

基于互信息的选择性集成核极端学习机

韩敏, 吕飞

(大连理工大学控制科学与工程学院, 辽宁大连116024)

摘要: 针对集成学习中的准确性和差异性平衡问题, 提出一种基于信息论的选择性集成核极端学习机. 采用具有结构简单、训练简便、泛化性能好的核极端学习作为基学习器. 引入相关性准则描述准确性, 冗余性准则描述差异性, 将选择性集成问题转化为变量选择问题. 利用基于互信息的最大相关最小冗余准则对生成的核极端学习机进行选择, 从而实现准确性和差异性的平衡. 基于UCI基准回归和分类数据的仿真结果验证了所提出算法的优越性.

关键词: 互信息; 选择性集成; 核方法; 极端学习机

中图分类号: TP273

文献标志码: A

Selective ensemble of extreme learning machine with kernels based on mutual information

HAN Min, LÜ Fei

(School of Control Science and Engineering, Dalian University of Technology, Dalian 116024, China. Correspondent: HAN Min, E-mail: minhan@dlut.edu.cn)

Abstract: Considering the accuracy and diverse balance problem of ensemble learning, a selective ensemble extreme learning machine with kernels based on mutual information is proposed. The extreme learning machine with kernels, which has characteristics of simple structure, fast training and good generalization, is chosen as the base learner. The correlation and redundancy criterion are introduced to describe the accuracy and diverse, and the selective ensemble problem is transformed as a variable problem. Then, the maximum correlation minimum redundancy criterion based on mutual information is used to select the extreme learning machine with kernels. The simulation results based on UCI benchmark regression and classification data show the advantages of the proposed algorithm.

Keywords: mutual information; selective ensemble learning; kernel methods; extreme learning machine

0 引言

极端学习机^[1]是一种新的前馈神经网络, 具有结构简单、训练简便的优点, 已被成功应用于模式识别、函数拟合、分类、回归、时间序列预测等领域^[2-3]. 极端学习机克服了传统神经网络固有的训练速度慢、易陷入局部最小等缺陷, 吸引了众多研究者的兴趣, 发表了一批有价值的研究成果^[3-4]. 特别值得关注的是, Huang等^[5]类比了极端学习机的随机隐含层特征映射和支持向量机的核函数特征映射, 提出了一种核极端学习机, 并指出核极端学习机可得到与支持向量机相近或更好的泛化能力.

神经网络集成算法最早由Hansen等^[6]提出, 他们的研究表明, 采用多个神经网络进行集成可以提高神经网络的泛化性能. 目前, 最流行的两种神经网络

集成方法为Bagging算法和Boosting算法^[7]. Zhou等^[8]的研究表明, 从基分类器中选择出部分进行集成可获得更好的泛化性能, 并提出了选择性集成学习算法. 选择出的基学习器通常需要较高的预测精度, 且相互间要保持较大的差异性. 张春霞等^[9]对现有的选择性集成算法进行了综述, 并指出现有的选择性集成算法的研究主要集中于差异性评测方法的选择上.

许多学者将极端学习机模型与集成算法相结合^[10-11]. 其中, 赵立杰等^[11]将极端学习机作为基学习器, 提出一种选择性极端学习机集成算法. 基于以上分析, 本文提出一种基于互信息的选择性集成核极端学习机算法. 采用训练简捷且泛化能力强的核极端学习机作为基学习器, 引入一种基于互信息的最大相关和最小冗余准则^[12]对基学习器进行选择集成, 保证

收稿日期: 2014-09-18; **修回日期:** 2015-01-13.

基金项目: 国家自然科学基金项目(61374154); 国家重点基础研究发展规划项目(2013CB430403).

作者简介: 韩敏(1959—), 女, 教授, 博士生导师, 从事神经网络、混沌时间序列等研究; 吕飞(1987—), 男, 博士生, 从事机器学习、遥感图像分类的研究.

参与集成的核极端学习机与实际输出之间的相关性最大,而相互间的冗余性最小,从而较好地平衡准确性和差异性。

1 核极端学习机

极端学习机^[1]具有简单的3层结构:输入层,隐含层和输出层.连接输入层和隐含层的输入权值和隐含层内部偏置值随机生成并在训练过程中保持不变.对于给定的训练样本集 $\{x_i, c_i\}$, $i = 1, 2, \dots, N$, 极端学习机的数学方程为

$$f(x_j) = \sum_{i=1}^H \beta_i g(w_i \cdot x_j + b_i), \quad j = 1, 2, \dots, N. \quad (1)$$

其中: w_i 表示连接输入节点与第 i 个隐含节点的输入权值, b_i 表示第 i 个隐含节点的偏置值, β_i 表示连接第 i 个隐含层节点与输出节点的输出权值, $g(\cdot)$ 表示激活函数. 令

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_L \cdot x_1 + b_L) \\ g(w_1 \cdot x_2 + b_1) & \cdots & g(w_L \cdot x_2 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}, \quad (2)$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_L \end{bmatrix}_{L \times 1}, \quad T = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_L \end{bmatrix}_{L \times 1}. \quad (3)$$

相应的解可以解析地表示为

$$\hat{\beta} = H^\dagger T, \quad (4)$$

其中 H^\dagger 表示矩阵 H 的 Moore-Penrose 氏广义逆.

将极端学习机的隐含层替换为核方法中支持向量机的核函数映射, 发展出一种核极端学习机^[5], 优化问题可以写为

$$\begin{aligned} \min P_{\text{ELM}} &= \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2; \\ \text{s.t. } h(x_i)\beta &= c_i - \xi_i. \end{aligned} \quad (5)$$

核极端学习机的输出函数可以写为

$$f(x) = h(x)\beta = h(x)H^\top \left(\frac{I}{C} + HH^\top \right)^{-1} T. \quad (6)$$

定义极端学习机核矩阵为

$$\begin{aligned} \Omega_{\text{ELM}} &= HH^\top; \\ \Omega_{\text{ELM}i,j} &= h(x_i) \cdot h(x_j) = K(x_i, x_j). \end{aligned} \quad (7)$$

相应的输出函数函数可以重写为

$$f(x) = \begin{bmatrix} K(x, x_1) \\ K(x, x_2) \\ \vdots \\ K(x, x_N) \end{bmatrix}^\top \left(\frac{I}{C} + \Omega_{\text{ELM}} \right)^{-1} T. \quad (8)$$

在该极端学习机的特殊核实现中, 隐层特征映射 $h(x)$ 可以是未知的, 但通常给出相应的核 $K(u, v)$ (例如 $K(u, v) = \exp(-\gamma \|u - v\|^2)$), 因此, 不再需要确定隐层节点数(隐层特征空间的维数).

核极端学习机综合了极端学习机与支持向量机的优势, 是一种十分有效的学习器.

2 基于信息论的选择性集成核极端学习机

2.1 最大相关最小冗余信息论准则

在信息论中, 互信息是一种衡量两个随机变量相互依赖关系的准则. 两个离散随机变量 X 和 Y 的互信息 $I(X; Y)$ 定义为

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \left(\frac{p(x, y)}{p(x)p(y)} \right). \quad (9)$$

其中: $p(x, y)$ 表示变量 X 和 Y 的联合概率密度函数, $p(x)$ 和 $p(y)$ 分别表示 X 和 Y 的边缘概率密度函数.

韩敏等^[13]基于互信息提出了一种同时对输入变量和极端学习机模型结构进行优化的学习算法. 本文基于互信息的概念, 首先引出最大相关最小冗余信息论准则^[12], 然后利用该准则对核极端学习机进行选择集成.

选择出的用于集成的分类器集合 S 与类标 c 的相关性 $D(S, c)$ 可以定义为单个分类器 f_i 与类标 c 的互信息的平均值

$$D(S, c) = \frac{1}{|S|} \sum_{f_i \in S} I(f_i; c), \quad (10)$$

其中 $|S|$ 表示集合 S 中包含的分类器个数.

集合 S 中包含的分类器之间的冗余性定义为集合中分类器 f_i 和分类器 f_j 的所有互信息的平均值

$$R(S) = \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j). \quad (11)$$

最大相关最小冗余信息论准则可以定义为

$$\begin{aligned} \text{mRMR} &= \\ \max_S &\left[\frac{1}{|S|} \sum_{f_i \in S} I(f_i; c) - \frac{1}{|S|^2} \sum_{f_i, f_j \in S} I(f_i; f_j) \right]. \end{aligned} \quad (12)$$

2.2 基于信息论的选择性集成核极端学习机

根据上述分析, 给出基于互信息的选择性集成核极端学习机算法的具体步骤.

Step 1: 将样本集分为两部分, 80% 用于训练, 20% 作为检验样本, 对训练样本采用 Bootstrap 方法产生 n 组规模为 L 的训练样本集

$$\begin{aligned} S &= \{x_{i,k}, y_{i,k}\}, \\ x_{i,k} &\in R^d, y_{i,k} \in R, \\ i &= 1, 2, \dots, n, j = 1, 2, \dots, L. \end{aligned}$$

Step 2: 给定样本集 S , 确定核函数类型并初始化正则化参数, 计算得到极端学习机核矩阵 (7).

Step 3: 利用得到的核矩阵, 求解核极端学习机 (8).

Step 4: 重复 Step 2 和 Step 3, 得到 n 个子核极端学习机模型.

Step 5: 计算 n 个子核极端学习机在检验样本上的预测输出.

Step 6: 计算 n 个预测输出与实际值之间的最大相关最小冗余信息准则 (12).

Step 7: 按照计算得出的最大相关最小冗余信息准则, 对 n 个子核极端学习机进行排序.

Step 8: 逐个增加集成个数, 对排序好的 n 个子核极端学习机进行集成, 采用平均加权算法或者相对多数投票算法求解最终输出, 当集成输出的误差变大时停止.

Step 9: 从 n 个子核极端学习机中选出 n^* 个子核极端学习机, 采用平均加权算法或者相对多数投票算法求解最终输出.

基于互信息的最大相关最小冗余准则可以衡量出子核极端学习机与预测输出之间的相关性和各子核极端学习机之间的冗余性, 采用准则 (12) 使得相关性最大而冗余性最小, 即选择预测准确性高而相互间差异性较大的基核极端学习机进行集成, 因此在减小集成规模的同时可以提高最终的预测精度.

3 仿真实验

为了验证本文所提方法的优越性, 本节在 UCI 基准数据 (<http://archive.ics.uci.edu/ml>) 上, 对极端学习机 (ELM) 和核极端学习机算法 (ELMK) 进行比较. 在实验中, 所有的输入变量均归一化到 $[-1, 1]$, 所有的输出变量均归一化到 $[0, 1]$.

第 1 组仿真实验采用 10 组 UCI 回归数据. 在实际的应用中, 这些数据集合的分布情况是未知的, 且多数是有噪声的. 对于每一种情况, 在每次仿真之前, 训练样本和测试样本都从整个数据集合中随机产生.

核极端学习机的核函数选择高斯核 $K(u, v) = \exp(-\gamma \|u - v\|^2)$, 核宽设置为 10, 正则化参数也设置为 10. 训练样本的 80% 用于训练模型, 剩下的 20% 作为检验样本, 用于确定集成的核极端学习机个数. 每次生成 20 个基核极端学习机进行选择集成.

图 1 所示为核极端学习机在 Abalone 数据上集成个数与测试 RMSE 的关系曲线.

从图 1 可以看出, 采用本文所提的互信息方法对生成的基核极端学习机进行排序, 测试 RMSE 随着集成个数的增加整体上呈现先下降后上升的趋势, 分析原因为: 首先参与集成的是与输出相关较大且相互间冗余最小的核极端学习机, 因此测试 RMSE 呈下降趋

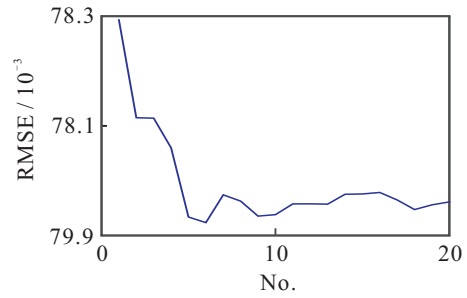


图 1 Abalone 数据的集成 KELM 的个数与 RMSE 势; 随着集成个数的增加, 与输出相关较小或相互间冗余性较大的核极端学习机加入到集成中, 导致泛化误差出现恶化, 这说明了选择性集成的优越性.

表 1 列出了所提出的基于互信息的集成核极端学习机 (SEKELM) 与极端学习 (ELM)、支持向量回归 (SVR) 及集成核极端学习机 (KELM-all) 在 12 组 UCI 回归数据上的仿真结果比较.

表 1 UCI 回归问题测试根均方差 (RMSE) 比较

数据集	ELM[1]	SVR[1]	KELM-all	SEKELM	No.
Abalone	0.0824	0.0784	0.0780	0.0779	6
Delta ailerons	0.0431	0.0429	0.0384	0.0384	8
Delta elevators	0.0568	0.0540	0.0533	0.0533	6
Computer activity	0.0382	0.0470	0.0318	0.0318	4
Census(house8L)	0.0660	0.0746	0.0684	0.0683	6
Auto price	0.0994	0.0937	0.0767	0.0745	4
Triazines	0.2002	0.1829	0.2104	0.1880	8
Machine CPU	0.0539	0.0811	0.0731	0.0661	7
Servo	0.1196	0.1177	0.1146	0.1098	6
Bank	0.0366	0.0467	0.0404	0.0403	9

从表 1 可以看出, 所提基于互信息的选择性集成核极端学习机在所比较的几种算法中取得了最好的预测结果, 与 ELM、SVR 和 KELM-all 相比, 预测 RMSE 更小. 表 1 的最后一列给出了基于互信息的最大相关最小冗余准则选择出的进行集成的基核极端学习机的个数. 可以看出, 选择的集成个数在 6~9 之前, 与生成的全部 20 个基学习器相比, 选择性集成效果明显.

为了进一步验证所提算法的优越性, 将所提算法应用于 UCI 基准数据集中的分类数据集中. 表 2 给出了几种算法在 UCI 分类数据集上的测试分类精度比较结果.

表 2 UCI 分类问题测试分类准确率比较

数据集	ELM[1]	SVR	KELM-all	SEKELM	No.
Diabetes	77.57	76.50	84.03	87.85	4
Breast-cancer	96.61	96.49	99.02	99.61	4
Live-disorders	72.01	83.90	87.21	87.98	3
Sonar	70.29	70.15	74.36	76.28	7

从表 2 可以看出, 所提 SEKELM 算法与 ELM、SVM 和 KELM-all 算法相比, 在 4 组数据上均取得了最好的分类精度. 表 2 的最后一列给出了基于互信息

的最大相关最小冗余准则选择的集成基和极端学习机个数,与原始的 20 个基核极端学习机相比,选择集成的基核极端学习机个数更少。

4 结 论

为了平衡神经网络集成学习中的准确性和差异性,提出了一种基于信息论优化的集成核极端学习机算法。首先,采用 bootstrap 方法生成不同的训练集,在此基础上得到不同的核极端学习机基学习器;然后,采用基于信息论的最大相关最小冗余策略选择与期望输出最相关而相互之间冗余最小的核极端学习机进行选择性集成;针对回归问题,采用简单平均方法进行选择性集成,针对分类问题,采用绝大多数投票方法进行选择性集成。基于实际的 UCI 回归和分类数据的仿真结果表明,所提出的基于信息论的选择性集成核极端学习机具有更好的泛化能力。

参考文献(References)

- [1] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1): 489-501.
- [2] 邓万宇, 郑庆华, 陈琳, 等. 神经网络极速学习方法研究[J]. *计算机学报*, 2010, 33(2): 279-287.
(Deng W Y, Zheng Q H, Chen L, et al. Research on extreme learning of neural networks[J]. *Chinese J of Computers*, 2010, 33(2): 279-287.)
- [3] Huang G B, Wang D H, Lan Y. Extreme learning machines: A survey[J]. *Int J of Machine Learning and Cybernetics*, 2011, 2(2): 107-122.
- [4] Huang G, Huang G B, Song S, et al. Trends in extreme learning machines: A review[J]. *Neural Networks*, 2015, 61: 32-48.
- [5] Huang G B, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification[J]. *IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics*, 2012, 42(2): 513-529.
- [6] Hansen L K, Salamon P. Neural network ensembles[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1990, 12(10): 993-1001.
- [7] Zhou Z H. Ensemble methods: Foundations and algorithms[M]. Boca Raton: CRC Press, 2012: 25-66.
- [8] Zhou Z H, Wu J, Tang W. Ensembling neural networks: Many could be better than all[J]. *Artificial Intelligence*, 2002, 137(1): 239-263.
- [9] 张春霞, 张讲社. 选择性集成学习算法综述[J]. *计算机学报*, 2011, 34(8): 1399-1410.
(Zhang C X, Zhang J S. A survey of selective ensemble learning algorithms[J]. *Chinese J of Computers*, 2011, 34(8): 1399-1410.)
- [10] 陆慧娟, 安春霖, 马小平, 等. 基于输出不一致测度的极限学习机集成的基因表达数据分类[J]. *计算机学报*, 2013, 36(2): 341-348.
(Lu H J, An C L, Ma X P, et al. Disagreement measure based ensemble of extreme learning machine for gene expression data classification[J]. *Chinese J of Computers*, 2013, 36(2): 341-348.)
- [11] 赵立杰, 汤健, 柴天佑. 基于选择性极限学习机集成的磨机负荷软测量[J]. *浙江大学学报: 工学版*, 2011, 45(12): 2088-2092.
(Zhao L J, Tang J, Chai T Y. Soft sensor of mill load based on selective extreme learning machine ensemble[J]. *J of Zhejiang University: Engineering Science*, 2011, 45(12): 2088-2092.)
- [12] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1226-1238.
- [13] 敏韩, 刘晓欣. 一种基于互信息变量选择的极端学习机算法[J]. *控制与决策*, 2014, 29(9): 1576-1580.
(Han M, Liu X X. An extreme learning machine based on mutual information variable selection[J]. *Control and Decision*, 2014, 29(9): 1576-1580.)

(责任编辑: 齐 霖)