

DOI:10.11918/202308070

基于三支决策的表情识别集成剪枝算法

唐玉梅, 李丹杨, 陈星, 吴义青, 黄仕松

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 集成剪枝通过去除弱学习器和冗余学习器, 能显著提升集成系统的表情识别效果。然而, 现有方法聚焦于对分类器的接受或拒绝, 导致评估信息缺乏可靠性或完整性时, 会保留弱分类器或剔除关键分类器。此外, 依赖准确率或多样性来评估分类器的优劣, 难以全面反映分类器的真实表现。因此, 提出基于三支决策的表情识别集成剪枝算法(3WDEP), 引入延迟决策解决分类器评估的不确定性问题。同时, 提出“预测偏好”概念, 综合预测结果与实际标签的相关性度量、准确率及召回率指标, 构建集成剪枝信息系统, 全面评估分类器性能。基于熵权法确定指标权重, 结合三支决策权衡分类器在不同决策选项下的损失, 选出对集成系统贡献最大的分类器进行集成。将召回率作为收益和代价属性, 优化集成剪枝效果。实验结果表明, 3WDEP能够有效提升表情识别效果, 相较于初始集成系统, 在 FER2013、JAFPE、CK+ 和 KDEF 上的准确率分别提高了 3.32%、9.39%、1.26% 和 4.9%。

关键词: 人脸表情识别; 集成剪枝; 三支决策; 预测偏好; 综合评估

中图分类号: TP3 **文献标志码:** A **文章编号:** 0367-6234(2026)03-0064-10

A three-way decision-based ensemble pruning algorithm for facial expression recognition

TANG Yumei, LI Danyang, CHEN Xing, WU Yiqing, HUANG Shisong

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

Abstract: By removing weak and redundant learners, ensemble pruning can significantly enhance the efficacy of ensemble system-based facial expression recognition. However, existing methods primarily focus on either accepting or rejecting classifiers, which results in the retention of weak classifiers or the exclusion of pivotal ones when evaluation information is unreliable or incomplete. Additionally, relying on accuracy or diversity to evaluate the merits of the classifier is difficult to fully reflect the true performance of the classifier. Consequently, this paper proposed a three-way decision-based ensemble pruning algorithm (3WDEP) for facial expression recognition, which introduced a delayed acceptance strategy to address uncertainties in classifier assessment. Simultaneously, the concept of “predictive preference” was introduced, integrating the correlation measurement between prediction results and actual labels, as well as accuracy and recall metrics, so as to construct an ensemble pruning information system and comprehensively evaluate the classifier performance. The entropy weight method was used to determine the weight of the indicators, and combined with a three-way decision, the loss of classifiers under different decision options was considered to select the classifiers that contributed the most to the ensemble system for integration. Recall was utilized as both a benefit attribute and a cost attribute to optimize the ensemble pruning effect. Experimental results show that 3WDEP effectively improves facial expression recognition performance, and the accuracy improves by 3.32%, 9.39%, 1.26%, and 4.9% compared to the initial ensemble system on FER2013, JAFPE, CK+, and KDEF, respectively.

Keywords: facial expression recognition; ensemble pruning; three-way decision; predictive preference; comprehensive evaluation

表情识别在舆情分析^[1]、道路驾驶^[2]、虚拟现实^[3]等场景扮演着重要角色。然而, 受光照、姿势等复杂因素影响^[4], 准确识别表情极具挑战。研究者常采用集成多个分类器的方法提升表情识别性

能^[5-6], 但集成系统中的弱分类器和冗余分类器会增加预测系统的存储和计算负担。因此, 学者提出集成剪枝方法对集成系统中的学习器进行筛选, 保留性能优良的分类器子集, 以弥补集成学习的不

收稿日期: 2023-08-23; 录用日期: 2023-11-08; 网络首发日期: 2024-06-25

网络首发地址: <https://link.cnki.net/urlid/23.1235.t.20240624.1508.004>

基金项目: 贵州省科技计划项目(黔科合平台人才[2018]5781)

作者简介: 唐玉梅(1997—), 女, 硕士研究生; 李丹杨(1991—), 女, 副教授, 硕士生导师

通信作者: 李丹杨, danyangcl@163.com

足^[7]。例如, Li等^[8]应用 Relief-F 算法评估分类器性能,并结合基于聚类和排序的集成剪枝方法提升子集的多样性;Huang等^[9]提出基于聚类软标签的集成剪枝方法,有效地提高了表情的识别性能和系统实时性。但现有研究侧重于直接接受或拒绝分类器,当评估信息不可靠或不完整时,易导致弱分类器被保留或关键分类器被剔除,降低集成系统性能。另外,主流集成剪枝方法在评估分类器优劣时依赖准确率或多样性指标,难以全面准确地反映分类器的真实表现。三支决策^[10]是一种解决带有不确定性决策问题的粒计算方法。其核心是在做决策时,除接受和拒绝外,还引入了延迟决策作为第三种中立决策选项,可更好地处理信息不完整、风险较高或不确定性较大的决策问题。该方法通过评估成本、效益等多个属性或指标,为每个选择赋予相应值,每个决策选项可用包含多个属性值的向量表示,再采用粒计算方法进行决策分析,综合属性间的相互影响得出最佳决策结果。

为此,本文提出基于三支决策的表情识别集成剪枝算法(3WDEP)。该算法通过划分分类器的接受域、暂缓接受域与拒绝域,有效避免了对潜在有用分类器的盲目剔除或对弱分类器的错误保留。同时,构建包含“预测偏好”在内的多维度评价体系,结合熵权法确定各指标权重,结合三支决策理论综合权衡不同决策选项下的分类器损失,选出对集成系统贡献最大的分类器子集,以优化表情识别系统性能。通过将召回率视为收益和代价属性进行双向分析,进一步优化决策权衡过程,提升识别效果。

1 基本理论

1.1 概率粗糙集

定义1 设 U 是论域, R 是 U 上的等价关系,其构成的等价类为 $[x]_R$ 。设 $0 \leq \beta \leq \alpha \leq 1$, 对 $\forall X \subseteq U$, 其概率下近似和上近似分别为:

$$\underline{\text{Apr}}(X) = \{x \in U | P(X|[x]_R) \geq \alpha\} \quad (1)$$

$$\overline{\text{Apr}}(X) = \{x \in U | P(X|[x]_R) > \beta\} \quad (2)$$

式中, $P(X|[x]_R) = \frac{|[x]_R \cap X|}{|[x]_R|}$ 为概率粗糙集,表示已知对象 x 被描述为等价类 $[x]_R$ 的情况下,属于类别 X 的置信度。概率粗糙集将论域 U 划分为正域 $\text{POS}(X)$ 、边界域 $\text{BND}(X)$ 和负域 $\text{NEG}(X)$, 分别对应决策过程的接受、弃权和拒绝行动,表示为:

$$\text{POS}(X) = \underline{\text{Apr}}(X) = \{x \in U | P(X|[x]_R) \geq \alpha\} \quad (3)$$

$$\text{BND}(X) = \overline{\text{Apr}} - \underline{\text{Apr}}(X) = \{x \in U | \beta < P(X|[x]_R) < \alpha\} \quad (4)$$

$$\text{NEG}(X) = U - \overline{\text{Apr}}(X) = \{x \in U | P(X|[x]_R) \leq \beta\} \quad (5)$$

1.2 三支决策模型

Yao^[11]引入贝叶斯决策理论为上述3个区域提供语义解释。设状态集 $\Omega = \{X, \neg X\}$ 表示对象属于 X 或不属于 X ; $A = \{A_P, A_B, A_N\}$ 为行动集,其中 A_P 、 A_B 和 A_N 分别对应正域 $\text{POS}(X)$ 、边界域 $\text{BND}(X)$ 和负域 $\text{NEG}(X)$ 的决策动作。表1为上述状态与动作对应的经典损失函数, $\lambda_{P, \in}$ 和 $\lambda_{P, \notin}$ 分别表示 $x \in X$ 和 $x \notin X$ 时采取决策行动 A_P 的成本。

表1 损失函数

Tab.1 The loss function

动作	X	$\neg X$
A_P	$\lambda_{P, \in}$	$\lambda_{P, \notin}$
A_B	$\lambda_{B, \in}$	$\lambda_{B, \notin}$
A_N	$\lambda_{N, \in}$	$\lambda_{N, \notin}$

设 $L(A_*|[x]_R)$ ($*$ = P、B、N) 为采取决策行动 A_* 的预期损失,由贝叶斯决策理论可得:

$$L(A_P|[x]_R) = \lambda_{P, \in} P(X|[x]_R) + \lambda_{P, \notin} P(\neg X|[x]_R) \quad (6)$$

$$L(A_B|[x]_R) = \lambda_{B, \in} P(X|[x]_R) + \lambda_{B, \notin} P(\neg X|[x]_R) \quad (7)$$

$$L(A_N|[x]_R) = \lambda_{N, \in} P(X|[x]_R) + \lambda_{N, \notin} P(\neg X|[x]_R) \quad (8)$$

根据贝叶斯决策理论,最佳决策代价最小,有:

R_P : 若 $L(A_P|[x]_R) \leq L(A_B|[x]_R)$ 且 $L(A_P|[x]_R) \leq L(A_N|[x]_R)$, 则 $x \in \text{POS}(X)$

R_B : 若 $L(A_B|[x]_R) \leq L(A_P|[x]_R)$ 且 $L(A_B|[x]_R) \leq L(A_N|[x]_R)$, 则 $x \in \text{BND}(X)$

R_N : 若 $L(A_N|[x]_R) \leq L(A_P|[x]_R)$ 且 $L(A_N|[x]_R) \leq L(A_B|[x]_R)$, 则 $x \in \text{NEG}(X)$

若 $\lambda_{P, \in} \leq \lambda_{B, \in} < \lambda_{N, \in}$, $\lambda_{N, \notin} \leq \lambda_{B, \notin} < \lambda_{P, \notin}$, $P(X|[x]_R) + P(\neg X|[x]_R) = 1$, 则决策规则可简化为:

R_P : 若 $P(X|[x]_R) \geq \alpha$ 且 $P(X|[x]_R) \geq \eta$, 则 $x \in \text{POS}(X)$

R_B : 若 $P(X|[x]_R) \leq \alpha$ 且 $P(X|[x]_R) \geq \beta$, 则 $x \in \text{BND}(X)$

R_N : 若 $P(X|[x]_R) \leq \beta$ 且 $P(X|[x]_R) \leq \eta$, 则 $x \in \text{NEG}(X)$

其中, 阈值 α, β, η 表示损失函数的相对值, 分别为:

$$\alpha = \frac{\lambda_{P, \notin} - \lambda_{B, \notin}}{(\lambda_{P, \notin} - \lambda_{B, \notin}) + (\lambda_{B, \in} - \lambda_{P, \in})} \quad (9)$$

$$\beta = \frac{\lambda_{B, \notin} - \lambda_{N, \notin}}{(\lambda_{B, \notin} - \lambda_{N, \notin}) + (\lambda_{N, \in} - \lambda_{B, \in})} \quad (10)$$

$$\eta = \frac{\lambda_{P, \varepsilon} - \lambda_{N, \varepsilon}}{(\lambda_{P, \varepsilon} - \lambda_{N, \varepsilon}) + (\lambda_{N, \varepsilon} - \lambda_{B, \varepsilon})} \quad (11)$$

若 $(\lambda_{P, \varepsilon} - \lambda_{B, \varepsilon})(\lambda_{N, \varepsilon} - \lambda_{B, \varepsilon}) > (\lambda_{B, \varepsilon} - \lambda_{P, \varepsilon})(\lambda_{B, \varepsilon} - \lambda_{N, \varepsilon})$, 则 $0 \leq \beta < \eta < \alpha \leq 1$, 决策规则简化为:

$R_{P'}$: 若 $P(X|[x]_R) \geq \alpha$, 则 $x \in \text{POS}(X)$

$R_{B'}$: 若 $\beta < P(X|[x]_R) < \alpha$, 则 $x \in \text{BND}(X)$

$R_{N'}$: 若 $P(X|[x]_R) \leq \beta$, 则 $x \in \text{NEG}(X)$

2 3WDEP 模型

本文所提 3WDEP 模型包括基分类器池构建、分类器选择及分类器集成 3 部分。分类器选择作为核心环节, 首先定义预测偏好 (PP), 并综合预测结果与实际标签之间的互信息相关性 (Cor)、准确率 (Acc) 以及召回率 (Recall) 作为分类器评价指标, 构建集成剪枝信息系统; 其次, 结合熵权法计算各个指标的权重 w_j ; 然后, 利用权重计算分类器 C_i 在不同决策 (接受、暂缓接受、拒绝) 下的相对损失函数和决策阈值 (θ_i, ξ_i) , 并根据其模糊 ε -邻域类 $[C_i]_{S_\alpha}^\varepsilon$ 与模糊集 X 之间的隶属度关系, 评估分类器性能; 最后, 根据模糊概率与决策阈值之间的关系, 将分类器划分为 POS(X)、BND(X) 和 NEG(X) 3 个区域, 由于接受 NEG(X) 区域的分类器会造成较大的决策损失, 因此, 直接剔除 NEG(X) 中的分类器子集, 再计算剩余分类器选择的代价并排序, 选择出代价最小的 top- k 个分类器进行集成。总体流程如图 1 所示。

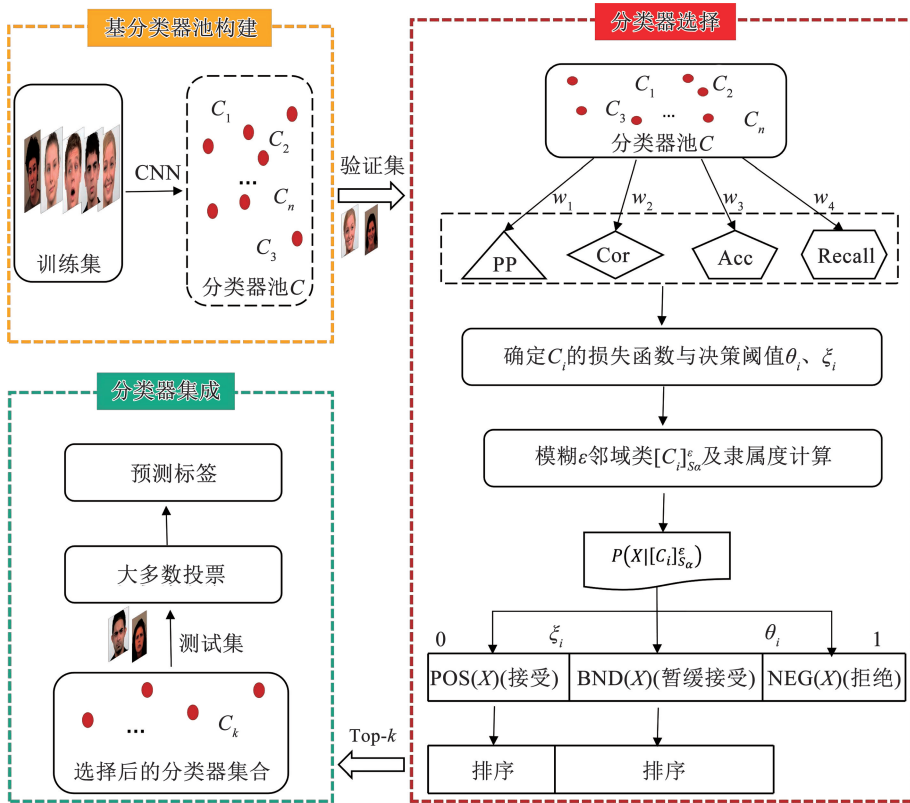


图 1 基于三支决策的表情识别集成剪枝方法流程图

Fig. 1 Flowchart of facial expression recognition ensemble pruning method based on three-way decision

2.1 基分类器池构建

卷积神经网络^[12] (CNN) 能有效提取图像深层次信息, 在大型图像处理中表现优异。因此, 本文选择 CNN 作为基分类器拟合表情数据。

如图 2 所示, CNN 基本结构包括输入层、卷积层、池化层、全连接层和输出层。本文共构造了 231 种不同 CNN 模型, 通过改变网络层数、卷积核大小等参数增加分类器多样性, 具体参数设置将在实验部分讨论。

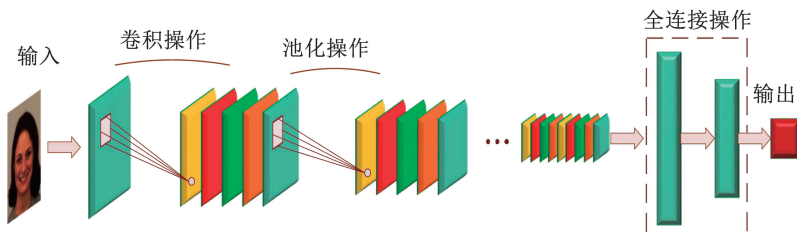


图 2 CNN 的结构框图

Fig. 2 Structural block diagram of CNN

2.2 分类器选择

1) 分类器评估标准

设 $C = \{C_i | i \in N\}$ ($N = \{1, 2, \dots, n\}$) 代表 n 个分类器; $A = \{A_j | j \in M\}$ ($M = \{1, 2, \dots, m\}$) 为分类器的评估标准, 包括预测偏好 (PP)、分类器与标签之间的互信息相关性 (Cor)、准确率 (Acc)、召回率 (Recall), 即 $m=4$; $W = \{w_j | j \in M\}$ ($0 \leq w_j \leq 1, \sum_{j=1}^m w_j = 1$) 为各标准的权重。

a. 预测偏好 (PP)

数据不平衡时, 模型会倾向于预测多数类别而忽略少数类别, 使用准确率评估分类器可能会产生误导。因此, 本文提出分类器“预测偏好”的概念, 用于衡量分类器在各个类别上的预测倾向, 即使样本分布不平衡时也能有效进行分类器评估。表2为分类器在验证集上的预测混淆矩阵。

表2 分类器预测混淆矩阵

Tab. 2 Classifier prediction confusion matrix

	l_1	l_2	...	l_k
pre ₁	num ₁₁	num ₁₂	...	num _{1k}
pre ₂	num ₂₁	num ₂₂	...	num _{2k}
⋮	⋮	⋮	⋮	⋮
pre _k	num _{k1}	num _{k2}	...	num _{kk}

其中, l_k 为样本的真实标签类别; pre_k 为分类器的预测标签类别; num_{ij} 为分类器将第 j 类样本预测为第 i 类的数量。则分类器对 l_i 类样本的偏好为

$$PP_i = \frac{\sum_{j=1}^k \text{num}_{ij} - \sum_{j=1}^k \text{num}_{ji}}{\sum_{j=1}^k \text{num}_{ji}} \quad (12)$$

由式(12)可知, 当分类器具有较好识别效果时, 预测为第 l_i 类样本的数量与第 l_i 类样本的真实数量应接近一致, 即 PP_i 趋近于 0; 且对第 l_i 类样本, 真实样本数量越多, 预测数量与真实数量之间的差值变化对偏好值的影响越小。则分类器对所有类别的总偏好为

$$PP = \sum_{i=1}^k PP_i \quad (13)$$

b. 分类器与标签的互信息相关性 (Cor)

互信息通常用于衡量任意两个随机变量之间的相关性和统计依赖性。设有 T 个验证样本, 则分类器 C_i 可以表示为 $C_i = (h_{i1}, h_{i2}, \dots, h_{iT})$, h_{it} 为 C_i 对第 t 个样本的预测标签; 真实标签 $L = (l_1, l_2, \dots, l_T)$, l_t 表示第 t 个样本的真实标签。由互信息可得分类器 C_i 与标签 L 的相关性为

$$\text{Cor} = \sum_{h_{it} \in C_i, l_t \in L} P(h_{it}, l_t) \log \frac{P(h_{it}, l_t)}{P(h_{it})P(l_t)} \quad (14)$$

式中: $P(h_{it}, l_t)$ 为 C_i 和 L 的联合概率分布; $P(h_{it})$ 和 $P(l_t)$ 分别为 C_i 和 L 的边缘概率分布。相关性值越大, 表示分类器预测结果越可靠。

c. 准确率 (Acc)

设真阳性 (TP) 和真阴性 (TN) 分别为正确分类的阳性和阴性样本, 假阳性 (FP) 和假阴性 (FN) 分别是错误分类的阳性样本和阴性样本。则准确率 Acc 表示为

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (15)$$

d. 召回率 (Recall)

召回率能够衡量分类器对于正例的查全率, 即正确识别的正例与所有实际正例的比例

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

综上, 可得表3所示的集成剪枝信息系统, 更全面地评估基分类器, 从而对分类器进行排序和择优。其中, c_{ij} 表示分类器 C_i 在评价指标 A_j 上的值。

表3 集成剪枝信息系统

Tab. 3 Ensemble pruning information system

	A_1 (PP)	A_2 (Cor)	A_3 (Acc)	A_4 (Recall)
C_1	c_{11}	c_{12}	c_{13}	c_{14}
⋮	⋮	⋮	⋮	⋮
C_n	c_{n1}	c_{n2}	c_{n3}	c_{n4}
W	w_1	w_2	w_3	w_4

为保证数据的统一性, 对数据集进行如下处理: 若 A_j 为收益属性, 则:

$$c_{ij} = \frac{c_{ij}}{\max \left(\bigcup_{i=1}^N c_{ij} \right)}, j \in M \quad (17)$$

若 A_j 为代价属性, 则:

$$c_{ij} = 1 - \frac{c_{ij}}{\max \left(\bigcup_{i=1}^N c_{ij} \right)}, j \in M \quad (18)$$

2) 指标权重确定

首先, 使用熵权法对指标进行客观赋权, 衡量各分类器在指标数值间的差异。对于 $i \in N$ 且 $j \in M$, 第 i 个分类器在第 j 个指标下的特征比重为

$$p_{ij} = \frac{c_{ij}}{\sum_{i=1}^n c_{ij}} \quad (19)$$

根据指标的特征比重可知第 j 项指标的信息熵为

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^n p_{ij} \ln p_{ij} \quad (20)$$

当 $c_{ij} = 0$ 时, $p_{ij} \ln p_{ij} = 0$ 。第 j 项指标的权重为

$$w_j = \frac{1 - e_j}{\sum_{j=1}^m (1 - e_j)} \quad (21)$$

3) 决策阈值计算

在获得准则的权重后, 计算相对损失函数和决策阈值 (θ_i, ξ_i) 。由于经典决策规则在实际应用时存在损失函数需主观赋值的不足, Jia 等^[13] 提出了一种确定性阈值计算方法。据此可以得出分类器 C_i 在接受 (A_P) 、暂缓接受 (A_B) 、拒绝 (A_N) 3 种决策下的损失如表 4 所示。

表 4 分类器 C_i 在两种状态下采取不同行动的相对损失函数

Tab. 4 Relative loss function of classifier C_i taking different actions in two different states

	X	$\neg X$
A_P	0	$\sum_{j=1}^m w_j (1 - c_{ij})$
A_B	$\sum_{j=1}^m \sigma w_j c_{ij}$	$\sum_{j=1}^m \sigma w_j (1 - c_{ij})$
A_N	$\sum_{j=1}^m w_j c_{ij}$	0

其中, $\sigma \in [0, 0.5]$ 为风险规避系数, 代表决策者的风险偏好, σ 取值越大, 表明越偏好风险, 越小则越厌恶风险, 当 $\sigma = 0$ 时决策完全规避风险, $\sigma = 0.5$ 时表示完全追求风险。

根据三支决策模型, 分类器 C_i 的决策阈值为:

$$\theta_i = \frac{\sum_{j=1}^m (1 - \sigma) w_j (1 - c_{ij})}{\sum_{j=1}^m (1 - \sigma) w_j (1 - c_{ij}) + \sum_{j=1}^m \sigma w_j c_{ij}} \quad (22)$$

$$\xi_i = \frac{\sum_{j=1}^m \sigma w_j (1 - c_{ij})}{\sum_{j=1}^m \sigma w_j (1 - c_{ij}) + \sum_{j=1}^m (1 - \sigma) w_j c_{ij}} \quad (23)$$

4) 隶属度计算

Ye 等^[14] 构建了一种自反模糊 α -邻域算子, 并提出了模糊 ε -邻域类的概念, 将集合划分为不同的等价类。根据模糊 α -邻域算子, 为了确保每个分类器都有相同被选择的机会, 给定临界值 $\alpha \in (0, \min_i \max_j c_{ij})$ 使得每个分类器至少在一个评估指标上的值达到 α , 则分类器 C_i, C_l 之间的模糊相似关系为

$$S_\alpha(C_i)(C_l) = \bigwedge_{A_j \in Md_A^\alpha(c_i)} \mathcal{I}(c_{ij}, c_{lj}) \quad (24)$$

其中, $Md_A^\alpha(c_i)$ 为 C_i 的模糊 α -最小描述^[15]。则 C_i 的模糊 α -邻域 $S_\alpha(C_i)$ 为分类器集合 C 上的模糊集, 即

$$S_\alpha(C_i) = \frac{S_\alpha(C_i)(C_1)}{C_1} + \frac{S_\alpha(C_i)(C_2)}{C_2} +$$

$$\dots + \frac{S_\alpha(C_i)(C_n)}{C_n} \quad (25)$$

设 ε 为邻域相似度系数, 用于衡量分类器之间的相似程度。分类器 C_i 的模糊 ε -邻域类 $[C_i]_{S_\alpha}^\varepsilon$ 为

$$[C_i]_{S_\alpha}^\varepsilon = \{C_l \in C \mid S_\alpha(C_i)(C_l) \geq \varepsilon\} \quad (26)$$

通过分类器集合的模糊集 $X = \frac{X(C_1)}{C_1} + \frac{X(C_2)}{C_2} + \dots + \frac{X(C_n)}{C_n}$, 评估分类器的优劣程度。其中, $X(C_i)$ 为分类器 C_i 关于 X 的隶属度, 表示为

$$X(C_i) = \sum_{j=1}^m w_j c_{ij} \quad (27)$$

$X(C_i)$ 的值越接近 1, 表明分类器性能越好。令

$$P(X \mid [C_i]_{S_\alpha}^\varepsilon) = \frac{\sum_{C_l \in [C_i]_{S_\alpha}^\varepsilon} X(C_l)}{|[C_i]_{S_\alpha}^\varepsilon|} \quad (28)$$

表示在 $[C_i]_{S_\alpha}^\varepsilon$ 中 C_l 属于模糊集 X 的条件概率。

5) 集成剪枝规则

根据三支决策模型, 对任意 $i \in N, 0 \leq \xi_i < \theta_i \leq 1$, 模糊集 X 的下、上近似为:

$$\underline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) = \{C_l \in C \mid P(X \mid [C_l]_{S_\alpha}^\varepsilon) \geq \theta_i\} \quad (28)$$

$$\overline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) = \{C_l \in C \mid P(X \mid [C_l]_{S_\alpha}^\varepsilon) > \xi_i\} \quad (29)$$

若 $\underline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) \neq \overline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X)$, X 为概率粗糙模糊集; 否则, X 为概率精确模糊集。根据三支决策模型将分类器集合 C 划分为接受域 $\text{POS}(X)$ 、暂缓接受域 $\text{BND}(X)$ 和拒绝域 $\text{NEG}(X)$, 分别表示为:

$$\text{POS}(X) = \underline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) = \{C_l \in C \mid P(X \mid [C_l]_{S_\alpha}^\varepsilon) \geq \theta_i\} \quad (30)$$

$$\text{BND}(X) = \overline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) - \underline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) = \{C_l \in C \mid \xi_i < P(X \mid [C_l]_{S_\alpha}^\varepsilon) < \theta_i\} \quad (31)$$

$$\text{NEG}(X) = C - \overline{\text{Apr}}_{|(\xi_i, \theta_i) \mid i \in N|}(X) = \{C_l \in C \mid P(X \mid [C_l]_{S_\alpha}^\varepsilon) \leq \xi_i\} \quad (32)$$

根据贝叶斯决策理论, 可以得到不同区域下分类器的代价为

$$\text{cost}(C_i) = \begin{cases} L(A_P \mid [C_i]_{S_\alpha}^\varepsilon), & C_i \in \text{POS}(X); \\ L(A_B \mid [C_i]_{S_\alpha}^\varepsilon), & C_i \in \text{BND}(X); \\ L(A_N \mid [C_i]_{S_\alpha}^\varepsilon), & C_i \in \text{NEG}(X) \end{cases} \quad (33)$$

$L(A_* \mid [C_i]_{S_\alpha}^\varepsilon)$ ($*$ = P、B、N) 表示分类器 C_i 采取行动 A_* 的期望损失。显然, 当 C_i 与 C_l 属于同一

区域且 $\text{cost}(C_i) < \text{cost}(C_l)$ 时,分类器 C_i 效果优于分类器 C_l ,记为 $C_i > C_l$ 。结合 $\text{POS}(X) > \text{BND}(X) > \text{NEG}(X)$,可以得到所有分类器的最终排名^[16]。

2.3 分类器集成

使用大多数投票法将排名 top- k 的分类器子集 $C_{\text{new}} = \{C_1, C_2, \dots, C_k\}$ ($k < N$) 的预测结果融合,直到达到最优识别准确率。设 $P_l(x)$ 为大多数投票后,测试样本 x 的最终识别结果,则

$$P_l(x) = \arg \max_{l \in L} \sum_{i=1}^k I(C_i(x) = l) \quad (34)$$

式中: $I(\cdot)$ 是指示函数; L 代表所有样本类别; $C_i(x)$ 是第 i 个分类器对测试样本 x 的预测标签。若分类器对样本的预测标签为 l ,即 $C_i(x) = l$,则 $I(\cdot) = 1$,否则为0。

2.4 时间复杂度分析

本文涉及 n 个分类器和 m 个评价指标。收益



图3 各数据集部分表情示例

Fig. 3 Partial examples of facial expressions in each dataset

表5 各数据集的详细信息

Tab. 5 Details of each dataset

数据集	生气	厌恶	恐惧	快乐	悲伤	惊喜	中性	总计
FER2013	4 953	547	5 121	8 989	6 077	4 002	6 198	35 887
JAFFE	30	29	32	31	31	30	30	213
CK +	135	177	75	207	84	249	593	1520
KDEF	70	70	70	70	70	70	70	490

2) 初始集成系统构建

将表情数据集划分为训练集、测试集和验证集三个部分。其中,FER2013中的训练图作为训练集,用于训练分类器;公共验证图作为验证集,用于验证分类器性能,并基于验证集数据选择分类器;测试集包括FER2013中的私有验证图、JAFFE、CK+和KDEF,用于测试整个识别系统的性能。

将图像的大小统一为 48×48 像素,使用OpenCV人脸识别算法去除背景区域,突出显示人脸区域,并将所有图像转换为灰度图。基于TensorFlow

属性与代价属性归一化的时间复杂度为 $O(m)$,指标权重计算的时间复杂度为 $O(mn)$,而相对损失函数和阈值计算的时间复杂度为 $O(n)$ 。此外,模糊 ε -邻域类计算的时间复杂度为 $O(m^2n + n^2m)$,确定集成剪枝规则的时间复杂度为 $O(n)$ 。因此,分类器选择过程的时间复杂度为 $O(m^2n + n^2m)$ 。

3 实验结果和分析

3.1 实验准备

1) 数据集

使用公开数据集可以有效节约收集数据的时间,也能更公平地评价分类器以及模型的性能。本实验选用FER2013^[17]、JAFFE^[18]、CK+^[19]、KDEF^[20]4个经典的人脸表情数据集进行实验,部分表情如图3所示。表5为各个数据集的详细信息。

框架,共生成231个卷积神经网络分类器形成分类器池。调整的神经网络参数类型及范围如表6所示。

表6 生成的CNN模型参数范围

Tab. 6 The range of parameters of the generated CNN model

参数类型	参数范围	参数类型	参数范围
网络层数	3~9	随机失活率	0.4~1
卷积核	[4,4]~[8,8]	激活函数	ReLU, Leaky_ReLU, Sigmoid
池化层核	[2,2]~[5,5]	特征图数量	16, 32, 48, 64, 80, 96
池化步幅	2	全连接层维度	7, 1 024, 2 048, 3 072

图 4 展示了基分类器池中的分类器在不同数据集上的预测准确率分布情况。可以发现,分类器对 JAFFE 识别准确率上限明显低于其余数据集。这是由于 FER2013 包含了西方人表情,而 JAFFE 为亚洲人表情,不同文化背景和社会习惯会导致表情表达方式存在差异。例如,在表达愤怒时,亚洲人倾

向于微微皱眉或者紧抿嘴唇,而西方人则可能表现出夸张的表情。因此,基于 FER2013 数据集训练的分类器在 JAFFE 数据集上的测试效果较差,说明单一模型的鲁棒性较弱。总体而言,基分类器池的识别准确率区间范围大,表明基分类器池具有多样性。

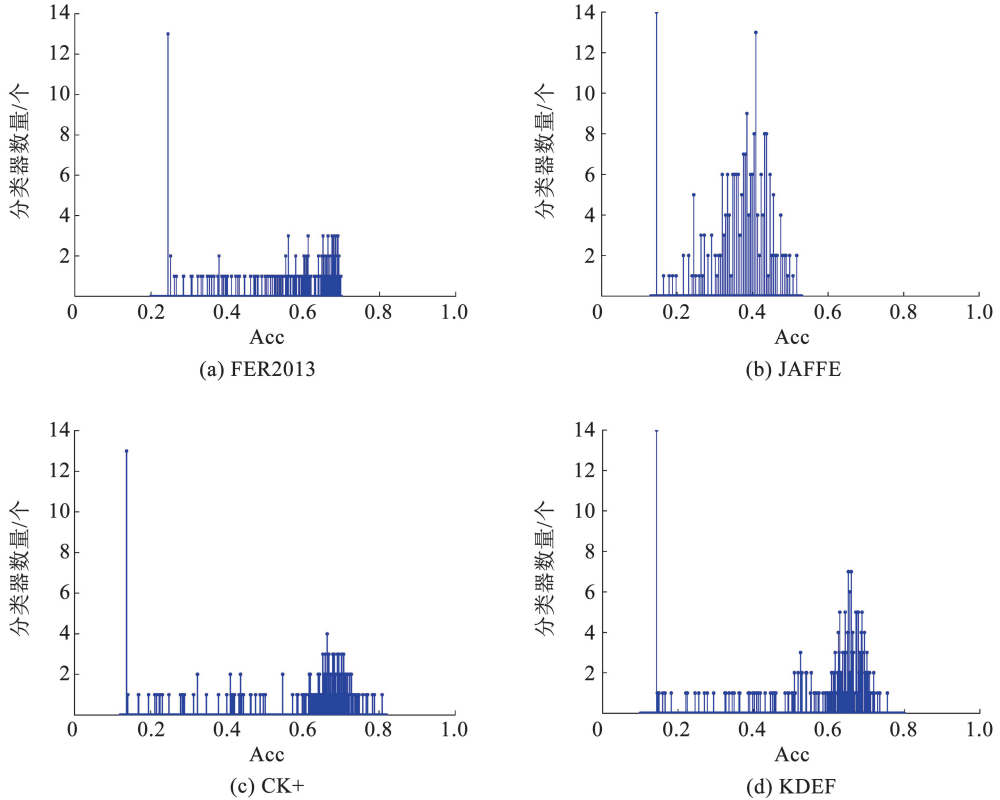


图 4 分类器在不同数据集上的预测结果分布情况

Fig. 4 The distribution of classifiers predictions on different datasets

3.2 召回率影响分析

本文涉及 3 个参数:风险规避系数 $\sigma \in [0, 0.5]$, 临界值 $\alpha \in (0, 0.5]$, 邻域相似度系数 $\varepsilon \in (0, 1]$; 参数设置步长均为 0.1。通过网格搜索方法确定最佳参数,所有实验结果均在最佳参数下分析得出。

在三支决策中,收益属性有助于了解每个分类器的积极结果,代价属性则揭示负面影响。综合考虑这两个属性,可以更全面了解分类器的对集成系统的潜在影响。3WDEP 的 4 个评价指标中,高准确率能有效识别样本,属于收益属性;分类器与实际标签相关性高表明预测结果可靠,属于收益属性;而预测偏好大可能导致分类器偏向预测某一类别,属于代价属性;高召回率有利于减少漏报但可能增加误报。因此,实验分别探索召回率作为收益属性 Recall⁺ 和代价属性 Recall⁻ 两种情况对集成剪枝效果的影响。

由表 7 可知,将召回率视为代价属性对于提升表情识别效果更为显著。这是因为随着召回率的增

加,分类器更容易将负样本错误地分类为正样本,导致误报增加,从而降低了分类器的整体预测效果。反之,系统则更倾向于降低误报的影响,从而提高准确率。

表 7 使用召回率作为收益属性和代价属性的集成剪枝准确率

Tab. 7 Leveraging recall as profit and cost attributes for ensemble pruning accuracy

Dataset	Baseline		Recall ⁺		Recall ⁻	
	准确率	数量	准确率	数量	准确率	数量
JAFFE	0.422 5	231	0.511 7	131	0.516 4	134
CK +	0.757 8	231	0.769 7	142	0.770 4	149
KDEF	0.716 3	231	0.761 2	39	0.765 3	49
FER2013	0.703 5	231	0.735 3	32	0.736 7	27

3.3 灵敏度分析

为了研究本文算法对参数的敏感性,改变 σ 和 ε 的值,当 $\alpha = 0.5$ 、召回率视为代价属性时,在各个数据集上的调优结果如图 5 所示。

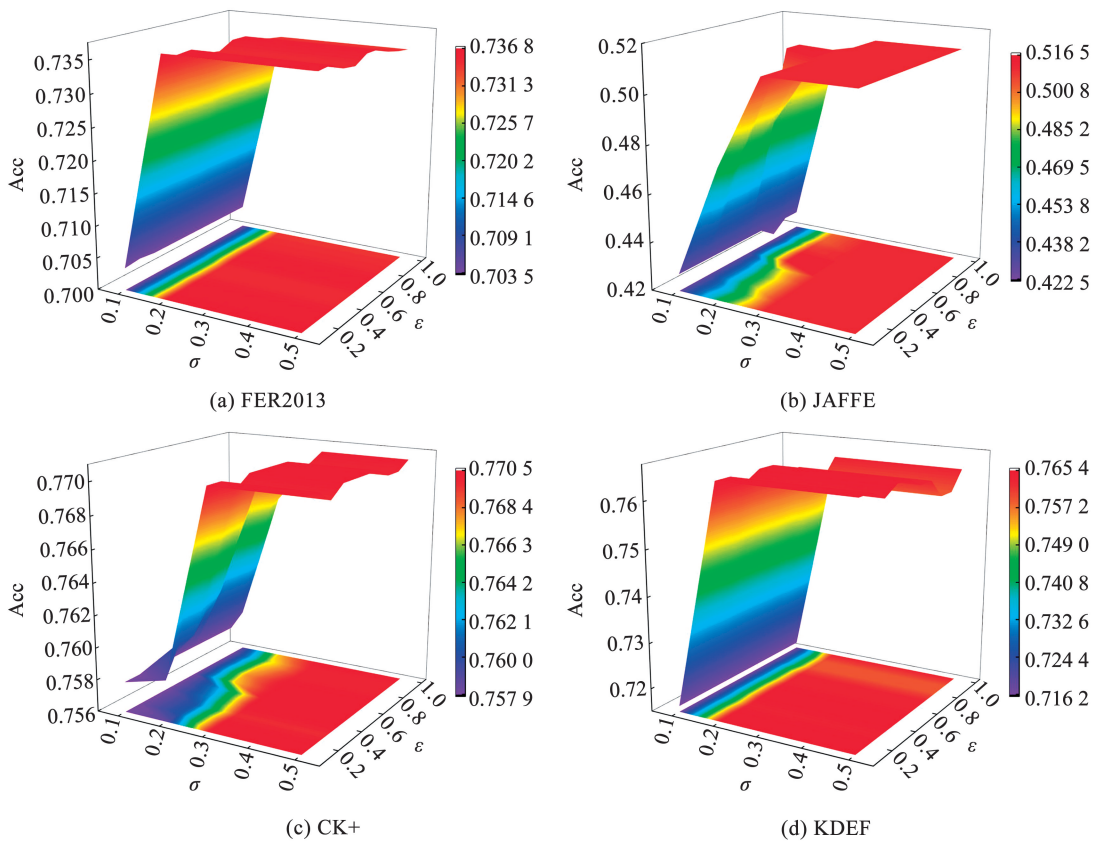


图5 当 $\alpha=0.5$ 时,在不同 σ 和 ε 下 3WDEP 的准确率

Fig. 5 Accuracy of 3WDEP with different σ and ε values when $\alpha=0.5$

由图5可知,随着风险规避系数 σ 的增大,4个数据集上的识别准确率逐渐增大并最终趋于稳定。这表明在分类器选择过程中,不同的风险态度会导致集成剪枝效果的波动,但当风险规避系数达到一定阈值时,算法会进入稳定状态。此外,当 ε 从0.1增加到1时,总体上准确率变化较小,说明本文提出的集成剪枝模型具有良好的稳定性。

图6展示了集成剪枝前后分类器的准确率分布情况,宽度和长度分别表示当前区间内分类器的数量及准确率分布范围。结果显示,剪枝后,图形变得更为紧凑,且集成系统的准确率的均值增大,表明本文方法能够剔除大多数冗余和弱分类器,优化集成系统性能。

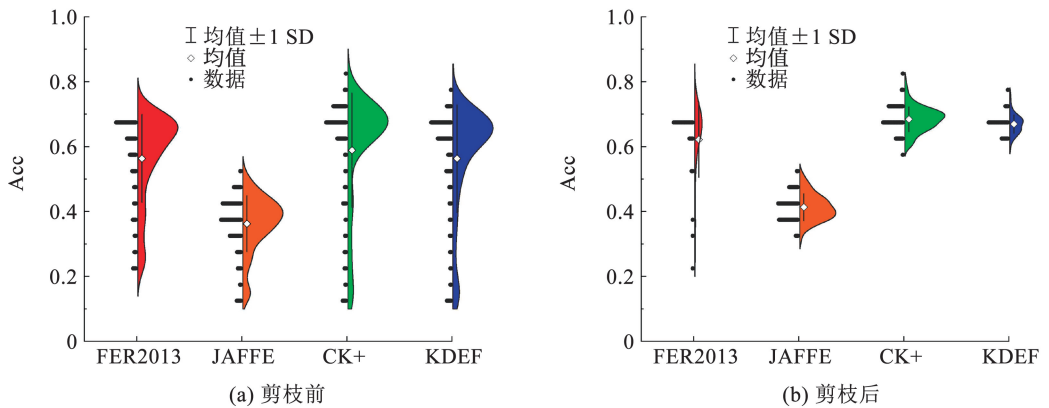


图6 集成剪枝前后的基分类器分布图

Fig. 6 Base classifiers distribution map before and after ensemble pruning

3.4 消融实验

为了验证预测偏好对实验效果的影响,图7展示了具有预测偏好 with_PP 时,表情识别准确率随分类器数量的变化趋势。4个数据集上,随着分类

器数量的增加,表情识别准确率总体上呈先上升后下降的趋势。这说明增加分类器的数量并不一定能持续提升集成系统的识别效果,验证了本文提出的集成剪枝方法的重要性。

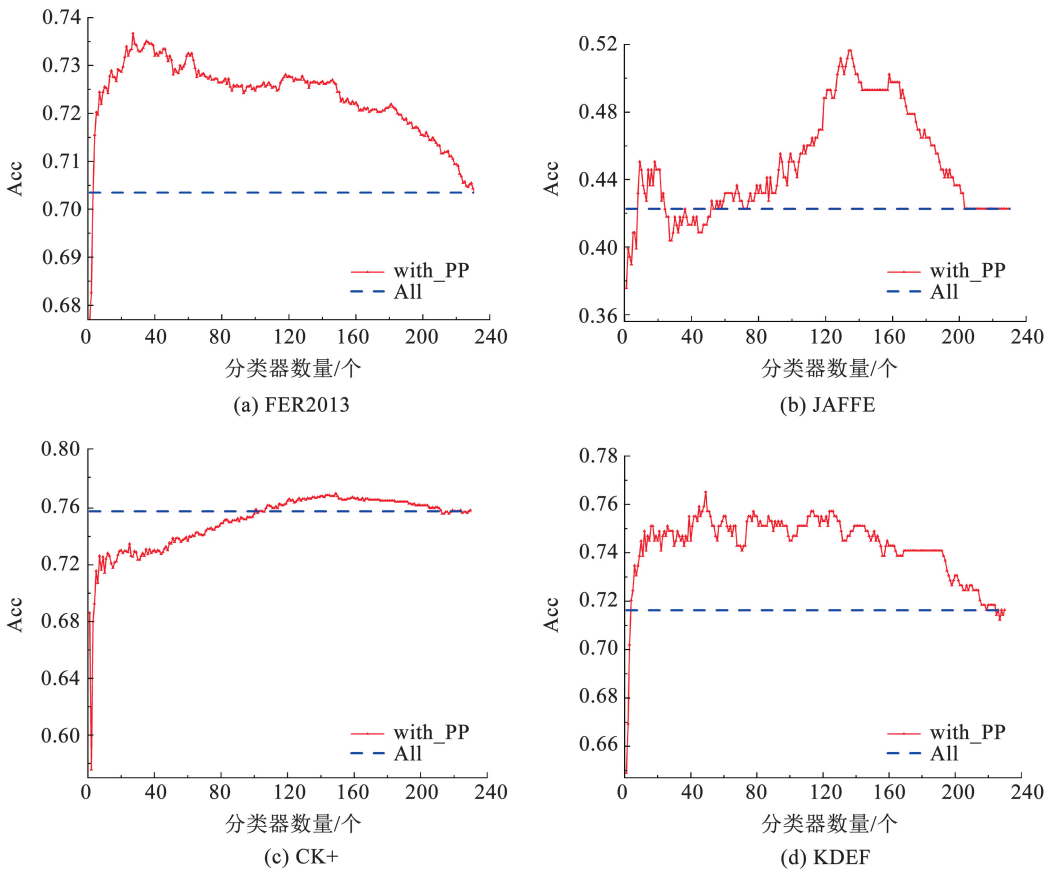


图 7 预测准确率随分类器数量的变化趋势

Fig. 7 The trend of the prediction accuracy with the number of classifiers

表 8 可知,相比没有预测偏好(no_PP),引入预测偏好后,表情识别的准确率得到提升。这表明在决策权衡中考虑预测偏好的影响有助于提高表情识别的准确性。因此,预测偏好作为一项新的评价指标,对于选择合适的分类器进行集成具有积极影响,进一步优化了表情识别系统的性能。

3.5 整体分类性能

表 9 呈现了 3WDEP 与 8 种剪枝算法在各个数据集上多次实验的最佳效果。结果表明,集成剪枝方法能以较小的分类器规模实现更优的表情识别效果。相比初始集成系统(Baseline),本文方法在 FER2013、JAFFE、KDEF、CK+ 上的准确率分别提高了 3.32%、9.39%、4.9%、1.26%。3WDEP 的独特

之处在于可以全面、多维度地评估分类器,从而更有效地选择贡献最大且互补性更好的分类器子集,因此表情识别效果优于其他集成剪枝方法。

表 8 有预测偏好和没有预测偏好对实验效果的影响
Tab. 8 The impact of PP and absence of PP on the experimental results

Dataset	Baseline		no_PP		with_PP	
	准确率	数量	准确率	数量	准确率	数量
JAFFE	0.4225	231	0.5117	134	0.5164	134
CK+	0.7578	231	0.7711	142	0.7704	149
KDEF	0.7163	231	0.7571	55	0.7653	49
FER2013	0.7035	231	0.7347	20	0.7367	27

表 9 3WDEP 与其他剪枝方法对比

Tab. 9 Comparisons between 3WDEP and other ensemble pruning methods

	FER2013		JAFFE		KDEF		CK+	
	准确率	数量	准确率	数量	准确率	数量	准确率	数量
Baseline	0.7035	231	0.4225	231	0.7163	231	0.7578	231
comEP	0.7108	209	0.5070	3	0.7204	209	0.7592	224
CSLSEP	0.7359	29	0.5117	133	0.7612	50	0.7717	143
RE ^[21]	0.7314	53	0.5023	101	0.7571	9	0.7592	151
DREP ^[22]	0.7320	29	0.4695	116	0.7551	105	0.7645	114

表9(续)

	FER2013		JAFFE		KDEF		CK +	
	准确率	数量	准确率	数量	准确率	数量	准确率	数量
Kappa ^[23]	0.713 0	196	0.469 5	192	0.716 3	231	0.758 6	230
QS ^[24]	0.709 4	206	0.446 0	174	0.716 3	231	0.758 6	230
SDAcc ^[25]	0.732 5	18	0.511 7	129	0.757 1	57	0.769 7	142
3WDEP	0.736 7	27	0.516 4	134	0.765 3	49	0.770 4	149

4 结 语

本研究提出基于三支决策的表情识别集成剪枝算法(3WDEP),借鉴三支决策理念引入延迟决策,以解决分类器评估的不确定性问题。同时提出预测偏好概念,通过综合考虑预测结果与实际标签的相关性、准确率和召回率,建立分类器选择信息系统,全面评估分类器性能。结合熵权法确定指标权重,权衡各分类器在不同决策选项下的损失,筛选出对集成系统贡献最大的分类器进行集成,有效提升表情识别系统性能。此外,还将召回率视为收益和代价属性进行综合考虑,优化集成剪枝效果。本文提出的集成剪枝方法具有较强的通用性,除了表情识别领域,在医疗和金融等领域也具有应用潜力。

参 考 文 献

- [1] 赵晓芳,金志刚. 融合表情符号与短文本的微博多维情感分类[J]. 哈尔滨工业大学学报, 2020, 52(5): 113
ZHAO Xiaofang, JIN Zhigang. Multi-dimensional sentiment classification of microblog based on emoticons and short texts[J]. Journal of Harbin Institute of Technology, 2020, 52(5): 113. DOI: 10.11918/201907004
- [2] SUDHA S S, SUGANYA S S. On-road driver facial expression emotion recognition with parallel multi-verse optimizer (PMVO) and optical flow reconstruction for partial occlusion in internet of things (IoT)[J]. Measurement, 2023, 26: 100711. DOI: 10.1016/j.measen.2023.100711
- [3] WEI S. Dynamic facial expressions on virtual humans for virtual reality (VR) mental health therapy[C]//2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 2022: 926. DOI: 10.1109/VRW55335.2022.00314
- [4] LI Xi, XIAO Zhenhua, LI Chao, et al. Facial expression recognition network with slow convolution and zero-parameter attention mechanism[J]. Optik, 2023, 283: 170892. DOI: 10.1016/j.ijleo.2023.170892
- [5] GAN Yanling, CHEN Jingying, XU Luhui. Facial expression recognition boosted by soft label with a diverse ensemble[J]. Pattern Recognition Letters, 2019, 125: 105. DOI: 10.1016/j.patrec.2019.04.002
- [6] PEI Ercheng, HU Zhanxuan, HE Lang, et al. An ensemble learning-enhanced multitask learning method for continuous affect recognition from facial images[J]. Expert Systems with Applications, 2023: 121290. DOI: 10.1016/j.eswa.2023.121290
- [7] LI Danyang, WEN Guihua. MRMR-based ensemble pruning for facial expression recognition[J]. Multimedia Tools and Applications, 2018, 77(12): 15251. DOI: 10.1007/s11042-017-5105-z
- [8] LI Danyang, WEN Guihua, HOU Zhi, et al. RTCRelief-F: An effective clustering and ordering-based ensemble pruning algorithm for facial expression recognition[J]. Knowledge and Information Systems, 2019, 59(1): 219. DOI: 10.1007/s10115-018-1176-z
- [9] HUANG Shisong, LI Danyang, ZHANG Zhuhong, et al. CSLSEP:

- An ensemble pruning algorithm based on clustering soft label and sorting for facial expression recognition[J]. Multimedia Systems, 2023, 29(3): 1463. DOI: 10.1007/s00530-023-01062-5
- [10] JIN Chengyong, HU Baoqing. Three-way decisions based on hesitant sets over three-way decision spaces[J]. Information Sciences, 2023: 119365. DOI: 10.1016/j.ins.2023.119365
 - [11] YAO Yiyu. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180(3): 341. DOI: 10.1016/j.ins.2009.09.021
 - [12] TEJA B, ANITA C, RAJALAKSHMI D, et al. A CNN based facial expression recognizer[J]. Materials Today: Proceedings, 2021, 37: 2578. DOI: 10.1016/j.matpr.2020.08.501
 - [13] JIA Fan, LIU Peide. A novel three-way decision model under multiple-criteria environment[J]. Information Sciences, 2019, 471: 29. DOI: 10.1016/j.ins.2018.08.051
 - [14] YE Jin, ZHAN Jianming, XU Zeshui. A novel decision-making approach based on three-way decisions in fuzzy information systems[J]. Information Science, 2020, 541: 362. DOI: 10.1016/j.ins.2020.06.050
 - [15] YANG Bin, HU Baoqing. Fuzzy neighborhood operators and derived fuzzy coverings[J]. Fuzzy Sets and Systems, 2019, 370: 1. DOI: 10.1016/j.fss.2018.05.017
 - [16] ZHANG Kai, DAI Jianhua, ZHAN Jianming. A new classification and ranking decision method based on three-way decision theory and TOPSIS models[J]. Information Sciences, 2021, 568: 54. DOI: 10.1016/j.ins.2021.03.039
 - [17] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: A report on three machine learning contests[J]. Neural Networks, 2015, 64: 59. DOI: 10.1016/j.neunet.2014.09.005
 - [18] LYONS M, AKAMATSU S, KAMACHI M, et al. Coding facial expressions with Gabor wavelets[C]//Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998: 200. DOI: 10.1109/AFGR.1998.670949
 - [19] LUCEY P, FCOHN J, KANADE T, et al. The extended cohn-kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2010: 94. DOI: 10.1109/CVPRW.2010.5543262
 - [20] GOELEN E, DERAEDT R, Leyman L, et al. The Karolinska directed emotional faces: A validation study[J]. Cognition & Emotion, 2008, 22(6): 1094. DOI: 10.1080/02699930701626582
 - [21] DAI Qun, ZHANG Ting, LIU Ningzhong. A new reverse reduce-error ensemble pruning algorithm[J]. Applied Soft Computing, 2015, 28: 237. DOI: 10.1016/j.asoc.2014.10.045
 - [22] LI N, YU Y, ZHOU Z H. Diversity regularized ensemble pruning[C]//Proceedings of the 2012 European Conference on Machine Learning and Knowledge Discovery in Databases - Part I. Berlin: Springer, 2012: 330. DOI: 10.1007/978-3-642-33460-3_27
 - [23] WANG Zhenlei, ZHAO Suyun, LI Zheng, et al. Ensemble selection with joint spectral clustering and structural sparsity[J]. Pattern Recognition, 2021, 119: 108061. DOI: 10.1016/j.patcog.2021.108061
 - [24] LI Danyang, WEN Guihua, LI Xu, et al. Graph-based dynamic ensemble pruning for facial expression recognition[J]. Applied Intelligence, 2019, 49(9): 3188. DOI: 10.1007/s10489-019-01435-2
 - [25] DAI Qun, YE Rui, LIU Zhuan. Considering diversity and accuracy simultaneously for ensemble pruning[J]. Applied Soft Computing, 2017, 58: 75. DOI: 10.1016/j.asoc.2017.04.058

(编辑 丁晓清)