

基于类权重和最小化预测熵的测试时集成方法

宋辉,张轶哲*,张功萱,孟元

(南京理工大学计算机科学与工程学院,江苏南京210094)

摘要:针对传统集成学习方法忽略不同样本需使用不同模型权重的问题,提出一种基于类权重和最小化预测熵(class and entropy weights, CEW)的测试时集成方法。类权重为模型预测结果与验证集上各类概率对错分布的相似度,利用欧氏距离计算相识度;在最小化熵过程中,线性组合模型预测经过类权重模块加权后的输出,寻找最小预测熵对应的线性组合作为熵权重,提高集成模型预测能力。试验结果表明:在4个公开医学图像数据集上,CEW方法与最优单一模型相比,平均召回率提高0.23%~2.81%,准确率提高0.5%~2.54%;与DS方法相比,CEW方法平均召回率最多提高1.25%,准确率最多提高1.1%。基于CEW的测试时集成方法能够在测试时(无标签情况下)动态调整模型权重,比同类方法的预测精度更高。

关键词:测试时集成方法;医学图像分类;类权重;最小化熵;深度学习

中图分类号:TP391 **文献标志码:**A

引用格式:宋辉,张轶哲,张功萱,等. 基于类权重和最小化预测熵的测试时集成方法[J].山东大学学报(工学版),2024,54(3):36-43.

SONG Hui, ZHANG Yizhe, ZHANG Gongxuan, et al. A test-time ensemble method based on class weights and prediction entropy minimization[J]. Journal of Shandong University (Engineering Science), 2024, 54(3):36-43.

A test-time ensemble method based on class weights and prediction entropy minimization

SONG Hui, ZHANG Yizhe*, ZHANG Gongxuan, MENG Yuan

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China)

Abstract: To address the issue of traditional ensemble learning methods overlooking the necessity for different model weights for varied samples, a test-time ensemble approach based on class and entropy weights (CEW) was proposed. Class weights were determined by the similarity between the model's predictive results and the distribution of correct and incorrect probabilities for each class on the validation set, calculated using Euclidean distance. During the entropy minimization process, the output from the linear combination of model predictions was weighted by the class weight module. The linear combination corresponding to the minimum predictive entropy was identified as the entropy weight, enhancing the predictive capability of the ensemble model. Experimental results showed that on four public medical image datasets, compared to the optimal single model, the CEW method improved the average recall rate by 0.23% to 2.81%, and accuracy by 0.5% to 2.54%. Compared to the DS method, the CEW method improved the average recall rate by up to 1.25% and accuracy by up to 1.1%. The test-time ensemble method based on CEW proved capable of dynamically adjusting model weights during testing (in an unlabeled situation), achieving higher prediction accuracy than similar methods.

Keywords: test-time ensemble method; medical image classification; class weight; entropy minimization; deep learning

0 引言

集成学习是机器学习领域中一种用来提升模

型性能的技术,通过融合多个模型的预测,得到一个更为精准且稳定的结果^[1-2]。在实际应用中,集成学习已经取得了广泛的成功,例如,在Kaggle数据科学竞赛中,集成学习方法常常是获胜方案关

收稿日期:2023-07-03

基金项目:国家自然科学基金资助项目(62272232,62201263)

第一作者简介:宋辉(1998—),男,江苏盐城人,硕士研究生,主要研究方向为集成学习. E-mail:1677279299@qq.com

*通信作者简介:张轶哲(1987—),男,江苏南京人,副教授,硕士生导师,博士,主要研究方向为医学图像分析,机器学习和算法设计。

E-mail:zhangyizhe@njst.edu.cn

键。随着深度学习技术发展,在计算机视觉领域,集成学习已用于解决诸如图像分类、目标检测和语义分割等任务^[3-4]。

集成学习在许多领域都取得了成功,实际应用中,集成学习的方法仍然面临一些挑战。经典的集成学习采用预测平均方法,对基模型预测结果的和进行平均^[5-7]。与单一模型相比,预测平均可以提升模型性能并降低过拟合风险^[8],等比例权重分配可能导致忽视弱模型对集成模型性能的影响,导致集成模型性能下降。

为克服这个问题,近年来许多研究人员提出不同集成方法来调整强弱模型间的权重比例。文献[9]提出一种基于高斯函数的方法来估计集合中每个模型的权重,通过累加一组样本预测概率与预测平均概率的高斯距离为每个模型分配权重;文献[10]提出基于模型在验证集中基模型准确性的加权方法,该方法会给验证集上准确率高的模型分配更高的权重。这两种方法都着力降低弱模型的影响,成功提高集成模型性能。这种集成方法仍存在一些不足之处,尤其在面对新样本时,弱模型也可能预测正确一些强模型预测错误的样本,如果模型权重保持不变,会导致集成学习模型的性能一直偏向于某个强模型的性能,而忽略不同样本需要不同模型权重的需求。动态选择权重是解决这一问题的有效方法,这是一种根据每个新样本动态为每个分类器分配权重的技术。文献[11]提出一个在线动态选择系数(dynamic weights, DS)方案,估计模型预测的狄利克雷分布,应用马氏距离来确定每个样本权重。DS方法计算过于复杂且权重更新时间较长,在一些时效性强的情况下不予考虑。如何基于样本信息快速地动态调节基分类器间的权重,以提高集成模型的预测能力,是目前集成学习的主要挑战,针对上文所提出的局限性,本研究提出一种基于类权重和最小化预测熵的测试时集成方法。

1 理论基础

信息熵用于衡量信息的不确定性和随机性^[12-13],熵值越高,随机变量包含的信息越多,不确定性和随机性也越高^[14]。

信息熵与错误率密切相关,文献[15]指出了熵与错误率有一定的正相关性,如图1所示,它是模型在CIFAR-100-C上结果。从图1中可以看出,当错误率较低时,对应的熵值也相对较低。降低模型输

出的熵值可能会让模型预测概率分布更加集中、更加确定,即减小模型预测的不确定性,使得模型更加有信心地进行正确预测。基于这一理论基础,本研究旨在降低集成模型的预测熵,提高集成模型的预测准确性。

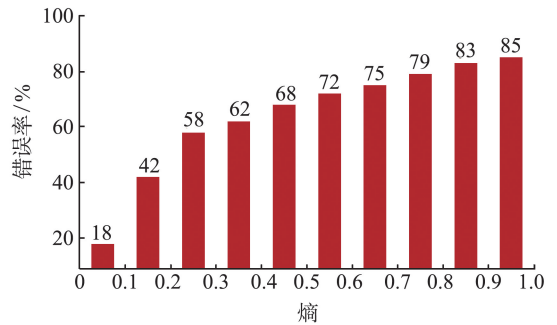


图1 熵和错误率的关系

Fig.1 The relationship between entropy and error rate

2 本研究方法

本研究将分别介绍保存概率集合模块、类权重模块和最小化熵权重模块,CEW方法的总体流程如图2所示。

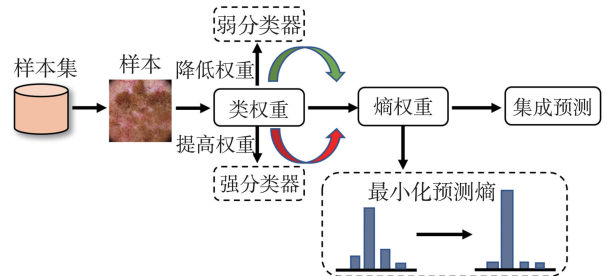


图2 CEW方法的总体流程

Fig.2 The overall flow of the CEW method

在测试时,新样本经过一个类权重模块,每个模型会得到一个经过类权重加权修正的输出,这个模块针对每个新样本信息来动态调整弱分类器的权重比例,提高强分类器对于集成模型的影响。在熵权重值域范围内,动态组合类权重模块修正后的预测结果,以最小化集成模型输出的熵。

2.1 保存概率集合

在基模型训练结束后,得到一组深度学习模型的集合 $G = \{f_1, f_2, \dots, f_k, \dots, f_g\}$ ($k \in \{1, 2, \dots, g\}$), g 为基模型的数量。

对验证集上所有标签为 l 的样本,保存模型 f_k 预测正确时的概率集合 P_l^h 和预测错误时概率的集合 P_l^m 。如图3所示,当所有概率保存完毕后,分别

对这两个概率集合取平均值,记为概率向量 μ_l^h 和概率向量 μ_l^m :

$$\mu_l^h = \frac{\sum_{j=1}^{|\mathbf{P}_l^h|} \mathbf{P}_l^h(j)}{|\mathbf{P}_l^h|}, \quad (1)$$

式中, $\mathbf{P}_l^h(j)$ 为 \mathbf{P}_l^h 集合中的一个概率向量, $|\mathbf{P}_l^h|$ 为 \mathbf{P}_l^h 集合中元素的数量, L 为标签的数量。

$$\mu_l^m = \frac{\sum_{j=1}^{|\mathbf{P}_l^m|} \mathbf{P}_l^m(j)}{|\mathbf{P}_l^m|}. \quad (2)$$

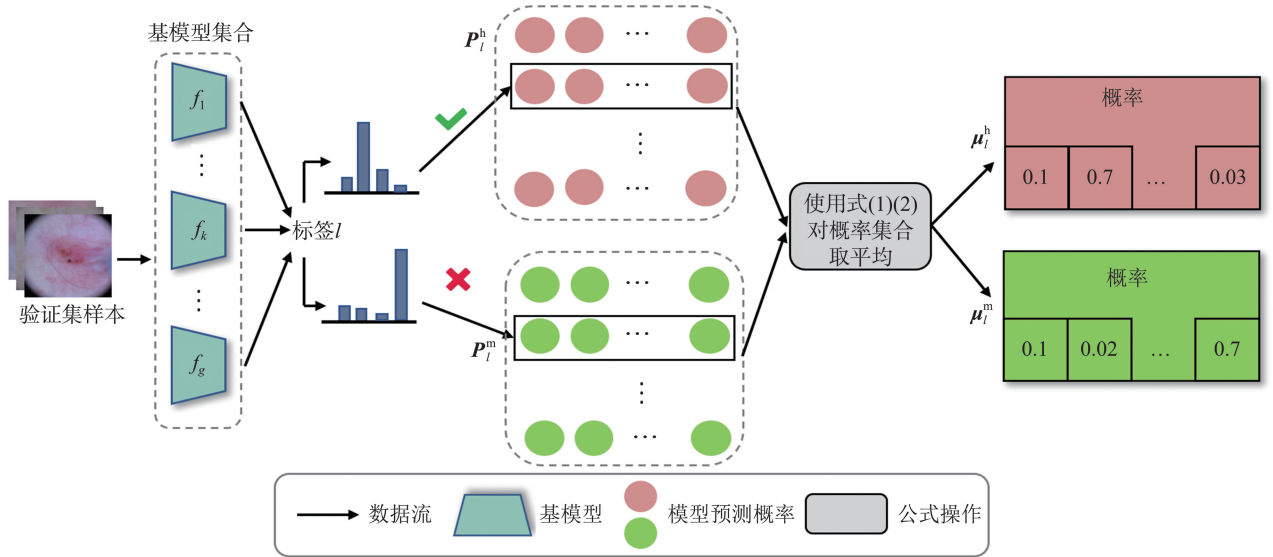


图3 保存验证集概率分布流程

Fig.3 Process of preserving the probability distribution of the validation set

2.2 计算类权重

2.1节中保存模型在验证集上的概率分布特征,认为模型在验证集上预测概率分布是可靠的,当模型在对目标域样本进行预测时,就会参考模型在验证集上预测的特征。如图4所示,对于新样本,若模型 f_k 的预测结果为绿色小点 p_k ,并认为当前样本的标签为2,此时样本有3个标签, \mathbf{P}_2^h 为红色点集, \mathbf{P}_2^m 为橘黄色点集。 p_g 为预测正确时的概率,应当接近 \mathbf{P}_2^h 分布,远离 \mathbf{P}_2^m 分布,这种预测结果对集成模型预测结果正确性有着更大作用,应当赋予模

型更大权重系数。本节提出类权重(class weights, CW)的概念,它表示两个概率向量中每个类的概率相似度。这里以分布均值点值点和当前预测概率的相似度来计算类权重,欧式距离可以很好表示概率向量之间的相似度:

$$d(x, y) = \left[\sum_{i=1}^L (x_i - y_i)^T (x_i - y_i) \right]^{\frac{1}{2}}, \quad i \in \{1, \dots, L\}, \quad (3)$$

式中 x 和 y 都为 L 维向量。

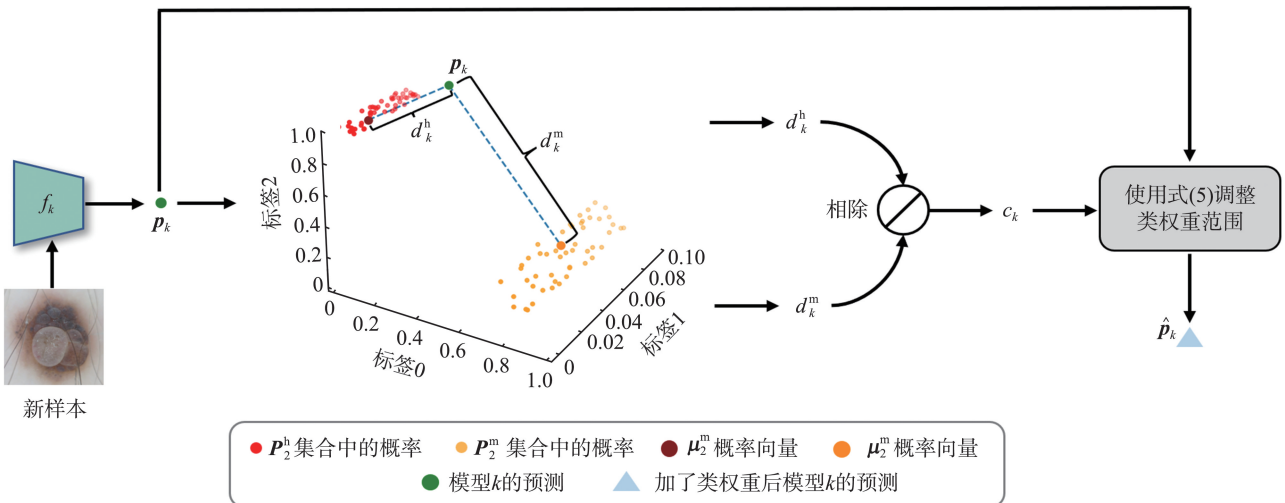


图4 计算类权重流程

Fig.4 Process of calculating class weights

若 \mathbf{p}_k 认为当前样本的标签为 l , 记 $c_k = d_k^m / d_k^h$ 为模型 f_k 的类权重, 即:

$$c_k = \frac{d_k^m}{d_k^h} = \frac{d(\mathbf{p}_k, \boldsymbol{\mu}_l^m)}{d(\mathbf{p}_k, \boldsymbol{\mu}_l^h)}, \quad (4)$$

式中, d_k^h 为 \mathbf{p}_k 和 $\boldsymbol{\mu}_l^h$ 的欧式距离, d_k^m 为 \mathbf{p}_k 和 $\boldsymbol{\mu}_l^m$ 的欧式距离。

类权重可以根据模型对于不同样本预测特性, 动态赋予模型权重, 提高强弱分类器之间权重比例。弱分类器也可能学习到强分类器未能捕捉的特征, 弱分类器的权重不应太小。应当对类权重的值域进行调整, 让强弱分类器的权重比例保持在合理的区间, 调整后的预测结果

$$\hat{\mathbf{p}}_k = \mathbf{p}_k \cdot \left(1 + \frac{c_k}{\sum_{j=1}^g c_j}\right), \quad (5)$$

式中 $g = |G|$ 为模型的集合 G 中基模型的数量。

2.3 最小化预测熵

本研究引入熵权重 (entropy weights, EW) 模

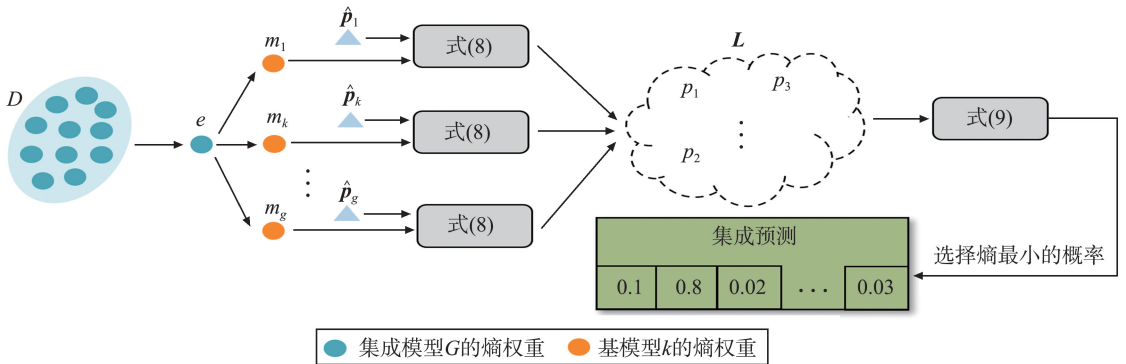


图5 最小化预测熵

Fig.5 Minimize the prediction entropy

集成模型的预测结果为集合 L 中熵最小的概率, 最小化熵模型为:

$$\min_{\mathbf{p} \in L} H(\mathbf{p}) = \min_{\mathbf{p} \in L} \left(-\sum_{i=1}^L \mathbf{p}(i) \log_2(\mathbf{p}(i))\right), \quad (9)$$

式中 $\mathbf{p}(i)$ 为概率向量 \mathbf{p} 中标签 i 的概率。

3 样本更新优化算法

在更新模型权重的过程中, 主要的耗时环节为从 L 集合内选择预测熵最小的概率 \mathbf{p} 。对类权重, 只需计算每个基模型的 c_k , 时间复杂度为 $O(|G|)$ 。而熵权重的更新, 须从包含 $r^{|G|}$ (这里 r 为 50) 个概率元素的集合中, 选择熵最小的、所对应的权重, 其时间复杂度为 $O(r^{|G|})$, 远超 $O(|G|)$ 。在时间和复杂度至关重要场景下, 模型面对大量的新样本, 那

么串行权重更新方式就无法做出实时处理。需要对权重计算方式进行改进, 让多个样本权重进行并行更新。

$$D = \left\{ (m_1, m_2, \dots, m_k, \dots, m_g) \mid m_k = \frac{e^{t_k}}{\sum_{j=1}^g e^{t_j}} \right\}, \quad (6)$$

式中, t_k 为向量集合 T 中某个向量的元素, t_k 取值是 0~1 之间的小数:

$$T = \left\{ (t_1, t_2, \dots, t_k, \dots, t_g) \mid t_k = 0.02 * i, \right. \\ \left. k \in \{1, 2, \dots, g\}, \right. \\ \left. i \in \mathbf{Z}, 1 \leq i \leq 50 \right\}. \quad (7)$$

基模型 f_k 的熵权重为 m_k , 集成模型对于单个样本的预测结果为所有基模型的熵权重 m_k 与预测概率 $\hat{\mathbf{p}}_k$ 乘积的和, 得到一组预测概率 \mathbf{p} 的集合 L ,

$$\mathbf{p} = \sum_{k=1}^g m_k \cdot \hat{\mathbf{p}}_k, \quad (8)$$

式中 m_k 为熵权重 e 的第 k 列元素。

基于模型对于每个测试样本 m_k 范围和数量是固定的, 样本间的系数更新具有独立性。本研究提出样本更新优化算法以空间换时间为出发点, 将所有基模型的预测拼接成矩阵, 利用 GPU 对矩阵加速运算。

图 6 展示了并行更新单个样本流权重的设计方案。长方体为模型对于样本流的预测结果, 高度为样本数量 B , 长度为标签数量 L 。按照 2.2 节步骤, 计算模型对于每个样本的类权重, 对原预测结果进行矫正。将所有模型的预测张量沿着第三个维度进行连接, 记为 $T, T \in \mathbf{R}^{B \times L \times g}$ 。熵权重矩阵为 E, E 是固定的, 在算法执行流程中不会改变:

图 6 展示了并行更新单个样本流权重的设计方案。长方体为模型对于样本流的预测结果, 高度为样本数量 B , 长度为标签数量 L 。按照 2.2 节步骤, 计算模型对于每个样本的类权重, 对原预测结果进行矫正。将所有模型的预测张量沿着第三个维度进行连接, 记为 $T, T \in \mathbf{R}^{B \times L \times g}$ 。熵权重矩阵为 E, E 是固定的, 在算法执行流程中不会改变:

$$E = [e_1 \ e_2 \ \cdots \ e_r]_{g \times r \times g} = \begin{bmatrix} m_1^1 & m_2^1 & \cdots & m_{r \times g}^1 \\ \vdots & \vdots & \ddots & \vdots \\ m_1^g & m_2^g & \cdots & m_{r \times g}^g \end{bmatrix}_{g \times r \times g}。$$

对 T 和 E 进行矩阵乘法,得到的矩阵重新赋值给 T ,这样 D 中的每个熵权重元素都可以作用于基

模型的预测上。为了计算所有预测的熵,还需转换矩阵 T 的第二个维度和第三个维度。通过 argmax 函数确定最小预测熵对应的权重索引,还原出每个样本的熵权重。下面给出并行优化权重计算的算法伪代码, P 为整个样本流的预测结果。

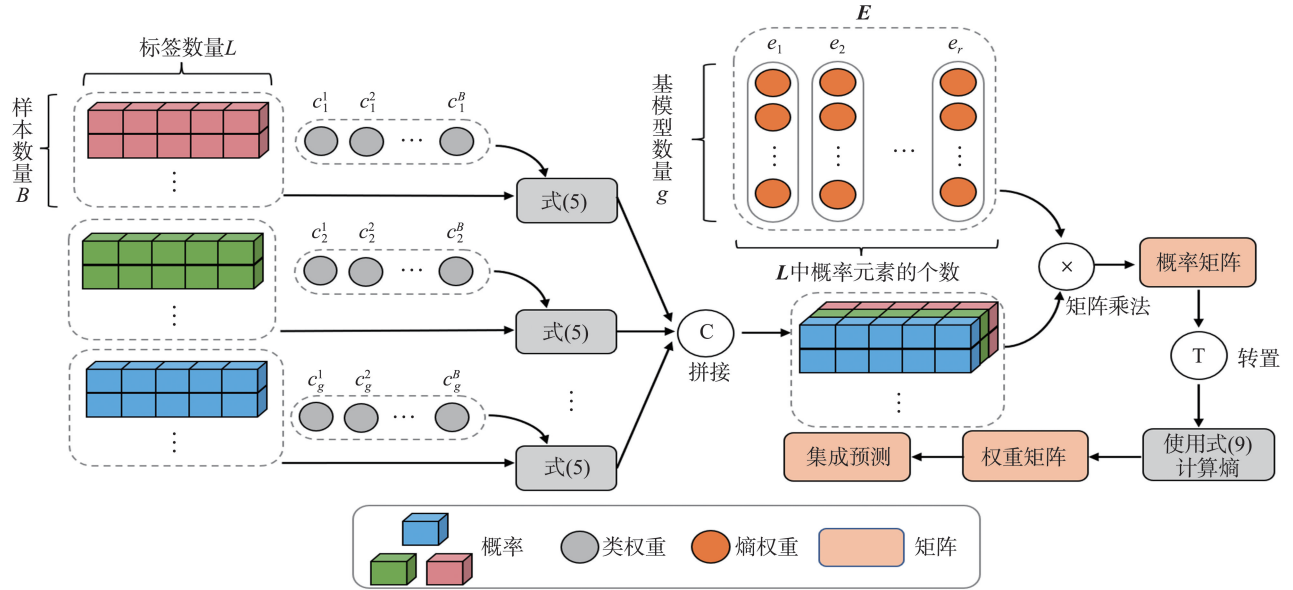


图6 并行更新单个样本流权重的流程

Fig.6 Process of updating individual sample stream weights in parallel

算法1 并行更新算法

输入 数据流 $S_i \in \{S_1, S_2, \dots, S_n\}$, 集成模型集合 $G = \{f_1, f_2, \dots, f_k, \dots, f_g\}$, 熵权重矩阵 E ;

输出 n 批样本流的预测矩阵 P ;

初始化 样本批次大小 $B=64$, $P = \{\}$;

(1) For $S_i \in \{S_1, \dots, S_n\}$ do

(2) 计算每个基模型 f_k 对于样本流 S_i 的概率矩阵: $O_k = f_k(S_i)$, $O_k \in \mathbf{R}^{B \times L}$

(3) For $b \in \{1, 2, \dots, B\}$ do

(4) For $f_k \in G$ do

(5) 计算类权重 c_k^b : 使用式(3)和(4)

(6) For $f_k \in G$ do

(7) 使用式(5)计算加了类权重后模型 k 对样本 b 的概率 O_k^b

(8) 从第三个维度对所有模型的概率矩阵 O_k 进行拼接, 记为 T

(9) 对 T 和 E 进行矩阵乘法, 结果重新赋值给 T , $T \in \mathbf{R}^{B \times L \times r \times g}$

(10) 对矩阵 T 的第二和第三维度进行转置, 此时 $T \in \mathbf{R}^{B \times r \times s \times L}$

(11) 使用式(9)从第三维度计算 T 的熵矩阵

(12) 先用 argmax 函数获得最小熵的索引,

然后再还原样本流 S_i 的熵权重向量 w_i

(13) 线性组合 w_i 和所有模型预测得到样本流 S_i 的集成预测, 然后将预测加入到 P 中

(14) 返回 P

4 试验及分析

在本章中共进行3组试验, 试验一展示每个基模型在测试集上的效果; 试验二将 CEW 与其他4个集成方法进行比较; 试验三验证 CW 和 EW 两个模块的有效性。分析了 CEW 算法的时间复杂度和局限性。

4.1 数据集

为验证 CEW 模型的有效性, 选取4个医学图像分类任务进行试验。ISIC 2019 是一个用于皮肤病分类和分割的数据集, 包含25 331张皮肤病图像。这些图像来自世界各地的皮肤病医院和诊所, 包括恶性黑色素瘤、非恶性黑色素瘤和良性皮肤病等; DermaMNIST 是一个基于真实皮肤病患者图像创建的数据集, 包含7个不同的皮肤病类型, 共计10 015张图像; OCTMNIST 是一个用于视网膜光学相干断层扫描图像分类的数据集, 包含了109 309张 OCT 图像, 共计涉及4种类别; OrganSMNIST 是

一个用于医学图像分类的数据集,包含8个不同的器官类型,共计25 221张CT扫描图像。

4.2 试验环境及设置

集成模型集合由 ResNet 18^[16-18]、MobileNet_v2^[19]、ShuffleNet_v2_x1_0^[20]组成,并使用两块 TITAN 24G 训练模型,CPU 为 AMD Ryzen 75 800 H,内存大小为 16 GB。在训练阶段,采用了 Adam 优化器,设置初始学习率为 0.001,衰减系数为 0.1。每个样本的批次大小为 64,训练时的最大迭代次数为 200 次。为避免过拟合,设置了早停方法。训练集、验证集和测试集的数量比例为 3:1:1。为衡量性能,本研究使用的图像分类标准为多分类准确率 A_c (accuracy) 和平均召回率 A_r (average recall),其中平均召回率也称平衡准确率 (balanced accuracy, BACC):

$$A_c = \frac{\sum_{i=1}^c T_i}{N},$$

$$A_r = \frac{1}{C} \sum_{i=1}^c \frac{T_i}{T_i + F_i},$$

式中: T_i 为标签 i 中被正确分类的样本数, F_i 为标签 i 中被错误分类的样本数, N 为测试集的总样本数。

将本研究方法与常见的集成方法进行了比较:预测平均(predictions average, PAvg)^[10]、最大预测(maximum prediction, PMax)^[10]、预测乘积(product of predictions, PProd)^[10]和动态权重(dy-

namic weights, DS)^[11]。PAvg 方法计算基模型预测的平均值作为集成结果;PMax 方法选择各个基模型中预测值最大的结果作为集成输出。PProd 方法将基模型的预测值对于类别概率相乘作为集成输出;DS 方法通过估计模型预测的狄利克雷分布,并应用马氏距离确定每个模型的权重。

4.3 试验结果及分析

4.3.1 基模型性能表现

表 1 展示的是基模型在不同数据集上的性能表现。基模型在 OCTMNIST 和 OrganSMNIST 数据集上的准确率和平均召回率均在 71.77% 以上,且对于同一个数据集,两个指标之间的数值相近,这表明基模型对于每个类的识别能力也相近。在 DermaMNIST 和 ISIC 2019 数据集上,基模型的准确率和平均召回率数值差距较大,其中 ResNet 18 在 DermaMNIST 数据集上,两指标数值相差 28.77%,发生这种情况的主要原因有两个:(1) 数据集不平衡,在 DermaMNIST 和 ISIC 2019 数据集上,样本数量最多类别和最少类别的比超过 40:1,这可能导致模型过度关注于数量较多的类别,忽略了数量较少的类别,少数类别的预测效果较差。准确率会因多数类预测正确的更多而提升,平均召回率会因多数类和少数类的召回率差别大而降低。(2) 图像特征不足,表现为 DermaMNIST 和 ISIC 2019 数据集上个别病变特征之间的强烈相似性。例如,在 ISIC 2019 数据集上,模型很难区分黑素瘤和黑痣。

表 1 基模型的性能表现
Table 1 Performance of the base model

Method	DermaMNIST		OCTMNIST		ISIC 2019		OrganSMNIST	
	A_r	A_c	A_r	A_c	A_r	A_c	A_r	A_c
ResNet18	48.24	77.01	76.1	76.1	71.94	82.69	74.50	78.84
MobileNet_v2	55.44	76.86	75.6	75.6	73.72	83.42	73.61	78.53
ShuffleNet_v2_x1_0	56.07	76.76	72.5	72.5	70.70	81.05	71.77	76.18

注:黑体数字为最佳性能指标值。

4.3.2 CEW 与其他方法对比

DermaMNIST:所有集成方法的效果在最优基模型上均有提升,如表 2 所示,其中本研究方法 CEW 提升幅度最大,平均召回率为 58.88%,准确率为 79.55%。相比于最优基模型,CEW 方法的平均召回率提高 2.81%,准确率提高 2.54%。相比于 DS 方法,CEW 方法的平均召回率提高 1.25%,准确率提高 0.25%。

OCTMNIST:CEW 的准确率和平均召回率均为 76.6%,相比于其他集成方法,平均召回率提高 0.3%~1%,准确率提高 0.3%~1%。

ISIC-2019:在准确率上,各个集成方法表现均

相近,而 CEW 的平均召回率最高,为 75.56%。所有集成方法中最低平均召回率和最低准确率分别为 74.51% 和 84.70%,均高于最优基模型性能。

OrganSMNIST:在这个数据集上,CEW 展现出不错的性能,与次优集成方法相比,平均召回率为 74.73%,提高 0.78%;准确率为 79.42%,提高 1.1%。

4.3.3 消融试验

为进一步探究 CEW 方法中 CW 和 EW 方法对集成模型的提升效果,本节设计了多组特征消融试验,其中 Baseline 为所有模型预测概率的平均,试验结果如表 3 所示。

表2 CEW 与其他集成方法的性能比较
Table 2 The comparison between CEW and other ensemble method

Method	DermaMNIST		OCTMNIST		ISIC 2019		OrganSMNIST	
	A_r	A_c	A_r	A_c	A_r	A_c	A_r	A_c
PAvg	57.55	79.15	76.30	76.30	74.51	84.84	72.50	78.09
PMax	57.09	78.95	75.90	75.90	74.92	84.70	72.42	78.03
PProd	57.15	79.15	75.60	75.60	74.51	84.84	72.42	78.03
DS	57.63	79.30	75.60	75.60	75.52	84.76	73.95	78.32
CEW	58.88	79.55	76.60	76.60	75.56	84.76	74.73	79.42

注:黑体数字为最佳性能指标值。

表3 消融试验
Table 3 Ablation experiments

Method	DermaMNIST		OCTMNIST		ISIC 2019		OrganSMNIST	
	A_r	A_c	A_r	A_c	A_r	A_c	A_r	A_c
BaseLine	57.55	79.35	75.60	75.60	74.99	84.70	74.35	79.13
BaseLine+CW	58.03	79.40	76.00	76.00	74.99	84.68	74.52	79.29
BaseLine+EW	58.68	79.50	76.40	76.40	75.20	84.70	74.57	79.34
BaseLine+CW+EW	58.88	79.55	76.60	76.60	75.56	84.76	74.73	79.42

注:黑体数字为最佳性能指标值;+代表模块融合。

在 DermaMNIST 数据集和 OCTMNIST 数据集验证了 CW 方法的有效性,与基线方法相比较,CW 方法性能有明显提高,准确率最高提高 0.4%,平均召回率最高提高 0.48%,在 ISIC 2019 数据集和 OrganSMNIST 上提高幅度较小,在 ISIC 2019 数据集上,CW 方法的准确率降低 0.02%,说明 CW 模块在一定程度上提升集成模型的分类效率。

从表 3 中可以看出,EW 方法在 4 个数据集上的准确率和平均召回率均高于基线方法,且提升幅度较大,平均召回率最高提高 1.13%,准确率最高提高 0.8%。对比 CW 方法和 EW 方法,EW 模块比 CW 模块提高更明显,在所有数据集上提升的幅度均高于 CW 模块。

虽然 CW 对于基模型的提升能力有限,但这并不意味着 CW 模块和 EW 模块间存在协调作用,集成两个模块的 CEW 方法在 4 个数据集上均表现最高性能,与 CW 方法相比较,平均召回率高出 0.16%~0.36%,其中 CEW 在 OCTMNIST 数据集上的准确率比 CW 方法高出 0.2%。

4.3.4 更新时间分析

图 7 给出 CEW 和同类方法 DS 在 DermaMNIST 测试集上更新单个样本的平均时间,样本数量为 2 005。图 7 总共给出 5 组批次大小的对比试验结果,从图 7 中可以看出 DS 方法更新单个样本权重的时间为 11.23~13.70 ms,CEW 方法更新权重的时间不到 DS 方法的十分之一,为 0.84~1.11 ms。

CEW 方法无论是计算复杂度,还是时间复杂度都优于 DS 方法。

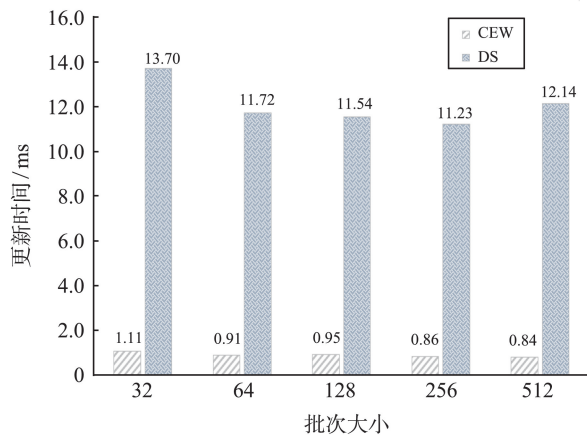


图7 CEW 与 DS 更新时间比较

Fig.7 Comparison of update times between CEW and DS

4.3.5 局限性

本研究的结果具有一定局限性。相较于基线集成方法,采用样本更新算法会增加计算资源和显存的消耗,在一些性能较低的边缘设备上可能难以发挥性能优势。该集成方法适用于少量模型的集成,若存在大量的基模型需要集成,则本方法在熵权重模块会消耗大量显存资源,这一问题可以通过调节熵权重的值域得到缓解。以基模型 ResNet 18、MobileNet_v2 和 ShuffleNet_v2_x1_0 为例,输入一张 $3 \times 224 \times 224$ Px 的图片,CEW 方法需要的计算量为 4.58 GFlops,参数量为 14.67 兆。

本研究未在实际场景的边缘设备中进行试验,

可能存在某些未知的干扰变量,需要在未来的研究中进一步探讨。

5 结论

本研究以提高集成模型性能为目标,解决样本权重的动态更新问题。提出 CEW 集成方法,一方面根据类权重来调节强弱分类器间的权重比值;另一方面在模型增加了类权重的基础上,使用熵权重动态组合模型预测;试验结果表明,本研究方法在医学图像分类问题上取得较好的表现,在削弱弱分类器对集成模型负面影响的同时,又降低集成模型预测的不确定性和错误率,进一步提升集成模型预测的性能。该方法还引入样本系数并行更新设计方案,这在时间和复杂度至关重要的场景中具有重要的应用价值。下一步将研究应用于医疗模型中,辅助医生分析病情。

参考文献:

- [1] DONG X, YU Z, CAO W, et al. A survey on ensemble learning[J]. *Frontiers of Computer Science*, 2020, 14: 241-258.
- [2] YANG Y, LV H, CHEN N. A survey on ensemble learning under the era of deep learning[J]. *Artificial Intelligence Review*, 2023, 56(6): 5545-5589.
- [3] XIAO Y, WU J, LIN Z, et al. A deep learning-based multi-model ensemble method for cancer prediction[J]. *Computer Methods and Programs in Biomedicine*, 2018, 153: 1-9.
- [4] SHARIFANI K, AMINI M. Machine learning and deep learning: a review of methods and applications[J]. *World Information Technology and Engineering Journal*, 2023, 10(7): 3897-3904.
- [5] XUE D, ZHOU X, LI C, et al. An application of transfer learning and ensemble learning techniques for cervical histopathology image classification[J]. *IEEE Access*, 2020, 8: 104603-104618.
- [6] MIENYE I D, SUN Y. A survey of ensemble learning: concepts, algorithms, applications, and prospects [J]. *IEEE Access*, 2022, 10: 99129-99149.
- [7] ZHANG W, LI H, HAN L, et al. Slope stability prediction using ensemble learning techniques: a case study in Yunyang County, Chongqing, China[J]. *Journal of Rock Mechanics and Geotechnical Engineering*, 2022, 14(4): 1089-1099.
- [8] QUMMAR S, KHAN F G, SHAH S, et al. A deep learning ensemble approach for diabetic retinopathy detection [J]. *IEEE Access*, 2019, 7: 150530-150539.
- [9] KRAWCZYK B, WOZNIAK M. Untrained weighted classifier combination with embedded ensemble pruning [J]. *Neurocomputing*, 2016, 196: 14-22.
- [10] HARANGI B. Skin lesion classification with ensembles of deep convolutional neural networks[J]. *Journal of Biomedical Informatics*, 2018, 86: 25-32.
- [11] PACHECO A G C, TRAPPENBERG T, KROHLING R A. Learning dynamic weights for an ensemble of deep models applied to medical imaging classification[C]//2020 International Joint Conference on Neural Networks (IJCNN). Glasgow, UK: IEEE, 2020: 1-8.
- [12] GU R. Multiscale Shannon entropy and its application in the stock market [J]. *Physica A: Statistical Mechanics and its Applications*, 2017, 484: 215-224.
- [13] 姜茸, 廖鸿志, 杨明. 信息熵在软件领域中的应用研究现状[J]. *自动化技术与应用*, 2015 (4): 1-6.
JIANG Rong, LIAO Hongzhi, YANG Ming. The current research of information entropy in software domain [J]. *Techniques of Automation and Applications*, 2015 (4): 1-6.
- [14] ZHOU X, WANG X, HU C, et al. An analysis on the relationship between uncertainty and misclassification rate of classifiers [J]. *Information Sciences*, 2020, 535: 16-27.
- [15] WANG D, SHELLHAMER E, LIU S, et al. Fully test-time adaptation by entropy minimization [EB/OL]. (2020-06-18) [2023-10-16]. <https://arxiv.org/abs/2006.10726>.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [17] HUANG B, LIU J, ZHANG Q, et al. Identification and classification of aluminum scrap grades based on the Resnet18 model[J]. *Applied Sciences*, 2022, 12(21): 11133.
- [18] LI B, LIMA D. Facial expression recognition via ResNet-50[J]. *International Journal of Cognitive Computing in Engineering*, 2021, 2: 57-64.
- [19] HARJOSEPUTRO Y, YUDA I, DANUKUSUMO K P. MobileNets: efficient convolutional neural network for identification of protected birds [J]. *IJASEIT (International Journal on Advanced Science, Engineering and Information Technology)*, 2020, 10(6): 2290-2296.
- [20] REZAEI K, MOUSAVIRAD S J, KHOSRAVI M R, et al. An autonomous UAV-assisted distance-aware crowd sensing platform using deep ShuffleNet transfer learning [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 23(7): 9404-9413.