

文章编号: 2095-2163(2021)07-0024-08

中图分类号: TP181

文献标志码: A

样本线性化与数据去重的极限学习机

张 灿, 代子彪, 安 鑫, 李建华

(合肥工业大学 计算机与信息学院, 合肥 230009)

摘 要: 对于多标签分类中存在非线性的数据样本和重复的样本数据问题, 本文提出了一种基于在线顺序极限学习机(Online Sequential Extreme Learning Machine, OS-ELM)的改进算法—样本线性化和数据预处理极限学习机(PDC-ELM)。PDC-ELM 算法对线性不可分的数据样本先利用核函数进行处理, 使数据样本具有线性可分的特征, 对于处理后的数据样本, 利用在线顺序极限学习机(OS-ELM)在计算之前对分类数据进行预处理, 即从训练和测试数据集中查找不一样的特征标签并保存类标签中, 实验中新生成的标签组将不具有重复的特征标签, 大大减少了训练的对比次数。实验表明, 相比于其他没有样本线性化和数据预处理的极限学习机模型, 计算的准确度得到很大的提升, 计算时间也有所降低。

关键词: 极限学习机; 核函数; 多标签分类; 多标签数据; 支持向量机

Extreme learning machine for sample linearization and data deduplication

ZHANG Can, DAI Zibiao, AN Xin, LI Jianhua

(School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009, China)

[Abstract] For the problem of nonlinear data samples and duplicate data in multi-label classification, this paper proposes an improved algorithm based on Online Sequential Extreme Learning Machine (OS-ELM) for sample linearization and data preprocessing extreme learning machine (PDC-ELM). The PDC-ELM algorithm first uses the kernel function to process the linearly inseparable data samples, so that the data samples have linearly separable characteristics. For the processed data samples, the Online Sequential Extreme Learning Machine (OS-ELM) is used to preprocess the classified data before calculation. That is, different feature labels from the training and test data sets are searched and saved in the class labels, the newly generated label group in the experiment will not have repeated feature labels, which greatly reduces the number of training comparisons. Experimental results show that compared with other extreme learning machine models without sample linearization and data preprocessing, the calculation accuracy is significantly improved and the calculation time is also reduced.

[Key words] extreme learning machine; kernel function; multi-label classification; multi-label data; support vector machine

0 引 言

极限学习机是 Huang 提出的一种训练单隐层前馈神经网络^[1]的新型机器学习算法。在过去的 20 多年里, 很多学者^[2-5] 已对单隐层前馈神经网络展开深入研究。单隐层前馈神经网络有 2 种主要架构。一种是带有加性隐藏节点的架构, 另一种是带有径向基函数隐藏节点的架构。对于使用单隐层前馈神经网络的许多应用程序来说, 训练方法通常是批处理学习类型的。多数情况下, 批处理学习是一个耗时的事情, 因为其中可能涉及到对训练数据的多次迭代。在大多数应用程序中, 这可能需要几分钟到几个小时, 并且必须正确地选择学习参数, 以确保收敛。此外, 每当接收到新数据时, 批处理学习都会将历史数据和新数据一起使用, 并进行再训练, 因

此会消耗大量时间。在许多工业应用中, 在线顺序学习算法比批处理学习算法更受欢迎, 因为顺序学习算法在接收新数据时不需要再训练。

与此同时, 许多研究者将经典的极限学习机扩展到核学习, 证明了极限学习机不仅适用于 Sigmoid 网络和 RBF 网络^[6] 等经典的特征映射, 还适用于多种类型的特征映射。在对类别分布不平衡的数据进行分类时, Priya 等人^[7-8] 提出了一种统一的加权极限学习机方法(weighted ELM)来处理复杂的数据分布, 即在其中对每个样本点额外赋予一个权重。然而, 上文论述的极限学习机方法并没有将样本线性化和数据预处理数据添加到操作中。

极限学习机算法在泛化性能上也有较高的效率和精度^[9]。在线顺序极限学习机^[10-12] (OS-ELM) 算法是基于极限学习机框架的在线学习版本。在线

基金项目: 国家自然科学基金青年基金(61402145, 61673156)。

作者简介: 张 灿(1995-), 男, 硕士研究生, 主要研究方向: 数据挖掘; 代子彪(1994-), 男, 硕士研究生, 主要研究方向: 情感语音合成; 安 鑫(1986-), 男, 博士, 副教授, 主要研究方向: 情感计算、语音合成; 李建华(1985-), 男, 博士, 副研究员, 主要研究方向: 键值存储、预取。

通讯作者: 李建华 Email: jhli@hfut.edu.cn

收稿日期: 2021-04-07

顺序极限学习机不仅可以逐个地学习训练数据,而且可以逐块地学习,丢弃已经训练过的数据。在以下意义上,这就是一个通用的顺序学习算法。将训练观察结果按顺序,逐个地或块接块,呈现给学习算法。在任何时候,只有新到达的单个或大块的观察,而不是整个过去的的数据被看到和学习。一旦某一特定、单个或组块,观察的学习过程结束,就丢弃单个或一组训练观察。在线顺序极限学习机起源于批量学习极限学习机^[13-16],为具有加性和径向基函数节点的单隐层前馈神经网络开发。在函数回归和分类领域的一些基准问题上对极限学习机的性能进行了评估。结果表明,与其他基于梯度下降的学习算法(包括反向传播^[17-18]算法)相比,极限学习机在更高的学习速度下提供了更好的泛化性能,许多应用的学习阶段在秒内即可完成^[19]。

需要一提的是,确定标签子集时,在线顺序极限学习机却忽略了样本的特征。一方面,由标签子集转换而来的类可能具有较低的可分离性;另一方面,一些标签组合会导致分布高度不平衡^[20],导致现有的单标签学习算法难以应用。这对这些不足,需要解决2个问题:

(1)如何挖掘样本特征,设计有效的指标来评价标签子集的质量。

(2)如何基于质量度量从大量候选子集中实现快速选择。

针对问题(1),研究中采用基于线性判别比的测度来评价主标签类的可分性,再使用联合熵来描述数据的不平衡程度。由于核技术在特征空间中更容易分析可分性,因此可采用核支持向量机(kernel support vector machine, SVM)作为基础学习器^[21-22]。对于问题(2),应在极限学习机中预先从训练和测试数据集中查找不同的标签并保存类标签中,如此一来,在时间和效率上都得到了很大的提升。

极限学习机算法在单标签分类^[23]中表现出更优异的性能。基于极限学习机算法的多标签分类^[24-25]研究相对较少。针对多标签分类问题, Sun等人^[25]提出了一种基于阈值法的极限学习机算法。此外,文献^[7]也提出了一种多标签分类的核极限学习机算法。然而,现有的基于极限学习机的多标签分类方法并没有考虑到样本线性化和样本数据预处理。本文中介绍了一种新的基于在线顺序极限学习机来处理多标签分类问题的改进算法,即样本线性化和数据预处理极限学习机(PDC-ELM)。本文

对多标签分类方案的主要贡献如下:

(1)针对线性不可分的数据样本,先采用核函数进行处理,使之线性可分。

(2)对于处理后的数据样本,为了减少计算量,预先从训练数据集和测试数据集中查找不一样的标签并保存类标签中。

(3)将多标签分类问题分解为单独的单标签分类问题,就是将多标签问题转化为一组单标签问题。

(4)在极限学习机算法中,输出权值由最小二乘法确定。

本文组织如下:首先,介绍了相关背景工作,包括多标签分类和极限学习机的探讨。其次,研究相关工作有在线顺序极限学习机与核函数。然后,详细分析了PDC-ELM相关操作与处理,包括动机、优化的预处理、算法分析和处理过程。接下来,给出了实验相关数据和比较结果。最后,是本文的结论部分。

1 背景

1.1 多标签分类

传统的监督学习是著名的机器学习范式之一,其中每个例子由一个实例和一个标签组成。然而,在多标签应用程序中,每个例子也由一个单独的实例表示,同时与一组标签相关联,多标签问题更符合现实世界的应用,同时也给监督学习领域带来了挑战^[26]。

多标签分类的目的是将一个实例 $x \in R^d$ 映射到标签集 $Y \subseteq L = \{1, 2, \dots, k\}$,其中 k 是类的数量。多标签分类的任务是根据训练集 $D = \{(x_i, y_i) \mid i = 1, 2, \dots, n\}$ 构造一个函数 $f(\cdot): f(X) \rightarrow Y$ 。对于任何看不见的实例 $x \in X$,多标签分类器 $f(\cdot)$ 将 $f(x) \subseteq y$ 预测为 x 的正确标签集。

1.2 极限学习机

一般情况下,给定一个由 N 个任意不同样本组成的训练集 $s = \{(x_i, t_i) \mid x_i \in R^n, t_i \in R^m, i = 1, 2, \dots, N\}$,具有 \tilde{N} 个隐节点的单隐层前馈网络函数可以表示为:

$$f_{\tilde{N}} = \sum_{j=1}^{\tilde{N}} \beta_j G(a_j, b_j, x_i) = t_i, \quad i = 1, 2, \dots, N \quad (1)$$

其中, a_j 和 b_j 表示随机确定的学习参数; β_j 表示输出权重矩阵连接第 j 个隐藏输出节点; x_i 和 t_i 分别是第 i 个观测的特征和输出; $G(a_j, b_j, c_j)$ 是一个满足极限学习机通用逼近能力定理的非线性分段连续函数。

2 在线顺序极限学习机

在初始阶段,假设有 N_0 个任意的训练样本 (X_i, t_i) , 其中 $X_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in R^n$, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$ 。利用基本的极限学习机算法的思想,希望求得满足 $\|H_0\beta - T_0\|$ 最小的 β_0 , 运算时需用到:

$$H_0 = \begin{pmatrix} g(a_1, b_1, X_1) & \cdots & g(a_{N_0}, b_{N_0}, X_1) \\ \vdots & \ddots & \vdots \\ g(a_1, b_1, X_{N_0}) & \cdots & g(a_{N_0}, b_{N_0}, X_{N_0}) \end{pmatrix}_{N_0 \times N_0} \quad (2)$$

$$T_0 = \begin{pmatrix} t_1^T \\ \vdots \\ t_{N_0}^T \end{pmatrix}_{N_0 \times m} \quad (3)$$

此时,根据广义逆的计算方法,可以计算出 β_0 。

3 PDC-ELM

3.1 动机

相比于代表性的极限学习机模型(包括 ELM, L21-ELM^[27]),在线顺序极限学习机有比较明显的优势,在线顺序极限学习机不仅可以逐个地学习训练数据,而且可以逐块地学习,丢弃已经训练过的数据。在以下操作上,这就是一个通用的顺序学习算法。对此可做详述如下。

(1)将训练观察结果按顺序,逐个或块接块,呈现给学习算法。

(2)在任何时候,只有新到达的单个或大块的观察,而不是整个过去的的数据,被看到和学习。

(3)如果某一特定单个或组块的观察学习过程完成,就丢弃单个或一组训练观察。

在线顺序极限学习机虽然有以上的优势,但却未能很好地去处理非线性样本。与之相对应,核函数的优势恰在于此,对其特点可概述如下。

(1)核函数的引入避免了复杂维数,大大减小了计算量。而输入空间的维数 n 对核函数矩阵并无影响,因此,核函数方法可以有效处理高维输入。

(2)无需知道非线性变换函数 ϕ 的形式和参数。

(3)核函数的形式和参数的变化会隐式地改变从输入空间到特征空间的映射,如此即会对特征空间的性质产生影响,从而改变各种核函数方法的性能。

(4)核函数方法能和不同的算法相结合,形成多种不同的基于核函数技术的方法,且这2部分的设计可以单独进行,并可以为不同的应用选择不同的核函数和算法。

综上所述可知,若是将核函数的自身优势与在线顺序极限学习机相结合,不仅能够减少计算量,而且在时间效率上也会获得提升。

3.2 预处理分析

对于非线性数据样本通过核函数处理使样本线性化,降低后续在线顺序极限学习机处理的计算量。通常训练样本数据较大,实验样本中很多处理后的特征值相同,文中采用比较和统计的方法,对分类的数据进行预处理,即从训练集和测试集中找到不同的标签,存储到类标签中,这样操作即可把所有不同的标签类存储到一起,成为一个标签组,而数量远低于原来的数据样本容量。

3.3 算法

为了解决样本数据中比较重复的问题,研究提出一种新的基于在线顺序极限学习机的算法来处理多标签分类问题的改进算法、即样本线性化和数据预处理极限学习机(PDC-ELM),以适应多标签的应用。对于线性不可分的数据样本,PDC-ELM 算法先利用核函数进行处理,使数据样本具有线性可分的特征,而对于处理后的数据样本,可利用在线顺序极限学习机在计算前对分类数据进行预处理,也就是从训练和测试数据集中查找不同的特征标签并保存类标签中。形成一个新的和没有重复的样本标签组,为后续的比较实验提供便利。PDC-ELM 算法的研发实现步骤可参考算法1。

算法1 PDC-ELM 算法

输入 训练集 $\{(x_i, y_i) | x_i \in \mathfrak{R}^n, y_i \in \mathfrak{R}^n, i = 1, 2, \dots, N\}$, 测试集 $\{(x_j, y_j) | x_j \in \mathfrak{R}^n, y_j \in \mathfrak{R}^n, j = 1, 2, \dots, N\}$, 激活函数 $g(\cdot)$, 和隐藏层节点数 p

输出 训练和测试的分类准确度和时间

Step 1 对样本数据使用相关核函数算法。

Step 2 预处理分类数据,

for 训练集中第1个到第 N 个实例执行。

Step 3 从训练和测试数据集找到不同标签并保存在类标签中。

Step 4 处理训练目标和测试目标,生成隐藏层节点的输入权重 w_i 和偏差 b_i 。

Step 5 计算隐神经元输出矩阵 H 。

Step 6 end for

对于实验中的数据样本处理,共设置了5个参

数。其中,2个参数是从数据样本中挑选出来的测试集和训练集。数据样本需要进行多次训练、多次试验,增加实验的可靠性。再有一个参数是极限学习机训练类型。分类则设置为1,在实验中,0表示回归,1表示分类,考虑到实验就是处理分类问题,故设定为1;另外,还有一个参数是隐藏层节点数目,即 p 。隐藏层节点数目的不同会影响到训练时间、测试时间、训练准确度和测试准确度,在不断的调试中找到最佳的隐藏层节点数目,使得实验结果最好;最后一个参数是激活函数,即 $g(\cdot)$,根据数据样本需要设置好最恰当的激活函数。

4 实验

为了更好地评价 PDC-ELM 模型的性能,本节从实验环境、数据集、代表性的 ELM 模型比较、性能评估和实验结果对比进行分析。为了更好地呈现模型的数据比较,在表与图上分别从相关数据对比上进行研究。

4.1 运行环境

试验运行环境为 Windows 10 企业版,处理器为 Inter(R) Core (TM) i5-8400 CPU@ 2.80 GHz,内存 16.0 GB,64 位操作系统。实验模型调试、运行、结果对比和图形绘制在 Matlab 2016 中。

4.2 实验数据集

本实验在扩展 iris、abalone、wine 和 yeast 4 个数据集上进行。4 组数据集将用于本文的优化模型 PDC-ELM 与代表性的 ELM 模型(包括 ELM、L21-ELM^[25]和 OS-ELM^[10])的效果比较。

4.3 比较模型

本实验拟与代表性的 ELM 模型(包括 ELM、L21-ELM 和 OS-ELM)进行比较。其中,ELM 是一种具有竞争力的机器学习方法,也是基于最小二乘的学习算法,可为分类和回归应用提供高效统一的学习解决方案。ELM 就是一种针对单隐含层前馈神经网络的神经网络的算法,训练时的目的旨在寻得输出层的权值求解。ELM 的显著特点就是输入权值和隐含节点的偏置都是在给定范围内随机生成的,现已证得其学习效率高、且泛化能力强,并已广泛应用于分类、回归、聚类、特征学习等问题中。但是 ELM 鲁棒性较差,稳定性也差。L21-ELM 是一种低秩正则化极限学习机,将提取的特征与 ELM 上的相关特征进行低秩约束。首先,在低秩重构过程中通过对重构矩阵施加具有旋转不变性的 $L_1(2,1)$

模约束,可在挖掘目标域数据的关键特征的同时提高算法对不同姿态图片分类的鲁棒性。其次,在目标函数中引入结构的正则化,使得迁移时数据中的局部几何结构信息得以充分利用,进一步提高了分类性能。最后,为解决源域数据较少带来的欠完备特征空间覆盖问题,在公共子空间中利用源域数据和目标域数据联合构造字典,保证了重构的鲁棒性。但却很容易过拟合,求解效果欠佳。OS-ELM 是一种快速、准确的在线序列学习算法,可以递增地添加隐藏节点,并可按块大小变化或固定的块大小逐块地输入训练数据。OS-ELM 分为 2 个部分。一部分为通过少量的训练样本,利用 ELM 算法计算并初始化输出权重;另一部分开始在线学习,每当一个新的数据样本到来时,通过一个递推公式得到新的输出权重,从而实现在线且快速的训练。OS-ELM 具备了 ELM 的速度和泛化能力上的优点,并且可以随着新数据的到来不断更新模型,而不是重新训练模型。但是对于样本重复数据较多和非线性化样本处理却并未臻至完善。此外,为了与传统方法进行仿真对比,还对 SVM^[21-22]模型进行了研究。其主要思想即为找到空间中的一个能够将所有数据样本划开的超平面,并且使得样本集中所有数据到这个超平面的距离最短,虽然通过使用核函数可以向高维空间进行映射,使用核函数对于解决非线性的分类是有着长足优势的,不仅思想简单,分类效果也较好,但是对于大规模数据训练却比较困难,也无法直接支持多分类,而只能使用间接的方法来做。

4.4 实验结果与分析

在本节中,研究给出了 PDC-ELM 算法在多标签分类应用中的仿真结果。研发所得数据详见表 1~表 4。

图 1 表示实验模型 PDC-ELM 分别在 iris 和 abalone 数据集下的训练准确度和测试准确度的变化。图 1 中,红线表示训练结果,蓝线表示测试结果。由图 1 可以看出,随着隐藏层节点数目的增加,PDC-ELM 的训练准确性也有所增加。隐藏层节点数越多,就能够提取更多的特征,使得准确度更高。ELM 模型对任何非线性函数都具有普遍的逼近能力。随着隐藏层节点数目的增加,预测误差趋于零。图 1 给出了随着 ELM 隐含层隐藏节点数增加分类精度的变化情况。在本实验中,隐藏节点的数量从 100 增加到 1 000,步长增幅为 100。分析图 1 可知,随着隐藏层节点数目的增加,预测精度也逐渐提高。

表1 平均准确度 (%) \pm abalone 数据集上不同数量隐藏层节点的标准差Tab. 1 The average accuracy (%) \pm standard deviation with different number of hidden neurons on the abalone dataset

隐藏层节点数	SVM	ELM	L21-ELM	OS-ELM	PDC-ELM
100	60.02 \pm 1.66	50.12 \pm 1.35	51.01 \pm 1.27	51.12 \pm 2.96	51.93 \pm 2.26
200	60.02 \pm 1.66	50.51 \pm 0.57	51.99 \pm 1.74	50.51 \pm 1.99	52.81 \pm 2.72
300	60.02 \pm 1.66	50.73 \pm 0.90	52.31 \pm 1.65	51.73 \pm 2.35	53.20 \pm 2.04
400	60.02 \pm 1.66	46.76 \pm 0.77	57.45 \pm 1.98	47.76 \pm 2.63	58.37 \pm 2.00
500	60.02 \pm 1.66	45.65 \pm 0.70	57.66 \pm 1.73	46.65 \pm 2.05	58.98 \pm 1.64
600	60.02 \pm 1.66	46.09 \pm 0.59	50.23 \pm 1.96	49.09 \pm 2.54	59.39 \pm 1.79
700	60.02 \pm 1.66	46.65 \pm 1.72	40.11 \pm 1.14	47.66 \pm 1.53	59.95 \pm 1.06
800	60.02 \pm 1.66	50.42 \pm 1.05	30.26 \pm 1.04	51.42 \pm 2.03	60.33 \pm 1.12
900	60.02 \pm 1.66	46.97 \pm 1.74	20.11 \pm 1.65	47.97 \pm 1.85	60.87 \pm 1.14
1 000	60.02 \pm 1.66	47.76 \pm 1.47	16.55 \pm 0.52	48.78 \pm 2.55	62.46 \pm 1.61

表2 平均准确度 (%) \pm iris 数据集上不同数量隐藏层节点的标准差Tab. 2 The average accuracy (%) \pm standard deviation with different number of hidden neurons on the iris dataset

隐藏层节点数	SVM	ELM	L21-ELM	OS-ELM	PDC-ELM
100	98.32 \pm 0.43	95.41 \pm 3.45	95.21 \pm 3.94	95.32 \pm 3.88	95.44 \pm 3.76
200	98.32 \pm 0.43	97.68 \pm 3.80	97.34 \pm 3.32	97.53 \pm 3.24	97.41 \pm 3.44
300	98.32 \pm 0.43	97.79 \pm 2.95	97.76 \pm 3.23	97.91 \pm 3.35	97.85 \pm 3.24
400	98.32 \pm 0.43	98.12 \pm 2.02	97.81 \pm 3.10	97.76 \pm 3.39	98.18 \pm 3.09
500	98.32 \pm 0.43	97.91 \pm 3.25	98.01 \pm 3.36	98.06 \pm 3.26	98.12 \pm 2.27
600	98.32 \pm 0.43	98.15 \pm 3.28	97.99 \pm 3.24	97.82 \pm 3.16	98.18 \pm 2.01
700	98.32 \pm 0.43	98.21 \pm 2.20	98.07 \pm 3.10	97.94 \pm 3.05	98.26 \pm 2.11
800	98.32 \pm 0.43	99.59 \pm 0.48	98.87 \pm 1.98	99.47 \pm 0.44	99.76 \pm 0.33
900	98.32 \pm 0.43	99.53 \pm 0.44	99.01 \pm 1.45	99.59 \pm 0.42	99.76 \pm 0.32
1 000	98.32 \pm 0.43	99.65 \pm 0.54	99.16 \pm 0.44	99.35 \pm 0.33	99.71 \pm 0.42

表3 平均准确度 (%) \pm yeast 数据集上不同数量隐藏层节点的标准差Tab. 3 The average accuracy (%) \pm standard deviation with different number of hidden neurons on the yeast dataset

隐藏层节点数	SVM	ELM	L21-ELM	OS-ELM	PDC-ELM
100	79.46 \pm 2.01	71.75 \pm 1.38	69.55 \pm 0.75	70.56 \pm 1.33	70.04 \pm 1.55
200	79.46 \pm 2.01	72.89 \pm 1.65	70.19 \pm 1.17	71.75 \pm 1.51	72.92 \pm 1.44
300	79.46 \pm 2.01	71.28 \pm 1.69	72.67 \pm 1.10	69.40 \pm 1.22	73.15 \pm 1.01
400	79.46 \pm 2.01	69.17 \pm 1.48	74.78 \pm 1.06	65.70 \pm 1.17	76.38 \pm 1.54
500	79.46 \pm 2.01	59.56 \pm 1.58	76.83 \pm 1.01	56.77 \pm 0.99	79.33 \pm 1.67
600	79.46 \pm 2.01	57.91 \pm 0.98	59.66 \pm 2.10	54.17 \pm 1.19	78.56 \pm 0.89
700	79.46 \pm 2.01	56.60 \pm 0.74	40.35 \pm 0.89	53.13 \pm 1.43	80.69 \pm 0.65
800	79.46 \pm 2.01	55.59 \pm 1.57	30.13 \pm 0.56	57.29 \pm 0.88	79.50 \pm 0.57
900	79.46 \pm 2.01	54.53 \pm 1.66	20.34 \pm 0.67	56.25 \pm 0.46	79.58 \pm 0.44
1 000	79.46 \pm 2.01	55.65 \pm 1.83	17.26 \pm 1.02	54.77 \pm 1.02	79.71 \pm 0.37

表 4 平均准确度 (%) ±wine 数据集上不同数量隐藏层节点的标准差

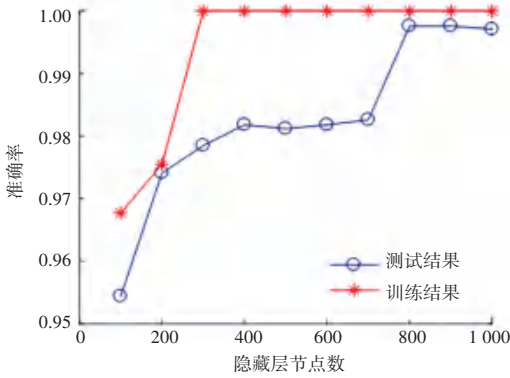
Tab. 4 The average accuracy (%) ±standard deviation with different number of hidden neurons on the wine dataset

隐藏层节点数	SVM	ELM	L21-ELM	OS-ELM	PDC-ELM
100	67.12±1.01	52.66±1.85	55.78±1.15	52.86±1.96	62.64±1.49
200	67.12±1.01	53.30±1.62	57.89±1.35	53.85±2.24	61.54±1.27
300	67.12±1.01	55.28±1.71	59.34±1.67	57.25±1.64	63.86±1.85
400	67.12±1.01	59.17±0.89	60.66±1.37	59.34±2.05	61.65±1.07
500	67.12±1.01	55.56±1.52	63.32±1.44	50.55±1.49	66.04±1.28
600	67.12±1.01	49.91±1.05	64.01±0.98	54.17±1.69	67.45±1.64
700	67.12±1.01	45.60±1.47	50.11±0.67	46.26±2.07	67.14±2.07
800	67.12±1.01	42.59±1.90	40.52±0.41	43.69±1.01	67.25±0.82
900	67.12±1.01	44.53±1.95	30.01±0.26	50.44±1.45	67.96±1.63
1 000	67.12±1.01	45.65±2.05	19.23±1.24	49.15±0.83	67.86±0.44

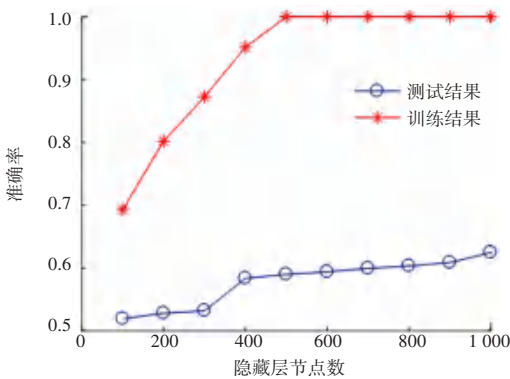
表 5 在 1 000 个隐藏层节点下 4 个数据集的耗时比较

Tab. 5 The time-consuming comparison on four datasets with 1 000 neurons

数据集名称	SVM	ELM	L21-ELM	OS-ELM	PDC-ELM
iris	0.70	0.07	3.12	0.15	0.09
abalone	2.45	0.68	3.89	0.46	0.37
wine	2.97	0.47	4.76	0.41	0.33
yeast	3.20	0.34	4.81	0.56	0.45



(a) iris 数据集



(b) abalone 数据集

图 1 基于不同数据集的隐藏节点数量的分类准确率

Fig. 1 The classification accuracy based on the number of hidden nodes in different data sets

耗时结果对比见表 5 所示。由表 5 可以看出,在这种情况下,PDC-ELM 算法的学习速度比其他方法要快。分析可知,在预处理时进行了去重,也就是过滤了重复的特征标签,减少了比较次数,但在数据较少时其优势并不明显。或者重复数据较少时,优势也并不突出。在今后的工作中需继续提高 PDC-ELM 的训练性能。

本次研究中,在选取的各个数据集上进行实验,仿真后得到的平均识别率如图 2~图 5 所示。由图 2~图 5 可知,PDC-ELM 明显优于其他代表 ELM 模型,证明了 PDC-ELM 采用预处理标签数据的方法应用在多标签分类上是有效的,也是有优势的。随着隐藏层节点数目的增加,PDC-ELM 的准确性也在增加,一定程度上隐藏层节点数越多,越多的特征即被隐藏层提取在图 2 和图 3 中。当隐藏层节点数量小于 300 时,PDC-ELM 的结果比其他 ELM 方法比较接近,究其原因即在于较少的神经元不能有效地提取良好的特征。PDC-ELM 要优于 OS-ELM 方法,根本原因在于:一方面,PDC-ELM 进行了样本线性化处理;另一方面将特征选择在测试集和训练集进行了预处理,有明显提升的效果。L21-ELM 优于原始的 ELM 方法,根本原因在于 L21-ELM 将特征选择融入到线性回归模型中,这样一来分类提取特征就比较方便,可以更好地描述输入数据。但从图 4 和图 5 可以看出,当神经元数量超过 600 个时,L21-ELM 的性能急剧下降,分析原因则可能是隐藏层节点太多导致了算法的过拟合。由图 2 还可以看出,也许是因为数据及样本容量的影响,与其他 3 组数据集比较而言,iris 数据集的样本容量相对来说较小,可能造成 PDC-ELM 模型无法充分地发挥自身

优势,数据样本容量较少,样本的特征较为分散,预处理的训练集和测试集的特征效果也未得到很好的呈现,造成的后果就是多组训练模型的测试结果较为接近,无法更好地突出自身的优势。在隐藏层神经元较少的情况下,SVM 优于大多数 ELM 方法,然而,随着隐藏层神经元数量的增加,本文提出的 PDC-ELM 优于 SVM,由图 4 和图 5 可以看出,当隐藏神经元数量大于为 700 时,只有本文研发的模型获得了最好的结果。

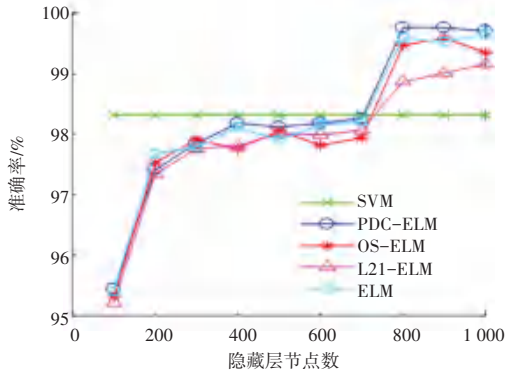


图 2 在 iris 数据集,实验的平均识别率

Fig. 2 Average experimental recognition rate in iris dataset

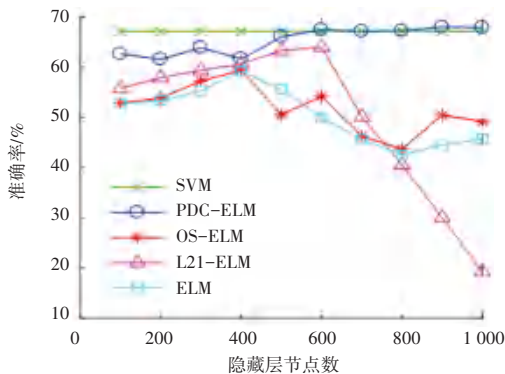


图 3 在 wine 数据集,实验的平均识别率

Fig. 3 Average experimental recognition rate in wine dataset

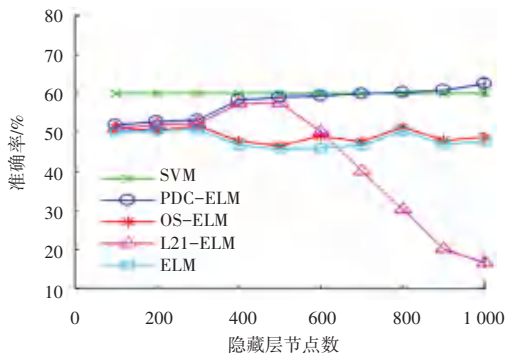


图 4 在 abalone 数据集,实验的平均识别率

Fig. 4 Average experimental recognition rate in abalone dataset

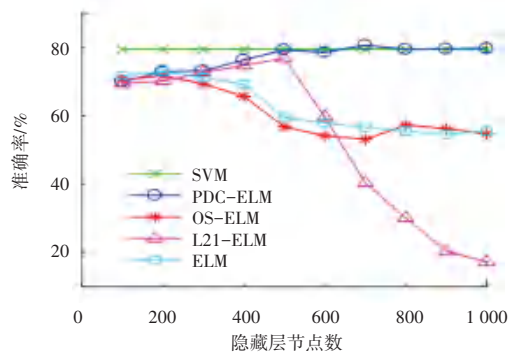


图 5 在 yeast 数据集,实验的平均识别率

Fig. 5 Average experimental recognition rate in yeast dataset

5 结束语

在本文中,提出了一种基于在线顺序极限学习机的多标签分类改进方法,即 PDC-ELM,对非线性样本通过核函数进行线性处理,再进行样本特征数据预处理,与其他 ELM 方法相比,不仅降低了计算量,还能提高准确率,在多标签分类实验中有明显的效果。

参考文献

- [1] SHEN Guorui, YUAN Ye. On theoretical analysis of single hidden layer feedforward Neural Networks with relu activations [C]// 2019 34th Youth Academic Annual Conference of Chinese Association of Automation (YAC). Jinzhou, China; IEEE, 2019: 706-709.
- [2] LEE D, KIM S, TAK S, et al. Real-time feed-forward Neural network-based forward collision warning system under cloud communication environment[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(12): 4390-4404.
- [3] ENE A, STIRBU C. A Java application for the failure rate prediction using feed forward neural networks [C]// 2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). Ploiesti, Romania; IEEE, 2016: 1-4.
- [4] ENE A, STIRBU C. A Java simulation software for the study of the effects of the short-circuit faults in a feed forward neural network [C]// 2015 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). Bucharest, România; IEEE, 2015: 1-4.
- [5] VENKADESAN A, BHAVANA G, HANEESHA D, et al. Comparison of feed forward and cascade neural network for harmonic current estimation in power electronic converter [C]// 2017 International Conference on Innovative Research in Electrical Sciences (IICIRES). Nagapattinam, India; IEEE, 2017: 1-5.
- [6] VIPANI R, HORE S, BASAK S, et al. Gait signal classification tool utilizing Hilbert transform based feature extraction and logistic regression based classification [C]// 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). Kolkata, India; IEEE, 2017: 57-61.
- [7] PRIYA S, MANAVALAN R. A hybrid classifier using weighted ELM with biogeography based optimization for hepatitis diagnosis

- [C]//2019 International Conference on Intelligent Computing and Control Systems (ICCS). Madurai, India; IEEE, 2019; 1463–1467.
- [8] PRIYA S, MANAVALAN R. Optimum parameters selection using ACOR algorithm to improve the classification performance of weighted extreme learning machine for hepatitis disease dataset [C]// 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). Coimbatore; IEEE, 2018; 986–991.
- [9] MA Liancai, LI Jun. Short-term wind power prediction based on multiple kernel extreme learning machine method [C]// 2020 7th International Forum on Electrical Engineering and Automation (IFEAA). Hefei, China; IEEE, 2020; 871–874.
- [10] CHUPONG C, PLANGKLANG B. Comparison study on artificial Neural Network and online sequential extreme learning machine in regression problem [C]//2019 7th International Electrical Engineering Congress (iEECON). Hua Hin, Thailand; IEEE, 2019; 1–4.
- [11] LIN Jun, ZHANG Qinqin, SHENG Gehao, et al. Prediction system for dynamic transmission line load capacity based on PCA and online sequential extreme learning machine [C]//2018 IEEE International Conference on Industrial Technology (ICIT). Lyon, France; IEEE, 2018; 1714–1717.
- [12] CHEN Yita, CHUANG Yuchuan, WU A Y. Online extreme learning machine design for the application of federated learning [C]//2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS). Genova, Italy; IEEE, 2020; 188–192.
- [13] VANLI N D, SAYIN M O, DELIBALTA I, et al. Sequential nonlinear learning for distributed multiagent systems via extreme learning machines [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(3): 546–558.
- [14] LU Xinjiang, LIU Chang, HUANG Minghui. Online probabilistic extreme learning machine for distribution modeling of complex batch forging processes [J]. IEEE Transactions on Industrial Informatics, 2015, 11(6): 1277–1286.
- [15] TANG Jiexiong, DENG Chenwei, HUANG Guangbin. Extreme learning machine for multilayer perceptron [J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(4): 809–821.
- [16] CILIMKOVIC M. Neural networks and back propagation algorithm [D]. North Dublin: Institute of Technology Blanchardstown, 2015.
- [17] PRIANDANA K, ABIYOGA I, WULANDARI, et al. Development of computational intelligence-based control system using backpropagation neural network for wheeled robot [C]// 2018 International Conference on Electrical Engineering and Computer Science (ICECOS). Pangkal, Indonesia; IEEE, 2018; 101–106.
- [18] MOGHADAM M H, SAADATMAND M, BORG M, et al. Learning-based response time analysis in real-time embedded systems; A simulation-based approach [C]//2018 IEEE/ACM 1st International Workshop on Software Qualities and their Dependencies (SQUADE). Gothenburg, Sweden; ACM, 2018; 21–24.
- [19] CAO Jiuwen, WANG Tianlei, SHANG Luming, et al. A novel distance estimation algorithm for periodic surface vibrations based on frequency band energy percentage feature [J]. Mechanical Systems and Signal Processing, 2018, 113: 222–236.
- [20] WU Xiaohe, ZUO Wangmeng, LIN Liang, et al. F-SVM: Combination of feature transformation and SVM learning via convex relaxation [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(11): 5185–5199.
- [21] YAPP E K Y, LI X, LU W F, et al. Comparison of base classifiers for multi-label learning [J]. Neurocomputing, 2020; 394: 51–60.
- [22] CAHYANI D E, NUZRY K A P. Trending topic classification for single-label using multinomial Naive Bayes (MNB) and multi-label using K-Nearest Neighbors (KNN) [C]//2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). Yogyakarta, Indonesia; IEEE, 2019; 547–557.
- [23] 郑伦川, 邓亚平. 一种获取标签相关信息的多标签分类方法 [J]. 计算机工程与应用, 2016, 52(14): 156–160.
- [24] 李锋, 杨有龙. 基于标签特征和相关性的多标签分类算法 [J]. 计算机工程与应用, 2019, 55(4): 48–55.
- [25] SUN Xia, XU Jingting, JIANG Changmeng, et al. Extreme learning machine for multi-label classification [J]. Entropy, 2016, 18(6): 225.
- [26] JIANG Mingchu, LI Na, PAN Zhisong. Multi-label text categorization using L_{21} -norm minimization extreme learning machine [M]//CAO J, MAO K, WU J, et al. Proceedings of ELM-2015 Volume 1. Proceedings in Adaptation, Learning and Optimization. Cham: Springer, 2015, 6: 121–133.
- [27] HUANG Gao, HUANG Guangbin, SONG Shiji, et al. Trends in extreme learning machines: A review [J]. Neural Networks; The Official Journal of the International Neural Network Society, 2015, 61: 32–48.

(上接第23页)

- [15] 孙力娟, 陈小东, 韩崇, 等. 一种新的数据流模糊聚类方法 [J]. 电子与信息学报, 2015, 37(7): 1620–1625.
- [16] 于爽, 李淑梅. 图书电子资源信息分布式多模块检索仿真 [J]. 计算机仿真, 2020, 37(6): 439–442.
- [17] LIU Jinliang, XIAO Liang, LIU Guolong, et al. Active authentication with reinforcement learning based on ambient radio signals [J]. Multimedia Tools and Applications, 2017, 76(3): 3979–3998.
- [18] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述 [J]. 计算机学报, 2018, 41(1): 1–27.
- [19] 姚哲, 陶剑文. 多源适应多标签分类框架 [J]. 计算机工程与应用, 2017, 53(7): 88–96, 170.
- [20] 赵国锋, 黎军, 王新恒, 等. 基于SDN的WLAN负载感知切换方案实现 [J]. 计算机应用研究, 2019, 36(5): 180–184.
- [21] ROWSHANRAD S, NAMVARASL S, ABDI V, et al. A survey on SDN, the future of networking [J]. Journal of Advanced Computer Science and Technology, 2014, 3(2): 232–248.
- [22] 田保军, 刘爽, 房建东. 融合主题信息和卷积神经网络的混合推荐算法 [J]. 计算机应用, 2020, 40(7): 1901–1907.
- [23] SHU Jiangbo, SHEN Xiaoxuan, LIU Hai, et al. A content-based recommendation algorithm for learning resources [J]. Multimedia Systems, 2018, 24(2): 163–173.