

# 控制与决策

Control and Decision

## 基于主动样本精选与跨模态语义挖掘的图像情感分析

张红斌, 石焱炜, 熊其鹏, 侯婧怡

引用本文:

张红斌, 石焱炜, 熊其鹏, 侯婧怡. 基于主动样本精选与跨模态语义挖掘的图像情感分析[J]. *控制与决策*, 2022, 37(11): 2949–2958.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.0622>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### [联合SENet异构层特征融合与集成学习的材质图像识别](#)

Material image recognition combining heterogeneous-layer feature fusion of SENet and ensemble learning  
*控制与决策*. 2022, 37(6): 1632–1642 <https://doi.org/10.13195/j.kzyjc.2020.1559>

#### [基于深度情感唤醒网络的多模态情感分析与情绪识别](#)

Deep emotional arousal network for multimodal sentiment analysis and emotion recognition  
*控制与决策*. 2022, 37(11): 2984–2992 <https://doi.org/10.13195/j.kzyjc.2021.0782>

#### [一种基于多层语义特征的图像理解方法](#)

An image understanding method based on multi-level semantic features  
*控制与决策*. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

#### [基于卷积神经网络的云雾遮挡舰船目标识别](#)

Obscured ship target recognition based on convolutional neural network  
*控制与决策*. 2021, 36(3): 661–668 <https://doi.org/10.13195/j.kzyjc.2019.0781>

#### [基于联合知识表示学习的多模态实体对齐](#)

Multi-modal entity alignment based on joint knowledge representation learning  
*控制与决策*. 2020, 35(12): 2855–2864 <https://doi.org/10.13195/j.kzyjc.2019.0331>

# 基于主动样本精选与跨模态语义挖掘的图像情感分析

张红斌<sup>†</sup>, 石皞炜, 熊其鹏, 侯婧怡

(华东交通大学 软件学院, 南昌 330013)

**摘要:** 图像情感分析是机器视觉领域的研究热点,它面临的关键问题是:标注者的主观差异导致情感标签明确的高质量样本匮乏,且异构图像特征间跨模态语义未有效利用.为此,提出基于主动样本精选与跨模态语义挖掘的图像情感分析模型 ASRF<sup>2</sup>(active sample refinement & feature fusion):融合主动学习与样本精选思想,设计主动样本精选策略,优选情感标签明确的样本;对异构图像特征执行判别相关分析,生成能准确刻画图像情感内容的低维跨模态语义;采用跨模态语义训练 Catboost 模型,实现图像情感分析.在 Twitter I 与 FI 数据集上验证 ASRF<sup>2</sup> 模型,识别准确率分别达 90.06% 和 75.77%,优于主流基线且实时效率良好.与基线相比,ASRF<sup>2</sup> 模型仅需两类特征,参数调制简单,更易复现.ASR 策略还具备一定的泛化性,可为基线模型提供优质训练样本,以改善识别性能.

**关键词:** 主动学习; 样本精选; 跨模态语义; 图像情感分析; 判别相关分析; Catboost

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.0622

开放科学(资源服务)标识码(OSID):



**引用格式:** 张红斌,石皞炜,熊其鹏,等.基于主动样本精选与跨模态语义挖掘的图像情感分析[J].控制与决策,2022,37(11):2949-2958.

## Image sentiment analysis via active sample refinement and cross-modal semantics mining

ZHANG Hong-bin<sup>†</sup>, SHI Hao-wei, XIONG Qi-peng, HOU Jing-yi

(School of Software, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** Image sentiment analysis is a research focus in the field of computer vision. However, we are faced with the following key problems: First, owing to the subjective differences of different annotators, high-quality samples with definite sentimental annotations are very scarce. Second, the implicit cross-modal semantics among heterogeneous features has not been fully explored. To address these two problems, we propose an active sample refinement & feature fusion (ASRF<sup>2</sup>) via active sample refinement and cross-modal semantics mining: an active sample refinement strategy is designed by fusing the active learning and sample refinement ideas. High-quality samples with definite sentimental annotations are obtained in turn. Then, the state-of-the-art discriminant correlation analysis (DCA) algorithm is employed to fully mine the cross-modal correlations among the heterogeneous features. Low-dimensional but more discriminant cross-modal semantics that can better depict the key sentimental contents of images are generated. The cross-modal semantics is used to train a Catboost classifier and complete image sentiment analysis. We validate the proposed ASRF<sup>2</sup> model on the Twitter I and FI datasets. The corresponding accuracies reach about 90.06% and 75.77%, respectively, which outperform other state-of-the-art baselines as well as the real-time efficiency. Compared with the baselines, the proposed model only needs two image features, and it is easy to tune and reproduced the ASRF<sup>2</sup> model. Moreover, the ASR strategy is robust, which can offer many more high-quality samples for the baselines to improve the final recognition performance.

**Keywords:** active learning; sample refinement; cross-modal semantics; image sentiment analysis; discriminant correlation analysis; Catboost

## 0 引言

伴随社交网络日益流行,各类社交平台(如知乎、微信、微博等)应运而生.人们在社交平台上上传图片

以表达个人情感,因此平台拥有海量图像数据,对图像进行有针对性地情感分析能更好地了解人们的内在情绪或对特定事件的真实看法,进而实现更多有

收稿日期: 2021-04-12; 录用日期: 2021-07-30.

基金项目: 国家自然科学基金项目(61762038, 61861016); 江西省研究生创新专项项目(YC2020-S352).

责任编辑: 柴利.

<sup>†</sup>通讯作者. E-mail: zhanghongbin@whu.edu.cn.

价值的应用<sup>[1-3]</sup>。例如,预测人们是否存在心理障碍并及时干预、预测某类商品在用户心目中的满意度、预测某个物品的流行趋势以及跟踪时下的舆情倾向等。所以,图像情感分析的研究对于学术界和工业界都具有十分重要的意义。

早期的图像情感分析多采用手工特征,例如,Lu等<sup>[4]</sup>通过分析图像中形状特征如何影响人的情感来量化情感程度。然而,单一特征很难完整地描述图像中的情感内容。由于深度学习特征具有更强的表征能力<sup>[5]</sup>,它们被证明优于手工特征<sup>[6]</sup>。晁浩等<sup>[7]</sup>提取深度置信网络特征并采用判别式受限玻尔兹曼机完成脑图情感分析。Peng等<sup>[8]</sup>训练一个全卷积网络(fully convolutional networks with euclidean loss, FCNEL),用人工标记的真值预测情绪刺激图。Yang等<sup>[9]</sup>通过标签分布解决图像中的情感模糊问题,并联合分类和预测分布设计多任务学习框架。Zhao等<sup>[10]</sup>设计一个带权重的多模态条件概率神经网络,实现图像情感分析。为充分利用图像中局部区域所呈现的情感表征。Sun等<sup>[11]</sup>生成对象建议并使用神经网络进行评分,以发现图像中情感语义强烈的区域。She等<sup>[12]</sup>提出弱监督耦合卷积网络(weakly supervised coupled networks, WSCNet),该网络在给出弱标签后自动选择相关区域,以减少情感标注代价。Zhu等<sup>[13]</sup>提出软建议(soft proposal, sp)组件,其可插入到任意标准的卷积架构中完成区域定位。Wu等<sup>[14]</sup>提出一种增强卷积神经网络(convolutional neural network, CNN)提升情感分类效果。Durand等<sup>[15]</sup>设计WILDCAT(weakly supervised learning of deep convNets)模型对图像区域进行对齐,以获得空间不变性和局部特征。基于跨模态数据的情感分析逐渐成为近年的热点。Fan等<sup>[16]</sup>研究注意力对图像情感分析的影响,以探寻图像区域的相对重要性。You等<sup>[17]</sup>提出跨模态连贯性回归(cross-modality consistency regression, CCR)模型,以提取两种异构模态间的隐含情感语义。

虽然深度学习模型给图像情感分析带来积极影响,但它们通常需要大量标注样本。现有图像情感数据集均由人工标注,标注者的主观差异导致图像情感语义模糊,进而制约样本数量。由于高质量样本匮乏,深度学习模型易过拟合,进而影响模型分类性能,有研究者提出采用领域适配缓解样本稀缺问题<sup>[18-19]</sup>,主要基于生成对抗网络(generative adversarial networks, GAN)生成新的图像样本。然而,GAN模型训练过程较为复杂,且需要大量计算资源。

为降低模型复杂性,Zhang等<sup>[20]</sup>提出静态样本精选算法,并挖掘异构图像特征间的相关性,完成图像情感预测。

综上,现有研究存在“情感标签明确的高质量样本匮乏”“异构图像特征间的跨模态相关性未有效利用”等关键问题。为此,本文提出ASRF<sup>2</sup>(active sample refinement & feature fusion)模型,致力于高质量样本精选与深度特征融合,以改善图像情感分析精度。同时,拟进一步降低模型对资源的依赖,从而提高模型实用性。本文主要贡献如下:

1) 提出主动样本精选(active sample refinement, ASR)策略,精选情感标签明确的高质量样本,为训练优良的图像情感分析模型奠定坚实的数据基础。该精选策略既是对现有数据增强方法的有益补充,又可配合主流图像情感分析模型,改善其分类精度,故ASR策略具备一定泛化性。

2) 提出基于主动样本精选与跨模态语义挖掘的图像情感分析模型ASRF<sup>2</sup>,模型主要特点是:①识别精度优于主流基线;②所需特征更少且参数调制便捷,故模型易于复现;③模型实时效率较好,可部署在相应终端以对接实际应用;④变种模型表现优异,模型鲁棒性较强。

## 1 模型描述

### 1.1 模型框架

ASRF<sup>2</sup>模型由3部分组成:主动样本精选、跨模态语义挖掘和图像情感分类。基于FI(flickr and instagram)数据集展示所提出模型框架,如图1所示。首先设计主动样本精选策略,对候选样本池执行样本精选。该策略分为粗精选和细精选两步,样本精选后将生成新的、更完备的数据集,细节参见第1.2.1节;其次提取图像特征,ASRF<sup>2</sup>模型仅需提取SIFT(S)和VGG19(V19)特征;然后对两个特征执行判别相关分析<sup>[21]</sup>(discriminant correlation analysis, DCA),生成跨模态语义,细节见第1.2.2节;最后采用跨模态语义训练Catboost<sup>[22]</sup>模型,细节见第1.2.3节,输出图像情感标签。

### 1.2 模型构成

#### 1.2.1 主动样本精选

现有图像情感分析数据集中缺少情感标签明确的高质量样本,故对待精选样本执行主动样本精选,以获得优质图像样本,为模型训练奠定坚实的数据基础。所提出主动样本精选(ASR)策略分为粗精选和细精选两步。

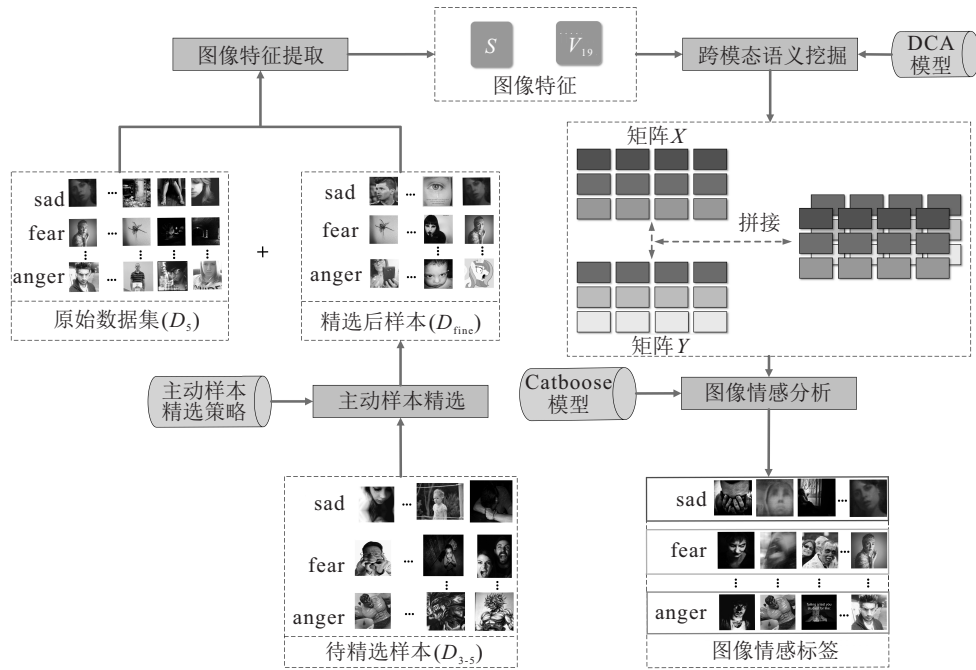


图 1 ASRF<sup>2</sup> 模型框架

step 1: 粗精选. 在公开数据集中,  $D_5$  表示有 5 名 AMT(amazon mechanical turk) 员工对同一张图像标记相同情感标签,  $D_3$  表示至少有 3 名 AMT 员工对同一张图像标记相同情感标签,  $D_5$  为  $D_3$  的高量子集.  $D_{3-5}$  表示从  $D_3$  中选出除  $D_5$  外的图像样本, 它是 ASRF<sup>2</sup> 模型的候选样本池. 当然,  $D_{3-5}$  中也存在情感模糊的图像, 主动样本精选的目的便是自动地从  $D_{3-5}$  中选出高质量的、情感语义明确的图像, 为训练模型奠定坚实数据基础. 故首先基于  $D_3$  和  $D_5$  生成待精选样本  $D_{3-5}$ ; 然后以  $D_5$  为训练集,  $D_{3-5}$  为测试集, 设计粗精选算法, 即从  $D_{3-5}$  中初选出优质样本. 为防止选择性偏倚, 综合  $N$  个互补的分类器评估  $D_{3-5}$  中的样本. 令  $N = 1, 2, \dots, 9$ , Twitter I 中使用 LR(linear regression)、RF(random forest) 等 9 个分类器, FI 中使用 KNN( $K$ -nearest neighbor)、DT(decision tree) 等 5 个分类器. 为提升精选精度和算法效率, 所有样本均用跨模态特征刻画,  $D_{3-5}$  和  $D_5$  数据的跨模态特征表示分别为  $F_{3-5}$ 、 $F_5$ . 当所有分类器预测值  $P_{3-5}$  与样本真实标签一致时, 粗选出该样本, 产生数据集  $D_{coarse}$  以备后续细精选.

step 2: 细精选. 对  $D_{coarse}$  数据集做进一步精化, 输出情感语义更明确的高质量样本. 以  $D_5$  为训练集,  $D_{coarse}$  为测试集, 在细精选中融入主动学习策略, 这是一个迭代过程, 每次迭代使用分类器对  $D_{coarse}$  中的样本进行预测, 得到  $D_{coarse}$ , 然后以排序批处理模式<sup>[23]</sup>(ranked batch-mode, RBM) 作为查询函数, 筛

选出样本, 样本的每个预测值会由下式进行评分:

$$score = \partial(1 - \Phi(x, D_5)) + (1 - \partial)U(x). \quad (1)$$

其中: 权重  $\partial = D_5 / (|D_{3-5}| + |D_5|)$ .  $U(x)$  为样本  $x$  预测值的不确定性;  $\Phi$  为相似度函数, 使用欧氏距离计算相似度.

式(1)前半部计算样本  $x$  与训练样本之间的相似度, 后半部评价特征空间在样本  $x$  附近的分布情况. 经过排序, 评分高的样本置于列表顶部. 每次迭代选取列表中评分最高的  $M$  个样本, 对比其预测值与真实标签, 若完全一致, 则将该样本选出并置入训练集中, 然后进入下一轮训练. 本文将参数  $M$  设为 3, 经过多轮迭代, 主动样本精选模型自适应地加入了许多有价值的知识, 使其更高效、鲁棒并具备自主“学习”能力. 为防止选择性偏倚, 采用  $Z$  个分类器完成主动样本精选(在 Twitter I 数据集中  $Z = 2$ , 在 FI 数据集中  $Z = 4$ . 具体见第 2.3.2 节), 生成一组候选的高质量样本集合, 对这些样本集合分别执行交集和并集处理, 输出高质量样本  $D_{fine}$ . 采用  $D_{fine}$  增强原数据集  $D_5$ , 训练全新的图像情感分析模型. 综上, 主动样本精选策略如算法 1 所示.

**算法 1** 主动样本精选(ASR).

输入: 原始数据集  $D_3$ 、 $D_5$ ;

输出: 精化后的样本集合  $D_{fine}$ .

1. 基于  $D_3$ 、 $D_5$  生成  $D_{3-5}$  数据集.

2. 分别提取  $D_{3-5}$  与  $D_5$  数据集的 SIFT、VGG19 特征.

3. 基于判别相关分析(第1.2.2节)完成特征融合,生成 $F_{3-5}$ 和 $F_5$ .

4. 用 $N$ 个分类器和 $F_5$ 训练模型并对 $F_{3-5}$ 测试,得到预测值 $P_{3-5}$ .

5. repeat

比较每个分类器预测值与真实标签,若所有预测值均一致,则将样本加入 $D_{\text{coarse}}$ .

6. until 遍历完 $P_{3-5}$ 中全部预测值.

7. 提取 $D_{\text{coarse}}$ 的SIFT和VGG19特征.

8. 基于判别相关分析(第1.2.2节)完成特征融合,生成 $F_{\text{coarse}}$ .

9. repeat

用一个分类器和 $F_5$ 训练模型,并对 $F_{\text{coarse}}$ 测试,得到 $P_{\text{coarse}}$ ,基于式(1)对 $P_{\text{coarse}}$ 排序,选择评分最高的 $M$ 个样本,比较每个样本预测值与真实标签,若一致则将它加入训练集,将当前 $M$ 个样本从测试集中移除.

10. until 到达循环次数或 $P_{\text{coarse}}$ 遍历完毕.

11. 重复第10~15步,基于 $Z$ 个分类器分别生成细精选样本集合.

12. 对各分类器选出的样本集合执行交集和并集运算,得到 $D_{\text{fine}}$ .

### 1.2.2 跨模态语义挖掘

如图1所示,ASRF<sup>2</sup>模型仅使用SIFT和VGG19这两个特征刻画图像情感内容,SIFT是描述图像中形态变化的手工特征,而VGG19属于深度学习特征,主要刻画图像中深层语义.由于提取方式、视觉角度等不同,它们彼此异构,但两者间蕴涵非常丰富的跨模态语义,相比单一特征,该跨模态语义能更准确、全面地描述图像情感内容.故在主动样本精选的基础上,ASRF<sup>2</sup>模型对SIFT和VGG19这两个特征执行判别相关分析(DCA),并完成特征融合.DCA通过最大化特征矩阵间的相关性生成中间层语义,有效挖掘异构特征间相关性,生成判别性更强且低维跨模态语义,为训练分类模型奠定重要基础.跨模态语义挖掘流程如图2所示.

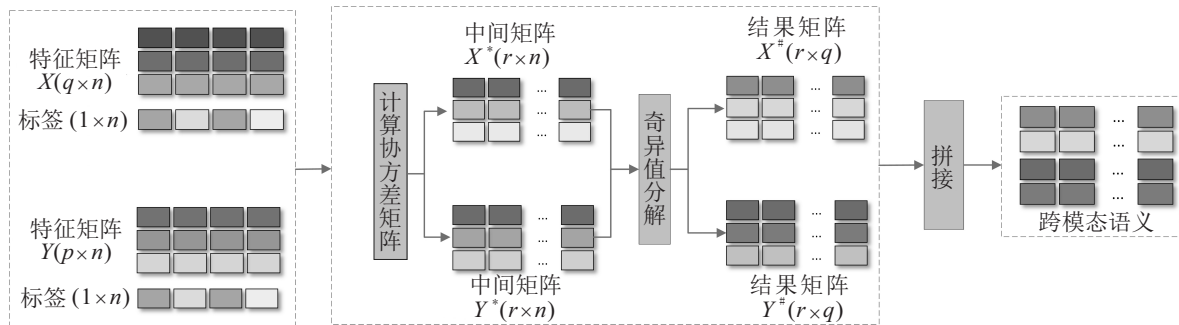


图2 基于判别相关分析挖掘两个特征间的跨模态语义

针对特征矩阵 $X$ 和 $Y$ ,首先根据样本标签计算协方差矩阵 $S_{bx}$ 和 $S_{by}$ ,利用 $S_{bx}$ 和 $S_{by}$ 求得特征向量 $P$ 并取其主分量.基于 $S_{bx}$ 和 $S_{by}$ 计算转换矩阵 $W_{bx}$ 和 $W_{by}$ ,并与原特征矩阵相乘,得到中间矩阵 $X^*$ 和 $Y^*$ .基于 $X^*$ 和 $Y^*$ 生成 $S_{xy}$ ,对其进行奇异值分解,得到转换矩阵 $W_{cx}$ 和 $W_{cy}$ .用这两个转换矩阵与原特征矩阵相乘,生成结果矩阵 $X^\#$ 和 $Y^\#$ .对 $X^\#$ 和 $Y^\#$ 执行拼接操作,生成跨模态语义.最终,结果矩阵维度为 $r$ ,具体值为 $c-1$ , $c$ 为情感类别数.以FI数据集为例,输入的SIFT与VGG19特征维度分别为500、4096,执行判别相关分析后,特征维度降为7维(类别数减1,即 $8-1=7$ ),特征复杂度大幅降低,有助于提高模型实时效率,进而增强其实用性.

### 1.2.3 Catboost模型

ASRF<sup>2</sup>模型使用Catboost分类器完成图像情感预测.CatBoost是提升类分类算法,它将对称树

作为基础预测器,这种树是平衡的且不易过拟合.Catboost还提供GPU接口并支持多GPU操控,大幅提高了计算效率.

## 2 实验结果及讨论

为验证模型的有效性,选择数据集Twitter I<sup>[24]</sup>和FI<sup>[25]</sup>完成图像情感分析实验.随机选取70%的数据作为训练集,剩余30%为测试集.电脑配置:CPU为Intel(R) Core(TM)i7-8550,显卡MX150,内存8G.

Twitter I由社交网站收集,You等<sup>[24]</sup>为每张候选图像招募5名AMT员工. $D_5$ 表示5名AMT员工对同一张图像给出相同情感标签, $D_4$ 表示至少有4名AMT员工对同一张图像给出相同情感标签, $D_3$ 表示至少有3名AMT员工对同一张图像给出相同情感标签,本文需要用到 $D_5$ 和 $D_3$ 数据.Twitter I数据集分正面(Positive)和负面(Negative)两类,它是一种粗粒度数据集,共57.5 MB.

FI 通过在社交网站上搜索 8 种情绪关键字: anger、amusement、awe、contentment、disgust、excitement、fear、sadness 得到, 各类对应图像数分别为: 1 168、4 635、2 851、5 056、1 562、2 687、946 和 2 603, 共计 21 508 张图像。FI 是一种细粒度数据集, 共约 1.5 G。

执行算法 1 对 Twitter I 和 FI 进行样本精选后的详细信息如表 1 所示。

表 1 Twitter I 和 FI 数据集相关信息

数据集	$D_5$	$D_4$	$D_3$	$D_{\text{fine}}(\text{I})$	$D_{\text{fine}}(\text{U})$
Twitter I	882	1 116	1 269	151	188
FI	5 238	12 644	21 508	301	610

表 1 中, I 和 U 分别表示对精选后的样本执行交集或并集操作。将  $D_{\text{fine}}$  加入对应的  $D_5$  中, 以增强原数据集, ASRF<sup>2</sup> 模型变种包括: ASRF<sup>2</sup>(I)、ASRF<sup>2</sup>(U)。将 ASRF<sup>2</sup> 模型与如下 3 类基线进行比较:

1) 微调后的深度学习网络: VGG16<sup>[26]</sup>、ResNet101<sup>[27]</sup>、AlexNet<sup>[28]</sup> 和 CAM-Res101<sup>[29]</sup>;

2) 特征融合类模型: 典型相关分析 (CCA)<sup>[30]</sup>、稀疏核典型相关分析 (gradKCCA)<sup>[31]</sup>、判别相关分析 (DCA)<sup>[21]</sup>、GS-XGB<sup>[32]</sup>、GS-Ada<sup>[32]</sup>、GS-LR<sup>[32]</sup>、GS-SVM<sup>[32]</sup> 和 GS-RF<sup>[32]</sup> 模型;

3) 主流图像情感分析模型: WILDCAT<sup>[15]</sup>、SPN<sup>[13]</sup>、WSCNet<sup>[12]</sup>、Sun 的模型<sup>[11]</sup>、Yang 的模型<sup>[9]</sup>、CycleEmotionGAN<sup>[18]</sup>、MSGAN<sup>[19]</sup>、ME2M(M)<sup>[20]</sup> 和 SR-w-DCA<sup>[20]</sup>。

## 2.1 定量实验结果及分析

使用准确率评估模型性能。首先, 将 ASRF<sup>2</sup> 与所有基线进行比较, 实验结果如表 2 所示。表 2 中: F 表示融合类模型, D 表示深度学习模型, B 表示主流模型。括号内标注  $D_5$  表示使用  $D_5$  评估, 标注  $D_5$ +ASR 表示使用  $D_5$  和 ASR 增强后的样本评估, 该类实验着力验证 ASR 策略的泛化性, 其余用  $D_3$  评估。如表 2 所示, ASRF<sup>2</sup> 模型在两个数据集上均表现优异。下面从 3 个角度进行对比分析:

1) 与微调后的深度学习网络相比, ASRF<sup>2</sup> 模型的性能在两个数据集上均有较大幅度提升。例如, 在 Twitter I 数据集上, ASRF<sup>2</sup> 相对最强深度学习网络 (VGG16 ( $D_5$ )) 提升  $90.06\% - 86.56\% = 3.5\%$ 。深度学习网络复杂且高质量样本匮乏, 故它们易过拟合, 尤其当执行细粒度图像情感分析时, 深度学习网络需更多优质样本, 过拟合问题也愈发严重。相反, ASRF<sup>2</sup> 模型能较好地应对上述问题。一方面, 它通过主动样本

精选获取情感语义明确的高质量样本, 以增强现有数据集; 另一方面, 它采用判别性更强且低维跨模态语义刻画图像情感内容, 从而更好地匹配样本数, 避免模型过拟合。此外, 如图 1 所示, ASRF<sup>2</sup> 模型的计算过程更透明且调制简单, 便于快速复现。若将 ASR 策略增强的数据与  $D_5$  相混合, 则会在 Twitter I 数据集上出现性能衰减, 而在 FI 数据集则有较大幅度性能提升。这表明: ASR 更多地从细粒度角度丰富现有数据, 而从粗粒度角度划分, 增强的样本与已有样本在情感语义上趋于一致, 故未能对性能提升有积极影响。因此, ASR 策略在 FI 数据集上的泛化性更好, 它能生成更多高质量的细粒度样本。

2) 与典型特征融合类模型相比, 在 Twitter I 数据集上, ASRF<sup>2</sup> 模型比最强的 GS-SVM 模型提高 1.34%。由于均采用特征融合思想, 主动样本精选策略在 ASRF<sup>2</sup> 模型中扮演更重要的角色: 它着力选出高质量样本, 以增强现有数据集, 这也有助于更好地提取 SIFT 和 VGG 特征, 以及挖掘它们之间蕴含的跨模态语义。相反, GS 类模型需在执行较复杂的特征中融合, CCA 模型生成的新特征维度偏高, 不利于实时预测, gradKCCA 模型则需要计算多个核矩阵, 对应复杂度偏高。ASRF<sup>2</sup> 仅需两类特征且特征融合方式更简单 (一次融合即可), 这有助于提升模型实用性。此外, 在 FI 数据集上, 为融合类模型引入 ASR 精选的样本, 可进一步改善分类性能, 这充分表明 ASR 策略在细粒度数据集上具有更强的泛化性, 即它能为融合类模型提供高质量的细粒度样本, 为训练优质模型奠定重要的数据基础, 这与深度学习模型的分析相吻合。

3) 与主流图像情感分析模型相比, ASRF<sup>2</sup> 模型也表现优异。在 Twitter I 数据集上, 它比最强基线提升 0.66%, 在 FI 数据集上略优于 SR-w-DCA 模型。ASRF<sup>2</sup> 模型的两大变种也表现优异, 其中并集操作优于交集操作 (见表 1, 并集操作精选出更多情感标签明确的样本, 为训练模型奠定重要的数据基础), 故 ASRF<sup>2</sup> 模型具有较强鲁棒性。相反, SPN、WILDCAT 等基线着力寻找图像中情感语义强烈的区域, 它们依赖额外人工标注, 即通过标注将粗粒度图像转为细粒度图像, 以提升分类性能。而 ASRF<sup>2</sup> 模型关注全局图像, 无需额外人工标注, 节省了人力成本。此外, SPN、WILDCAT 等模型基于深度学习网络, 需要更多的计算资源, 而 ASRF<sup>2</sup> 模型在普通配置的计算机上即可获取有竞争力的分类精度和实时性能。ASRF<sup>2</sup> 模型还优于 CycleEmotionGAN、MSGAN 等基于数据增强的基线, 不同于这些模型, ASRF<sup>2</sup> 不生成新样本, 模

表2 模型分类精度

数据集	模型	类型	准确率/%	模型	类型	准确率/%	模型	类型	准确率/%
Twitter I	VGG16	D	76.75	GS-LR( $D_5$ +ASR)	F	84.51	SPN	B	81.67
	AlexNet	D	73.24	GS-RF( $D_5$ +ASR)	F	82.39	Sun et al. ( $D_5$ )	B	88.94
	Res101	D	78.13	GS-XGB( $D_5$ +ASR)	F	83.10	ME <sup>2</sup> M(A)	B	85.30
	CAM-Res101	D	82.67	GS-Ada( $D_5$ +ASR)	F	84.51	ME <sup>2</sup> M(M)	B	87.15
	VGG16( $D_5$ )	D	86.56	GS-SVM( $D_5$ +ASR)	F	82.04	CycleEmotionGAN	B	62.38
	VGG16( $D_5$ +ASR)	D	82.69	CCA( $D_5$ )	F	80.08	MSGAN	B	63.58
	GS-LR( $D_5$ )	F	87.22	gradKCCA( $D_5$ )	F	77.07	WSCNet	B	84.25
	GS-RF( $D_5$ )	F	84.21	DCA( $D_5$ )	F	87.59	WSCNet( $D_5$ )	B	87.62
	GS-XGB( $D_5$ )	F	86.47	WILDCAT	B	79.53	WSCNet( $D_5$ +ASR)	B	89.40
	GS-Ada( $D_5$ )	F	85.34	WILDCAT( $D_5$ )	B	71.43	ASRF <sup>2</sup> (I)	/	88.65
GS-SVM( $D_5$ )	F	88.72	WILDCAT( $D_5$ +ASR)	B	73.85	ASRF <sup>2</sup> (U)	/	<b>90.06</b>	
FI	VGG16	D	63.75	GS-LR( $D_5$ +ASR)	F	75.30	SPN	B	66.57
	AlexNet	D	58.13	GS-RF( $D_5$ +ASR)	F	75.31	Zhu et al.	B	73.00
	Res101	D	66.16	GS-XGB( $D_5$ +ASR)	F	75.98	Yang et al. ( $D_5$ )	B	67.48
	CAM-Res101	D	68.54	GS-Ada( $D_5$ +ASR)	F	74.27	SR-w-DCA	B	75.72
	Res101( $D_5$ )	D	75.61	GS-SVM( $D_5$ +ASR)	F	74.79	CycleEmotionGAN	B	67.78
	Res101( $D_5$ +ASR)	D	<b>78.11</b>	CCA( $D_5$ )	F	50.29	MSGAN	B	70.63
	GS-LR( $D_5$ )	F	73.59	gradKCCA( $D_5$ )	F	61.84	WSCNet	B	70.07
	GS-RF( $D_5$ )	F	70.41	DCA( $D_5$ )	F	73.71	WSCNet( $D_5$ )	B	72.22
	GS-XGB( $D_5$ )	F	72.89	WILDCAT	B	67.03	WSCNet( $D_5$ +ASR)	B	74.72
	GS-Ada( $D_5$ )	F	72.51	WILDCAT( $D_5$ )	B	70.09	ASRF <sup>2</sup> (I)	/	74.32
GS-SVM( $D_5$ )	F	73.52	WILDCAT( $D_5$ +ASR)	B	72.23	ASRF <sup>2</sup> (U)	/	75.77	

型训练更简单. 虽然ME<sup>2</sup>M(M)、SR-w-DCA等模型也采用样本精选策略,但它们仅执行一次样本精选,精选过程是相对静态的,故选出的样本质量一般.相反,ASRF<sup>2</sup>模型采用主动样本精选策略,它更主动、灵活,即通过多轮迭代优选出潜在的高质量样本,进而增强现有数据集,为训练优质的图像情感分析模型奠定重要的数据基础.此外,若在主流图像情感分析模型中添加精选的样本,则分类性能也会有显著提升,且在两个数据集上表现一致,进一步表明ASR策略具备一定的泛化性,能为主流基线提供高质量训练数据.当然,整体而言,ASRF<sup>2</sup>模型仍优于经过ASR数据增强的主流基线模型.

除准确率外,本节还对ASRF<sup>2</sup>模型进行实时效率评估,并与微调的深度学习模型VGG16<sup>[26]</sup>、Res101<sup>[27]</sup>以及主流基线ME<sup>2</sup>M(M)<sup>[20]</sup>进行对比. ASRF<sup>2</sup>模型的测试过程为:特征提取、判别相关分析和情感预测,故其预测一张图片平均需0.54 s,其中提取SIFT和VGG特征平均花费0.47 s和0.07 s,而判别相关分析与情感预测时间极短,可忽略不计.如表3所示:相比ME<sup>2</sup>M(M)模型,ASRF<sup>2</sup>的预测时间有所提升,具备较好的实时效率.

表3 模型预测单张图片所需时间

模型	VGG16	Res101	ME <sup>2</sup> M(M)	ASRF <sup>2</sup>
时间/s	0.053	0.098	2.17	0.54

综上,ASRF<sup>2</sup>模型优于主流基线,且模型具有如

下重要特性:简单、易于复现、实时效率较高、可解释性更好,这些都有助于增强模型的实用价值.

2.2 定性实验结果及分析

除上述定量分析外,本节通过定性实例证明ASRF<sup>2</sup>模型的有效性,具体如图3~图5所示,它们均为执行主动样本精选策略后得到的样本.



图3 Twitter I中正面情感的图像



图4 Twitter I中负面情感的图像

如图3和图4所示:ASRF<sup>2</sup>模型精选出的图像均富有极强的情感倾向,正面情感的图像色彩鲜明、主题明确,或包含人物笑脸(图3中间分图中生日宴会的主角非常开心,对这类图像执行情感分析有助于预测人们的即时情绪),或包含积极类物品(图3左分图中一双新的运动鞋蕴涵了积极的情感:兴奋或开心,对这类图像执行情感分析有助于预测物品流行度



图 5 FI 数据集中的图像

并进一步挖掘用户满意度). 相反, 负面情感的图像整体颜色偏灰暗(图 4 左分图中的工厂正在制造污染, 这是非常负面的, 对这类图像的情感分析有助于实时舆情监测), 或人脸呈现恐吓表情(图 4 右分图中面目狰狞的小丑, 这也是较负面情绪, 对它们的情感分析有助于及时干预心理疾病患者).

如图 5 所示: 旋转木马与摩天轮能正确对应 amusement 标签, 这是一种欢快的情绪. 而 sad 和 anger 标签则对应较灰暗的图像, 情绪较低落. 人脸常作为评判情感极性的重要标准(与图 3、图 4 类似); contentment 和 awe 等标签对应颜色较舒缓的图像; fear 中包含令人恐惧的表情. 由于 FI 是细粒度数据集, 对图 5 所示图像执行情感分析可获取更丰富、细腻的情感语义, 从而完成细粒度舆情跟踪、心理干预、热点预测等实际应用. 综上, 定性实验结果表明: ASRF<sup>2</sup> 模型能优选出高质量样本, 从而有效地增强现有数据集.

### 2.3 参数调制

#### 2.3.1 特征选取

对单图像特征进行评估, 包括: SIFT、LBP、GIST、DenseNet、VGG、ResNeXT 以及 ResNet. 采用主成分分析对传统特征进行降维, 降为 500 维. 分别取 DenseNet 中的 121 层和 161 层、ResNet 的 50 层和 152 层、VGG 的 16 层和 19 层、ResNeXT 的 50 层作为深度学习特征. 表 4 为 Twitter I 和 FI 数据集中  $D_5$  数据上的分类性能. 如表 4 所示: 在 Twitter I 数据集中, VGG19 表现最优, SIFT 次之; 在 FI 数据集中, VGG16 表现最优, VGG19 次之; 在全部传统特征中, SIFT 性能最优, 它主要刻画图像中局部形态变化, 图像背景与前景之间梯度变化大, 不同局部区域蕴涵不同的情感语义, 而 SIFT 特征能准确捕获这些变化, 从而刻画图像情感内容. 例如, 在图 3~图 5 中, 人面部的表情正体现了关键的局部形态特性. 相反, GIST、LBP 特征分别

描述图像中全局、局部纹理, 虽然它们也包含情感语义, 但相对局部形态特性其情感倾向偏弱. VGG 着力描述图像中深层语义, 它更关注情感强烈的完整区域, 如摩天轮、笑脸、破旧的房屋等均蕴涵非常显著的情感信号, 这些信号被 VGG 特征捕获, 进而准确刻画情感内容. 相反, ResNet、DenseNet 等对应的特征图维度偏低, 深层次判别信息偏少, 影响了最终的分类性能.

表 4 单图像特征的分类准确率

数据集	特征	准确率 / %	特征	准确率 / %
Twitter I	DenseNet121	67.17	SIFT	83.02
	DenseNet161	65.66	LBP	83.02
	ResNet50	70.19	GIST	83.02
	ResNet152	69.06	ResNeXT50	83.02
	VGG16	82.26	VGG19	<b>83.77</b>
FI	DenseNet121	29.13	SIFT	55.47
	DenseNet161	42.29	LBP	46.85
	ResNet50	57.44	GIST	50.51
	ResNet152	58.72	ResNeXT50	46.06
	VGG16	<b>71.18</b>	VGG19	70.99

此外, 表 4 中原始特征的维度偏高, 当高质量样本匮乏时易出现过拟合, 从而影响分类性能, 故应在保留特征判别性的前提下尽量降低特征维度, 而异构图像特征间的跨模态语义挖掘(见图 2)能够满足这一需求.

综上, SIFT 和 VGG 属于各类(传统、深度学习)特征中的最优者. 同时, 其抽取方式有较大差异, 可形成较强互补, 为跨模态语义挖掘做好准备; 再次, 这两类特征所需计算资源相对较少, 便于快速提取. 因此, 综合分类性能、特征间互补性及模型复杂度, ASRF<sup>2</sup> 模型仅选取 SIFT 和 VGG19 完成跨模态语义挖掘, 为图像情感分析提供判别性更强且低维的新特征, 有助于改善实时效率. 此外, 算法 1 也采用该类跨模态语义 ( $F_3$ 、 $F_{3-5}$  和  $F_{coarse}$ ), 实现更高效的主动样本精选.

### 2.3.2 细精选中分类器调制

在主动样本精选策略中,设置不同数量、不同类型的分类器可获得不同的精选样本及识别精度,故需对分类器类型和数量进行适当调制,以获取最优识别性能.在 Twitter I 数据集中,选用逻辑回归(LR)与 K-近邻(KNN)两大分类器完成主动样本精选.在 FI 数据集中,选用 Catboost、自适应提升树(Adaboost)、随机森林(RF)和逻辑回归(LR)完成主动样本精选.对各分类器精选出的样本执行并集运算(见表1),以增强现有数据集.继而训练 Catboost 分类器,获取测试集准确率,实验结果如图6所示.由图6(a)可见:在 Twitter I 数据集上,使用全部分类器执行主动样本精选时模型最优,这表明采用多个异构且互补的分类器实施样本精选可防止选择性偏倚,当然 Twitter I 数据集出现训练集准确率更高的情况,可能的原因是 Twitter I 中图像样本偏少,或训练集与测试集划分不均匀,因此对于小样本数据集的训练是一大挑战,未来拟引入少样本学习方法应对该问题.由图6(b)可见:与 Twitter I 一致,使用全部分类器执行主动样本精选时模型表现更好,且未出现过拟合,多个分类器间的互补性得以充分利用.

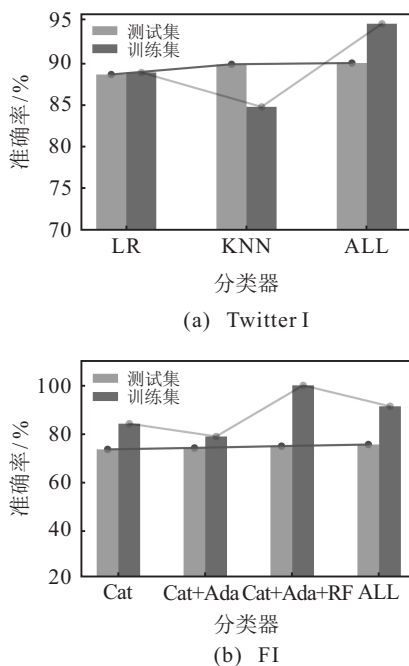


图6 主动样本精选策略中的分类器调制

### 2.3.3 Catboost参数调制

ASRF<sup>2</sup>模型使用 Catboost 分类器完成情感预测, Catboost 是一种鲁棒的分类器,包含若干参数,需对它们进行调制以获取最佳的图像情感分析性能.本文调节 Catboost 的4个参数: iteration(树的数量)、random\_seed(随机种子)、depth(树的深度)及 learning\_rate(学习率,用于减少梯度步长).其中 iteration、depth

更重要,故仅展示 iteration 与 depth 的调制过程.具体参数调制过程如下:采用网格搜索调制 iteration、depth 以及 learning\_rate,待其最优后再调制 random\_seed.实验结果如图7所示.为方便展示,图7中未采用全部 iteration,而是对其作量化再绘制对应图形.此时 learning\_rate 与 random\_seed 值分别为  
(Twitter I) learning\_rate = 0.9, random\_seed = 357;  
(FI) learning\_rate = 0.1, random\_seed = 336.

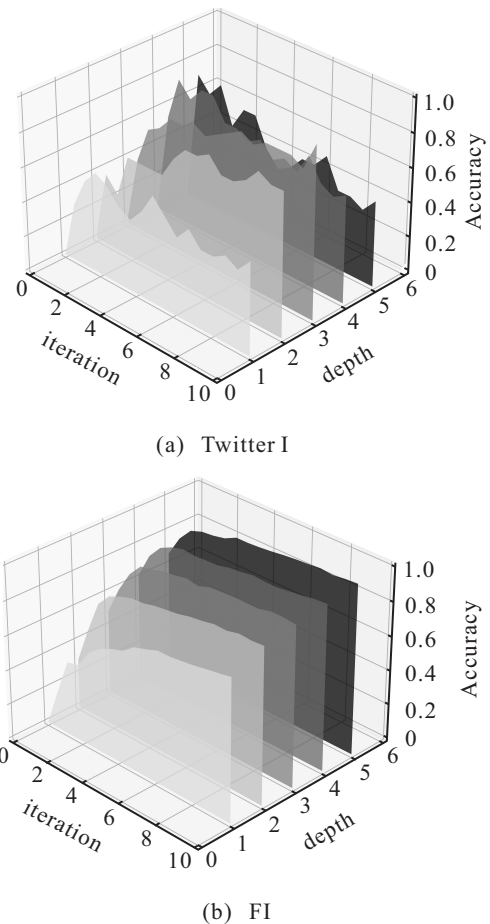


图7 Catboost参数调制的可视化展示

### 2.3.4 ASR策略的泛化性实证

为进一步论证 ASR 策略的泛化性,本节完成如下两类实验:

1) 选取微调的深度学习模型 VGG16 和 Res101 围绕 Twitter I 数据集的  $D_5$  数据依次执行传统数据增强和 ASR 策略.传统数据增强包括:旋转、裁剪、调整饱和度和等7种操作.实验过程如下:选取一种基底操作,从7种操作中随机挑选3种操作结合基底操作,共生成4张图像.故传统数据增强生成3528张图像,继续执行 ASR 策略,精选出2121张图像,以丰富  $D_5$  数据,基于新数据集完成图像情感分析,对应模型标注“传统+ASR”.

2) 在主流基线 WSCNet、WILDCAT 及融合类模

型 GS-XGB 的训练数据中加入 ASR 增强的新数据, 对应标注 “ $D_5 + ASR$ ”. 实验结果如表 5 所示.

表 5 ASR 泛化性论证实验结果

模型	准确率 / %	模型	准确率 / %
VGG16( $D_5$ )	86.56	WILDCAT( $D_5$ )	71.43
Res101( $D_5$ )	87.96	WILDCAT( $D_5 + ASR$ )	73.85
VGG16( $D_5 + ASR$ )	82.69	GS-XGB( $D_5$ )	86.47
Res101( $D_5 + ASR$ )	86.93	GS-XGB( $D_5 + ASR$ )	83.10
VGG16(传统)	88.75	WSCNet( $D_5$ )	87.62
Res101(传统)	88.37	WSCNet( $D_5 + ASR$ )	89.40
VGG16(传统 + ASR)	89.46	ASRF <sup>2</sup> (I)	88.65
Res101(传统 + ASR)	90.18	ASRF <sup>2</sup> (U)	90.06

由表 5 可见: VGG16 和 Res101 模型在执行传统数据增强后继续引入 ASR 策略, 对应识别精度分别提升 2.9%、2.22%, 显然, ASR 策略可与预训练深度学习模型配合使用, 它是对现有数据增强方法的有益补充. 将 ASR 策略精选出的样本加入到  $D_5$  数据中, 新训练出的主流基线能获得一定程度性能提升, 如在 WSCNet 模型上有 1.78% 提升. 由于 ASR 策略更倾向于从细粒度角度增强现有数据 (见表 2 及其对应分析, ASR 在 FI 上的泛化性更好), 它在深度学习、融合类模型上的泛化性一般. 当然, 由于既考虑主动样本精选, 又引入跨模态语义挖掘, ASRF<sup>2</sup> 模型的性能已逼近最优, 这充分表明了本文模型的有效性.

综上, ASR 策略具有一定的泛化性, 它既适用于 ASRF<sup>2</sup> 模型, 也能泛化到基线模型中, 以改善图像情感分析性能. 所以, ASR 策略既是对传统数据增强方法的有益补充, 以服务于对数据质量要求较高的机器视觉任务, 也可单独使用, 或配合主流图像情感分析模型, 获取更优分类性能. 总体上, 在相同数据样本情况下, ASRF<sup>2</sup> 模型表现更优.

### 3 结论

图像情感分析具有重要的研究价值. 高质量样本匮乏、未有效利用异构图像特征间的跨模态语义是制约情感预测精度的关键. 为此, 本文提出基于主动样本精选与跨模态语义挖掘的图像情感分析模型 ASRF<sup>2</sup>: 设计主动样本精选策略, 以增强现有数据集, 继而深入挖掘异构特征间的跨模态语义, 准确刻画图像情感内容. 实验表明: ASRF<sup>2</sup> 模型优于主流基线, 它能精选出更多情感标签明确的高质量样本, 进而丰富数据集, 是对现有数据增强方法的有益补充; 其次, ASRF<sup>2</sup> 模型输出判别性更强且低维的跨模态语义, 既有助于改善识别精度, 又能提高模型实时效率, 最终增强模型实用性; 此外, 相比主流基线<sup>[20,23]</sup>, ASRF<sup>2</sup> 模型仅需两类图像特征且参数调制

更便捷, 模型简单易于训练, 可部署在相应终端以对接实际应用; 最后, ASR 策略具备一定的泛化性, 既可单独使用, 也能配合主流图像情感分析模型, 获取更优的分类性能.

未来工作如下: 拟引入注意力或目标检测机制<sup>[33]</sup>, 探寻图像情感语义更强烈的局部区域, 基于“整体 + 局部”的思路进一步改善模型性能; 将主动样本精选策略与 AutoAugment<sup>[34]</sup> 相结合, 获取更多高质量样本, 继续增强现有数据集; 将 ASRF<sup>2</sup> 模型部署到服务器上完成在线图像情感预测, 缩小理论研究与实际应用间的“鸿沟”.

### 参考文献 (References)

- [1] Zhao Y Y, Qin B, Liu T, et al. Social sentiment sensor: A visualization system for topic detection and topic sentiment analysis on microblog[J]. *Multimedia Tools and Applications*, 2016, 75(15): 8843-8860.
- [2] Yang J F, She D Y, Lai Y K, et al. Weakly supervised coupled networks for visual sentiment analysis[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 7584-7592.
- [3] Zadeh A, Zellers R, Pincus E, et al. Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages[J]. *IEEE Intelligent Systems*, 2016, 31(6): 82-88.
- [4] Lu X, Suryanarayan P, Adams R B J, et al. On shape and the computability of emotions[C]. *Proceedings of the 20th ACM international conference on Multimedia*. Nara, 2012: 229-238.
- [5] Ge W F, Lin X R, Yu Y Z. Weakly supervised complementary parts models for fine-grained image classification from the bottom up[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 2019: 3029-3038.
- [6] Peng K C, Chen T, Sadovnik A, et al. A mixed bag of emotions: Model, predict, and transfer emotion distributions[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Boston, 2015: 860-868.
- [7] 晁浩, 刘永利, 连卫芳. EEG 情感识别中基于集成深度学习模型的多分析域特征融合[J]. *控制与决策*, 2020, 35(7): 1674-1680.  
(Chao H, Liu Y L, Lian W F. Multi-analysis domain feature fusion of EEG emotion recognition based on integrated deep learning model[J]. *Control and Decision*, 2020, 35(7): 1674-1680.)
- [8] Peng K C, Sadovnik A, Gallagher A, et al. Where do emotions come from? Predicting the emotion stimuli map[C]. *IEEE International Conference on Image Processing*. Phoenix, 2016: 614-618.
- [9] Yang J F, She D Y, Sun M. Joint image emotion classification and distribution learning via deep convolutional neural network[C]. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*.

- Melbourne, 2017: 3266-3272.
- [10] Zhao S C, Ding G G, Gao Y, et al. Learning visual emotion distributions via multi-modal features fusion[C]. Proceedings of the 25th ACM International Conference on Multimedia. New York, 2017: 369-377.
- [11] Sun M, Yang J F, Wang K, et al. Discovering affective regions in deep convolutional neural networks for visual sentiment prediction[C]. IEEE International Conference on Multimedia and Expo. Seattle, 2016: 1-6.
- [12] She D Y, Yang J F, Cheng M M, et al. WSCNet: Weakly supervised coupled networks for visual sentiment classification and detection[J]. IEEE Transactions on Multimedia, 2020, 22(5): 1358-1371.
- [13] Zhu Y, Zhou Y Z, Ye Q X, et al. Soft proposal networks for weakly supervised object localization[C]. IEEE International Conference on Computer Vision. Venice, 2017: 1859-1868.
- [14] Wu L F, Liu S, Jian M, et al. Reducing noisy labels in weakly labeled data for visual sentiment analysis[C]. IEEE International Conference on Image Processing. Beijing, 2017: 1322-1326.
- [15] Durand T, Mordan T, Thome N, et al. WILDCAT: Weakly supervised learning of deep convnets for image classification, pointwise localization and segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 5957-5966.
- [16] Fan S J, Jiang M, Shen Z Q, et al. The role of visual attention in sentiment prediction[C]. Proceedings of the 25th ACM International Conference on Multimedia. New York, 2017: 217-225.
- [17] You Q Z, Luo J B, Jin H L, et al. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia[C]. Proceedings of the 9th ACM International Conference on Web Search and Data Mining. New York, 2016: 13-22.
- [18] Zhao S C, Lin C, Xu P F, et al. CycleEmotionGAN: Emotional semantic consistency preserved CycleGAN for adapting image emotions[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 2620-2627.
- [19] Lin C, Zhao S C, Meng L, et al. Multi-source domain adaptation for visual sentiment classification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(3): 2661-2668.
- [20] Zhang H B, Wu J P, Shi H W, et al. Multidimensional extra evidence mining for image sentiment analysis[J]. IEEE Access, 2020, 8: 103619-103634.
- [21] Haghghat M, Abdel-Mottaleb M, Alhalabi W. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(9): 1984-1996.
- [22] Liudmila P, Gleb G, Aleksandr V, et al. Catboost: Unbiased boosting with categorical features[C]. Proceedings of Neural Information Processing Systems. Piscataway: IEEE, 2018: 6639-6649.
- [23] Cardoso T N C, Silva R M, Canuto S, et al. Ranked batch-mode active learning[J]. Information Sciences, 2017, 379: 313-337.
- [24] You Q Z, Luo J B, Jin H L, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[J]. CoRR, DOI: abs/1509.06041.
- [25] You Q Z, Luo J B, Jin H L, et al. Building a large scale dataset for image emotion recognition: The fine print and the benchmark[J/OL]. 2016, arXiv: 1605.02677.
- [26] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. 2014, arXiv: 1409.1556.
- [27] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.
- [28] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [29] Zhou B L, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 2921-2929.
- [30] Hotelling H. Relations between two sets of variates[M]. New York: Springer, 1992: 162-190.
- [31] Viivi U, Sahely B, Juho R. Large-scale sparse kernel canonical correlation analysis[C]. Proceedings of International Conference on Machine Learning. Long Beach, 2019: 6383-6391.
- [32] Zhang H B, Qiu D D, Wu R Z, et al. Novel framework for image attribute annotation with gene selection XGBoost algorithm and relative attribute model[J]. Applied Soft Computing, 2019, 80: 57-79.
- [33] You Q, Jin H, Luo J. Visual sentiment analysis by attending on local image regions[C]. Proceedings of 31st AAAI Conference on Artificial Intelligence. New York: AAAI, 2017: 231-237.
- [34] Cubuk E D, Zoph B, Mané D, et al. AutoAugment: Learning augmentation strategies from data[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 113-123.

### 作者简介

张红斌(1979—),男,副教授,博士,从事计算机视觉、自然语言处理、推荐系统等研究, E-mail: zhanghongbin@whu.edu.cn;

石皞炜(1996—),男,硕士生,从事图像情感分析、深度学习的研究, E-mail: 2297383338@qq.com;

熊其鹏(1997—),男,硕士生,从事计算机视觉、材质图像识别的研究, E-mail: danicaghost@163.com;

侯婧怡(1998—),女,硕士生,从事图像情感分析、机器学习的研究, E-mail: 914602773@qq.com.

(责任编辑: 郑晓蕾)