

灵敏度正则化极限学习机及其在数字识别中的应用^①

张松林, 李 雪

(河南工学院 电子通信工程系, 新乡 453002)

摘要: 针对正则化极限学习机(RELM)中隐节点数影响分类准确性问题, 提出一种灵敏度正则化极限学习机(SRELM)算法. 首先根据隐含层激活函数的输出及其相对应的输出层权重系数, 推导实际值与隐节点输出值残差相对于隐节点的灵敏度计算公式, 然后根据不同隐节点的灵敏度进行排序, 利用优化样本的分类准确率删减次要隐节点, 从而有效提高 SRELM 的分类准确率. MNIST 手写体数字库实验结果表明, 相比于传统的 SVM 和 RELM, SRELM 方法的耗时与 RELM 相差不大, 均明显低于 SVM, SRELM 对手写数字的识别准确率最高.

关键词: 灵敏度分析; 极限学习机; 模式识别; 分类; 手写数字

Sensitivity Regularized Extreme Learning Machine and Its Application in Digit Recognition

ZHANG Song-Lin, LI Xue

(Electronic Communications Engineering Department, Henan Institute of Technology, Xinxiang 453002, China)

Abstract: To solve the problem that the number of hidden nodes in regularized extreme learning machine(RELM) affects classification accuracy, sensitive regularized extreme learning machine(SRELM) algorithm is proposed. Firstly, based on the output of hidden layer activation function and its corresponding output layer weighting factor, the formula of computing the sensitivity for hidden node is deduced by residual between actual value and hidden nodes output. Then different hidden nodes are sorted according to sensitivity. And minor hidden nodes are deleted based on classification accuracy of optimization samples. As a result, SRELM classification accuracy is increased effectively. A case study of MNIST handwritten digit database shows that, compared with common SVM and RELM, time consuming of SRELM is almost the same as RELM, and is obviously lower than SVM. Meanwhile SRELM recognition accuracy for handwritten digit is the highest.

Key words: sensitivity analysis; extreme learning machine; pattern recognition; classification; handwritten digit

1 引言

手写数字识别技术作为模式识别中的一项重要应用, 引起了学者们的广泛关注^[1-6]. 数字识别可以分为动态识别问题和静态识别问题两种^[6], 动态识别问题由于可以观察到数字书写的顺序和笔画, 识别难度较小. 对于静态识别问题, 由于没有已知的书写信息, 而只已知二维的数字图像, 其识别难度要高于动态数字识别问题.

静态手写数字识别技术常用的方法包括主元分析法^[1]、K 最近邻方法^[2]、神经网络^[3,4]、支持向量机(SVM)^[5,6]等. 其中文献[1]采用增量主元分析法识别手写数字; 文献[2]比较了不同 K 最近邻方法在数字识别

中的性能; 文献[3]采用一种快速的神经网络方法实现数字识别技术; 文献[4]采用前馈神经网络识别数字; 文献[5]和[6]利用支持向量机对手写数字进行识别, 其中文献[6]提出基于概率测度 SVM 的数字识别方法, 该方法将概率分布情况作为均值, 插入建立可再生的希尔伯特空间, 实验结果证明了该方法的有效性. 但传统的数字识别方法存在一些不足, 比如 K 最近邻方法的识别准确率较低; 神经网络通常采用 BP 算法, 但是 BP 算法的学习速度较慢, 并且可能陷入局部最优, 影响分类精度; SVM 对大规模训练样本的建模存在计算量较大的问题等.

极限学习机(ELM)作为近年来发展起来的一种单

① 收稿时间:2016-09-21;收到修改稿时间:2016-11-21 [doi:10.15888/j.cnki.csa.005823]

隐层前馈神经网络学习算法^[7], 与传统学习算法需要训练网络中的所有参数不同, ELM 只需要训练网络的输出层权重系数, 因此 ELM 具有学习速度快、泛化能力强等优点^[8,9], 已经在时间序列预测^[10,11]、模式识别^[12-14]等领域得到了应用. 文献[15]将 λ 正则化因子引入 ELM 求解输出权重公式中, 提出了带正则化因子的极限学习机 RELM, 提高了 ELM 的预测准确性. 对于 RELM, 除了正则化因子外, 隐节点数目同样影响其分类准确率, 若能根据一定准则对隐节点的重要性进行排序, 删减不必要的隐节点, 则能进一步提高 RELM 分类的准确性. 为此, 本文将灵敏度分析法引入 RELM 隐节点重要性排序过程, 提出基于灵敏度的正则化极限学习机(SRELM)算法, SRELM 利用隐节点输出及其相应的输出层权重系数, 计算 RELM 输出值与真实值之间的残差相对于隐节点的灵敏度, 然后根据计算的灵敏度对不同隐节点的重要性进行排序, 利用优化样本删减次要的隐节点, 从而优化 SRELM 的网络结构, 提高 SRELM 的分类准确性. 实验结果证明了本文提出方法的有效性.

2 正则化极限学习机

ELM 的数学表达式为:

$$f_L(x) = \sum_{i=1}^L \beta_i G(a_i, b_i, x) \quad (1)$$

式中, L 表示隐节点数目, $a_i \in R^n$ 和 $b_i \in R$ 表示隐含层参数, $\beta_i \in R^m$ 表示第 i 个隐节点对于输出节点的权重系数. $G(a_i, b_i, x)$ 为第 i 个隐节点的输出函数, 其中加性隐节点函数可以表示为:

$$G(a_i, b_i, x) = g(a_i \cdot x + b_i) \quad (2)$$

式中, a_i 表示第 i 个隐含层节点的输入权重, b_i 表示第 i 个隐含层节点的偏差, $g(\cdot)$ 表示激活函数.

假设已知由 N 组输入输出数据对构成的训练样本 $\{(x_l, y_l)\}_{l=1}^N$, 其中 $x_l \in R^n$ 表示输入变量, $y_l \in R^m$ 表示相应的实际输出, 则:

$$y_l = \sum_{i=1}^L \beta_i G(a_i, b_i, x_l), l=1, 2, \dots, N \quad (3)$$

将式(3)变为矩阵形式, 则:

$$H\beta = Y \quad (4)$$

式中:

$$H = \begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_L, b_L, x_1) \\ \vdots & & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_L, b_L, x_N) \end{bmatrix}_{N \times L} = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_L \end{bmatrix}_{L \times m}, Y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}_{N \times m} \quad (5)$$

对于式(3), 参数 a_i 和 b_i 利用随机生成方式获得, 只有参数 β 根据 $\{(x_l, y_l)\}_{l=1}^N$ 计算得到. 在解决分类问题时, 为了提高 ELM 的分类准确性, 文献[15]引入正则化因子 λ , 从而可以将求解 β 看成如下优化问题:

$$\min_{\beta} : V_{ELM} = \frac{1}{2} \|\beta\|^2 + \frac{\lambda}{2} \sum_{l=1}^N \|\varepsilon_l\|^2 \quad (6)$$

s.t.: $h(x_l)\beta = y_l - \varepsilon_l, l=1, 2, \dots, N$

式中, 训练偏差 ε_l 可以有效避免过拟合问题, 依照 Karush-Kuhn-Tucker 定理^[16], 式(6)的求解问题可以转化成双重优化问题:

$$V_{ELM} = \frac{1}{2} \|\beta\|^2 + \frac{\lambda}{2} \sum_{l=1}^N \|\varepsilon_l\|^2 - \sum_{l=1}^N \sum_{j=1}^m a_{l,j} (h(x_l)\beta_j - y_{l,j} + \varepsilon_{l,j}) \quad (7)$$

根据式(7)计算 β 的最优解为^[16]:

$$\beta = H^T \left(\frac{I}{\lambda} + HH^T \right)^{-1} Y \quad (8)$$

由式(8)可以看出, RELM 中的 β 主要由式(5)中的矩阵 H 、矩阵 Y 以及正则化因子 λ 决定, 其中 H 的维数与训练样本数 N 和隐节点数 L 有关, Y 为训练样本对应的输出. N 和 Y 随训练样本已经确定, 但隐节点数 L 和正则化因子 λ 需要人为确定, 其对 RELM 的性能影响较大.

若为二分类问题, RELM 的决策方程按下式求得:

$$f(x) = \text{sign} \left(h(x)H^T \left(\frac{I}{\lambda} + HH^T \right)^{-1} Y \right) \quad (9)$$

若为多分类问题, RELM 的决策方程按下式求得:

$$\text{label}(x) = \arg \max_{i \in \{1, \dots, m\}} f_i(x) \quad (10)$$

式中, $f_i(x)$ 为第 i 个输出节点的值, 且有

$$f(x) = [f_1(x), \dots, f_m(x)]^T$$

3 灵敏度正则化极限学习机

对于 RELM, 除了正则化因子 λ 外, 隐节点数目 L 同样影响着 RELM 的分类准确率, 为此, 本节首先给出能够反映 RELM 不同隐节点重要性的灵敏度计算公式, 然后根据不同隐节点的灵敏度值对隐节点进行排序, 从而删除次要的隐节点, 确定最优的 RELM 隐节点数目 L , 优化网络结构, 提高分类准确率.

3.1 灵敏度定义

假设删除第 j 个隐节点, $1 \leq j \leq L$, 则式(3)变为:

$$y_l' = k_{1l}\beta_1 + \dots + k_{(j-1)l}\beta_{j-1} + \dots + k_{Ll}\beta_L \quad (11)$$

式中, k_{pi} 对应于式(3)中的 $G(a_p, b_p, x_i)$.

将 y_i 和 y_i' 相减, 可得残差:

$$\|y_i - y_i'\| = \|k_{ji}\beta_j\| = |k_{ji}|\|\beta_j\| \quad (12)$$

其中, $\|\cdot\|$ 为 2-范数. 对于第 i 个样本, 去掉第 j 个隐节点产生的误差为 k_{ji} 的绝对值与 β_j 模的乘积. 因此, 对于所有 N 个样本, 定义残差相对于第 j 个隐节点的灵敏度为:

$$R_j = \frac{1}{N} \sum_{i=1}^N |k_{ji}|\|\beta_j\| \quad (13)$$

其中, R_j 越大表明第 j 个隐节点引起的残差越大, 即第 j 个隐节点的重要性越强.

3.2 SRELM 算法流程

给定训练样本集 $\{(x_i, y_i)\}_{i=1}^N$ 和参数优化样本 $\{(x_i, y_i)\}_{i=1}^M$, 其中 $x_i \in R^n$ 为输入向量, $y_i \in \{1, \dots, D\}$ 为类别标签, 根据式(13)给出的不同隐节点的灵敏度定义公式, 整理得灵敏度正则化极限学习机算法步骤如下:

步骤 1. 确定初始的 RELM 模型参数, 其中选择足够大的初始隐节点数 L , 根据训练样本 $\{(x_i, y_i)\}_{i=1}^N$, 生成由 L 个隐节点组成的 H 矩阵, 如式(5)所示, $L < N$.

步骤 2. 根据式(13)计算所有 L 个隐节点的灵敏度, 按照灵敏度由大到小顺序, 重新对 L 个隐节点进行排序, $R_1' \geq R_2' \geq \dots \geq R_L'$.

步骤 3. 令 L' 在 $1 \sim L$ 范围内变化, 对于变化的隐节点个数 L' , 根据步骤 2 中对 L 个隐节点的排序, 选取前 L' 个隐节点组成新的 H' 矩阵, 如式(14)所示. 将 H' 代入式(8)计算输出权重 $\hat{\beta}'$, 利用参数优化样本计算分类准确率, 选取分类准确率最高时对应的 L' 为 SRELM 的隐节点个数.

$$H' = \begin{bmatrix} G(a_1, b_1, x_1) & \dots & G(a_{L'}, b_{L'}, x_1) \\ \vdots & & \vdots \\ G(a_1, b_1, x_N) & \dots & G(a_{L'}, b_{L'}, x_N) \end{bmatrix}_{N \times L'} \quad (14)$$

步骤 4. 将最优 L' 对应的 H' 矩阵和训练样本 $\{(x_i, y_i)\}_{i=1}^N$ 中的 $\{y_i\}_{i=1}^N$ 代入式(8)中, 计算输出层权重 $\hat{\beta}'$.

步骤 5. 当有待分类输入向量 x_k 到来时, 计算对应的 $h_k' = [G(a_1, b_1, x_k) \dots G(a_{L'}, b_{L'}, x_k)]$, 代入式(10)中, 计算得到其属于的类别.

4 实验验证

4.1 实验数据

利用 MNIST 手写体数字库验证本文提出方法的

有效性^[17]. 该数据库是由美国国家标准和技术研究所提供的一个专门用于手写体数字识别研究的数据库. MNIST 中包括 6 万个训练样本和 1 万个测试样本, 每个样本是一个 28×28 像素的 bmp 格式图片, 其中部分样本如图 1 所示.



图 1 MNIST 中部分手写体数字

在本实验中, 随机选取 1000 个 MNIST 训练样本作为本实验的训练样本集, 选取 500 个 MNIST 训练样本作为参数优化样本, 用于优化模式识别算法的参数, 在 MNIST 测试样本中选取 1000 个样本作为本实验的测试样本集.

4.2 参数选择

为了证明本文提出方法的有效性, 利用支持向量机(SVM)和正则化 ELM(RELM)作为对比方法, 其中 SVM 采用 libsvm 工具箱提供的 SVM 函数. 实验运行在 Windows XP 系统, 酷睿 T6600 CPU(主频 2.2GHz), 2GB 内存, 仿真软件为 MATLAB R2011b.

对于 SVM, 惩罚参数 C 和核参数 γ 影响着数字识别的准确率, 对于 RELM 和本文提出的 SRELM, 正则化因子 λ 和隐节点数 L 影响着识别准确性. 为此, 对于 SVM, 分别令 C 和 $\gamma = \{2^{-24}, 2^{-23}, \dots, 2^{24}, 2^{25}\}$, 根据网格搜索法求出每一对 C 和 γ 对应的用于优化模型参数样本的识别准确率. 同理, 对于 RELM 和 SRELM, 分别令 $\lambda = \{2^{-24}, 2^{-23}, \dots, 2^{24}, 2^{25}\}$, $L = \{100, 200, \dots, 3000\}$, 计算识别准确率. 则 SVM、RELM 和 SRELM 三种方法对参数优化样本的识别准确率如图 2 所示, 参数寻优结果如表 1 所示. 可以看出, SVM 的寻优耗时最长,

RELM 和 SRELM 的寻优耗时相差不大, 由于 SRELM 需要根据式(13)计算不同隐节点的灵敏度, 所以 SRELM 的耗时略高于 RELM.

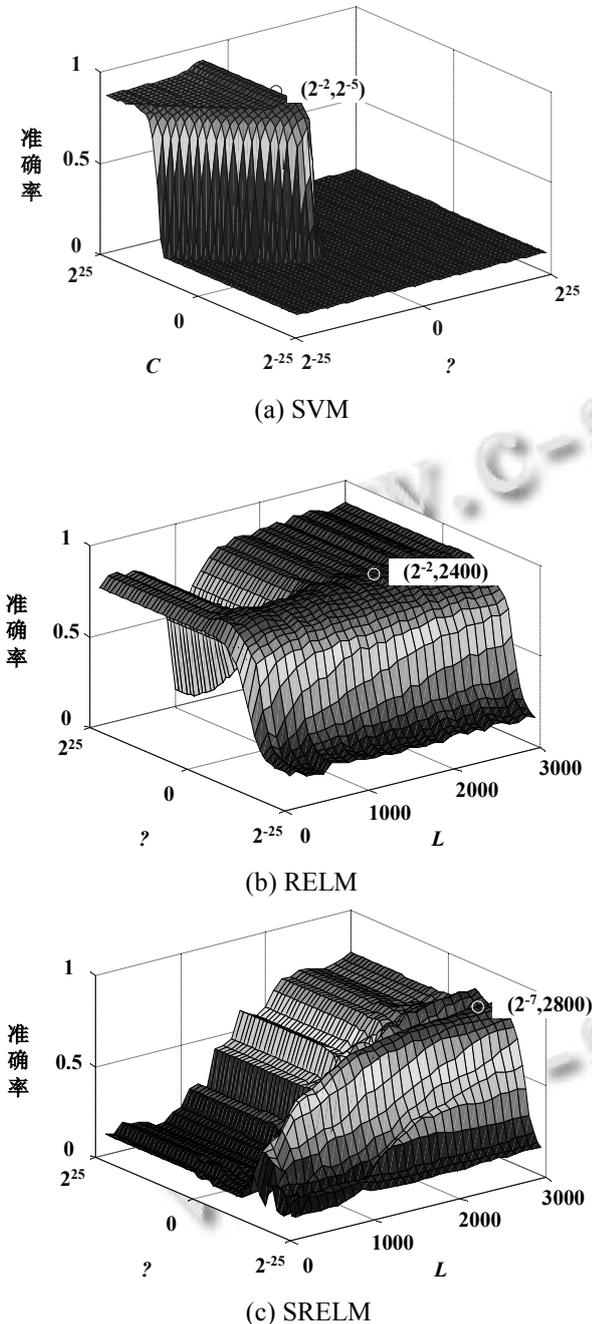


图 2 不同参数的识别准确率

表 1 参数寻优结果和寻优耗时

SVM			RELM			SRELM		
C	γ	耗时/s	λ	L	耗时/s	λ	L	耗时/s
2^{-2}	2^{-5}	4636	2^{-2}	2400	1853	2^{-7}	2800	1872

4.3 实验结果

对于 SVM、RELM 和 SRELM 三种方法, 利用训练样本和表 1 所示的参数建立识别模型, 对测试样本进行测试, 进行 20 次蒙特卡洛仿真, 不同方法的单次仿真结果对比如图 3 所示, 识别准确率、训练耗时和测试耗时如表 2 所示. 其中, 训练耗时为根据训练样本和表 1 所示的优化参数建立数字识别模型的时间, 测试耗时为根据训练好的模型识别测试样本所需的时间.

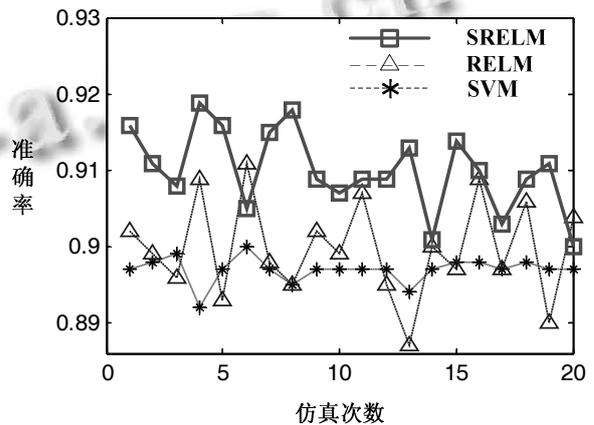


图 3 不同方法的测试准确率

表 2 测试准确率和耗时

	准确率	训练耗时/s	测试耗时/s
SVM	89.70%	11.34	5.62
RELM	90.00%	3.23	0.94
SRELM	91.02%	3.94	1.16

可以看出, 三种方法均能准确地识别出大部分测试样本, 相比之下, SVM 的识别准确率较低, 与 RELM 相比, 本文提出的 SRELM 方法的识别准确率高于 RELM, 说明根据灵敏度对 ELM 隐节点的重要性进行排序的方法, 能够有效地提高对手写数字识别的准确率. 此外, 可以看出, SVM 方法的训练耗时和测试耗时最长, SRELM 方法的训练耗时和测试耗时与 RELM 相差不大, 说明 SRELM 方法的计算量小于 SVM, 与 RELM 方法相当.

5 结论

为了解决静态手写数字识别问题, 本文提出了一种灵敏度正则化极限学习机(SRELM)算法. 该算法结合了正则化极限学习机(RELM)能提高 ELM 泛化能力

的优点, 以及灵敏度分析法能够有效删除 ELM 中的次要隐节点的特性, 使得 SRELM 具有更高的分类准确性. 利用 MNIST 手写数字数据库的样本进行了实验验证, 结果表明, 传统的 SVM 的耗时最长, 而 RELM 和 SRELM 的耗时相对较少. 相比于 SVM 和 RELM, SRELM 的识别准确性最高, 从而证明了本文提出方法的有效性.

参考文献

- 1 Zhao WD, Yan T, Liu C, Li D, Liu YH. Incremental tensor principal component analysis for handwritten digit recognition. *Mathematical Problems in Engineering*, 2014, (1): 1–10.
- 2 Roy K, Chaudhuri C, Kundu M, Nasipuri M, Basu DK. Comparison of the multilayer perceptron and the nearest neighbor classifier for handwritten digit recognition. *Journal of Information Science & Engineering*, 2015, 21(6): 1247.
- 3 Jagtap VN, Mishra SK. Fast efficient artificial neural network for handwritten digit recognition. *International Journal of Computer Science & Information Technology*, 2014, 5(2): 2302–2306.
- 4 Medhi K, Kalita SK. Assamese digit recognition with feed forward neural network. *International Journal of Computer Applications*, 2015, 109(1): 34–40.
- 5 Ebrahimzadeh R, Jampour M. Efficient handwritten digit recognition based on histogram of oriented gradients and SVM. *International Journal of Computer Applications*, 2014, 104(9): 10–13.
- 6 方向, 陈思佳, 贾颖. 基于概率测度支持向量机的静态手写数字识别方法. *微电子学与计算机*, 2015, 32(4): 107–110.
- 7 Huang GB, Zhu QY, Kheong SC. Extreme learning machine: Theory and applications. *Neurocomputing*, 2006, (70): 489–501.
- 8 Fossaceca JM, Mazzuchi TA, Sarkani S. MARK-ELM: Application of a novel Multiple Kernel learning framework for improving the robustness of Network Intrusion Detection. *Expert Systems with Applications*, 2015, 42(8): 4062–4080.
- 9 Akusok A, Veganzones D, Miche Y, Björk KM, Du Jardin P, Severin E, Lendasse A. MD-ELM: Originally mislabeled samples detection using OP-ELM model. *Neurocomputing*, 2015, (159): 242–250.
- 10 Gu Y, Liu JF, Chen YQ, Jiang XL, Yu HC. TOSELM: Timeliness online sequential extreme learning machine. *Neurocomputing*, 2014, (128): 119–127.
- 11 张弦, 王宏力. 具有选择与遗忘机制的极端学习机在时间序列预测中的应用. *物理学报*, 2011, 60(8): 68–74.
- 12 Luo XZ, Liu F, Yang SY, Wang XD, Zhou ZG. Joint sparse regularization based Sparse Semi-Supervised Extreme Learning Machine (S3ELM) for classification. *Knowledge-Based Systems*, 2014, (73): 149–160.
- 13 Ahila R, Sadasivam V, Manimala K. An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. *Applied Soft Computing*, 2015, (32): 23–37.
- 14 Luo X, Chang XH. A novel data fusion scheme using grey model and extreme learning machine in wireless sensor networks. *International Journal of Control, Automation, and Systems*, 2015, 13(3): 539–546.
- 15 Martinez-Martinez JM, Escandell-Montero P, Soria-Olivas E, et al. Regularized extreme learning machine for regression problems. *Neurocomputing*, 2011, (74): 3716–3721.
- 16 Huang GB, Zhou HM, Ding XJ, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics*, 2012, 42(2): 513–529.
- 17 MNIST 手写数字库. <http://yann.lecun.com/exdb/mnist>.