beta^k) d^i)}$.

Step 4: 更新 $\bar{s}^{i+1} = \bar{s}^i + \alpha_i d^i$.

Step 5: 判断 $\|\bar{s}^{i+1}\| \geq \Delta_k$ 是否满足. 若满足, 则计算 τ 满足

$$\|\bar{s}^i + \tau d^i\| = \Delta_k, \quad (9)$$

返回 Newton 方向的最优近似解 $s^k = \bar{s}^i + \tau d^i$, 算法结束; 否则, 转 Step 6.

Step 6: 更新 $r^{i+1} = r^i - \alpha_i \nabla^2 f(\beta^k) d^i$.

Step 7: 计算 $\beta_i = \frac{\|r^{i+1}\|^2}{\|r^i\|^2}$.

Step 8: 更新 $d^{i+1} = r^{i+1} + \beta_i d^i$, 转 Step 2.

由上可以看出, 求解信赖域子问题 (7) 的最优解 s^k 的内层共轭梯度算法的主要计算量, 是代价函数 Hessian 阵和向量的乘积 $\nabla^2 f(w^k) d^i$. 然而实际上, 在算法的每次迭代中仅需计算一次 Hessian 阵和向量的乘积.

因为

$$r^i = -\nabla f(\beta^k) - \nabla^2 f(\beta^k) \bar{s}^i, \quad (10)$$

所以停止准则 (8) 与下式等价:

$$\|-\nabla f(\beta^k) - \nabla^2 f(\beta^k) \bar{s}^i\| \leq \xi_k \|\nabla f(\beta^k)\|. \quad (11)$$

这表示满足停止准则 (8) 的 \bar{s}^i 是信赖域子问题的一个最优近似解. 在信赖域 Newton 法的内层共轭梯度法中, 需要仔细考虑信赖域约束条件 $\|s\| \leq \Delta$, 因此其不同于一般的共轭梯度法. 当 $\bar{s}^0 = 0$ 时 (参见文献中的定理 2.1), 有

$$\|\bar{s}^i\| \leq \|\bar{s}^{i+1}\|, \forall i.$$

因此, 在有限步的共轭梯度迭代中, 或者停止准则 (8) 成立, 或者 \bar{s}^{i+1} 超过信赖域约束条件. 在后者情形下, 利用式 (9) 可以在信赖域边界上找到一个点满足

$$q_k(\bar{s}^i + \tau d^i) < q_k(\bar{s}^i).$$

因此, 整个信赖域 Newton 算法是经过严密设计的, 保证了求得的近似 Newton 方向 s^k 足够好, 同时能够使信赖域方法收敛.

3.2 参数设置

设置信赖域约束的初始值为 $\Delta_0 = \|\nabla f(\beta^0)\|$, 对于用于更新输出权值的式 (5) 取 $\eta_0 = 0.0001$, 对于用于更新信赖域约束 Δ_k 的式 (6) 取 $\eta_1 = 0.25$, $\eta_2 = 0.75$; $\sigma_1 = 0.25$, $\sigma_2 = 0.5$, $\sigma_3 = 4.0$. 停止准则 (8) 中设置 $\xi_k = 0.1$.

4 仿真实例

为了验证本文方法的有效性, 将其应用于 SinC 函数以及 8 组基准数据的仿真实验中^[12], 并与采用 MP 伪逆输出权值求解算法的 ELM 方法的训练时间和训练精度进行了比较.

4.1 SinC 函数

在本次实验中, 两种方法 (TRON-ELM 和 ELM)

均用来逼近 SinC 函数

$$y(x) = \begin{cases} \frac{\sin(x)}{x}, & x \neq 0; \\ 1, & x = 0. \end{cases}$$

其中: x 在 $[-10, 10]$ 上随机均匀分布, 训练样本和测试样本分别含有 5 000 个数据. 为了使逼近问题更加真实, 将分布在 $[-0.2, 0.2]$ 的高斯白噪声加入到训练样本数据中, 测试数据是无噪声的.

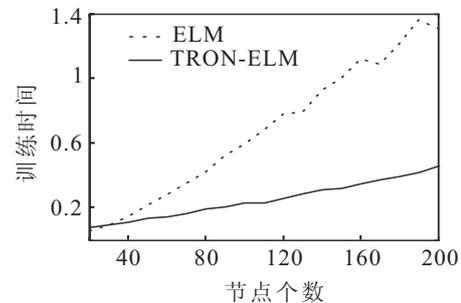


图 2 两种方法在隐含节点变化时的仿真结果比较

储备池内部节点从 20 递增至 200, 每种方法均仿真 50 次, 取仿真结果的平均值, 结果见图 2. 从图 2 中可以看出, 随着储备池规模的增大, 基于伪逆输出权值计算方法的 ELM 训练时间快速的增加, 而基于信赖域 Newton 算法的 ELM 训练时间增加的较为缓慢, 而这种训练时间上的优势会随储备池规模的增大而更加明显, 因此验证了本文方法的有效性.

4.2 基准数据回归

基于 8 组覆盖不同领域的基准数据对 ELM 和 TRON-ELM 进行了比较, 数据见表 1. 在实际应用中, 这些数据集的分布情况通常是未知的, 并且其中多数数据含有噪声. 训练样本和测试样本均从数据中随机选择, 每组数据均仿真 50 次取结果的平均值.

表 1 基准数据列表

数据类型	输入个数		属性维数		节点个数
	训练	测试	连续	名义	
Stocks	450	500	10	0	150
Abalone	2,000	2,177	7	1	100
Del ailerons	3,000	4,129	6	0	100
Del elevators	4,000	5,517	6	0	150
Com activity	4,000	4,190	8	0	150
Census(house8L)	10,000	12,784	8	0	200
Bank	4,500	3,692	8	0	200
Cal housing	8,000	12,460	8	0	100

表 2 给出了两种方法的训练和测试的均方根误差的比较, 当两个相比较的结果相差大于 0.0005 时, 用加粗的数字表示有优势的结果. 可以看出, 两种方法获得了几乎一致的训练精度, 但 TRON-ELM 的测试精度一般要高于 ELM.

表 3 给出了两种方法的训练和测试时间的比较.

表 2 两种方法的训练和测试根均方差 (RMSE) 比较

数据类型	ELM		TRON-ELM	
	训练	测试	训练	测试
Stocks	0.021 2	0.037 1	0.021 6	0.035 2
Abalone	0.070 7	0.090 1	0.071 1	0.083 5
Del ailerons	0.036 5	0.039 9	0.036 9	0.039 1
Del elevators	0.051 0	0.054 1	0.051 4	0.053 6
Com activity	0.031 3	0.042 6	0.031 5	0.042 6
Census(house8L)	0.060 4	0.071 1	0.060 6	0.069 5
Bank	0.040 6	0.043 4	0.040 5	0.043 6
Cal housing	0.123 8	0.133 3	0.123 8	0.130 7

表 3 两种方法的训练和测试时间比较

数据类型	ELM		TRON-ELM	
	训练	测试	训练	测试
Stocks	0.141 9	0.020 3	0.140 0	0.018 3
Abalone	0.291 9	0.051 3	0.185 6	0.044 7
Del ailerons	0.398 8	0.086 6	0.153 4	0.070 6
Del elevators	0.976 6	0.159 1	0.305 0	0.150 3
Com activity	0.995 3	0.140 0	0.301 3	0.120 5
Census(house8L)	2.854 4	0.452 5	1.317 5	0.424 7
Bank	1.278 4	0.136 6	0.446 6	0.132 8
Cal housing	0.926 3	0.228 8	0.443 4	0.221 9

可以看出, 8 组数据 TRON-ELM 均取得了较 ELM 更快的训练速度和测试速度. 结合表 2 和表 3 可以看出, 随着训练样本个数和储备池规模的增加, 基于信赖域 Newton 法的 ELM 的训练速度的优势也将更加明显. 因此, 说明了 TRON-ELM 更适用于较大规模数据, 验证了所提出方法的有效性.

5 结 论

基于信赖域 Newton 算法设计一种优化的 ELM 网络, 随机确定输入权值和储备池内部权值, 并采用信赖域 Newton 算法计算网络的输出权值, 在训练样本较大和隐含层节点较多的情况下, 不仅保证了训练精度, 而且具有更快的训练速度. 将其应用于人工函数和基准数据的实验仿真中, 通过比较, 验证了本文方法的有效性. 因此, 本文所提出方法为具有较大样本的时间序列预测和复杂系统辨识提供了一条有效的途径.

参考文献(References)

[1] Rumelhart D E, McClelland J L. Parallel distributed processing: Explorations in the microstructure of cognition[M]. Cambridge: MIT Press, 1986.

[2] Haykin S. Neural networks: A comprehensive foundation[M]. New Jersey: Prentice Hall, 1999.

[3] 叶军, 张新华. 多层前向神经网络的快速学习算法及其应用[J]. 控制与决策, 2002, 17(增): 817-819. (Ye J, Zhang X H. High speed learning algorithm for a multilayer feedforward neural network and its application[J]. Control and Decision, 2002, 17(S): 817-819.)

[4] 王俊年, 申群太, 周少武, 等. 基于种群小生境微粒群算法的前向神经网络设计[J]. 控制与决策, 2005, 20(9): 981-985. (Wang J N, Shen Q T, Zhou S W, et al. Evolutionary design of feed-forward neural network based on species niching particle swarm optimizer[J]. Control and Decision, 2005, 20(9): 981-985.)

[5] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006, 70(1/2/3): 489-501.

[6] Feng G R, Huang G B, Lin Q P, et al. Error minimized extreme learning machine with growth of hidden nodes and incremental learning[J]. IEEE Trans on Neural Networks, 2009, 20(8): 1352-1357.

[7] Huang G B, Chen L, Siew C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. IEEE Trans on Neural Networks, 2006, 17(4): 879-92.

[8] 史志伟, 韩敏. ESN 岭回归学习算法与混沌时间序列预测[J]. 控制与决策, 2007, 22(3): 258-261. (Shi Z W, Han M. Ridge regression learning in ESN for chaotic time series prediction[J]. Control and Decision, 2007, 22(3): 258-261.)

[9] Lin C J, Weng R C, Keerthi S S. Trust region Newton method for large-scale logistic regression[J]. J of Machine Learning Research, 2008, 9: 627-650.

[10] Nash S G. A survey of truncated-Newton methods[J]. J of Computational and Applied Mathematics, 2000, 124(1/2): 45-59.

[11] Lin C J, More J J. Newton's method for large-scale bound constrained problems[J]. SIAM J on Optimization, 1999, 9(4): 1100-1127.

[12] Asuncion A, Newan D. UCI machine learning repository[EB/OL]. Irvine: University California Irvine, (2009-04-15). Http://archive.ics.uci.edu/ml/.