

用于表示级特征融合与分类的相关熵融合极限学习机

吴超^① 李雅倩^{*②} 张亚茹^① 刘彬^①

^①(燕山大学信息科学与工程学院 秦皇岛 066004)

^②(燕山大学工业计算机控制工程河北省重点实验室 秦皇岛 066004)

摘要: 在极限学习机(ELM)网络结构和训练模式的基础上, 该文提出了相关熵融合极限学习机(CF-ELM)。针对多数分类方法中表示级特征融合不充分的问题, 该文将核映射与系数加权相结合, 提出了能够有效融合表示级特征的融合极限学习机(F-ELM)。在此基础上, 用相关熵损失函数替代均方误差(MSE)损失函数, 推导出用于训练F-ELM各层权重矩阵的相关熵循环更新公式, 以增强其分类能力与鲁棒性。为了检验方法的可行性, 该文分别在数据库Caltech 101, MSRC和15 Scene上进行实验。实验结果证明, 该文所提CF-ELM能够在原有基础上进一步融合表示级特征, 从而提高分类正确率。

关键词: 极限学习机; 表示级特征融合; 相关熵; 分类

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2020)02-0386-08

DOI: 10.11999/JEIT190186

Correntropy-based Fusion Extreme Learning Machine for Representation Level Feature Fusion and Classification

WU Chao^① LI Yaqian^② ZHANG Yaru^① LIU Bin^①

^①(School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China)

^②(Key Laboratory of Industrial Computer Control Engineering of Hebei Province, Yanshan University, Qinhuangdao 066004, China)

Abstract: Based on the network structure and training methods of the Extreme Learning Machine (ELM), Correntropy-based Fusion Extreme Learning Machine (CF-ELM) is proposed. Considering the problem that the fusion of representation level features is insufficient in most classification methods, the kernel mapping and coefficient weighting are combined to propose a Fusion Extreme Learning Machine (F-ELM), which can effectively fuse the representation level features. On this basis, the Mean Square Error (MSE) loss function is replaced by the correntropy-based loss function. A correntropy-based cycle update formula for training the weight matrices of the F-ELM is derived to enhance classification ability and robustness. Extensive experiments are performed on Caltech 101, MSRC and 15 Scene datasets respectively. The experimental results show that CF-ELM can further fuse the representation level features to improve the classification accuracy.

Key words: Extreme Learning Machine (ELM); Representation level feature fusion; Correntropy; Classification

1 引言

极限学习机(Extreme Learning Machine, ELM)^[1]是一种单隐层前馈神经网络, 其输出层权重可由随机赋值的输入层权重和偏置来计算, 能够在没有迭代的情况下快速有效地完成网络训练。在此基础上, Huang等人^[2]在训练过程中加入核函数

来实现隐层特征映射, 避免了随机赋值引起的不稳定性。Kasun等人^[3]提出了ELM自编码器(ELM AutoEncoder, ELM-AE), 以实现特征编码和ELM的深层发展。

ELM使用均方误差(Mean Square Error, MSE)作为损失函数, 而MSE依赖于高斯性和线性假设, 对实际数据中的异常值和脉冲噪声非常敏感。当训练集中存在噪声或异常值时, ELM会产生较差的泛化性能^[4]。为解决这一问题, Xing等人^[4]引入相关熵度量作为损失函数, 利用半2次优化求解满足最大化相关熵准则的ELM。Chen等人^[5]提出了基于相关熵损失函数和相关熵稀疏性惩罚项的

收稿日期: 2019-03-27; 改回日期: 2019-09-03; 网络出版: 2019-09-12

*通信作者: 李雅倩 yaqianli@126.com

基金项目: 国家自然科学基金(51641609)

Foundation Item: The National Natural Science Foundation of China (51641609)

多层极限学习机(Multilayer Extreme Learning Machine, MELM), 并利用快速迭代阈值收缩算法对目标函数进行求解。Luo等人^[6]通过将广义相关熵准则引入堆叠极限学习机(Stacked Extreme Learning Machine, SELM)来提高预测性能, 并将其应用于风速预测。上述改进算法仅提高了ELM的通用近似能力, 并没有改善它们提取图像特征的能力。且这些算法只能使用逐列连接图像矩阵生成的向量来训练网络, 因此不能用于处理大尺寸图像分类问题^[7]。

Li等人^[8]将提取出的空间金字塔词袋模型(Spatial Pyramid Matching, SPM)^[9]特征输入ELM进行分类, 使得ELM能够处理大尺寸图像分类问题。但是, 在处理大尺寸图像分类问题时, 单一的图像特征无法充分考虑所有影响因素, 具有明显的局限性, 如: SPM, 局部聚合描述符(Vector of Locally Aggregated Descriptors, VLAD)^[10], 费希尔向量(Fisher Vector, FV)^[11]等。因此, 很多分类方法^[12-15]分别从图像的全局与局部两方面提取表示级特征进行融合。但是, 这些方法仅通过经验加权的方式来融合表示级特征, 没有充分考虑特征之间的相互影响。

虽然ELM具有结构简单、学习速度快与泛化性能高的优点, 但是并不具备有效融合多个表示级特征的能力。因此, 本文基于ELM与相关熵提出了相关熵融合极限学习机(Correntropy-based Fusion Extreme Learning Machine, CF-ELM)。为了能够充分融合与利用多种特征, 本文在输入层中引入核映射来代替权重的随机赋值, 利用ELM-AE训练加权系数。通过核映射与系数加权相结合, 提出了用于融合表示级特征的融合极限学习机(Fusion Extreme Learning Machine, F-ELM)。在此基础上, 本文引入一种相关熵损失函数来替换F-ELM中的MSE损失函数。通过推导出新的相关熵循环更新公式来构造CF-ELM, 以实现增强分类能

力与鲁棒性的目的。由于提出CF-ELM的初衷是实现表示级特征的融合, 因此将文献^[12]中的SPM特征、局部位置特征、全局轮廓特征作为一组输入特征组合, 将SPM, FV与VLAD作为另一组, 并在Caltech 101, MSRC和15 Scenes数据库上进行大量实验来验证CF-ELM的有效性。

2 基于相关熵的融合极限学习机

针对表示级特征融合不充分的问题, 本文提出一种能够融合与分类表示级特征的融合极限学习机。同时, 引入相关熵损失函数, 推导出相关熵ELM与ELM-AE的循环更新公式来训练网络权重, 以进一步提高网络的分类能力与鲁棒性。

2.1 融合极限学习机

本文提出了如图1所示的网络框架, 对输入特征组合进行核映射与系数加权, 以进一步融合表示级特征。该网络包含 W_1, W_2, W_3 与 W_4 4个权重矩阵, 训练过程即为求解各个权重矩阵的过程。

设输入特征为 $A = [a_1, a_2, \dots, a_N]^T \in \mathbf{R}^{N \times n}$, 与之对应的标签特征为 $T \in \mathbf{R}^{N \times l}$ 。为避免随机赋值 W_1 导致的不稳定性, 利用直方图交叉核(Histogram Intersection Kernel, HIK)函数将 $A \in \mathbf{R}^{N \times n}$ 映射为特征 $B \in \mathbf{R}^{N \times N}$, 则 $B_{ij} = k(a_i, a_j)$, $k(\cdot)$ 为映射函数。

为增强特征中类别信息的辨识度与降低混淆信息的干扰度, 对 B 进行系数加权。本文利用 W_2 与 W_3 将 B 编码为加权系数矩阵, 编码过程如图1中虚线的系数加权部分所示。首先, 利用ELM求出权重矩阵 $W_2 \in \mathbf{R}^{N \times l}$, 如式(1)所示, K 为惩罚系数。用 W_2 将 B 降维成特征 $C \in \mathbf{R}^{N \times l}$, 则 $C = BW_2$ 。

$$W_2 = \left(B^T B + \frac{I}{K} \right)^{-1} B^T T \quad (1)$$

然后, 利用ELM-AE来训练稀疏编码矩阵 $W_3 \in \mathbf{R}^{l \times N}$ 。令 C 作为ELM-AE的输入, 随机赋值输入权重 $w \in \mathbf{R}^{l \times N}$ 、偏置 $\zeta \in \mathbf{R}^{N \times N}$, 激活函数为

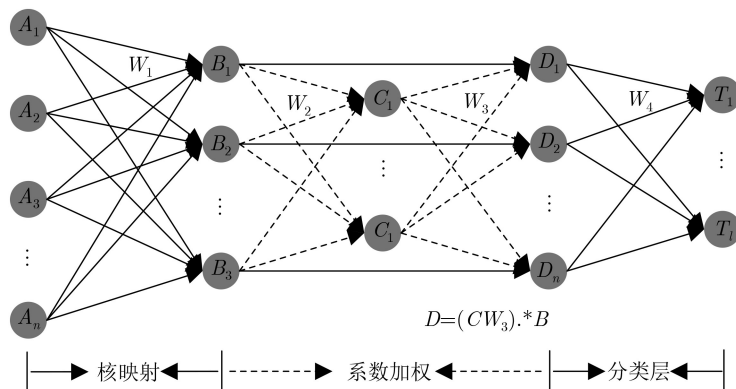


图1 融合极限学习机的网络结构

$g(\cdot)$, 则隐层输出特征 $\mathbf{H} = g(\mathbf{C}\mathbf{w} + \boldsymbol{\zeta}) \in \mathbf{R}^{N \times N}$. 让 \mathbf{C} 作为目标特征来训练网络输出权重 $\boldsymbol{\theta} \in \mathbf{R}^{N \times l}$, 如式(2)所示. 令 $\mathbf{W}_3 = \boldsymbol{\theta}^T$, \mathbf{B} 的加权系数为 $\mathbf{C}\mathbf{W}_3$, 加权后的特征为 $\mathbf{D} \in \mathbf{R}^{N \times N}$, 则 $\mathbf{D} = (\mathbf{C}\mathbf{W}_3) \cdot * \mathbf{B}$. 其中, $\cdot *$ 符号代表矩阵 $(\mathbf{C}\mathbf{W}_3)$ 与 \mathbf{B} 对应位置元素相乘.

$$\boldsymbol{\theta} = \left(\mathbf{H}^T \mathbf{H} + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{H}^T \mathbf{C} \quad (2)$$

最后, 令 \mathbf{D} 代替式(1)中的 \mathbf{B} , 来训练最终的分类权重矩阵 $\mathbf{W}_4 \in \mathbf{R}^{N \times l}$. 至此, F-ELM 训练完成.

传统 ELM 与 ELM-AE 使用 MSE 作为损失函数, 因此式(1)与式(2)训练的权重矩阵具有对噪声干扰敏感的缺点. 为解决这一问题, 故引入相关熵损失函数来替换 MSE 损失函数, 推导出相关熵 ELM 与 ELM-AE. 通过将其用于训练权重矩阵以增强 F-ELM 对图像中嘈杂背景干扰的鲁棒性, 从而继续提升分类正确率.

2.2 基于相关熵的训练方法

首先, 引入相关熵损失函数并推导出相关熵 ELM. 设输入 N 个 n 维特征 $\mathbf{X} \in \mathbf{R}^{N \times n}$, 与之对应的标签特征为 $\mathbf{T} \in \mathbf{R}^{N \times l}$. 随机赋值 ELM 的输入权重 $\mathbf{w} \in \mathbf{R}^{n \times n_h}$, 偏置 $\mathbf{b} \in \mathbf{R}^{N \times n_h}$, 则隐层输出特征 $\mathbf{H} = g(\mathbf{X}\mathbf{w} + \mathbf{b}) \in \mathbf{R}^{N \times n_h}$, 其中 n_h 代表隐层特征的维数. 为了便于利用拉格朗日乘子法求解基于相关熵的 ELM, 本文改变文献[4]中相关熵的约束条件, 将求取最大值改变为式(3)中的求取最小值.

$$\left. \begin{aligned} \text{Min } L &= \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{K}{2} \sum_{i=1}^N \xi_i^2 \\ \text{s.t. } \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}\|^2}{2\sigma^2}\right) &= 1 - \xi_i, \\ i &= 1, 2, \dots, N \end{aligned} \right\} \quad (3)$$

其中, ξ_i 为相关熵误差. 根据 Karush-Kuhn-Tucker^[3] 定理, 式(3)的对偶优化公式如式(4)所示

$$\left. \begin{aligned} L &= \frac{1}{2} \|\boldsymbol{\beta}\|^2 + \frac{K}{2} \sum_{i=1}^N \xi_i^2 \\ &\quad - \sum_{i=1}^N \mu_i \left[\exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}\|^2}{2\sigma^2}\right) - 1 + \xi_i \right] \end{aligned} \right\} \quad (4)$$

其中, μ_i 为拉格朗日算子. 分别对式(4)中 $\boldsymbol{\beta}$, ξ_i 与 μ_i 求偏导, 并令 $\mathbf{G} \in \mathbf{R}^{N \times N}$ 为对角矩阵, 主对角线上元素为

$$\begin{aligned} G_{ii} = \alpha_i &= \frac{1}{\sigma^2} \left(1 - \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}\|^2}{2\sigma^2}\right) \right) \\ &\quad \cdot \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}\|^2}{2\sigma^2}\right) \end{aligned} \quad (5)$$

可将式(4)解为

$$\boldsymbol{\beta} - K(\mathbf{H}^T \mathbf{G} \mathbf{T} - \mathbf{H}^T \mathbf{G} \mathbf{H} \boldsymbol{\beta}) = \mathbf{0} \quad (6)$$

当 $n_h \leq N$ 时,

$$\boldsymbol{\beta} = \left(\mathbf{H}^T \mathbf{G} \mathbf{H} + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{H}^T \mathbf{G} \mathbf{T} \quad (7)$$

将 $\boldsymbol{\beta} = \mathbf{H}^T \boldsymbol{\eta}$ ($\boldsymbol{\eta} \in \mathbf{R}^{N \times l}$) 代入式(6)中化简可得, 当 $n_h > N$ 时,

$$\boldsymbol{\beta} = \mathbf{H}^T \left(\mathbf{G} \mathbf{H} \mathbf{H}^T + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{G} \mathbf{T} \quad (8)$$

利用循环迭代算法来求解 \mathbf{G} 与 $\boldsymbol{\beta}$. 在第 1 次迭代时, 令 $G_{ii}^0 = 1$, $i = 1, 2, \dots, N$ 求出 $\boldsymbol{\beta}^1$. 在第 τ 次迭代时, 令

$$\begin{aligned} G_{ii}^{\tau-1} &= \frac{1}{\sigma^2} \left(1 - \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}^{\tau-1}\|^2}{2\sigma^2}\right) \right) \\ &\quad \cdot \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}^{\tau-1}\|^2}{2\sigma^2}\right) \end{aligned} \quad (9)$$

由于 $e^x - 1 \sim x$ ($x \rightarrow 0$), 则 $G_{ii}^{\tau-1}$ 可以等价变换为

$$G_{ii}^{\tau-1} = \frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}^{\tau-1}\|^2}{2\sigma^4} \exp\left(-\frac{\|\mathbf{t}_i - \mathbf{h}_i \boldsymbol{\beta}^{\tau-1}\|^2}{2\sigma^2}\right) \quad (10)$$

则 $\boldsymbol{\beta}^\tau$ 可由式(11)求出

$$\boldsymbol{\beta}^\tau = \begin{cases} \left(\mathbf{H}^T \mathbf{G}^{\tau-1} \mathbf{H} + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{H}^T \mathbf{G}^{\tau-1} \mathbf{T}, & n_h \leq N \\ \mathbf{H}^T \left(\mathbf{G}^{\tau-1} \mathbf{H} \mathbf{H}^T + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{G}^{\tau-1} \mathbf{T}, & n_h > N \end{cases} \quad (11)$$

其次, 本文利用式(10)与式(11)来构造相关熵 ELM-AE. 为了训练能够对输入特征等效编码的权重矩阵, 让 \mathbf{X} 作为目标特征来训练输出权重 $\boldsymbol{\beta}$. 则式(10)不变, 将式(11)变为式(12).

$$\boldsymbol{\beta}^\tau = \begin{cases} \left(\mathbf{H}^T \mathbf{G}^{\tau-1} \mathbf{H} + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{H}^T \mathbf{G}^{\tau-1} \mathbf{X}, & n_h \leq N \\ \mathbf{H}^T \left(\mathbf{G}^{\tau-1} \mathbf{H} \mathbf{H}^T + \frac{\mathbf{I}}{K} \right)^{-1} \mathbf{G}^{\tau-1} \mathbf{X}, & n_h > N \end{cases} \quad (12)$$

最后, 为了增强 2.1 节中 F-ELM 的分类能力与鲁棒性, 将 F-ELM 的网络结构与相关熵的循环训练方法相结合, 利用式(10)–式(12)训练网络中各层权重矩阵.

在训练阶段, 输入特征为 $\mathbf{A} \in \mathbf{R}^{N \times n}$, 与之对应的标签特征为 $\mathbf{T} \in \mathbf{R}^{N \times l}$, 利用 HIK 将特征 \mathbf{A} 映射为特征 \mathbf{B} . 将 \mathbf{B} 代入式(10)与式(11)中替换 \mathbf{H} 来训练权重矩阵 \mathbf{W}_2 , 则 $\mathbf{C} = \mathbf{B}\mathbf{W}_2$. 令 \mathbf{C} 作为输入, 利用相关熵 ELM-AE 来训练编码矩阵 \mathbf{W}_3 , 则

$D = (CW_3) \cdot B$ 。最后，令 D 代替式(10)与式(11)中的 H 来训练最终的分类权重 W_4 。

在测试阶段，输入测试特征 A^{test} ，则 $B^{\text{test}} = k(A^{\text{test}}, A)$, $C^{\text{test}} = B^{\text{test}}W_2$, $D^{\text{test}} = (C^{\text{test}}W_3) \cdot B^{\text{test}}$ ，最终的预测输出为 $T^{\text{test}} = D^{\text{test}}W_4$ 。

3 实验结果与分析

本文从Caltech 101^[9], MSRC^[12]和15 Scenes^[9]数据库的图像中提取出SPM特征、全局轮廓特征、局部位置特征、FV特征和VLAD特征，将这些特征组合后输入CF-ELM，以验证CF-ELM融合特征的能力。为了能够公平比较，所有数据库上的实验均参照文献[12]的实验步骤与参数进行设置。

CF-ELM中各层权重矩阵的训练需要3个参数，即惩罚系数 K 、循环次数与相关熵参数 $d = 2\sigma^2$ 。因此，本文在每个数据集上运行大量实验以确定最适合的参数。本文的实验环境为：Intel Xeon Silver 4110 @2.1 GHz 8核CPU, 64 GB RAM, MATLAB 2014a。

3.1 物体分类

(1) Caltech 101数据库：本节实验使用Caltech 101数据库中101类物体图像，从每一类图像中随机抽取30幅用于训练，再从每一类的剩余图像中随机抽取50幅用于测试，共选出5948幅图像进行实验。如果剩余图像不足50幅，则将剩余部分全部选中。将提取出的SPM特征(8400维)、FV特征(9840维)和VLAD特征(4920维)串联为特征组合1(23160维)，将SPM特征(8400维)、局部位置特征(3030维)和全局轮廓特征(3030维)串联为特征组合2(14460维)。

(2) MSRC数据库：本实验从MSRC和MSRC-21数据库中共选出21类物体图像进行实验，其中每个类别随机选择30幅图像进行训练，在剩下的图像中再选择30幅图像进行测试，共使用1260幅图像进行实验。将提取出的SPM特征(8400维)、FV特征(9840维)和VLAD特征(4920维)串联为特征组合

1(23160维)，将SPM特征(8400维)、局部位置特征(630维)和全局轮廓特征(630维)串联为特征组合2(9660维)。文献[12]在分类前用系数[1, 12, 0.2]对SPM特征、局部位置特征和全局轮廓特征进行加权，本文将加权后的这三种特征串联为特征组合3。

针对Caltech 101与MSRC数据库，需要为CF-ELM确定合适的循环次数、参数 K 与 d 。其中，循环次数的范围为{5, 10, 20, 30, 40, 50}， K 与 d 的取值范围均为 $\{2^{-5}, 2^{-4}, \dots, 2^{14}, 2^{15}\}$ 。将Caltech 101上的组合2与MSRC上的组合3分别作为输入，CF-ELM在各个循环次数能够出现的最高正确率如图2所示。在两个数据库中，当循环次数为10次时均可在耗时较少的情况下获得较高正确率。因此，将CF-ELM中循环次数设置为10次。

在两个数据库上，将不同的特征组合分别输入CF-ELM得到的实验结果如图3和图4所示。在Caltech 101中组合2上，当 $K = 0.5$ 与 $d = 16$ 时可以达到83.65%的最高分类正确率。在MSRC中组合3上，当 $K = 2$ 与 $d = 0.5$ 时可以达到95.76%的最高正确率。

将不同特征组合分别输入SVM, KELM和F-ELM中，并将实验结果与CF-ELM进行比较，以验证本文提出的网络结构与基于相关熵的训练方法的有效性。其中，SVM与KELM的核函数均为

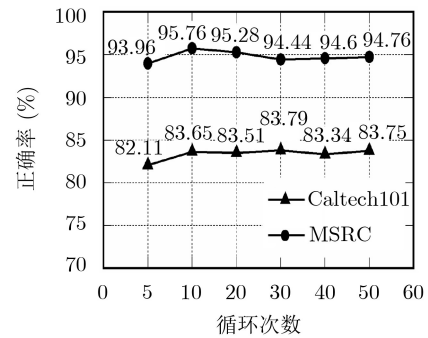


图2 CF-ELM中循环次数对Caltech 101与MSRC数据库上正确率的影响曲线

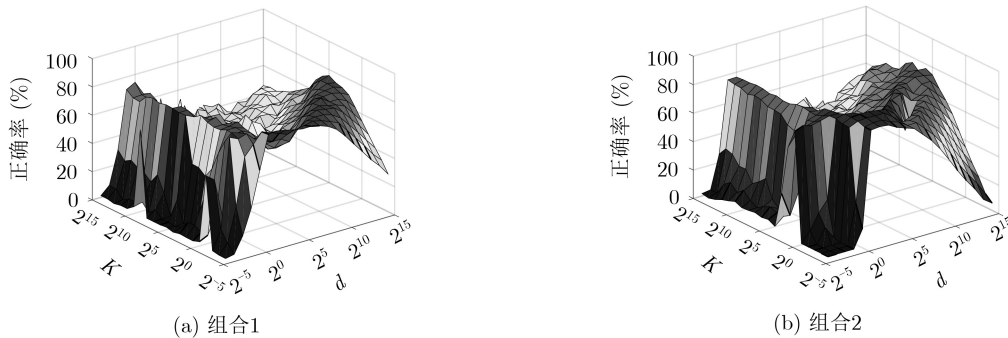


图3 Caltech 101数据库上2种组合的正确率曲面图

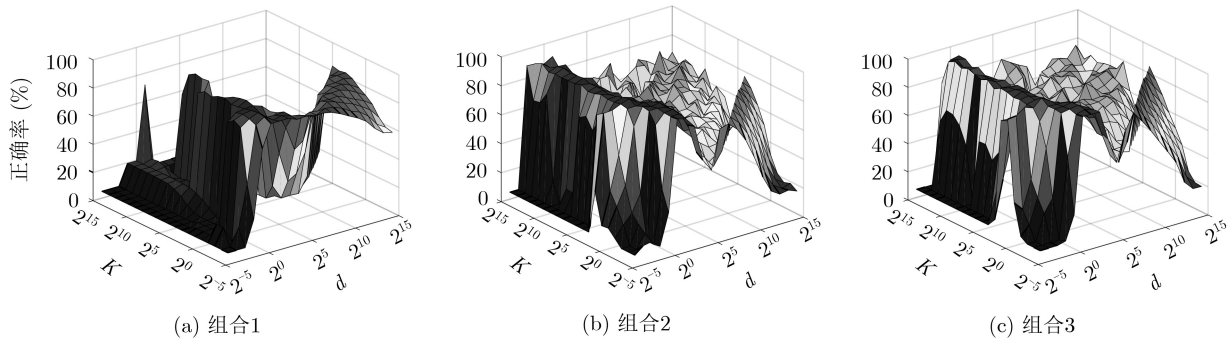


图4 MSRC数据库上3种组合的正确率曲面图

HIK, F-ELM中各层权重的训练方法为传统ELM方法。表1中列出了4种算法对两种数据库上不同特征组合的训练时间与分类正确率

首先, 将KELM与F-ELM进行比较。与KELM相比, F-ELM用微小的计算代价获得了分类正确率的提高。例如在MSRC数据库上, F-ELM对组合3比KELM多消耗了0.17%的训练时间, 提高了1.75%的正确率。在KELM的基础上, F-ELM利用增加的时间进行加权系数的训练。通过对映射特征进行系数加权有效地提升了分类正确率, 由此可以证明F-ELM网络结构的有效性。

其次, 将F-ELM与CF-ELM进行比较。CF-ELM比F-ELM消耗了更多的训练时间, 正确率也相对提高。两个数据库中部分物体图像含有嘈杂的背景, 可以将图中嘈杂的环境视为对物体起到干扰作

用的噪声因素。在Caltech 101中含有嘈杂背景的4种类别上, 利用两种方法对组合2进行分类。在MSRC中4种类别上, 对组合3进行分类。如图5所示, 将F-ELM与CF-ELM对8类图像的分类正确率用柱状图进行表示, 横线填充代表F-ELM, 网线填充代表CF-ELM。比较得出, CF-ELM对含有嘈杂干扰的物体图像具有更高的分类正确率。在F-ELM的基础上, CF-ELM利用增加的时间循环训练各层相关熵权重, 提高了CF-ELM对嘈杂环境的鲁棒性, 使得正确率再次得到提升, 由此可以证明循环训练相关熵权重的有效性。

再次, 将SVM与CF-ELM进行比较。在两个数据库上, CF-ELM对各个特征组合均有提升。综合比较可知, CF-ELM在训练时间少量增加的情况下, 显著提升了物体图像分类正确率, 由此可以证

表1 Caltech 101与MSRC的正确率与训练时间

组合方法	SVM			KELM			F-ELM			CF-ELM		
	组合1	组合2	组合3	组合1	组合2	组合3	组合1	组合2	组合3	组合1	组合2	组合3
Caltech 101 (%)	72.04	77.93	-	79.43	78.84	-	80.31	80.19	-	80.59	83.65	-
训练时间(s)	862.16	540.30	-	860.37	538.56	-	861.76	539.77	-	902.13	569.80	-
MSRC (%)	90.26	88.57	94.13	91.42	90.42	91.74	91.74	90.58	93.49	90.95	92.06	95.76
训练时间(s)	79.59	17.11	17.11	79.51	17.01	17.01	79.54	17.04	17.04	80.68	18.20	18.20

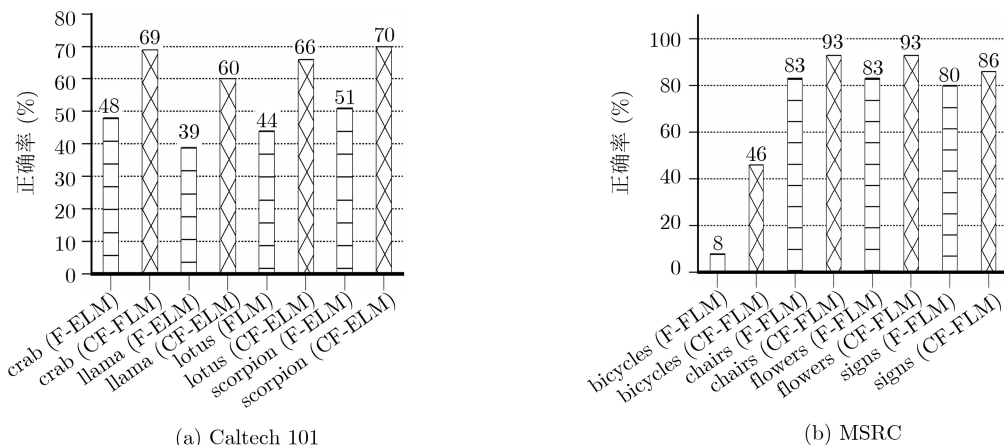


图5 F-ELM与CF-ELM对两种数据库中具有嘈杂背景图像的正确率

明CF-ELM中网络结构与相关熵训练方法相结合的有效性。

在Caltech 101数据库上, 将CF-ELM与组合2相结合, 与表2中其他方法的正确率进行比较。其中, 文献[16]的正确率略高于本文算法。考虑到文献[16]生成了180300维的图像特征, 特征组合2仅为

14460维。两者的特征维数相差12.4倍, 而文献[16]的正确率比本文方法高0.25%。由此可以反映出, 通过充分融合利用特征中的可辨识信息, CF-ELM有效弥补了特征维数上的劣势。在字典维数仅400维的情况下, CF-ELM与组合2即可获得较高物体图像分类正确率。

表2 Caltech 101的结果比较

方法	SPM ^[9]	LLC ^[17]	文献[18]	SDCD+PHOW ^[14]	文献[19]	ScSPM+DVM ^[20]	文献[12]	文献[8]	文献[21]	文献[16]	CF-ELM-组合2
字典维数	200	2048	1000	1024	-	-	400	1024	2048	600	400
正确率(%)	64.60	73.44	74.30	75.37	76.00	77.70	77.93	78.00	79.70	83.90	83.65

3.2 场景分类

本节使用场景图像数据库15 Scene来测试CF-ELM对场景图像的分类能力。使用全部15类场景图像进行实验, 从每一类图像中随机抽取100幅用于训练, 剩余图像用于测试, 总共使用4485幅图像进行实验。将提取出的SPM特征(8400维)、FV特征(9840维)和VLAD特征(4920维)串联为组合1(23160维), 将SPM特征(8400维)、局部位置特征(1500维)和全局轮廓特征(1500维)串联为组合2(11400维)。文献[12]在分类前用系数[1, 0.15, 0.35]对组合2中特征进行加权, 本文将加权后的这3种特征串联为特征组合4。

对于15 Scene数据库, CF-ELM中循环次数的取值范围为{5, 10, 20, 30, 40, 50}, K 的取值范围为 $\{2^{-10}, 2^{-9}, \dots, 2^9, 2^{10}\}$, d 的取值范围是 $\{2^{-5}, 2^{-4}, \dots, 2^{14}, 2^{15}\}$ 。在图6中, 当循环次数为10次可在耗时较短的情况下获得最高正确率。因此, 将循环次数设置为10次。同时, 将组合1、组合2与组合4分别输入CF-ELM, 实验结果如图7所示。在图7c中, 当 $K = 2^{-5}$ 与 $d = 8$ 时可以达到87.76%的最高正确率。

4种方法对15 Scene数据库的分类结果如表3所示。将KELM与F-ELM进行比较, F-ELM对3种组合的分类正确率均有提升, 由此可以证明网络结构

的有效性。将F-ELM与CF-ELM进行比较, CF-ELM在F-ELM的基础上对组合4提高了3.43%的正确率。在含有嘈杂背景的4种类别上, 利用两种方法对组合4进行分类, 并将分类正确率用图8中的柱状图表示。比较得出, CF-ELM对含有嘈杂干扰的场景图像具有更高的分类正确率, 由此可以证明基于相关熵的训练方法的有效性。将SVM与CF-ELM进行比较, CF-ELM对3种组合的分类正确率均有提升。综合比较可知, CF-ELM的网络结构与训练方法相结合有效提高了场景图像的分类正确率。

将CF-ELM与组合4相结合, 与表4中其他方法进行比较。其中, 文献[16]在高维特征优势下获得的正确率比本文方法高2.34%, 而CF-ELM通过自身特性有效弥补了特征维数上的劣势, 减小了正确

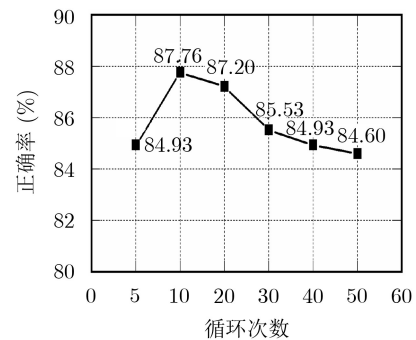


图6 CF-ELM中循环次数对15 Scene数据库上正确率的影响曲线

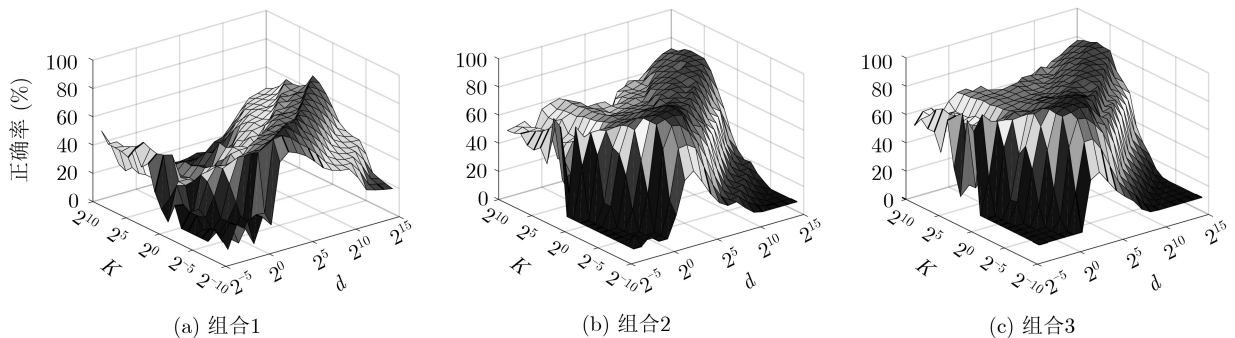


图7 15 Scene数据库上3种组合的正确率曲面图

表3 15 Scene的正确率与训练时间

组合方法	SVM			KELM			F-ELM			CF-ELM		
	组合1	组合2	组合4	组合1	组合2	组合4	组合1	组合2	组合4	组合1	组合2	组合4
15 Scene(%)	74.34	77.92	86.46	72.00	80.73	83.53	77.20	82.06	84.33	79.00	83.06	87.76
训练时间(s)	347.44	106.25	106.25	347.12	106.11	106.11	347.37	106.41	106.41	367.50	120.10	120.10

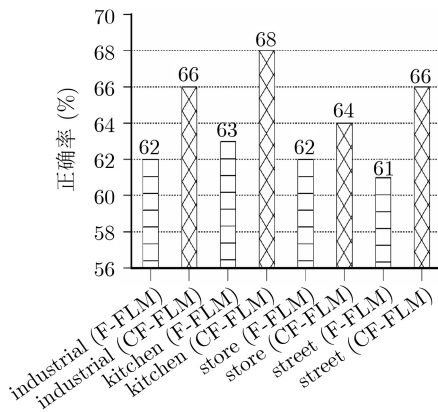


图8 F-ELM与CF-ELM对15 Scene的正确率

率之间的差距。将表4与表2中实验结果相对可知，CF-ELM对场景图像分类正确率的提升效果没有物体图像显著。这是因为不同场景图像之间包含有相似的局部物体，从而对CF-ELM的分类结果造成干扰。但是在字典维数仅有400维的情况下，

CF-ELM与组合4仍然可以获得较高场景图像分类正确率。

4 结束语

本文在ELM网络结构与训练模式的基础上进行改进，提出了相关熵融合极限学习机，并将其用于物体与场景图像分类。通过分析实验结果可以得到3个结论：(1)在网络结构上，通过核映射与加权组合实现了特征的进一步融合；(2)在训练模式上，利用基于相关熵的ELM与ELM-AE对特征融合结构中各层权重矩阵进行训练，进一步提高了网络的分类能力与鲁棒性；(3)相较于场景图像，相关熵融合极限学习机对物体图像具有更加显著的分类能力。因此，下一步将研究如何抵抗不同场景图像中相似局部物体对分类结果造成的干扰，以进一步提高相关熵融合极限学习机对场景图像的分类能力。

表4 15 Scene的结果比较

方法	SPM ^[9]	LLC ^[17]	SLC ^[22]	LSVQ ^[22]	文献[23]	LGF ^[15]	文献[12]	MFS ^[24]	文献[16]	CF-ELM-组合4
字典维数	200	1000	1024	1024	-	1024	400	-	600	400
正确率(%)	81.10	81.73	81.89	83.08	85.70	85.80	86.46	87.10	90.10	87.76

参考文献

- [1] HUANG Guangbin, ZHU Qinyu, and SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1/3): 489-501. doi: 10.1016/j.neucom.2005.12.126.
- [2] HUANG Guangbin, ZHOU Hongming, DING Xiaojian, et al. Extreme learning machine for regression and multiclass classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 2012, 42(2): 513-529. doi: 10.1109/TSMCB.2011.2168604.
- [3] KASUN L L C, ZHOU Hongming, HUANG Guangbin, et al. Representational learning with extreme learning machine for big data[J]. *IEEE Intelligent Systems*, 2013, 28(6): 31-34.
- [4] XING Hongjie and WANG Xinmei. Training extreme learning machine via regularized correntropy criterion[J]. *Neural Computing and Applications*, 2013, 23(7/8): 1977-1986. doi: 10.1007/s00521-012-1184-y.
- [5] CHEN Liangjun, HONEINE P, QU Hua, et al. Correntropy-based robust multilayer extreme learning machines[J]. *Pattern Recognition*, 2018, 84: 357-370. doi: 10.1016/j.patcog.2018.07.011.
- [6] LUO Xiong, SUN Jiankun, WANG Long, et al. Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy[J]. *IEEE Transactions on Industrial Informatics*, 2018, 14(11): 4963-4971. doi: 10.1109/TII.2018.2854549.
- [7] HAN Honggui, WANG Lidan, and QIAO Junfei. Hierarchical extreme learning machine for feedforward neural network[J]. *Neurocomputing*, 2014, 128: 128-135. doi: 10.1016/j.neucom.2013.01.057.
- [8] LI Qing, PENG Qiang, CHEN Junzhou, et al. Improving image classification accuracy with ELM and CSIFT[J]. *Computing in Science & Engineering*, 2019, 21(5): 26-34. doi: 10.1109/MCSE.2018.108164708.
- [9] LAZEBNIK S, SCHMID C, and PONCE J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]. 2006 IEEE Computer Society

- Conference on Computer Vision and Pattern Recognition, New York, USA, 2006: 2169–2178. doi: [10.1109/CVPR.2006.68](https://doi.org/10.1109/CVPR.2006.68).
- [10] JÉGOU H, DOUZE M, SCHMID C, *et al.* Aggregating local descriptors into a compact image representation[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 3304–3311. doi: [10.1109/CVPR.2010.5540039](https://doi.org/10.1109/CVPR.2010.5540039).
- [11] SÁNCHEZ J, PERRONNIN F, MENSINK T, *et al.* Image classification with the fisher vector: Theory and practice[J]. *International Journal of Computer Vision*, 2013, 105(3): 222–245. doi: [10.1007/s11263-013-0636-x](https://doi.org/10.1007/s11263-013-0636-x).
- [12] 李雅倩, 吴超, 李海滨, 等. 局部位置特征与全局轮廓特征相结合的图像分类方法[J]. 电子学报, 2018, 46(7): 1726–1731. doi: [10.3969/j.issn.0372-2112.2018.07.026](https://doi.org/10.3969/j.issn.0372-2112.2018.07.026).
LI Yaqian, WU Chao, LI Haibin, *et al.* Image classification method combining local position feature with global contour feature[J]. *Acta Electronica Sinica*, 2018, 46(7): 1726–1731. doi: [10.3969/j.issn.0372-2112.2018.07.026](https://doi.org/10.3969/j.issn.0372-2112.2018.07.026).
- [13] AHMED K T, IRTAZA A, and IQBAL M A. Fusion of local and global features for effective image extraction[J]. *Applied Intelligence*, 2017, 47(2): 526–543. doi: [10.1007/s10489-017-0916-1](https://doi.org/10.1007/s10489-017-0916-1).
- [14] MANSOURIAN L, ABDULLAH M T, ABDULLAH L N, *et al.* An effective fusion model for image retrieval[J]. *Multimedia Tools and Applications*, 2018, 77(13): 16131–16154. doi: [10.1007/s11042-017-5192-x](https://doi.org/10.1007/s11042-017-5192-x).
- [15] ZOU Jinyi, LI Wei, CHEN Chen, *et al.* Scene classification using local and global features with collaborative representation fusion[J]. *Information Sciences*, 2016, 348: 209–226. doi: [10.1016/j.ins.2016.02.021](https://doi.org/10.1016/j.ins.2016.02.021).
- [16] KONIUSZ P, YAN Fei, GOSSELIN P H, *et al.* Higher-order occurrence pooling for bags-of-words: Visual concept detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(2): 313–326. doi: [10.1109/TPAMI.2016.2545667](https://doi.org/10.1109/TPAMI.2016.2545667).
- [17] WANG Jinjun, YANG Jianchao, YU Kai, *et al.* Locality-constrained linear coding for image classification[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 3360–3367.
- [18] ZHU Qihai, WANG Zhezhen, MAO Xiaojiao, *et al.* Spatial locality-preserving feature coding for image classification[J]. *Applied Intelligence*, 2017, 47(1): 148–157. doi: [10.1007/s10489-016-0887-7](https://doi.org/10.1007/s10489-016-0887-7).
- [19] XIONG Wei, ZHANG Lefei, DU Bo, *et al.* Combining local and global: Rich and robust feature pooling for visual recognition[J]. *Pattern Recognition*, 2017, 62: 225–235. doi: [10.1016/j.patcog.2016.08.006](https://doi.org/10.1016/j.patcog.2016.08.006).
- [20] GUI Jie, LIU Tongliang, TAO Dacheng, *et al.* Representative vector machines: A unified framework for classical classifiers[J]. *IEEE Transactions on Cybernetics*, 2016, 46(8): 1877–1888. doi: [10.1109/TCYB.2015.2457234](https://doi.org/10.1109/TCYB.2015.2457234).
- [21] GOH H, THOME N, CORD M, *et al.* Learning deep hierarchical visual feature coding[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 25(12): 2212–2225. doi: [10.1109/TNNLS.2014.2307532](https://doi.org/10.1109/TNNLS.2014.2307532).
- [22] 肖文华, 包卫东, 陈立栋, 等. 一种用于图像分类的语义增强线性编码方法[J]. 电子与信息学报, 2015, 37(4): 791–797. doi: [10.11999/JEIT140743](https://doi.org/10.11999/JEIT140743).
XIAO Wenhua, BAO Weidong, CHEN Lidong, *et al.* A semantic enhanced linear coding for image classification[J]. *Journal of Electronics & Information Technology*, 2015, 37(4): 791–797. doi: [10.11999/JEIT140743](https://doi.org/10.11999/JEIT140743).
- [23] LI Lijia, SU Hao, LIM Y, *et al.* Object bank: An object-level image representation for high-level visual recognition[J]. *International Journal of Computer Vision*, 2014, 107(1): 20–39. doi: [10.1007/s11263-013-0660-x](https://doi.org/10.1007/s11263-013-0660-x).
- [24] SONG Xinhang, JIANG Shuqiang, and HERRANZ L. Multi-scale multi-feature context modeling for scene recognition in the semantic manifold[J]. *IEEE Transactions on Image Processing*, 2017, 26(6): 2721–2735. doi: [10.1109/TIP.2017.2686017](https://doi.org/10.1109/TIP.2017.2686017).
- 吴超: 男, 1990年生, 博士生, 研究方向为计算机视觉.
李雅倩: 女, 1982年生, 副教授, 研究方向为计算机视觉.
张亚茹: 女, 1995年生, 博士生, 研究方向为计算机视觉.
刘彬: 男, 1953年生, 教授, 研究方向为计算机视觉.