中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION

融合深度特征与多核学习的LSTWSVM及其工业应用

刘颖, 刘德彦, 吕政, 赵珺, 王伟

引用本文:
刘颖,刘德彦,吕政,赵,王伟. 融合深度特征与多核学习的LSTWSVM及其工业应用[J]. 控制与决策, 2024, 39(8): 2622−2630.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.1746

您可能感兴趣的其他文章

Articles you may be interested in

l_p−范数约束下MKL−OC−ELM的装备故障检测

MKL and OC−ELM fault detection based on l_p−norm constraint

控制与决策. 2021, 36(10): 2379−2388 https://doi.org/10.13195/j.kzyjc.2020.0443

基于深度强化学习与迭代贪婪的流水车间调度优化

Scheduling optimization for flow−shop based on deep reinforcement learning and iterative greedy method

控制与决策. 2021, 36(11): 2609−2617 https://doi.org/10.13195/j.kzyjc.2020.0608

基于迁移学习灰支持向量回归机的交互式进化计算

Interactive evolutionary computation based on transfer learning grey support vector regression

控制与决策. 2021, 36(10): 2399−2408 https://doi.org/10.13195/j.kzyjc.2020.0420

基于深度时序特征迁移的轴承剩余寿命预测方法

Remaining useful life prediction of bearing based on deep temporal feature transfer

控制与决策. 2021, 36(7): 1699−1706 https://doi.org/10.13195/j.kzyjc.2019.1809

基于改进堆叠自动编码器的循环冷却水系统工艺介质温度预测控制方法

Predictive control method of process medium temperature in circulating cooling water system based on improved stacked auto encoders

控制与决策. 2020, 35(12): 2835−2844 https://doi.org/10.13195/j.kzyjc.2019.0694

融合深度特征与多核学习的LSTWSVM及其工业应用

刘 颖, 刘德彦, 吕 政†, 赵 珺, 王 伟

(大连理工大学 控制科学与工程学院，辽宁 大连 116024)

摘 要: 为了提高多核学习(MKL)的表示能力同时降低其计算成本,提出一种融合深度特征与多核学习的最小二乘孪生支持向量机(LSTWSVM)算法. 针对支持向量机等核分类器在多核学习中高计算复杂度的问题,提出一种基于边缘错误最小化原则的多核LSTWSVM框架,利用分类器优势提高多核学习的性能. 针对高斯多核浅层结构的问题,采用MKL法设计一种基于深度神经网络多层信息的高鲁棒性深度映射核,将此深度核与多尺度高斯基核以核矩阵哈达玛积方式相融合,构造一组新的具有高度表达能力的改进核. 最后,将基于LSTWSVM的多核训练算法与改进的多核结构进行高度集成,通过大量基准数据集与工业数据实验表明,其能有效结合深度学习与多核学习的优势,且以较低的计算成本提高分类精度与泛化能力.

关键词: 多核学习；深度学习；最小二乘孪生支持向量机；复杂工业数据建模

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1746

引用格式: 刘颖, 刘德彦, 吕政, 等. 融合深度特征与多核学习的LSTWSVM及其工业应用[J]. 控制与决策, 2024, 39(8): 2622-2630.

LSTWSVM fusion of deep feature and multiple kernel learning and its industrial applications

LIU Ying, LIU De-yan, LV Zheng†, ZHAO Jun, WANG Wei

(School of Control Science and Engineering，Dalian University of Technology，Dalian 116024，China)

Abstract: In order to improve the representation ability of multi-kernel learning and reduce its computational cost, this paper proposes a least squares twin support vector machine (LSTWSVM) algorithm that combines depth feature and multi-kernel learning. Aiming at the problem of high computational complexity of kernel classifiers such as support vector machine in multi-kernel learning, a multi-kernel LSTWSVM framework based on the principle of edge error minimization is proposed. The cost-sensitive learning idea is adopted to improve the performance of multi-kernel learning by using the advantages of classifiers. Aiming at the problem of Gauss multi-kernel shallow structure, a highly robust depth mapping kernel based on depth neural network multi-layer information is designed using the MKL method. The depth kernel and multi-scale Gaussian basis kernel are fused in the form of kernel matrix Hadamard product to construct a new set of improved cores with high expressiveness, which contains the deep feature information of data. Finally, this paper highly integrates the multi-kernel training algorithm based on the LSTWSVM with the improved multi-kernel structure. Through benchmark datasets and industrial experiments, it shows that it can combine the advantages of deep learning and multi-kernel learning, and improve the classification accuracy and generalization ability at a lower computational cost.

Keywords: multiple kernel learning；deep learning；least squares twin support vector machine；complex industrial data modeling

0 引 言

在复杂流程工业问题中,采集到的数据往往具有高维、非线性、不确定性的特点[1],给数据建模方法带来了精准程度、稳定性和自适应性方面的挑战[2]. 基于核的模式分析法能够高效处理系统中的非线性特征,被广泛应用于工业数据的建模问题中. 尤其在多工况等复杂信息源问题中,多核学习(multiple kernel learning, MKL)[3]能通过预定义一组基核,并以权重优化方式来适应不同数据特征的要求.

然而,多核学习的浅层架构不能充分捕获数据

收稿日期: 2022-10-07；**录用日期:** 2023-06-04.

基金项目: 国家自然科学基金项目(61873048, 62003072)；国家科技部重点研发计划项目(2017YFA0700300)；中央高校基本科研业务费专项资金项目(DUT22JC16)；辽宁省应用基础研究计划项目(2023JH2/101600043).

责任编委: 胡清华.

†通讯作者. E-mail: lvzheng@dlut.edu.cn.

关联信息,近年来多种新颖的数据驱动核架构被提出. 文献[4]提出了非线性成对层学习框架,包含一个固定的高斯映射层和一个非线性成对层,充分考虑了数据之间的耦合信息. 文献[5]进一步将此架构扩展到多核学习,通过采用快速交替方向乘子法来训练数据驱动核,使用此核去驱动高斯多核的优化训练,得到更好的权重结果. 然而,利用该方法训练数据驱动核需要额外考虑训练数据与测试数据的分布差异性,不利于样本对外扩展. 相比之下,基于深度神经网络的深度映射核具有非线性映射的显式表达,能够以解析解的方式由训练数据扩展到测试数据. 此外,深度核还能够有效提取数据中蕴含的特征,去除数据中的噪声与冗余信息[6-7]. 但传统深度映射核对网络超参数的鲁棒性低,且未能很好地解决深度学习易过拟合的问题[8]. 因此,针对多核浅层特征结构与深度核表示方面不足的问题,探索深度学习与多核学习的协同学习方案具有重要意义.

目前,已有学者研究多核学习与深度学习的相互结合[9]. 文献[10] 提出了一种多模态情感分析的集成方法,将使用卷积神经网络提取的文本特征与从视觉和听觉模式提取的其他类型特征通过MKL融合,是对标准架构的重大改进. 文献[11]提出了对卷积长短期记忆网络的一种增强方法,它支持容纳多个卷积核,为多核扩展设计了一种基于注意力的机制,关键思想是用一组卷积核代替网络中的单个卷积核. 文献[12]提出了kernet法,通过MKL将神经网络的所有中间隐藏层信息进行优化组合,在一定程度上保留了特征映射中的局部性,异构层的融合有利于性能的提高. 文献[13]介绍了深核机器优化方法的多核变体,以有效地执行MKL,其中多个核可以来自不同的核参数化或对应于多个特征源. 文献[14]提出了多参数多核深度神经网络,利用MKL提高回归领域问题的预测精度. 以上方法在提升深度学习表达性方面取得良好效果,但是当网络层数较多时,鉴于待优化基核数量会相应增加且基核之间相似性较高,对各种MKL具有不同程度的影响,会降低学习效果[15]. 深度多核学习[16]也是一种重点联合方式,与深度映射核不同,深度多核的“深度”体现在核架构上,每个核代表深度多核网络的一个节点,核与核之间的联结方式为网络形式. 基于对偶目标函数的交替优化方法、基于跨度界的算法与带梯度上升法的反向传播算法[17-19]被提出用来训练优化深度多核网络,然而,这些模型需要较多参数,尤其当网络层数增加时,计算量迅速增大,而且它们的网络结构固定,对时变数据适应性低[20]. 因此,本文考虑通过深度学习与多核学习双向促进的方式来提高核学习的表示能力,同时兼顾核学习的计算成本.

由于多核学习的计算复杂度主要取决于其训练方法及其基础分类器的计算复杂度,目前关于提高多核学习效率的研究,没有充分发挥出基础分类器的潜力,可考虑通过使用不同的分类器来提高多核学习的效率,进一步提高分类精度.

支持向量机(support vector machine, SVM)是目前最流行的核分类器之一,其基于结构风险最小化原则与VC维理论,具有良好的泛化能力. 文献[21]在此基础上提出了孪生支持向量机(twin support vector machine, TWSVM),通过求解一对二次规划问题来构造一对非平行超平面,区别于SVM单分类超平面,计算复杂度为SVM的1/4. 为了进一步减小计算复杂度,文献[22]借鉴最小二乘的思想,提出了最小二乘孪生支持向量机(least squares twin support vector machine, LSTWSVM),通过求解一对线性方程组代替二次规划问题,使得求解速度更快. 鉴于TWSVM与LSTWSVM在计算效率和处理不平衡数据方面的良好性能,多年来受到广泛研究. 然而,它们在训练过程中的核矩阵形式为非半正定的,而多数多核SVM算法要求训练矩阵为半正定形式,故针对SVM的多核方法难以直接迁移到TWSVM与LSWTSVM中,有必要设计基于具体孪生模型的高效多核算法,以体现它们的特性,从而实现从分类器的角度提升多核学习的性能.

本文提出一种基于LSTWSVM的融合深度学习与多核学习优势的联合核架构,主要工作如下。

1) 根据边缘损失误差最小化原则设计一种适应于LSTWSVM的多核权重训练方法,凭借分类器自身优势与低成本参数寻优提升多核学习的精度与效率,将LSTWSVM纳入到多核理论体系的研究中.

2) 采用MKL思想设计一种高鲁棒性的深度映射核,组合优化深度神经网络的输入层,最高隐藏层与输出层网络表示,提高了深度核的信息丰富度.

3) 针对高斯多核浅层结构的问题,将深度核与多尺度高斯基核以核矩阵哈达玛积形式相融合,构造一组新的具有高度表达能力的改进多核,其蕴含数据深度特征.

最后,将基于LSTWSVM的多核权重训练方法与融入深度核信息的高斯多核结构相统一,得到最终的核学习分类器,并在大量基准数据集与实际工业数据上进行多维度实验与论证.

1 多核学习

核学习是处理非线性问题的强大工具,而核函数的选取是其重要环节. 当输入空间是欧氏空间或离散集,特征空间是希尔伯特空间时,核函数表示输入空间映射到特征空间得到的向量之间的内积,从而可以避免高维空间的内积运算. 对于含n个训练样本的二元分类问题,$\varphi(x)$表示从原始空间到高维希尔伯特空间的特征映射,$k(x_i,x_j)=\varphi(x_i)\varphi(x_j)$为相应的核函数. 常用核函数包括线性核、高斯核、多项式核,这些都是对数据进行单核训练,存在着过于依赖核函数的选取以及在异构信息上灵活性差的问题. 多核学习能有效解决此类问题,其形式为基本内核的线性或非线性组合. 本文的研究形式为基核的凸组合,即权向量具有固定1范数的线性非负组合

$$k_\beta(x_i,x_j)=\sum_{r=1}^{M}\beta_r k_r(x_i,x_j)=$$

$$\sum_{r=1}^{M}\beta_r\langle\varphi_r(x_i),\ \varphi_r(x_j)\rangle,\ \beta_r\geqslant 0,\ \sum_{r=1}^{M}\beta_r=1. \quad (1)$$

其中:k_r代表第r个核,β_r代表第r个核权重向量.

高斯核的局部泛化能力强,其表达式为

$$k(x_i,x_j)=\exp\left(-\frac{\|x_i-x_j\|^2}{2\sigma^2}\right). \quad (2)$$

核宽度参数σ对其性能影响较大,由此定义一组不同核宽度的高斯核,可捕获不同的数据相似性信息,通过对其进行有效组合,以适应于多源信息.

2 融合深度特征与多核学习的LSTWSVM

为了提升多核学习的表示能力并降低计算成本,本文从多核训练算法与核函数自身结构两个角度进行优化:针对SVM等分类器高计算复杂度的问题,同时考虑到其与孪生模型的具体区别,提出一种基于LSTWSVM的多核权重训练算法,凭借分类器自身优势与低成本参数寻优提升多核学习的精度与效率. 针对高斯多核浅层结构的问题,采用MKL法设计了一种鲁棒性高的基于深度神经网络多层表示的深度映射核,进而计算该深度核与多尺度高斯基核的哈达玛积,得到一组新的蕴含深度特征的改进核(deep-Gauss-kernel, DGk). 通过集成这些思想,得到统一的分类器.融入深度特征的多核结构如图1所示.

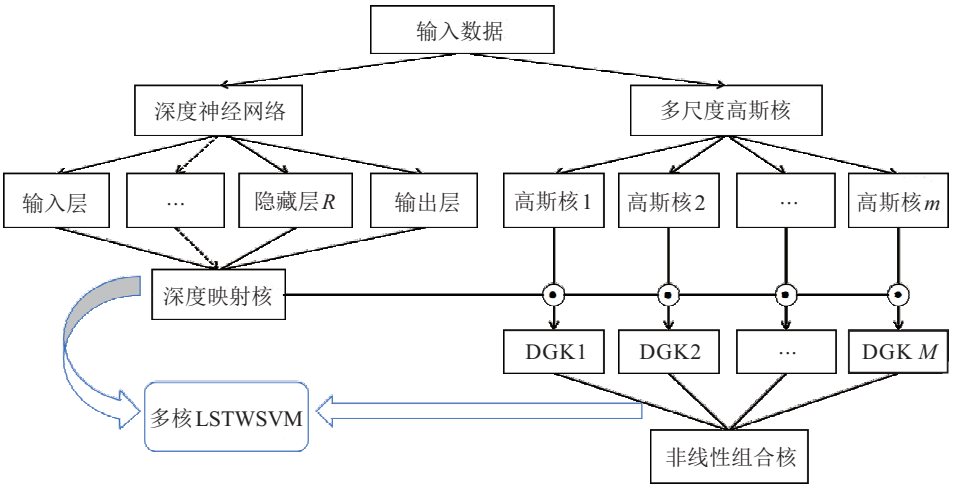

图 1 融入深度特征的多核结构

2.1 基于边缘错误最小化的多核LSTWSVM

本文采用代价敏感学习思想,设计一种基于边缘错误最小化原则的多核LSTWSVM框架,通过最小化正负类样本的加权边缘损失误差来优化多核权重,同时在整体上执行分类器参数与多核权重交替迭代的联合优化策略,以充分发挥LSTWSVM的优势,是一种高效的MKL方案. 其意义还在于,可以通过设置不同的正负类惩罚参数来调节目标损失函数的加权系数,对少数类样本错分误差施以更高的惩罚参数,在优化过程中更加突出关键类的影响,对数据不均衡问题有一定的处理能力.

对于LSTWSVM的非线性分类问题,本文额外增加了正则化项以避免求解过程中的矩阵奇异性问题,由此引入多核机制后,本文方法的目标优化函数为

$$\min\ \frac{\|K(A,X^{\mathrm{T}})w_1+eb_1\|^2}{2}+\frac{C_1\|\xi_1\|^2}{2}+\frac{\|w_1\|^2}{2};$$

$$\text{s.t. }-(K(B,X^{\mathrm{T}})w_1+eb_1)+\xi_1=e. \quad (3)$$

$$\min\ \frac{\|K(B,X^{\mathrm{T}})w_2+eb_2\|^2}{2}+\frac{C_2\|\xi_2\|^2}{2}+\frac{\|w_2\|^2}{2};$$

$$\text{s.t. }(K(A,X^{\mathrm{T}})w_2+eb_2)+\xi_2=e. \quad (4)$$

其中:w_1,b_1,w_2,b_2分别为两个分类面的权重与偏置;$X=[A;B]$,K为核函数,A为正类样本矩阵,B为负类样本矩阵;C_1,C_2为两个分类面的惩罚参数;ξ_1,ξ_2

为对应松弛向量; e 为合适维度的元素均为1的列向量. 进而,根据凸优化理论可以求得两组权重与偏置

$$\begin{bmatrix} w_1 \\ b_1 \end{bmatrix} = -[C_1 Q^{\mathrm{T}} Q + P^{\mathrm{T}} P + I]^{-1} C_1 Q^{\mathrm{T}} e,$$

$$\begin{bmatrix} w_2 \\ b_2 \end{bmatrix} = [C_2 P^{\mathrm{T}} P + Q^{\mathrm{T}} Q + I]^{-1} C_2 P^{\mathrm{T}} e. \tag{5}$$

其中: $Q = [K(B, X^{\mathrm{T}}), e], P = [K(A, X^{\mathrm{T}}), e]$, $K(A, X^{\mathrm{T}})$ 代表 $\sum_{r=1}^{M} \beta_r K_r(A, X^{\mathrm{T}})$, $K(B, X^{\mathrm{T}})$ 代表 $\sum_{r=1}^{M} \beta_r K_r(B, X^{\mathrm{T}})$, β_r 为第 r 个核权重.

该多核框架采用两步法学习思想,为了便于优化问题的求解,通过预定义一组平均多核权重,继而训练得到LSTWSVM的两组参数后,运用它们去更新原本的多核权重,所提出的代价敏感目标优化函数为

$$\min \mathrm{Loss} = \min \left[\frac{C_1}{C_1 + C_2} \|\xi_1\|^2 + \frac{C_2}{C_1 + C_2} \|\xi_2\|^2 \right]. \tag{6}$$

根据式(5)与(6), ξ_1 与 ξ_2 可表示为

$$\begin{aligned} \xi_1 &= K(B, X^{\mathrm{T}}) w_1 + e b_1 + e, \\ \xi_2 &= -(K(A, X^{\mathrm{T}}) w_2 + e b_2) + e. \end{aligned} \tag{7}$$

该约束条件要求每个超平面到另一类样本点的距离为单位1,故 ξ_1 和 ξ_2 分别用来记录针对于本类超平面的另一类样本点的距离违反量. 因此,该目标函数为两个类别的超平面对应另一类样本距离违反量的二范数平方加权和,优化问题等同于最小化边界错误.

结合加权1范数权重约束条件,求解多核权重 β_r 的约束优化问题为以下形式:

$$\begin{cases} \min\limits_{\beta} \left[\dfrac{C_1}{C_1 + C_2} \|e + K(B, X^{\mathrm{T}}) w_1 + e b_1\|^2 + \right. \\ \qquad \left. \dfrac{C_2}{C_1 + C_2} \|e - K(A, X^{\mathrm{T}}) w_2 - e b_2\|^2 \right], \\ K = \sum\limits_{r=1}^{M} \beta_r K_r, \ \sum\limits_{r=1}^{M} \beta_r = 1, \\ 0 \leqslant \beta_r \leqslant 1, \ r = 1, 2, \ldots, M. \end{cases} \tag{8}$$

本文通过有效集法求解该优化问题. 另一方面,交替迭代的停止标准设置为迭代次数 t 达到最大标准 T 或者损失函数收敛 $|\mathrm{loss}^{t+1} - \mathrm{loss}^{t}| \leqslant \mathrm{eps}$,交替迭代完成后按照下式进行分类:

$$i = \arg\min \left| \sum_{r=1}^{M} \beta_r K_r(x, X^{\mathrm{T}}) w_k + b_k \right|, \ k = 1, 2. \tag{9}$$

多核LSTWSVM框架具体步骤如下.

step 1: 初始化加权多核,设置 M 个核函数,每个核函数的权重 $\beta^{(0)}$ 为 $1/M$. 定义惩罚参数 C_1 与 C_2,最大迭代次数 T,损失函数收敛阈值eps.

step 2: 求解式(5)得到两组参数 $w^{(0)}, b^{(0)}$.

step 3: 利用两组权重与偏重求解优化问题(8),得到更新的多核权重 $\beta^{(1)}$,同时获取该多核权重下的损失函数 $\mathrm{loss}^{(1)}$.

step 4: 判断损失函数loss求解得到的收敛性(loss初值设为0),若收敛则训练完成,采用的 $\beta^{(0)}$, $w^{(0)}$, $b^{(0)}$ 按照式(9)进行分类;若不收敛则使用新的多核权重 $\beta^{(1)}$ 重复step 2、step 3, 直到满足停止标准为止.

2.2　基于DNN多层网络表示的深度映射核

针对传统深度核表示方面不足的问题,设计一种高鲁棒性的深度映射核,通过MKL对深度神经网络(deep neural network, DNN)的输入层、隐藏层及输出层信息进行组合优化,从而蕴含更为全面的信息. 其采用深度神经网络逼近核映射思想来构造,具有显式的函数表达形式. 核映射以神经网络的方式将原始输入空间变换到特征空间,而不是使用核方法诱导的隐式形式. 此核函数是一种数据驱动形式的非参数核,可直接从数据中学习更具表达性和灵活性的核. 与其他数据驱动核的不同之处在于,训练过程中的核矩阵无需限制为半正定形式,因此能够适用于TWSVM与LSTWSVM等核分类器. 具体的深度神经网络本文采用多层感知器(multilayer perceptron, MLP),基于MKL的网络架构如图2所示.

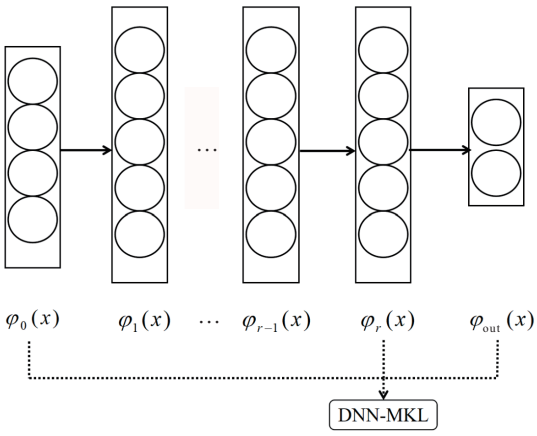

图 2　基于MKL的深度网络结构

对于多层感知器结构,隐藏层之间的非线性映射函数为 $f(x) = \sigma(Wx + b)$,故图2中输入数据 x 在每一层网络中的表示为如下形式:

$$\begin{cases} \varphi_0(x) = x; \\ \varphi_j(x) = \sigma(W^j \varphi_{j-1}(x) + b^j), \ 1 \leqslant j \leqslant r; \\ \varphi_{\mathrm{out}}(x) = \sigma(W^{\mathrm{out}} \varphi_j(x) + b^{\mathrm{out}}). \end{cases} \tag{10}$$

其中: W 为层与层之间的权重矩阵, b 为偏置, σ 为激活函数, r 为隐藏层个数. 鉴于神经网络的隐藏层能

有效地进行特征提取,蕴含着数据深层特征;输出层能够直接影响分类的结果,蕴含着精简关键的信息;输入层能够起到避免过拟合的作用. 故本文提取具有代表性的非线性迭代映射的最高隐藏层、输出层和输入层信息通过MKL方式进行组合,有

$$\begin{cases} k_0(x_i,x_j)=\langle x_i,x_j\rangle;\\ k_1(x_i,x_j)=\langle \varphi_r(x_i),\varphi_r(x_j)\rangle;\\ k_2(x_i,x_j)=\langle \varphi_{\text{out}}(x_i),\varphi_{\text{out}}(x_j)\rangle;\\ k_{\text{DNN}}=\beta_0k_0+\beta_1k_1+\beta_2k_2,\beta_0+\beta_1+\beta_2=1. \end{cases} \tag{11}$$

对于以上的基于MKL的深度映射核,虽然多层网络表示能够提供不同层次的异构信息,但是组合的方式很关键,因此本文采用2.1节提出的多核算法来优化集成这些表示,也是首次构造一种深度核函数形式的LSTWSVM.

2.3 融入深度特征的多核LSTWSVM

针对高斯核浅层结构的不足而不能充分挖掘数据关键特征的问题,本文构造一种融入数据深层信息的改进高斯核,以便以更充分的方式建模输入与输出之间的关系,其形式为深度映射核函数与高斯核函数的乘积,对应核矩阵为深度映射核矩阵与高斯核矩阵之间的哈达码积(Hadamard product). 新核为半正定的,满足MERCER定理,其优势在于深度核将复杂非线性输入的先验知识结合到训练中,当应用时,新核已经映射了多变量输入与输出向量之间关系的主要部分,达到更好的分类效果. 因此,本文对多尺度高斯核实施此方法,预定义M个不同核宽度的高斯基核$g_1,\dots,g_M$,深度核为k_{DNN},可构造M个融入深层信息的非线性基核. 为了便于LSTWSVM算法与所提出多核方法的运算,本文进行核融合时以核矩阵为基础元素,结合矩阵哈达玛积$\odot$的对应位置元素相乘的原理,M个非线性核矩阵形式为

$$\begin{cases} K_1=G_1\odot k_{\text{DNN}},\\ K_2=G_2\odot k_{\text{DNN}},\\ \quad\vdots\\ K_M=G_M\odot k_{\text{DNN}}. \end{cases} \tag{12}$$

最后,通过MKL对式(12)中的M个核矩阵进行优化组合. 值得注意的是,虽然使用改进的高斯多核对式(5)进行分类器参数求解和对式(9)执行分类预测,但是在多核权重求解过程中仍然采用基于高斯核的方式,即权重优化问题(8)中的矩阵仅有高斯核,而非基于改进高斯核的方式. 由于改进后的核仍以高斯核为主导,该方式是可行的,能使优化得到的多核权重稀疏性更强,且效率更高.

本文最终提出的融合深度特征与多核学习的LSTWSVM总体流程如下所示.

step 1: 配置深度神经网络的层数、神经元个数、激活函数等网络超参数,进而训练网络.

step 2: 根据式(10)提取深度神经网络不同层次的核映射,并根据式(11)将其构造成显式核函数.

step 3: 采用所提出的多核LSTWSVM框架优化step 2所得到的3个显式核,求解出式(11)中的核权重,形成组合深度映射核.

step 4: 预定义一组多尺度高斯基核,根据输入数据构造相应的高斯基核矩阵. 计算每一个高斯基核矩阵与深度映射核矩阵之间的哈达玛积,得到一组新的核矩阵.

step 5: 采用所提出的多核框架对step 4中得到的非线性核矩阵进行优化组合,其中多核权重的计算采用基于高斯核的方式,求解得到组合核矩阵与分类器权重偏置.

step 6: 根据定义与求解的超参数分类判别.

3 实验和分析

为了验证本文所提出方法的有效性,采用大量的基准数据集和实际工业数据集进行仿真验证. 本文在26个不同特征与样本量的基准数据集上进行实验,以分析所提出方法在不同环境下的效果. 这些数据集均来源于UCI网站,实验中数据已经过预处理与归一化. 分类准确率用来评估模型性能,数据集的样本数、维度等信息见表1. 分类准确率为随机进行20次实验的平均值,在每次实验中,随机选取50%样本为训练集,其余50%样本为测试集. 部分数据集的训练集与测试集已经固定划分,不再重新进行划分.

表1 26个UCI数据集信息

数据集	样本数	维度	数据集	样本数	维度
horse	300+68	25	cancer	286	9
monks1	124+432	6	ionosphere	351	34
monks2	169+432	6	voting	435	16
monks3	122+432	6	climate	540	18
promoters	106	55	diag	569	30
iris	150	3	indian	583	10
hepatisis	155	19	balance	625	4
parkinsons	195	23	credit	690	15
prog	198	33	wisconsin	699	9
sonar	208	60	pima	768	8
connection	208	60	oocytes	912	25
heart	270	13	banknote	1 372	4
haberman	306	3	cmc	1 473	9

3.1 基准数据集实验

本文将所提出的基于LSTWSVM的高斯多核方法简记为MLTSVM,将融入深层信息的改进高斯多核方法记为DMLTSVM. 令它们与BP神经网络及如下4种多核方法进行准确率对比实验: MEMO (兼顾最大化分类间隔与最小化模型复杂度)、easymkl (时间与空间高效的多核方法)、simplemkl (基于梯度下降与二次规划相结合的两步法优化策略)、kernet (深度学习与多核学习相结合的多核方法).

本文对MLP采用误差反向传播算法进行训练,即反向传播神经网络(back propagation neural network, BPNN). 实验中将其隐藏层数设置为2,每层节点数均为100,输出层节点数为2,激活函数设置为sigmoid函数. 网络学习率采用网格搜索法从$\{0.01, 0.05, 0.1, 0.5, 1\}$中选取. MEMO法的分类器采用SVM, 超参数$\theta$设置为0. easymkl法的分类器采用SVM,超参数λ采用网格搜索法从$\{0, 0.1, 0.4, 0.7, 1\}$中选取. kernet法的基础多核框架采用easymkl法,分类器采用KOMD,超参数λ的选取范围为$\{0, 0.1, 0.4, 0.7, 1\}$. MLTSVM方法的最大迭代次数设置为10,损失函数收敛阈值设置为0.01. DMLTSVM的网络部分超参数设置同BP神经网络,网络学习率固定为10. 同时为了避免过拟合与降低计算复杂度,DMLTSVM法的深度核部分的迭代次数固定为1次,改进高斯核部分的迭代次数阈值设置为2. 所有多核方法的高斯多核集核参数σ^2均设置为$\{5^{-2}, 5^{-1}, 1, 5^1, 5^2, 5^3, 5^4, 5^5\}$,惩罚参数$C$设为100. 本文两种方法的多数类惩罚参数$C_1$设为100,少数类惩罚参数$C_2 = C_1(m_1/m_2)^2$. 其中: m_1是多数类样本个数,m_2是少数类样本个数,通过类别不平衡率来设置C_1与C_2间的比例以调节数据不平衡的影响. 7种多核方法在26组基准数据上的分类准确率与标准差如表2所示. 其中: rank代表在所有数据集上的精度平均排名,win代表获得最高分类精度的数据集个数.

表2　多核方法分类精度对比

数据集	BPNN	MEMO	simplemkl	easymkl	kernet	MLTSVM	DMLTSVM
horse	78.53±1.75	86.76±0.00	85.29±0.00	83.82±0.00	**89.71±0.00**	85.29±0.00	86.76±0.00
monks1	**99.98±0.10**	86.57±0.00	83.56±0.00	93.06±0.00	66.20±0.00	81.02±0.00	91.60±0.00
monks2	78.10±1.10	84.95±0.00	76.16±0.00	**86.81±0.00**	67.13±0.00	80.79±0.00	85.42±0.00
monks3	81.49±3.39	92.59±0.00	90.05±0.00	91.90±0.00	83.33±0.00	91.90±0.00	**95.14±0.00**
promoters	78.30±4.17	76.79±5.03	53.77±11.5	**79.43±4.95**	76.23±5.18	75.38±8.73	78.77±6.36
iris	90.33±9.52	95.40±2.13	95.87±1.64	93.40±3.64	68.40±4.56	94.80±1.93	**95.93±1.59**
hepatitis	79.74±2.51	81.86±4.06	81.82±3.91	82.18±2.81	80.45±2.75	82.08±3.23	**82.47±3.33**
parkinson	84.23±3.26	90.36±3.25	90.26±4.65	85.51±5.83	84.44±1.88	89.85±4.02	**93.61±2.47**
prog	75.66±4.18	78.89±3.61	75.81±2.92	75.40±3.44	78.03±3.87	78.03±3.84	**79.19±3.38**
sonar	76.15±3.25	84.90±3.51	81.68±4.25	85.43±2.98	77.07±3.54	80.91±4.63	**85.67±2.68**
connection	74.71±3.46	83.32±4.05	82.60±5.10	84.86±3.48	77.36±4.76	81.83±3.70	**86.73±2.64**
heart	82.24±2.08	81.36±2.28	81.73±2.55	78.20±2.86	82.43±3.04	**82.62±2.31**	81.09±1.97
haberman	74.28±2.34	71.08±2.44	73.92±2.67	65.59±3.16	73.92±3.03	73.89±3.10	**74.35±2.69**
cancer	71.08±2.99	**72.27±3.93**	70.35±2.66	65.80±2.85	71.19±3.74	70.49±3.54	70.56±3.36
ionosphere	88.14±2.24	94.09±1.79	93.97±1.60	93.30±1.58	88.10±2.17	94.31±1.49	**94.83±1.47**
voting	59.17±2.13	59.68±3.09	60.41±2.84	60.09±3.58	60.96±3.23	60.65±3.08	**61.57±3.28**
climate	94.63±1.55	93.33±1.48	91.41±1.26	93.85±1.02	**94.76±1.21**	91.43±1.43	93.26±1.32
diag	96.30±1.14	97.01±0.82	95.69±1.08	95.33±1.27	97.07±0.88	97.13±0.69	**97.20±0.62**
indian	71.80±2.09	71.25±2.02	71.77±1.55	64.43±3.47	71.23±2.19	69.93±2.72	**72.20±1.57**
balance	95.45±0.95	95.39±1.06	94.28±1.67	95.63±1.39	94.90±1.46	**95.72±1.10**	94.46±1.42
credit	85.35±1.30	85.80±1.55	85.86±1.23	82.87±1.56	85.96±1.27	**86.26±1.43**	85.45±1.29
wisconsin	96.48±0.88	96.67±0.81	96.91±0.52	95.53±0.61	96.64±0.80	**97.09±0.73**	96.91±0.66
pima	75.54±1.41	64.53±1.78	70.09±4.01	70.16±3.05	**75.59±2.40**	74.52±1.85	74.53±2.41
oocytes	**83.16±2.15**	80.23±1.84	68.73±2.23	77.15±1.76	79.54±1.37	80.39±1.63	82.37±2.02
banknote	99.97±0.10	**100.0±0.00**	99.16±0.23	92.99±0.98	98.49±0.42	99.89±0.19	**100.0±0.00**
cmc	76.67±1.26	76.26±1.44	**77.30±1.20**	76.48±1.16	77.04±1.26	76.10±1.12	76.98±1.06
rank	4.54	3.69	4.58	4.73	4.27	3.76	2.19
win	2	2	1	2	3	4	13

3.1.1 多核方法总体对比实验

综合对比表2中的7种方法,所提出的MLTSVM与DMLTSVM法取得了优异的效果,分别在4组与13组数据集上取得最高分类精度,且平均精度排名远优于其他方法. 这表明本文多核框架具有良好的普适性与稳定性,充分利用了LSTWSVM的特点,通过最小化正负类两个超平面的边界错误之和来优化多核权重,对复杂非线性数据的适应性更强;通过采用类

别不平衡率来设置两个惩罚参数,可弥补数据不均衡的影响. 与其他多核算法相比,无需花费大量的时间去进行参数寻优,只需利用先验信息来设置固定的参数. 总之,其主要从基础分类器角度提升了多核学习性能,可在多种数据场景中获得更好的效果.

另一方面,与基于传统高斯多核的MLTSVM相比,利用深层信息改进后的DMLTSVM法在22组数据集上的精度有所提升,且在部分数据集中效果显著,验证了深度核的特征提取能力强. 将深层先验信息融入到高斯多核中,以提升其表达能力. 同时,也体现了哈达玛积核融合方式的优异性,表明了该核并不是简单地对两种体系下的核进行中和,而是集成了多核学习与深度学习的优势,达到最佳结合状态.

3.1.2 模型鲁棒性分析

实验过程中,本文方法对神经网络层数、神经元个数以及学习率等参数变化的鲁棒性强,是一种高效的深度学习方案. 具体地,以网络学习率为例进行说明. 图3可视化了BP神经网络与DMLTSVM两方法在monks3与horse数据集上精度随学习率变化情况.

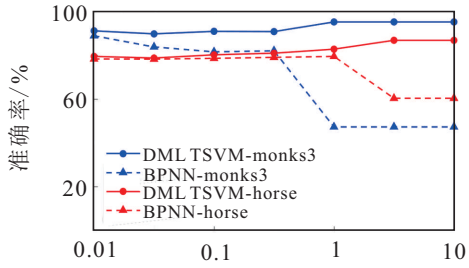

图 3 分类准确率与网络学习率的相关性

根据可视化结果,随着学习率的逐步提高,本文方法不仅没有BP神经网络所出现的精度大幅下降现象,反而有一定程度的提升,故本文方法适合更高的学习率超参数值,同时可以很快地完成迭代与训练. 而且,当有过拟合的趋势时多核权重会向输入层进行倾斜,故数据量较小时也不会出现过拟合现象,可以在中小数据集中发挥出优势. 由此,本文方法在神经网络超参数择优方面的计算成本远远低于直接利用神经网络进行分类,也是一种高鲁棒性的方案.

3.1.3 多核权重优化方式分析

DMLTSVM法求解改进高斯多核权重过程中,采用基于高斯核自身进行权重优化,而非基于改进高斯核进行优化,故通过精度比较对其合理性进行验证. 图4可视化了两种核权重训练方式在部分数据集上的分类精度. 根据可视化结果,基于高斯核的方式优于另一种训练方式,原因在于权重分配以高斯核为主导信息,从而导致权重优化的结果稀疏性更强,而后者的训练结果会趋向于平均核权重,优化效果不明显,进而具有不同结果. 同时,因为本文算法的权重初始状态为平均核权重,故实验结果从侧面表明其交替迭代过程是可行有效的,训练后的多核权重优于初始状态.

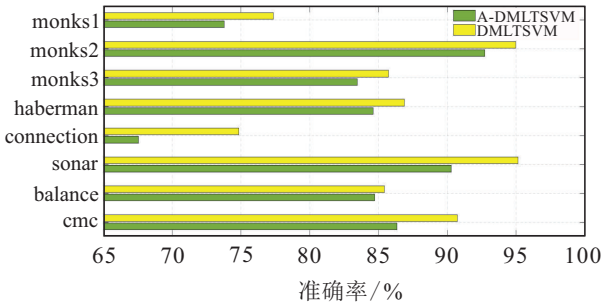

图 4 两种多核权重训练方式准确率对比

3.1.4 不同深度特征信息对高斯多核分类的影响

为了验证所提出的融入深度核(MKL)信息的改进高斯多核的有效性,本文从不同深度核信息对于高斯多核分类精度的影响程度进行实验与分析. 本节中只提取神经网络的单个网络层映射来生成深度映射核,进而与高斯多核相融合进行核学习,融合策略均采用2.3节所提出的哈达玛积方式,故可以再次建立3种不同层次的深度核-多核联合学习方法,简记为1-MLTSVM、2-MLTSVM和3-MLTSVM,分别对应BP网络的输入层、最高隐藏层和输出层,从而联同MLTSVM与DMLTSVM做精度对比实验. 同时,添加基于交叉验证的TWSVM与LSTWSVM做辅助实验,其高斯单核超参数选取范围同3.1节的多核集. 所有方法的其余参数设置同3.1节,分类精度如表3所示,它们的20次随机划分方式均一致.

由实验结果可知,由MKL组合得到的深度映射核对高斯核的改进效果最优,在26组数据集中12组取得最高分类精度,可以印证组合重构多层信息的有效性. 其通过神经网络的逐层变换,将样本数据映射到不同深度的特征空间,获取多角度不同信息,有效提高了深度映射核的表示能力与分类精度. 由最高隐藏层与输出层核映射得到的深度核的改进效果次之,分别在8组与7组数据上达到最高精度,印证了隐藏层特征提取的能力与输出层的良好结果. 其余两种多核方法分别在2组与1组数据上达到最高精度. 虽然重构多层网络信息并不总是取得比单层更优异的结果,但是其能够提高算法稳定性与泛化能力,在平均精度排名与数据普适性方面效果显著. 另外,该部分的实验结果突出体现了哈达玛积型核融合方式的良好特性,即使是效果次优的2-MLTSVM与3-MLTSVM,在数据集普适性方面也优于原始的MLTSVM法.

表3　深度映射核有效性实验

数据集	TWSVM	LSTWSVM	MLTSVM	1-MLTSVM	2-MLTSVM	3-MLTSVM	DMLTSVM
horse	78.46±1.67	79.41±0.00	85.29±0.00	83.82±0.00	**86.76±0.00**	83.82±0.00	**86.76±0.00**
monks1	**92.13±0.00**	89.51±0.50	81.02±0.00	86.57±0.00	91.67±0.00	78.70±0.00	90.74±0.00
monks2	**87.67±1.13**	87.27±0.00	80.79±0.00	84.26±0.00	86.59±0.00	83.10±0.00	85.42±0.00
monks3	91.25±0.21	93.98±0.00	91.90±0.00	90.28±0.00	92.07±0.00	92.82±0.00	**95.14±0.00**
promoters	81.70±4.02	82.17±3.27	75.38±8.73	**78.87±4.46**	77.55±5.53	72.08±9.80	78.77±6.36
iris	94.33±2.73	94.27±2.77	94.80±1.93	95.40±2.16	95.73±2.03	92.60±3.09	**95.93±1.59**
hepatitis	79.74±3.05	80.91±3.99	82.08±3.23	81.95±3.81	**82.47±3.78**	82.27±3.40	**82.47±3.33**
parkinson	88.76±3.75	88.66±4.20	89.85±4.02	91.49±3.09	**93.81±2.55**	88.25±4.01	93.61±2.47
prog	75.20±3.01	73.69±6.86	78.03±3.84	79.14±3.75	78.94±4.01	77.12±4.97	**79.19±3.38**
sonar	82.50±4.37	82.69±3.30	80.91±4.63	83.08±3.23	85.24±2.41	79.52±4.80	**85.67±2.68**
connection	83.99±3.66	84.71±3.76	81.83±3.70	84.71±2.99	86.44±2.87	79.23±3.19	**86.73±2.64**
heart	79.46±2.19	77.24±2.69	82.62±2.31	81.12±1.97	80.58±2.33	**83.50±1.70**	81.09±1.97
haberman	68.53±7.41	68.33±6.47	73.89±3.10	73.86±3.24	73.40±2.97	**74.93±2.53**	74.35±2.69
cancer	64.55±3.14	65.28±7.48	70.49±3.54	69.06±3.13	66.64±4.47	**71.15±3.61**	70.56±3.36
ionosphere	93.83±1.47	90.11±5.21	94.31±1.49	94.11±1.54	94.83±1.19	93.97±2.84	**94.89±1.47**
voting	54.77±4.89	56.24±7.11	60.65±3.08	60.53±3.06	60.69±2.85	60.83±2.62	**61.57±3.28**
climate	**93.76±1.35**	91.43±1.73	91.43±1.43	91.94±1.30	91.83±1.27	91.78±1.56	93.26±1.32
diag	96.30±0.92	94.84±1.97	97.13±0.69	96.90±0.65	97.06±0.68	96.37±0.82	**97.20±0.62**
indian	65.14±2.93	64.79±14.0	69.93±2.72	70.74±1.85	71.20±1.94	70.50±2.16	**72.20±1.57**
balance	94.44±1.80	91.81±2.10	95.72±1.10	95.45±1.56	93.48±2.29	**96.20±1.05**	94.46±1.42
credit	81.93±1.66	81.19±1.61	86.26±1.43	85.80±1.60	85.12±1.75	**86.65±1.47**	85.45±1.29
wisconsin	96.30±1.29	95.92±1.45	**97.09±0.73**	96.68±0.82	96.70±0.73	97.01±0.67	96.91±0.66
pima	72.28±2.21	70.12±1.92	74.52±1.85	74.77±2.37	72.60±2.42	**75.57±1.23**	74.53±2.41
oocytes	81.72±1.52	78.55±1.87	80.39±1.63	**83.50±1.58**	81.47±2.19	79.34±2.53	82.37±2.02
banknote	99.96±0.06	99.86±0.27	99.89±0.19	99.89±0.15	**100.0±0.00**	99.93±0.18	**100.0±0.00**
cmc	72.62±1.14	77.13±1.02	76.20±1.12	74.86±1.65	**77.32±1.19**	76.22±1.33	76.98±1.06
rank	5.04	5.58	4.12	3.85	3.11	4.07	2.03
win	3	0	1	2	5	6	12

3.2　工业案例

本节将所提出核方法应用到实际炼焦工业案例中. 炼焦过程是一个具有多参数、时变、非线性和不确定等特性的复杂生产过程,有必要基于焦炭质量数据的统计分析,建立合适的焦炭质量等级分类模型,从而实现优化配煤结构,在稳步提高焦炭质量的同时降低配煤成本. 焦炭质量根据硫分含量可分为不同的等级,本文建立焦炭质量二分类模型,按国家标准,硫分含量小于0.8%为正类,大于等于0.8%为负类. 所使用数据来自某钢铁企业炼焦系统2020年的1 000条数据,其中正负类样本数分别为686与314.

当生产稳定时,炼焦煤的灰分和硫分与焦炭的灰分和硫分之间有较好的线性关系,本文首先建立一种具有3个输入特征的模型: 配合煤灰分、配合煤硫分、配合煤挥发. 同时在实际工业中经常面临着生产不稳定的情况,为保证模型的适应性就需要考虑更多的因素. 因此,本文建立第2种具有10个输入特征的模型,其在前一模型的基础上增加了配合煤细度、配合煤G值、配合煤X值、配合煤Y值、平均推焦电流、平均结焦时间、机侧温度7个属性,以探究增加特征数量的影响以及本文方法对多维非线性特征与不稳定工况的处理能力.

训练数据与测试数据按照50%: 50%随机划分,分类准确率为随机进行20次实验的平均值. 所有分类方法的参数设置同3.1节. 两种生产特征情形下的焦炭质量分级精度与标准差如表4所示.

表4　焦炭质量分级精度

算法	三特征	十特征
BPNN	86.86±1.73	91.89±1.24
MEMO	85.62±1.54	91.40±2.32
simplemkl	85.78±1.56	91.80±1.60
easymkl	85.84±1.61	90.44±2.73
kernet	86.74±2.61	92.04±1.35
MLSTSVM	86.30±1.63	91.93±1.02
DMLTSVM	**87.28±1.91**	**93.42±0.68**

由实验结果可知,在工业生产不稳定的情况下,使用多个特征进行分类明显优于仅使用配合煤的硫分、灰分与挥发分3个特征来分类,表明在生产过程中额外采用炼焦工艺条件等多方面属性来分类是一种更加稳健的方式. 另外,在两种不同数据特征的情形下,本文的DMLTSVM法均取得最高分类精度,但在十特征情形下的领先效果优于三特征,表明在复杂的多维非线性特征中更能发挥其优势,能够自动从数据中挖掘出关键特征信息,而无需进行大量的机理分析与手工特征工作, 满足多工况情形下稳定性、精准

性与自适应性的需要.

4 结 论

本文提出了一种基于LSTWSVM的融合深度学习与多核学习优势的集成核架构,该架构利用多核学习来提高深度核的表示能力,进而将该深度核融入到高斯多核的训练中,是一种双向促进的联合学习方案. 同时,设计了一种适应于LSTWSVM的多核框架,可以有效地的将所提出核架构实例化.

最后,通过对所提出方法的分类精度、泛化能力、哈达玛积型核融合方式、模型鲁棒性、权重优化方式以及深度核信息对于高斯多核分类精度的改善程度等方面进行实验分析,验证了本文的多种观点.

参考文献(References)

[1] Luo L. Research on some problems of fault diagnosis of nonlinear process based on data drive[D]. Hangzhou: Zhejiang University, 2015.

[2] Chai R Z. Research on classification model and fitting model of industrial big data based on neural network[D]. Hangzhou: Zhejiang University, 2018.

[3] Mehmet Gonen, Ethem Alpaydin. Multiple kernel learning algorithms[J]. Journal of Machine Learning Research, 2011, 12(64): 2211-2268.

[4] Liu F H, Huang X L, Gong C, et al. Nonlinear pairwise layer and its training for kernel learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, 2018: 3659-3666.

[5] Alavi F, Hashemi S. FKNDT: A flexible kernel by negotiating between data-dependent kernel learning and task-dependent kernel learning[C]. The 25th International Computer Conference, Computer Society of Iran. Tehran, 2020: 1-7.

[6] Li Y J, Zhang T, Hu H H. Deep kernel mapping support vector machines based on multi-layer perceptron[J]. Journal of Beijing University of Technology, 2016, 42(11): 1652-1661.

[7] Wang Q, Lyu Z, Wang L Q, et al. Deep denoising kernel mapping-based long-term prediction model[J]. Control and Decision, 2019, 34(5): 989-996.

[8] Kouziokas G N. SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting[J]. Applied Soft Computing, 2020, 93: 106410.

[9] Wang T H, Zhang L, Hu W Y. Bridging deep and multiple kernel learning: A review[J]. Information Fusion, 2021, 67: 3-13.

[10] Poria S, Peng H Y, Hussain A, et al. Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis[J]. Neurocomputing, 2017, 261: 217-230.

[11] Agethen S, Hsu W H. Deep multi-kernel convolutional LSTM networks and an attention-based mechanism for videos[J]. IEEE Transactions on Multimedia, 2020, 22(3): 819-829.

[12] Lauriola I, Gallicchio C, Aiolli F. Enhancing deep neural networks via multiple kernel learning[J]. Pattern Recognition, 2020, 101: 107194.

[13] Song H, Thiagarajan J J, Sattigeri P, et al. Optimizing kernel machines using deep learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(11): 5528-5540.

[14] Kalaiarasi E, Anbarasi A. Multi-parametric multiple kernel deep neural network for crop yield prediction[J]. Materials Today: Proceedings, 2022, 62: 4635-4642.

[15] Gretton A, Bousquet O, Smola A, et al. Measuring statistical dependence with Hilbert-Schmidt norms[C]. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 63-77.

[16] Wu C F, Huang S C, Chiou C C, et al. A predictive intelligence system of credit scoring based on deep multiple kernel learning[J]. Applied Soft Computing, 2021, 111: 107668.

[17] Zhuang J, Tsang I W, Hoi S C H. Two-layer multiple kernel learning[C]. Proceedings of the 14th International Conference on Artificial Intelligence andStatistics. Fort Lauderdale, 2011: 909-917.

[18] Strobl E V, Visweswaran S. Deep multiple kernel learning[C]. The 12th International Conference on Machine Learning and Applications. Miami, 2014: 414-417.

[19] Rebai I, BenAyed Y, Mahdi W. Deep multilayer multiple kernel learning[J]. Neural Computing and Applications, 2016, 27(8): 2305-2314.

[20] Guo J N, Mao X L, Lin S Y, et al. Deep kernel supervised hashing for node classification in structural networks[J]. Information Sciences, 2021, 569: 1-12.

[21] Jayadeva, Khemchandani R, Chandra S. Twin support vector machines for pattern classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5): 905-910.

[22] Arun Kumar M, Gopal M. Least squares twin support vector machines for pattern classification[J]. Expert Systems with Applications, 2009, 36(4): 7535-7543.

作者简介

刘颖(1982−), 女, 副教授, 博士, 从事系统仿真建模智能优化算法及应用的研究, E-mail: liu_ying@dlut.edu.cn;

刘德彦(1998−), 男, 硕士生, 从事机器学习的研究, E-mail: 839852663@qq.com;

吕政(1987−), 男, 副教授, 博士, 从事机器学习、复杂系统调度等研究, E-mail: lvzheng@dlut.edu.cn;

赵珺(1981−), 男, 教授, 博士生导师, 从事流程工业生产计划与优化调度的研究, E-mail: zhaoj@dlut.edu.cn;

王伟(1955−), 男, 教授, 博士生导师, 从事复杂系统建模控制与优化、流程工业生产计划与优化调度等研究, E-mail: wangwei@dlut.edu.cn.