·人工智能与模式识别· 文章编号:1000-3428(2021)07-0067-07 文献标志码:A 中图分类号:TP389.1

基于多模态的在线序列极限学习机研究

李 琦[1],谢 珺[1],张 喆[2],董俊杰[1],续欣莹[2]
(1.太原理工大学 信息与计算机学院,山西 晋中 030600;2.太原理工大学 电气与动力工程学院,太原 030024)

摘 要: 单一模态包含的物体信息有限,导致在物体材质识别分类中表现不佳,而传统多模态融合方法在样本训练过程中需要输入所有数据。提出一种多模态的多尺度局部感受野在线序列极限学习机方法。对物体不同模态样本运用改进的特征提取框架,利用多尺度局部感受野感知样本信息提取特征,并将不同模态特征融合后通过在线序列极限学习机进行训练学习。在线序列极限学习机在训练过程中增量式地输入样本进行训练,当有新数据需要训练时无需对所有数据重新训练。在TUM触觉纹理数据库上进行验证,实验结果表明,多模态融合的分类精度高于单模态的分类精度,且改进的特征提取框架可以显著提升分类性能。

关键词: 多模态;RGB颜色三通道;局部感受野;在线序列极限学习机;物体材质分类

开放科学(资源服务)标志码(OSID):

中文引用格式: 李琦,谢珺,张喆,等.基于多模态的在线序列极限学习机研究[J].计算机工程,2021,47(7):67-73,80.
英文引用格式: LI Q,XIE J,ZHANG Z,et al.Research on online sequence extreme learning machine based on multi-modal[J].Computer Engineering,2021,47(7):67-73,80.

Research on Online Sequence Extreme Learning Machine Based on Multi-Modal

LI Qi[1],XIE Jun[1],ZHANG Zhe[2],DONG Junjie[1],XU Xinying[2]
(1.College of Information and Computer,Taiyuan University of Technology,Jinzhong,Shanxi 030600,China;
2.College of Electrical and Power Engineering,Taiyuan University of Technology,Taiyuan 030024,China)

【Abstract】 The object information that a single modality contains is limited,degrading the performance in object material recognition and classification.At the same time,the sample training of the traditional multi-modal fusion methods require all data to participate.To address the problem,a multi-modal online sequence extreme learning machine method with multi-scale Local Receptive Fields(LRF) is proposed.The method employs an improved feature to extract the framework of different modality samples of the objects,and then uses multi-scale local receptive fields to perceive sample information and extract the features.Different modality features are fused through the Online Sequence Extreme Learning Machine(OSELM) for training and learning.The online sequence extreme learning machine can be trained with incrementally input samples during the training process,and does not need to retrain all the data every time there is new data to be trained.The method is verified on the TUM tactile texture database.The experimental results show that the classification accuracy of fused multi-modal is higher than that of the single modality,and the improved feature extraction framework can significantly improve the classification performance.

【Key words】 multi-modal; RGB color three channels; Local Receptive Field(LRF); Online Sequence Extreme Learning Machine(OSELM); surface material classification
DOI: 10. 19678/j. issn. 1000-3428. 0058173

0 概述

机器学习是使计算机模拟或实现人类的学习行为从而获取新知识或技能的一种途径。人们在生活中的感知是多元的,识别一个物体不仅依靠视觉,还可以通过触觉、嗅觉、听觉等形式进行感知。任何感知能力的缺失都会造成生活能力减退。因此,在研究物体分类时,不仅可以依赖图像的视觉信息,还可以采集其真实的其他模态信息,通过多模态融合来为计算机提供更丰富的物体特征,使计算机充分感

基金项目: 国家自然科学基金(61503271,61603267);山西省自然科学基金(201801D121144,201801D221190)。
作者简介: 李 琦(1994—),女,硕士研究生,主研方向为计算机视觉、智能信息处理;谢 珺,副教授、博士;张 喆,讲师、博士;董俊杰,硕士研究生;续欣莹,教授、博士。
收稿日期: 2020-04-26 **修回日期:** 2020-06-12 **E-mail:** xiejun@tyut.edu.cn

计算机工程 www.ecice06.com

知物体信息，从而更好地实现物体识别与分类。例如，在物体材质分类研究中，由于不同材质的物体可能有相同的形状以及相似的纹理，在光照等因素的影响下，单纯依靠视觉信息可能无法对其进行有效分类，需要将不同模态的信息进行融合以实现物体识别与分类。

在多模态信息融合方面，研究者提出了较多方法。文献[1]以物体触觉加速度信号和相应的表面纹理图像为输入处理表面材料分类问题，有效地提高了分类精度。文献[2]研究表明，不同模态的特征对材料分类的性能具有不同的影响。文献[3]提出一种基于稀疏表示的多模态生物特征识别算法。文献[4]将视觉特征和触觉特征相融合以研究步态识别问题。文献[5]对RGB-D信息进行融合分类研究。文献[6]从不同的应用领域介绍多模态的研究现状。尽管上述研究取得了一定成果，但是如何将不同的模态信息进行有效融合仍具有较高难度。文献[7]建立一种新的投影字典学习框架，通过引入一个潜在的配对矩阵，同时实现了字典学习和配对矩阵估计，从而提高融合效果。文献[8]设计一个字典学习模型，该模型可以同时学习不同度量下的投影子空间和潜在公共字典。在多模态融合框架的研究中，分类器选择也是一个重点环节。

近年来，卷积神经网络(Conrolutional Neural Networks，CNN)在图像识别分类领域取得了较多成果。从最早的LeNet到AlexNet、Overfeat、VGG、GoogLeNet、ResNet以及DenseNet，网络越来越深，架构越来越复杂，虽然分类精度大幅提升，但是模型中的参数也成倍增加，对计算机内存的要求也越来越高[9-11]。文献[12]在极限学习机(Extreme Learing Machine，ELM)的基础上引入局部感受野的概念，提出基于局部感受野的极限学习机(ELM-LRF)[13]。ELM-LRF可以实现输入层与隐含层的局部连接，不仅能够发挥局部感受野的局部感知优势，还继承了ELM学习速率快、泛化性能高的优点[14-15]，在保证分类性能的同时，模型参数和训练时间均较CNN大幅减小。但ELM-LRF算法中局部感受野采用单一尺度的卷积核，对复杂图像难以取得较好的分类效果。文献[16]提出多尺度局部感受野的极限学习机算法(ELM-MSLRF)，ELM-MSLRF通过多个不同尺度的卷积核更充分地提取图像信息，使得分类效果更好。文献[17]在ELM-MSLRF的基础上进行改进，构建一种多模态融合框架，算法通过将物体材质视觉和触觉信息进行融合，大幅提高了分类性能。但是，ELM-MSLRF使用的ELM在训练数据时需要将所有数据输入到模型中，不能单纯地更新数据。在线序列极限学习机(Online Sequence Extreme Learning Machine，OSELM)[18-19]可以逐个或逐块(数据块)学习数据，因此，可以采用OSELM用于在线学习和网络更新。OSELM不仅具有ELM速度快、泛化能力强的优点，还可以随着新数据的输入而不断更新模型，无需重新再训练所有数据。

本文针对传统多模态框架ELM在训练过程中需要输入所有数据的问题，提出一种多模态融合的多尺度局部感受野在线序列极限学习机算法。在训练过程中，对样本分批次地进行增量式训练，且训练新数据时不再训练旧数据。在特征提取过程中，对传统的ELM框架进行改进，通过保留更多的特征图来提高算法的学习性能，从而提高分类精度。

1 在线序列极限学习机

OSELM由LIANG等[18]于2006年提出，该算法主要解决极限学习机无法实时动态地处理数据而花费时间过长的问题。OSELM可以逐个或者逐块地学习，并丢弃已经完成训练的数据，从而大幅缩短训练所需时间。OSELM的训练过程主要分成初始阶段和在线学习阶段两部分。

1)初始阶段

初始样本$D_0=\left\{\left(X_i,t_i\right)\right\}_{i=1}^{N_0}$，其中，$X$为输入，$t$为对应的标签。分别用$n$、$L$、$m$表示输入网络神经元的数量、隐含层节点数量和输出神经元数量，用$G(X)$表示网络的激活函数，ω_i和b_i分别表示随机产生的输入权值和隐含层的偏置。则初始隐含层输出矩阵$\boldsymbol{H}_0$如式(1)所示：

$$\boldsymbol{H}_0=\begin{bmatrix} G\left(\omega_1\cdot X_1+b_1\right) & \cdots & G\left(\omega_L\cdot X_1+b_L\right) \\ \vdots & & \vdots \\ G\left(\omega_1\cdot X_{N_0}+b_1\right) & \cdots & G\left(\omega_L\cdot X_{N_0}+b_L\right) \end{bmatrix}_{N_0\times L} \tag{1}$$

相应地，网络的初始输出权重$\boldsymbol{\beta}^{(0)}$如式(2)所示：

$$\boldsymbol{\beta}^{(0)}=\boldsymbol{P}_0^{-1}\boldsymbol{H}_0^{\mathrm{T}}\boldsymbol{T}_0 \tag{2}$$

其中：$\boldsymbol{H}_0^{\mathrm{T}}$是初始隐含层输出矩阵的转置；$\boldsymbol{P}_0=\boldsymbol{H}_0^{\mathrm{T}}\boldsymbol{H}_0$；$\boldsymbol{T}_0=\left[t_1,t_2,\cdots,t_{N_0}\right]^{\mathrm{T}}$。

2)在线学习阶段

令g表示数据块个数，设定初始值$g=0$。通过数据块$D_{g+1}=\left\{\left(X_j,t_j\right)\right\}_{j=1}^{N_{g+1}}$对网络的输出权重进行顺序更新。假设当前已有$g$个数据块输入到模型中，当加入新的训练数据块时，输出权重$\boldsymbol{\beta}^{(g+1)}$如式(3)所示[18]：

$$\boldsymbol{\beta}^{(g+1)}=\boldsymbol{\beta}^{(g)}+\boldsymbol{P}_{g+1}\boldsymbol{H}_{g+1}^{\mathrm{T}}\left(\boldsymbol{T}_{g+1}-\boldsymbol{H}_{g+1}\boldsymbol{\beta}^{(g)}\right) \tag{3}$$

其中：$\boldsymbol{P}_{g+1}=\boldsymbol{P}_g-\boldsymbol{P}_g\boldsymbol{H}_{g+1}^{\mathrm{T}}\left(\boldsymbol{I}+\boldsymbol{H}_{g+1}\boldsymbol{P}_g\boldsymbol{H}_{g+1}^{\mathrm{T}}\right)^{-1}\boldsymbol{H}_{g+1}\boldsymbol{P}_g$；$\boldsymbol{H}_{g+1}$为第$g+1$个块数据集隐含层的输出向量；$\boldsymbol{T}_{g+1}$

计算机工程 www.ecice06.com

为第$g+1$个块数据集样本对应的标签。

2　极限学习机多模态融合算法

基于多尺度局部感受野的极限学习机多模态融合算法（MM-MSLRF-ELM）于2018年由LIU等提出，是一种通过多模态融合进行物体材质识别的算法[17]。该算法不仅可以通过融合多模态信息完成分类任务，而且在提取模态信息的过程中采用了多尺度局部感受野，使算法可以学习到更完备的特征。MM-MSLRF-ELM算法具体步骤如下：

步骤1　对每种模态样本随机生成初始权重并进行正交。

设局部感受野有S个不同的尺度，每个尺度局部感受野的大小为$r_s\times r_s, s=1,2,\cdots,S$。每个尺度下生成$K$个不同的输入权重，即每个尺度下可生成$K$个不同的特征图。设输入图像的大小为$(d\times d)$，则第$s$个尺度的特征图大小为$(d-r_s+1)\times(d-r_s+1)$。

为了方便起见，使用上标v和h分别表示视觉和触觉模态。由式(4)随机生成第s个尺度的初始视觉和触觉权重矩阵$\hat{\boldsymbol{A}}_{\text{init}}^{\text{v}(s)}$、$\hat{\boldsymbol{A}}_{\text{init}}^{\text{h}(s)}$，并通过奇异值分解（Singular Value Decomposition，SVD）进行正交化，正交化结果中的每一列$\hat{\boldsymbol{a}}_k^{\text{v}(s)}$和$\hat{\boldsymbol{a}}_k^{\text{h}(s)}$都是$\hat{\boldsymbol{A}}_{\text{init}}^{\text{v}(s)}$、$\hat{\boldsymbol{A}}_{\text{init}}^{\text{h}(s)}$的正交基。

$$\begin{aligned}&\hat{\boldsymbol{A}}_{\text{init}}^{\text{v}(s)},\hat{\boldsymbol{A}}_{\text{init}}^{\text{h}(s)}\in\mathbb{R}^{r_s^2\times K}\\&\hat{\boldsymbol{a}}_k^{\text{v}(s)},\hat{\boldsymbol{a}}_k^{\text{h}(s)}\in\mathbb{R}^{r_s^2}\\&s=1,2,\cdots,S\\&k=1,2,\cdots,K\end{aligned}\tag{4}$$

步骤2　多尺度特征映射。

每种模态第s个尺度的第k个特征图卷积节点(i,j)的值$C_{i,j,k}^{(s)}$根据式(5)计算，其中，X^{v}、X^{h}分别为不同模态的输入样本，不同模态第s个尺度的第k个特征图的输入权重$\boldsymbol{a}_k^{\text{v}(s)}$、$\boldsymbol{a}_k^{\text{h}(s)}\in\mathbb{R}^{r_s^2}$分别由$\hat{\boldsymbol{a}}_k^{\text{v}(s)}$和$\hat{\boldsymbol{a}}_k^{\text{h}(s)}$逐列排成。

$$\begin{aligned}&C_{i,j,k}^{\text{v}(s)}(X)=\sum_{m=1}^{r_s}\sum_{n=1}^{r_s}\left(X_{i+m-1,j+n-1}^{\text{v}}\cdot\boldsymbol{a}_{m,n,k}^{\text{v}(s)}\right)\\&C_{i,j,k}^{\text{h}(s)}(X)=\sum_{m=1}^{r_s}\sum_{n=1}^{r_s}\left(X_{i+m-1,j+n-1}^{\text{h}}\cdot\boldsymbol{a}_{m,n,k}^{\text{h}(s)}\right)\\&s=1,2,\cdots,S\\&k=1,2,\cdots,K\\&i,j=1,2,\cdots,(d-r_s+1)\end{aligned}\tag{5}$$

步骤3　多尺度平方根池化。

在步骤2之后，对卷积特征进行池化操作，令池化图的大小与特征图的大小相同，均为$(d-r_s+1)\times(d-r_s+1)$。第s个尺度的第k个池化图中的组合节点(p,q)的值$h_{p,q,k}^{(s)}$可由式(6)计算，其中，e_s表示第s个尺度的池化大小。

$$\begin{aligned}&h_{p,q,k}^{\text{v}(s)}=\sqrt{\sum_{i=p-e_s}^{p+e_s}\sum_{j=q-e_s}^{q+e_s}C_{i,j,k}^{2(\text{v}(s))}}\\&h_{p,q,k}^{\text{h}(s)}=\sqrt{\sum_{i=p-e_s}^{p+e_s}\sum_{j=q-e_s}^{q+e_s}C_{i,j,k}^{2(\text{h}(s))}}\\&s=1,2,\cdots,S\\&k=1,2,\cdots,K\\&p,q=1,2,\cdots,(d-r_s+1)\end{aligned}\tag{6}$$

若节点(i,j)不在$(d\text{-}r_s\text{+}1)$范围内，则$C_{i,j,k}^{\text{v}(s)},C_{i,j,k}^{\text{h}(s)}=0$。

步骤4　对每种模态特征进行全连接得到对应每种模态的特征组合矩阵。

将每种模态所有组合节点的值组合成一个行向量，并把N个输入样本的行向量放在一起得到组合矩阵$\boldsymbol{H}^{\text{v}},\boldsymbol{H}^{\text{h}}\in\mathbb{R}^{N\times K\cdot\sum_{s=1}^{S}(d-r_s+1)^2}$。

步骤5　多模态融合。

将不同模态的特征矩阵组合得到混合网络矩阵$\boldsymbol{H}=\left[\boldsymbol{H}^{\text{v}},\boldsymbol{H}^{\text{h}}\right]$，矩阵大小设置为$d'\times d''$，其中，$d'$手动设定，$d''$由式(7)计算得到[17]：

$$d''=\frac{P\times N\times K\cdot\sum_{s=1}^{S}(d-r_s+1)^2}{d'}\tag{7}$$

其中：P表示模态数量；手动设定d'的取值范围为$1\leqslant d'\leqslant P\times N\times K\cdot\sum_{s'=1}^{S'}(d-r_{s'}+1)^2$。

步骤6　混合矩阵的卷积和池化。

此步骤的特征提取过程与步骤2、步骤3相同，此处特征图及池化图的大小变为$(d'-r_{s'}+1)\times(d''-r_{s'}+1)$。混合网络中设局部感受野有$S'$个不同的尺度，每个尺度局部感受野的大小为$r_{s'}\times r_{s'}, s'=1,2,\cdots,S'$。

步骤7　混合网络的特征全连接。

与步骤4相似，将混合网络所有组合节点的值组合成一个行向量，并把输入样本的所有行向量放在一起，得到组合矩阵$\boldsymbol{H}^{\text{hybrid}}\in\mathbb{R}^{N\times K'\cdot\sum_{s'=1}^{S'}(d'-r_{s'}+1)(d''-r_{s'}+1)}$。

步骤8　计算输出权重。

输出权重$\boldsymbol{\beta}$如式(8)所示：

$$\begin{cases}N\leqslant K'\cdot\sum_{s'=1}^{S'}(d'-r_{s'}+1)(d''-r_{s'}+1),\\\boldsymbol{\beta}=\left(\boldsymbol{H}^{\text{hybrid}}\right)^{\text{T}}\left(\dfrac{\boldsymbol{I}}{C}+\boldsymbol{H}^{\text{hybrid}}\left(\boldsymbol{H}^{\text{hybrid}}\right)^{\text{T}}\right)^{-1}\boldsymbol{T}\\N>K'\cdot\sum_{s'=1}^{S'}(d'-r_{s'}+1)(d''-r_{s'}+1),\\\boldsymbol{\beta}=\left(\dfrac{\boldsymbol{I}}{C}+\left(\boldsymbol{H}^{\text{hybrid}}\right)^{\text{T}}\boldsymbol{H}^{\text{hybrid}}\right)^{-1}\left(\boldsymbol{H}^{\text{hybrid}}\right)^{\text{T}}\boldsymbol{T}\end{cases}\tag{8}$$

其中：C为正则化参数；K'为混合网络中的特征图数量；$\boldsymbol{T}$为样本对应的标签。

计算机工程 www.ecice06.com

MM-MSLRF-ELM算法在实验过程中还对输入样本进行颜色R、G、B分离。在对输入样本进行颜色三通道分离后，在每个颜色通道设置S个尺度，且每个尺度生成K个随机权重，整个网络生成$(3\times S\times K)$个特征图。但是，该算法在卷积生成特征图的过程中又将3个颜色通道对应生成的特征图进行合并，实际后续用于池化操作的还是$(S\times K)$个特征图[20-21]。

3 本文算法

本文在MSLRF-OSELM[22]的基础上，结合基于多尺度局部感受野的极限学习机多模态融合算法，提出一种多模态融合的多尺度局部感受野在线序列极限学习机算法(MM-MSLRF-OSELM)。该算法将保留单模态执行过卷积操作生成的特征图，并对实际生成的$(3\times S\times K)$个特征图都进行池化操作，最后完成特征矩阵全连接。

多模态融合通过提取物体在不同模态下的信息，然后进行融合以用于物体识别和分类。该方法不仅利用多尺度局部感受野更充分地提取了特征，而且通过将不同模态下的特征进行融合，大幅提高了算法的测试精度，此外还可在线更新训练数据，在实际问题中具有更大的适用性。MM-MSLRF-OSELM算法整体架构如图1所示，其包含$(p+1)$个MM-MSLRF-NET，每个MM-MSLRF-NET包含多种模态信息，在线生成的块数据集依次输入相应的网络以更新输出权重$\boldsymbol{\beta}$。

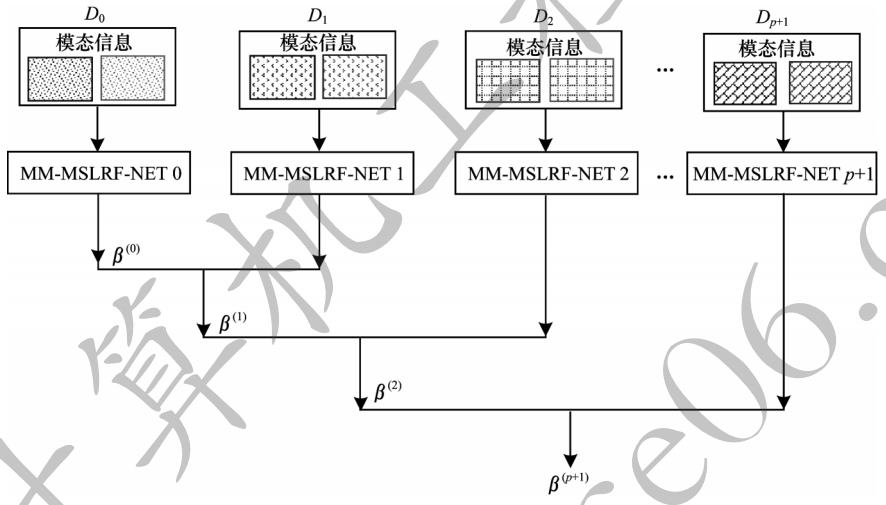

图1 MM-MSLRF-OSELM算法架构

Fig.1 The architecture of MM-MSLRF-OSELM algorithm

MM-MSLRF-OSELM算法具体步骤如下：

步骤1 初始阶段。

1)随机生成并正交化每种模态的初始权重。

设输入图像大小为$(d\times d)$，将输入图像分成R、G、B 3个颜色分量并送入对应的颜色通道中，每个颜色通道设置S个不同尺度的局部感受野，且在每个尺度下随机生成K个不同的初始权重。因此，整个网络可以生成$3\times S\times K$个特征图。记第s个尺度的局部感受野大小为$r_s\times r_s$，$s=1,2,\cdots,S$，则第s个尺度的特征图大小为$\left(d-r_s+1\right)\times\left(d-r_s+1\right)$。

为了方便起见，使用上标image、acceleration分别表示视觉模态和触觉加速度模态。根据式(9)，网络随机生成c颜色通道中第s个尺度的视觉图像与触觉加速度模态的初始权重矩阵$\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{acceleration}(s)}$。对初始权重矩阵$\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{acceleration}(s)}$通过SVD方法进行正交化操作，得到正交矩阵$\hat{\boldsymbol{A}}_{(c)}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{A}}_{(c)}^{\mathrm{acceleration}(s)}$。正交矩阵$\hat{\boldsymbol{A}}_{(c)}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{A}}_{(c)}^{\mathrm{acceleration}(s)}$中的每一列$\hat{\boldsymbol{a}}_{c,k}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{a}}_{c,k}^{\mathrm{acceleration}(s)}$都是初始权重矩阵$\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{acceleration}(s)}$的正交基。其中，$c$颜色通道中第$s$个尺度的第$k$个输入权重为$\boldsymbol{a}_{c,k}^{\mathrm{image}(s)}$、$\boldsymbol{a}_{c,k}^{\mathrm{acceleration}(s)}\in\mathbb{R}^{r_s\times r_s}$，对应于$\hat{\boldsymbol{a}}_{c,k}^{\mathrm{image}(s)}$、$\hat{\boldsymbol{a}}_{c,k}^{\mathrm{acceleration}(s)}\in\mathbb{R}^{r_s\times r_s}$。

$$\begin{aligned}
&\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{image}(s)}、\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{acceleration}(s)}\in\mathbb{R}^{r_s^2\times K}\\
&\hat{\boldsymbol{a}}_{\mathrm{init},c,k}^{\mathrm{image}(s)}、\hat{\boldsymbol{a}}_{\mathrm{init},c,k}^{\mathrm{acceleration}(s)}\in\mathbb{R}^{r_s^2}\\
&\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{image}(s)}=\left[\hat{\boldsymbol{a}}_{\mathrm{init},c,1}^{\mathrm{image}(s)},\hat{\boldsymbol{a}}_{\mathrm{init},c,2}^{\mathrm{image}(s)},\cdots,\hat{\boldsymbol{a}}_{\mathrm{init},c,K}^{\mathrm{image}(s)}\right]\\
&\hat{\boldsymbol{A}}_{\mathrm{init}(c)}^{\mathrm{acceleration}(s)}=\left[\hat{\boldsymbol{a}}_{\mathrm{init},c,1}^{\mathrm{acceleration}(s)},\hat{\boldsymbol{a}}_{\mathrm{init},c,2}^{\mathrm{acceleration}(s)},\cdots,\hat{\boldsymbol{a}}_{\mathrm{init},c,K}^{\mathrm{acceleration}(s)}\right]\\
&c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}\\
&k=1,2,\cdots,K\\
&s=1,2,\cdots,S
\end{aligned}\tag{9}$$

2)每种模态的多尺度特征映射。

视觉模态和触觉加速度模态在c颜色通道中第s个尺度的第k个特征图中卷积节点(i,j)值可由式(10)计算，其中，$\boldsymbol{X}^{(c)}$为不同模态样本进行R、G、B颜色三通道分离后对应的向量。

计算机工程 www.ecice06.com

$$C_{i,j,c,k}^{\text{image}(s)}\left(\boldsymbol{X}^{(c)}\right)=\sum_{m=1}^{r_s}\sum_{n=1}^{r_s}\left(X_{i+m-1,j+n-1}^{\text{image}(c)}\cdot\boldsymbol{a}_{m,n,c,k}^{\text{image}(s)}\right)$$

$$C_{i,j,c,k}^{\text{acceleration}(s)}\left(\boldsymbol{X}^{(c)}\right)=\sum_{m=1}^{r_s}\sum_{n=1}^{r_s}\left(X_{i+m-1,j+n-1}^{\text{acceleration}(c)}\cdot\boldsymbol{a}_{m,n,c,k}^{\text{acceleration}(s)}\right)$$

$$c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}$$

$$k=1,2,\cdots,K$$

$$s=1,2,\cdots,S$$

$$i,j=1,2,\cdots,\left(d-r_s+1\right)\tag{10}$$

3）每种模态的多尺度平方根池化。

视觉模态、触觉加速度模态在c颜色通道中第s个尺度的第k个池化图中组合节点(p,q)的池化特征计算如下：

$$h_{p,q,c,k}^{\text{image}(s)}=\sqrt{\sum_{i=p-e_s}^{p+e_s}\sum_{j=q-e_s}^{q+e_s}\left(C_{i,j,c,k}^{\text{image}(s)}\right)^2}$$

$$h_{p,q,c,k}^{\text{acceleration}(s)}=\sqrt{\sum_{i=p-e_s}^{p+e_s}\sum_{j=q-e_s}^{q+e_s}\left(C_{i,j,c,k}^{\text{acceleration}(s)}\right)^2}$$

$$c\in\{\mathrm{R},\mathrm{G},\mathrm{B}\}$$

$$k=1,2,\cdots,K$$

$$s=1,2,\cdots,S$$

$$p,q=1,2,\cdots,(d-r_s+1)\tag{11}$$

若节点(i,j)不在$(d-r_s+1)$范围内，则$C_{i,j,c,k}^{\text{image}(s)}=0$，$C_{i,j,c,k}^{\text{acceleration}(s)}=0$。

4）对每种模态进行特征全连接。

将视觉模态和触觉加速度模态输入样本对应的组合节点值分别连接成行向量，并将N_0个输入样本对应的行向量进行组合，得到2种模态的组合特征向量矩阵$\boldsymbol{H}^{\text{image}},\boldsymbol{H}^{\text{acceleration}}\in\mathbb{R}^{N_0\times K\cdot\sum_{s=1}^{S}\left(d-r_s+1\right)^2}$。

5）模态融合。

将2种模态的组合特征向量矩阵组合成1个混合矩阵$\boldsymbol{H}=\left[\boldsymbol{H}^{\text{image}},\boldsymbol{H}^{\text{acceleration}}\right]$，混合矩阵大小为$d'\times d''$，由式(7)得到。

6）多模态多尺度特征映射与平方根池化。

将2种模态融合后得到的混合矩阵输入到一个新的混合网络，该网络设有S'个尺度，每个尺度中产生K'个不同的输入权重，则该网络可以生成$S'\times K'$个特征图，记第s'个尺度的局部感受野大小为$r_{s'}\times r_{s'}$，则第s'个尺度的第k'个特征图的大小为$\left(d'-r_{s'}+1\right)\times\left(d''-r_{s'}+1\right)$。该网络的特征映射及平方根池化过程与第1步~第3步相似。

7）多模态特征向量全连接。

此时的特征全连接方法与第4步相似，得到混合网络的组合层矩阵$\boldsymbol{H}^{\text{hybrid}}\in\mathbb{R}^{N_0\times K'\cdot\sum_{s'=1}^{S'}\left(d'-r_{s'}+1\right)\left(d''-r_{s'}+1\right)}$。

8）计算初始输出权重$\boldsymbol{\beta}^{(0)}$。

根据式(8)计算初始输出权重$\boldsymbol{\beta}^{(0)}$。

步骤2　在线学习阶段。

1）设$g=0$，假设有N_{g+1}个新样本进入模型，该模型每个模态的特征提取以及特征全连接过程与步骤1初始阶段第2步~第4步相似，各步骤中的参数设置均相同。多模态融合及融合后的卷积、池化以及池化特征的全连接过程与步骤1初始阶段第5步~第7步相似，得到组合层矩阵$\boldsymbol{H}_{g+1}^{\text{hybrid}}\in\mathbb{R}^{N_{g+1}\times K'\cdot\sum_{s'=1}^{S'}\left(d'-r_{s'}+1\right)\left(d''-r_{s'}+1\right)}$。

2）由式(12)根据$\boldsymbol{H}_{g+1}^{\text{hybrid}}$更新输出权重$\boldsymbol{\beta}^{(g+1)}$。

$$\begin{aligned}\boldsymbol{\beta}^{(g+1)}&=\boldsymbol{\beta}^{(g)}+\boldsymbol{P}_{g+1}\left(\boldsymbol{H}_{g+1}^{\text{hybrid}}\right)^{\mathrm{T}}\left(\boldsymbol{T}_{g+1}-\boldsymbol{H}_{g+1}^{\text{hybrid}}\boldsymbol{\beta}^{(g)}\right)\\\boldsymbol{P}_{g+1}&=\boldsymbol{P}_g-\boldsymbol{P}_g\left(\boldsymbol{H}_{g+1}^{\text{hybrid}}\right)^{\mathrm{T}}\left(\boldsymbol{I}+\boldsymbol{H}_{g+1}^{\text{hybrid}}\boldsymbol{P}_g\left(\boldsymbol{H}_{g+1}^{\text{hybrid}}\right)^{\mathrm{T}}\right)^{-1}\\&\quad\boldsymbol{H}_{g+1}^{\text{hybrid}}\boldsymbol{P}_g\end{aligned}\tag{12}$$

其中：$\boldsymbol{P}_0=\boldsymbol{H}_0^{\mathrm{T}}\boldsymbol{H}_0$，$\left(\boldsymbol{H}_{g+1}^{\text{hybrid}}\right)^{\mathrm{T}}$为第$g+1$个块数据集在混合网络中组合层输出矩阵的转置；$\boldsymbol{T}_{g+1}$为第$g+1$个块数据集样本对应的标签。

3）令$g=g+1$，如果N_{g+1}是最后一个在线块数据集样本，则在线学习结束；否则，重复步骤2在线学习阶段的第1步~第2步，直到数据集是在线训练数据集的最后一个块数据集。最终根据式(13)更新输出权重：

$$\boldsymbol{\beta}=\boldsymbol{\beta}^{(g+1)}\tag{13}$$

4　实验验证

4.1　数据集

为了验证本文所提算法(MM-MSLRF-OSELM)的有效性，在TUM触觉纹理数据集上进行实验。TUM触觉纹理数据集是一个新型的多模态数据集，包含108种不同物体的触觉加速度、摩擦力、金属检测信号、反射率、声音和视觉图像信号，且TUM触觉纹理数据集每种信号均包含2组数据(有约束条件下记录的数据和无约束条件下的数据)，数据是由10个自由手(5个线性和5个圆形运动)记录组成。本文重组2组数据并随机从每组每个类别中选择一个样本作为测试集，其他数据作为训练集。每个模态设置(108×2)个测试样本和(108×18)个训练样本，并将(108×18)个静态训练样本转化为动态增量训练样本以训练在线网络。

4.2　实验设置

本文实验主要选取TUM数据集中的视觉图像信号和触觉加速度信号，输入样本预处理过程参考文献[17]。在实验中，分别通过单模态实验和两模态融合实验来验证算法的性能，具体实验设置如下：

1）单模态实验。将处理后得到的视觉图像和触觉加速度频谱图作为输入样本进行实验，本文局部感受野选择2个不同的尺度，且每个尺度通道设置

www.c-s-a.org.cn

2个特征图，为了验证块数据集大小对实验结果的影响以及本文算法是否可以使用新数据更新训练网络，设置数据块大小分别为162、243、486，具体设置如表1所示。

表1 单模态实验参数设置

Table 1 Parameters setting of single-modal experiment

尺度	r_s		e_s		K	C		数据块大小		
	r_1	r_2	e_1	e_2		视觉图像	触觉加速度			
2	3	6	2	5	2	1E-3	1E-1	162	243	486

2）两模态融合实验。本文通过将视觉模态和触觉加速度模态特征进行融合，形成混合网络进行实验以验证模态融合的有效性。在对每种模态分别提取特征时，本文采用2个不同尺度的局部感受野，感受野大小与单模态实验中的感受野大小相同。考虑计算机的内存问题，两模态融合后得到的混合网络进行特征提取时也选择2个不同尺度的局部感受野，每个尺度通道的特征图数量均设置为2。本文设置3组2个尺度的局部感受野，分别为{83,86}、{93,96}、{103,106}，然后进行实验以观察局部感受野大小对测试精度的影响。在实验过程中，设置块数据集大小为486，正则化参数$C=1E-6$。具体参数设置如表2所示。

表2 两模态融合实验参数设置

Table 2 Parameters setting of two-modal fusion experiment

尺度	模态融合前				模态融合后				K	C	数据块大小
	r_s		e_s		r_s'		e_s'				
	r_1	r_2	e_1	e_2	r_1'	r_2'	e_1'	e_2'			
2	3	6	2	5	83 93 103	86 96 106	82 92 102	85 95 105	2	1E-6	486

4.3 算法有效性验证

在2个不同尺度局部感受野的情况下，本文采用十折交叉验证统计实验结果。单模态实验中分批训练数据块大小对实验结果的影响如表3和表4所示。由表3和表4可以看出，块数据集越大，即训练样本越多，训练精度越高，整体训练时间越快，相对应的测试精度随着训练精度的不同也有所变化，由于测试数据大小无变化，因此测试时间几乎无变化。

两模态融合实验结果如表5所示，由表5可以看出，局部感受野大小对测试结果有明显影响，局部感受野越小，分类精度越高，局部感受野由小到大对应的测试精度分别为65.89%、59.63%、48.01%。通过对比表4和表5可以看出，两模态融合的分类精度远高于单模态，验证了模态融合的优势以及可行性。

表3 数据块大小不同时不同模态的训练精度及训练时间

Table 3 Training accuracy and training time of different modes corresponding to data block size

模态	数据块大小为162		数据块大小为243		数据块大小为486	
	训练精度/%	训练时间/s	训练精度/%	训练时间/s	训练精度/%	训练时间/s
视觉图像	88.63	2157.65	94.44	1 891.89	98.23	1 605.07
触觉加速度	88.17	2126.61	94.01	1 866.64	99.28	1 591.66

表4 数据块大小不同时不同模态的测试精度及测试时间

Table 4 Testing accuracy and testing time of different modes corresponding to data block size

模态	数据块大小为162		数据块大小为243		数据块大小为486	
	测试精度/%	测试时间/s	测试精度/%	测试时间/s	测试精度/%	测试时间/s
视觉图像	39.07	0.37	42.18	0.34	41.48	0.35
触觉加速度	41.56	0.35	42.55	0.33	43.58	0.33

表5 融合网络中不同局部感受野时的测试精度及测试时间

Table 5 Testing accuracy and testing time of different local receptive field sizes in fusion network

$r_{s'}$	测试精度/%	测试时间/s
83,86	65.89	31.91
93,96	59.63	25.81
103,106	48.01	19.58

为了更好地说明本文算法的有效性，将本文算法与MM-MSLRF-ELM[17]算法进行对比，结果如表6所示，单模态实验时两种对比算法的参数设置相同，MM-MSLRF-OSELM算法的测试精度在2种模态下均高于MM-MSLRF-ELM算法，同时时间消耗也都接近MM-MSLRF-ELM算法的3倍。因为本文实验的时间单位为s，所以3倍的时间换算法测试精度10%的提升（视觉图像）是值得的。在两模态融合的对比实验中，由表6可以观察到，虽然MM-MSLRF-OSELM的测试精度高于MM-MSLRF-ELM，但是提高幅度较低，这是由于局部感受野大小设置的原因，具体分析如下：

在模态融合网络局部感受野同样设置为2个尺度且大小分别为83×83和86×86时，MM-MSLRF-ELM两模态融合后的矩阵大小行小于本文设置的局部感受野大小，实验结果不可取。因此，本文对MM-MSLRF-ELM算法仿真时模态融合网络局部感受野2个尺度大小的设置分别为5×5和7×7，该感受野大小远小于本文算法仿真局部感受野的大小。从表5可以看出，局部感受野越小，分类精度越高，且分类精度变化明显。因此，本文的MM-MSLRF-OSELM在计算机内存满足的情况下精度提升空间很大，其具

计算机工程 www.ecice06.com

有可行性。虽然无论单模态实验还是模态融合实验,MM-MSLRF-OSELM耗时都比MM-MSLRF-ELM长,但精度明显提高,因此,MM-MSLRF-OSELM具有一定优势。

表6 不同模态时的测试精度与测试时间

Table 6 Testing accuracy and testing time in different modals

模态	测试指标	MM-MSLRF-ELM算法	MM-MSLRF-OSELM算法
视觉图像	测试精度/%	31.89	41.98
	测试时间/s	0.13	0.35
视觉加速度	测试精度/%	41.26	43.58
	测试时间/s	0.16	0.33
两模态融合	测试精度/%	62.59	65.89
	测试时间/s	0.59	31.39

5 结束语

本文提出一种MM-MSLRF-OSELM算法,选用TUM数据集中的视觉图像和触觉加速度信息进行实验,通过实验证明两模态融合后的分类精度明显高于单模态的分类精度,且通过与MM-MSLRF-ELM算法进行对比,进一步证明本文算法具有较好的分类性能。MM-MSLRF-OSELM在训练过程中仅对新数据进行在线更新训练,在实际中适用性更强。由于本文利用了不同模态的信息,而这些信息中可能存在一些冗余特征,因此下一步将采用属性约简算法对冗余特征进行约简。

参考文献

[1] ZHENG H T, FANG L, JI M Q, et al. Deep learning for surface material classification using haptic and visual information [J]. IEEE Transactions on Multimedia, 2016, 18(12): 2407-2416.

[2] STRESE M, SCHUWERK C, IEPURE A, et al. Multimodal feature-based surface material classification [J]. IEEE Transactions on Haptics, 2017, 10(2): 226-239.

[3] 王玉伟,董西伟,陈芸. 基于稀疏表示的多模态生物特征识别算法[J]. 计算机工程,2016,42(10):219-225.
WANG Y W, DONG X W, CHEN Y. Multimodal biometric recognition algorithm based on sparse representation [J]. Computer Engineering, 2016, 42(10): 219-225. (in Chinese)

[4] 杨楠. 基于视触觉多特征融合的步态识别方法研究[D]. 天津:河北工业大学,2015.
YANG N. Study on gait recognition method based on visual-tactile features fusion [D]. Tianjin: Hebei University of Technology, 2015. (in Chinese)

[5] 李凤雪. 基于局部感受野极限学习机的研究与应用[D]. 太原:太原理工大学,2017.
LI F X. Research and application based on ELM-LRF [D]. Taiyuan, Taiyuan University of Technology, 2017. (in Chinese)

[6] 冯德正,FRANCIS L. 多模态研究的现状与未来——第七届国际多模态会议评述[J]. 上海外国语大学学报,2015,38(4):108-113.
FENG D Z, FRANCIS L. State of the art and future directions of multimodal studies: a review of the 7th international conference on multimodality [J]. Journal of Foreign Languages, 2015, 38(4): 108-113. (in Chinese)

[7] LIU H P, WU Y P, SUN F C, et al. Weakly paired multimodal fusion for object recognition [J]. IEEE Transactions on Automation Science and Engineering, 2018, 15(2): 784-795.

[8] LIU H P, SUN F C, FANG B, et al. Multimodal measurements fusion for surface material categorization [J]. IEEE Transactions on Instrumentation and Measurement, 2018, 67(2): 246-256.

[9] 杨斌,钟金英. 卷积神经网络的研究进展综述[J]. 南华大学学报(自然科学版),2016,30(3):66-72.
YANG B, ZHONG J Y. Review of convolution neural network [J]. Journal of University of South China (Science and Technology), 2016, 30(3): 66-72. (in Chinese)

[10] 孙新胜. 基于多层卷积神经网络的研究与应用[D]. 杭州:杭州电子科技大学,2018.
SUN X S. Research and application of multi-layer convolution neural network [D]. Hangzhou: Hangzhou Dianzi University, 2018. (in Chinese)

[11] 杨楠,南琳,张丁一,等. 基于深度学习的图像描述研究[J]. 红外与激光工程,2018,47(2):9-16.
YANG N, NAN L, ZHANG D Y, et al. Research on image interpretation based on deep learning [J]. Infrared and Laser Engineering, 2018, 47(2): 9-16. (in Chinese)

[12] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks [C]//Proceedings of 2004 IEEE International Joint Conference on Neural Networks. Washington D. C., USA: IEEE Press, 2004: 985-990.

[13] HUANG G B, BAI Z, KASUN L L C, et al. Local receptive fields based extreme learning machine [J]. Computational Intelligence Magazine, 2015, 10(2): 18-29.

[14] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications [J]. Neurocomputing, 2006, 70(1/2/3): 489-501.

[15] HUANG G B, ZHOU H M, DING X J, et al. Extreme learning machine for regression and multiclass classification [J]. IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics), 2012, 42(2): 513-529.

[16] HUANG J H, YU Z L, CAI Z Q, et al. Extreme learning machine with multi-scale local receptive fields for texture classification [J]. Multidimensional Systems and Signal Processing, 2016, 28: 995-1011.

[17] LIU H P, FANG J, XU X Y, et al. Surface material recognition using active multi-modal extreme learning machine [J]. Cognitive Computation, 2018, 10: 937-950.

[18] LIANG N T, HUANG G B, SARATCHANDRAN P, et al. A fast and accurate online sequential learning algorithm for feedforward networks [J]. IEEE Transactions on Neural Networks, 2006, 17(6): 1411-1423.

(下转第80页)

计算机工程 www.ecice.com

（上接第73页）

[19] LAN Y, SOH Y C, HUANG G B. A constructive enhancement for online sequential extreme learning machine [C]// Proceedings of 2009 International Joint Conference on Neural Network. Washington D. C., USA: IEEE Press, 2009: 1708-1713.

[20] 方静. 基于LRF-ELM算法的研究及其在物体材质分类中的应用[D]. 太原: 太原理工大学, 2018.
FANG J. The research based on LRF-ELM algorithm and its application in object material classification [D]. Taiyuan, Taiyuan University of Technology, 2018. (in Chinese)

[21] FANG J, XU X Y, LIU H P, et al. Local receptive field based extreme learning machine with three channels for histopathological image classification [J]. International Journal of Machine Learning and Cybernetics, 2018, 10(7): 1437-1447.

[22] XU X Y, FANG J, LI Q, et al. Multi-scale local receptive field based online sequential extreme learning machine for material classification [C]//Proceedings of ICCSIP'18. Berlin, Germany: Springer, 2018: 37-53.

编辑　吴云芳　薛晋栋