中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION

联合多头数据增强与多粒度语义挖掘的图像情感分析

张红斌, 侯婧怡, 石皞炜, 吕敬钦, 李雄, 李广丽

引用本文:
张红斌,侯婧怡,石炜,吕敬钦,李雄,李广丽. 联合多头数据增强与多粒度语义挖掘的图像情感分析[J]. 控制与决策, 2024, 39(6): 2013−2021.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2022.1807

联合多头数据增强与多粒度语义挖掘的图像情感分析

张红斌[1†], 侯婧怡[1], 石晾炜[1], 吕敬钦[1], 李 雄[1], 李广丽[2]

(1. 华东交通大学 软件学院，南昌 330013；2. 华东交通大学 信息工程学院，南昌 330013)

摘 要: 图像情感分析是机器视觉领域热点问题,然而情感判断主观性较强,仅分析完整图像难以准确刻画图像中情感语义,且高质量图像情感数据不足. 为此,提出联合多头数据增强与多粒度语义挖掘的图像情感分析模型 M^2. 首先,设计多头数据增强方法,基于自动数据增强与主动样本精选策略构建递进式数据增强模型,从“质”与“量”两个角度提升数据集;其次,引入情感区域检测模型完成情感区域增强,深入挖掘图像中情感语义强烈的局部区域,进而联合局部区域与整幅图像构建多粒度图像;然后,基于深度互学习框架及局部区域完成模型预训练,充分挖掘异构SENet网络之间互补的情感语义,并以迁移学习方式指导多粒度图像情感分析;最后,设计自适应特征融合模块,融合异构SENet特征以完成多粒度语义挖掘,实现图像情感分析. 在TwitterI和FI数据集上验证 M^2 模型,其准确率分别达到90.97%和81.14%,优于主流基线. M^2 拥有泛化性更强的数据增强策略,可以为其训练提供坚实的数据基础,且对应的实证分析效果较好,模型具备一定的实用价值.

关键词: 多头数据增强；多粒度语义挖掘；图像情感分析；情感区域检测；深度互学习；SENet

中图分类号: TP391 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1807

引用格式: 张红斌, 侯婧怡, 石晾炜, 等. 联合多头数据增强与多粒度语义挖掘的图像情感分析 [J]. 控制与决策, 2024, 39(6): 2013-2021.

Image sentiment analysis via multi-head data augmentation and multi-granularity semantics mining

ZHANG Hong-bin[1†], *HOU Jing-yi*[1], *SHI Hao-wei*[1], *LV Jing-qin*[1], *LI Xiong*[1], *LI Guang-li*[2]

(1. School of Software，East China Jiaotong University，Nanchang 330013，China；2. School of Information Engineering，East China Jiaotong University，Nanchang 330013，China)

Abstract: Image sentiment analysis is a hot topic in the computer vision field. However, it is hard to accurately characterize the sentiment semantics by only analyzing the whole image since sentiment judgment is usually subjective. And the image samples with high-quality are scarce. To alleviate the two issues, we propose a multi-head data augment and multi-granularity semantics mining (M^2) model. First, a progressive data augmentation model is constructed based on automatic data augmentation and active sample refinement. We improve datasets from the perspectives of quality and quantity. Second, an affective region detection model is introduced for sentiment region augmentation. Intense sentiment semantics is deeply mined from these affective local regions. Then we combine local regions with the whole images to create multi-granularity image data. Third, we pretrain the model through the deep mutual learning framework and affective local regions. The complementary sentiment semantics between heterogeneous SENet networks are fully mined, which is transferred in turn to guide the sentiment analysis of multi-granularity images. Finally, an adaptive feature fusion module is proposed to fuse heterogeneous SENet features to complete multi-granularity semantics mining as well as realize image sentiment analysis. The accuracies of M^2 are 90.97% and 81.14% on TwitterI and FI, respectively, which outperform mainstream baselines. The M^2 contains a data augmentation strategy with powerful generalization ability, which builds a firm data basis for training. Meanwhile, the corresponding empirical analysis is satisfactory, indicating a certain practicality.

Keywords: multi-head data augmentation；multi-granularity semantics mining；image sentiment analysis；affective region detection；deep mutual learning；SENet

收稿日期: 2022-10-18；**录用日期:** 2023-03-21.

基金项目: 国家自然科学基金项目 (62161011, 62361027)；江西省自然科学基金项目 (20232BAB202004, 20202BABL202044)；江西省主要学科学术和技术带头人培养计划项目 (20204BCJL23035)；江西省社会科学基金项目 (22TQ01)；江西省重点研发计划重点项目 (20223BBE51036).

责任编委: 孙宗耀.

†通讯作者. E-mail: zhanghongbin@whu.edu.cn.

0 引 言

随着互联网的普及与发展,越来越多的人乐于在社交平台上发表观点并抒发情绪,图像因其视觉语义丰富已逐渐成为人们表达情绪的重要载体. 对图像进行情感分析,能准确了解人们的情绪,营造良好健康的网络环境. 例如,及早发现并清理含有暴力、恐怖等极端情绪的图像,准确评估人们对某种商品的满意度等. 因此,图像情感分析的研究在学术界与工业领域都具有十分重要的意义[1-2].

图像情感分析研究初期使用机器学习方法,但手工特征很难准确描述图像中蕴含的复杂情感语义[3]. 对此,Zhang等[4]采用判别相关分析模型来挖掘异构图像特征之间的判别相关性,更好地刻画图像中情感语义. Zhang等[5]设计了主动样本精选(active sample refinement, ASR)算法,并联合跨模态语义完成了图像情感分析. 最近,深度学习模型逐渐被用于图像情感分析,相比于传统特征,深度学习特征[3]具有更强的表征能力. Wu等[6]提出了增强卷积神经网络,以提升图像情感分类性能. 为充分利用图像中局部区域蕴含的情感信息,Sun等[7]提出了生成对象建议并使用神经网络对其进行评分,以发现图像中情感语义强烈的局部区域. She等[8]提出了弱监督耦合卷积网络,该网络基于弱标签自动选择相关局部区域,以减少标注代价. Zhu等[9]提出了软建议组件进行情感区域定位,基于定位结果完成图像情感分析. Rao等[10]提出了MldrNet模型,用于学习多层次网络表征,更准确地描述图像中情感语义. Zhang等[11]提出了DEAN模型,使用多个深度学习模型分析来自不同模态的数据,实现了多模态情感分析.

深度学习模型依赖大量标注数据. Zhu[12]等将每幅图像从源数据域翻译到目标数据域,为训练模型提供更丰富的样本. Lin等[13]提出了一种多源情感生成对抗网络,充分利用潜在的情感空间处理来自多个不同域的数据. 不同于上述生成式数据增强方法,Zhang等[4-5,14]根据图像情感数据集的特点设计样本精选方法以获得更多高质量图像,较好地应对了样本稀缺问题. 现有图像情感数据集均由人工标注,标注者的主观差异会导致数据集中图像的情感标注不明确,高质量图像情感数据稀缺. 此外,现有方法主要对完整图像进行分析,忽略了图像局部区域蕴含的细粒度情感信息. 综上,现有研究存在“缺少高质量图像情感数据”“图像中的多粒度语义未有效挖掘”等关键问题. 对此,本文提出M^2 (Multi-head data augment and Multi-granularity semantics mining)模型,设计多头数据增强方法以缓解高质量数据稀缺问题,继而充分挖掘图像中蕴含的多粒度语义,更全面、更准确地描述图像情感内容. 本文的主要贡献如下:

1) 提出多头数据增强方法,完成图像样本在“质”与“量”上的双重飞越,有效缓解高质量图像数据稀缺问题,为训练模型奠定坚实数据基础.

2) 自动检测情感语义强烈的局部区域,联合局部区域与完整图像构建多粒度图像. 基于深度互学习框架搭建局部区域情感分析模型,以迁移学习方式指导图像情感分析,实现多粒度语义挖掘.

3) 提出图像情感分析模型M^2,同时处理多种粒度的图像. 在TW (Twitter I)和FI两大基准数据集上,M^2模型的识别精度优于主流基线,且其实证分析效果良好.

1 M^2模型描述

1.1 模型框架

M^2模型由“多头数据增强”和“多粒度语义挖掘”两个部分组成,模型框架如图1所示. 首先,对图像情感数据集进行多头数据增强. 多头数据增强方法包括递进式数据增强(progressive data augmentation, PDA)和情感区域增强两部分. PDA输出增强后的整幅图像,而情感区域增强则生成情感语义强烈的图像局部区域. 其次,基于异构SENet构建深度互学习(deep mutual learing, DML)[15]框架,预训练出局部区域情感分析模型Net_{loc}. 然后,将情感区域增强模型产生的情感语义强烈的图像局部区域合并到PDA增强后的整幅图像集合中,构建多粒度图像集合. 最后,采用预训练好的局部区域情感分析模型Net_{loc}指导多粒度图像情感分析,即迁移局部区域情感分析模型Net_{loc}的全部参数,采用多粒度图像对Net_{loc}模型进行参数微调,完成多粒度语义挖掘并实现图像情感分析.

1.2 模型构成

1.2.1 多头数据增强

多头数据增强方法包括递进式数据增强和情感区域数据增强两部分,下面分别介绍其具体实现.

1) 递进式数据增强.

联合快速自适应数据增强模型(fast auto-augment, FAA)[16]与主动样本精选策略ASR[5]构建递进式数据增强模型PDA,其技术流程如图2所示.

图像情感数据包括D_5、D_4和D_3三大类子集. 其中:D_5表示有5名标注员工对同一张图像标记相同情感标签,D_4表示有4名及以上标注员工对同一张图像标记相同情感标签,D_3表示有3名及以上标注员工对同一张图像标记相同情感标签.

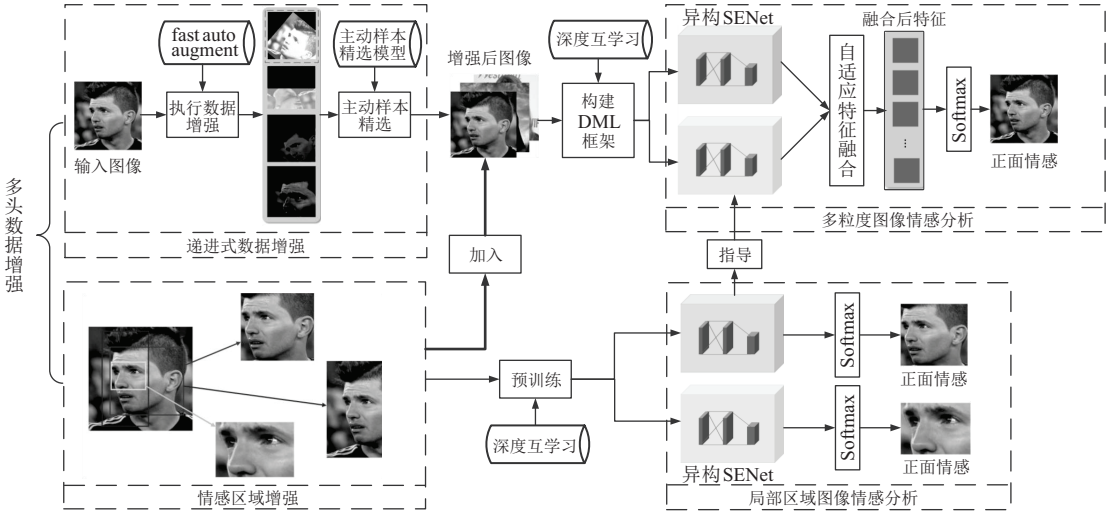


图 1　M^2 模型框架(以面向二元分类的TW数据集为例)

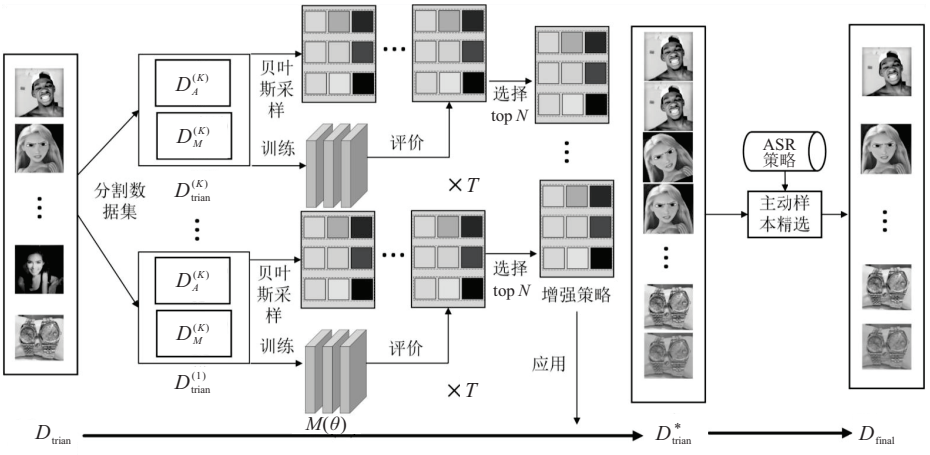


图 2　递进式数据增强模型PDA的执行过程

如图2所示: D_{train} 代表原始图像数据,本文使用 D_5; D^*_{train} 是经过FAA增强后的图像数据集,表示为 D(FAA); D_{final} 是对 D^*_{train} 继续执行ASR策略后生成的图像数据,表示为 D(PDA). 因此PDA模型分为两部分: 第1部分执行FAA模型,搜索符合图像情感数据的增强策略; 第2部分为ASR策略. 在第1部分中,首先,使用 K 折交叉验证分割数据集(本文 $K = 5$). 在每一折中,数据集被分为 D_A 和 D_M 两部分, D_A 用来执行贝叶斯优化器以寻找增强策略, D_M 用来评价增强策略优劣. 贝叶斯优化器会采样一组增强策略,对增强策略做出评估并将结果传递给优化器进行重新学习. FAA选择最优 N 次采样(本文 $N = 10$),然后在完整图像数据集上执行 T 次贝叶斯采样(本文 $T = 2$). 因此,FAA会生成100组策略. 将生成的100组策略随机应用到原图像数据集中1 : 1增强原数据集,得到经FAA增强后的数据集 D(FAA). 贝叶斯采样使用的采集函数为期望增量,即

$$\text{EI}(\Gamma) = E[\min(L(\theta|\Gamma(D_A)) - L^*, 0)]. \quad (1)$$

其中: E 表示期望值; L 表示策略得分; Γ 表示采样出的策略; θ 表示模型参数; L^* 为一个常数,它由上一次采样得到的策略计算分位数得到. FAA使用密度匹配法可大幅降低评估搜索策略所耗费的时间.

PDA模型的第2部分为ASR策略,分为粗精选和细精选两步. 首先执行粗精选,从 D(FAA)中选出情感语义明确的图像,为训练M^2奠定坚实的数据基础. 以 D_5 为训练集, D(FAA)为测试集,设计粗精选算法,从 D(FAA)中初选出优质样本. 为防止选择性偏倚,算法中综合 Q 个互补的分类器来评估 D(FAA)中样本的优劣. 为提升粗精选精度和算法效率,所有样本均基于DCA挖掘SIFT与VGG19之间的跨模态特征进行刻画. 当所有分类器预测值与样本真实标签一致时粗选出该样本,生成数据集 D_{coarse}(PDA). 然后执行细精选,对 D_{coarse}(PDA)进一步提炼,输出情感语义更明确的图像. 以 D_5 为训练集, D_{coarse}(PDA)为测

试集,在精选过程中融入主动学习策略. 在每次迭代中,使用分类器对D_{coarse}(PDA)中的样本进行预测,再以排序批处理模式作为查询函数,筛选出高质量图像样本,样本的每个预测值都会根据下式进行评分:

$$\text{score} = \partial(1-\Phi(x, D_5)) + (1-\partial)U(x). \quad (2)$$

其中:权重$\partial = \dfrac{D_5}{|D(\text{FAA})| + |D_5|}$,$U(x)$为样本$x$预测值的不确定性,$\Phi$为欧氏距离相似度函数. 式(2)前半部分计算样本与训练图像之间的相似度,后半部分计算特征空间在样本附近的分布. 每次迭代选取评分最高的W个图像,若预测值与真实标签完全一致,则将该图像选出并合并到训练集中. 经过若干轮迭代,细精选策略主动挖掘出关键情感知识,为模型训练提供优质图像. 为防止选择性偏倚,采用Z个分类器完成主动样本精选,每个分类器分别生成一组候选样本集合,由于样本来自不同分类器,它们具有较强的互补性,对这些样本集合执行交集操作或并集操作,输出高质量样本D(PDA),再将D(PDA)并入原数据集,以丰富原图像情感数据集.

执行PDA模型后,两大基准数据集的详细信息如表1所示. D(ASR)表示仅使用主动样本精选策略增强后的数据,D(FAA)表示仅执行快速自适应数据增强后的数据,D(PDA)表示执行本文递进式数据增强方法后的数据,D(AR)表示执行情感区域增强后的数据. 由表1可知,执行FAA数据增强模型后,数据集中的样本有了“量”的飞越. 在此基础上继续执行ASR策略,即以D_5为训练集,主动精选D(FAA)中的优质样本,数据集实现了“质”的飞越. 因此,递进式数据增强模型能生成更多优质图像样本,为提升图像情感分析精度奠定数据基础.

表1 原始数据集及执行多种数据增强后的数据集情况

数据集	D_5	D(ASR)	D(FAA)	D(PDA)	D(AR)
TW	882	62	616	154	4 928
FI	5 238	593	3 971	694	28 708

2) 情感区域数据增强.

图像局部区域中包含的物体或者目标对最终情感预测起着决定性作用,故识别图像中情感语义强烈的局部区域具有重要意义. 情感区域(affective regions, AR)[17]增强策略可以自适应地发现图像中有重要价值的情感区域,它分为两步:产生候选区域和筛选出情感语义强烈的局部区域.

step 1: 产生候选区域. 采用Edge Boxes[18]生成一组具有对象评分的候选边界框. 对于生成的大量候选边界框,首先筛除具有相同几何特征的候选框,同时过滤尺寸失调的候选框;然后计算余下边界框的交并比数值,基于该数值使用归一化切割算法将候选框分为h组,在每一组中选取对象分数最高的边界框,产生h个候选框.

step 2: 筛选出情感语义强烈的局部区域. 首先,将m个候选框输入到VGG16模型中预测候选框的情感精度. 若候选框的预测类别与情感图像数据集标签相似,则说明候选框中蕴含与该类别相同的情感语义,应保留这些候选框. 然后,设计判别标准AR_{score},由对象评分$\text{Obj}_{\text{score}}$和情感评分$\text{Senti}_{\text{score}}$共同决定. $\text{Obj}_{\text{score}}$是Edge Boxes输出候选框时计算的对象评分,$\text{Senti}_{\text{score}}$如下式所示:

$$\text{Senti}_{\text{score}_i} = \sum_{j=1}^{c} y_{ij} \times \log y_{ij} + 1, \quad (3)$$

其中c是情感类别数,y_{ij}表示第i个候选框在第j个类别的概率值. $\text{Senti}_{\text{score}}$反映图像中候选框的情感语义,$\text{Obj}_{\text{score}}$检测候选框中是否包含一个完整对象. 因此,$\text{AR}_{\text{score}}$由这两个互补的评分共同决定,其计算公式为

$$\text{AR}_{\text{score}_i} = \sqrt{(1-\beta) \times \text{Obj}^2_{\text{score}_i} + \beta \times \text{Senti}^2_{\text{score}_i}}, \quad (4)$$

其中β控制两个评分间的权重,本文取$\beta = 0.5$. 基于AR_{score}选择Top F(本文$F = 8$)的局部区域作为情感区域. 在对原数据集中的D_5数据执行基于AR策略的情感区域增强后,得到的图像信息如表1中D(AR)所示,两个数据集的训练图像相比增强之前增加8倍.

1.2.2 多粒度图像情感分析

如图1所示,在完成情感区域数据增强后,基于DML框架,采用AR策略增强出的情感区域图像预训练局部区域情感分析模型Net_{loc}. 联合局部情感区域与PDA输出的整幅图像构成多粒度图像. 采用多粒度图像微调Net_{loc}网络参数,完成多粒度语义挖掘及图像情感分析.

首先,基于异构的SENetXT50和SENetXT101网络构建局部区域情感分析模型. 每个网络均设计两个损失函数以完成协同训练,包括:监督损失函数和拟态损失函数. 监督损失指基于标签的损失,采用交叉熵;拟态损失指网络的后验类别,它要与另一个网络类别概率一致. 拟态损失选择KL散度,SENetXT101相对于SENetXT50的KL散度、SENetXT50相对于SENetXT101的KL散度公式分别为

$$D_{\text{KL}}(p_2 \| p_1) = \sum_{i=1}^{U} \sum_{m=1}^{M} p_2^m(x_i) \log \frac{p_2^m(x_i)}{p_1^m(x_i)}, \quad (5)$$

$$D_{\mathrm{KL}}(p_1\|p_2)=\sum_{i=1}^{U}\sum_{m=1}^{M}p_1^m(x_i)\log\frac{p_1^m(x_i)}{p_2^m(x_i)},\quad(6)$$

其中p_1和p_2分别表示Softmax层输出的概率.

SENetXT50网络与SENetXT101网络的监督损失函数分别为

$$L_{c1}=-\sum_{i=1}^{U}\sum_{m=1}^{M}I(y_i,m)\log(p_1^m(x_i)),\quad(7)$$

$$L_{c2}=-\sum_{i=1}^{U}\sum_{m=1}^{M}I(y_i,m)\log(p_2^m(x_i)).\quad(8)$$

因此,在局部区域情感分析模型中,单个网络的损失由KL散度和L_{c1}共同决定,两个网络的损失函数如下式所示:

$$L_1=L_{c1}+D_{\mathrm{KL}}(p_2\|p_1),\quad(9)$$

$$L_2=L_{c2}+D_{\mathrm{KL}}(p_1\|p_2).\quad(10)$$

其次,基于多粒度图像微调$\mathrm{Net_{loc}}$模型,完成对两个SENet网络的训练. 设计自适应特征融合模块,融合两个SENet网络最后一层输出的异构特征,挖掘来自异构网络的互补性判别信息,以进一步提升模型性能. 自适应特征融合流程为: 设计3层全连接层,分3步将两个SENet网络最后一层特征图映射为2 048维、1 024维以及与数据集类别数相同维度的一维向量,最后输入Softmax函数完成图像情感标签预测.

1.3　M^2模型算法描述

算法1　M^2模型.

输入: 原始数据集D_5;

输出: 图像情感标签.

step 1: 对原始数据集执行PDA模型.

step 1.1: 对D_5执行FAA操作,得到D(FAA);

step 1.2: 对D(FAA)执行ASR策略,得D(PDA).

step 2: 对原始数据集D_5执行AR策略.

step 2.1: 使用Edge Boxes在图像中生成候选区域集B;

step 2.2: 使用归一化切割算法将候选框分为h组;

step 2.3: 选取每组中对象分数最高的候选框,得到h个候选框;

step 2.4: 对h个候选框进行评分,选取Top F作为情感局部区域,得到D(AR).

step 3: 基于D(AR)预训练局部区域情感分析模型$\mathrm{Net_{loc}}$.

step 4: 基于多粒度图像"$D_5+D(\mathrm{AR})+D(\mathrm{PDA})$"微调$\mathrm{Net_{loc}}$模型.

step 5: 对$\mathrm{Net_{loc}}$模型中异构SENet进行自适应特征融合.

step 6: 基于特征融合结果预测图像情感标签.

2　实验结果与讨论

为验证M^2模型的有效性,选择TW和FI中的D_5数据完成全部实验. 随机选取70%数据为训练集,30%为测试集. TW[19]来自社交网站. TW数据集分正面(positive)和负面(negative)两类,是一种粗粒度数据集. FI[20]通过在社交网站上搜索8种情绪关键字: anger、amusement、awe、contentment、disgust、excitement、 fear和sadness得到,共计21 508张图像. 不同于TW, FI是一种细粒度数据集.

为防止选择性偏倚,在PDA的ASR粗精选中,TW使用逻辑回归、随机森林等9个分类器(即$Q=9$),FI使用K近邻、决策树等5个分类器(即$Q=5$)进行实验. 在ASR的细精选中$W=2$,对于TW数据集$Z=2$,对于FI数据集$Z=4$.

2.1　定量实验结果分析

使用准确率(accuracy)评估模型性能. 首先,将M^2与所有基线进行比较,实验结果如表2所示,其中粗体字表示最优值. 表2中: D表示微调后的深度学

表2　与基线模型的性能比较　　%

数据集	模型	类型	准确率	模型	类型	准确率	模型	类型	准确率
TW	VGG16[21]	D	76.75	DCA[26](D_5)	F	87.59	Sun et al[7](D_5)	B	88.94
	ResNet101[22]	D	78.13	GS-SVM[27](D_5)	F	88.72	SmileyNet[29]	B	89.16
	CAM-Res101[23]	D	82.67	MSGAN[13]	B	63.58	WSCNet[8](D_5+ASR)	B	89.40
	VGG16(D_5)	D	86.56	WILDCAT[28]	B	79.53	CCM[14]	B	89.90
	gradKCCA[24](D_5)	F	77.07	SPN[9]	B	81.67	ASRF^2[5]	B	90.06
	CCA[25](D_5)	F	80.08	ME^{2}M(M)[4]	B	87.15	M^2(本文)	/	**90.97**
FI	VGG16[21]	D	63.75	DCA[26](D_5)	F	73.71	WSCNet[8](D_5+ASR)	B	74.72
	ResNet101[22]	D	66.16	GS-XGB[27](D_5+ASR)	F	75.98	Rao et al[10]	B	75.46
	CAM-Res101[23]	D	68.54	Yang et al[2](D_5)	B	67.48	SR-w-DCA[4]	B	75.72
	ResNet(D_5+ASR)	D	78.11	MSGAN[13]	B	70.63	ASRF^2[5]	B	75.77
	gradKCCA[24](D_5)	F	61.84	Zhu et al[12]	B	73.00	CCM[14]	B	80.32
	CCA[25](D_5)	F	50.29	WILDCAT[28](D_5+ASR)	B	72.23	M^2(本文)	/	**81.14**

习模型,F表示特征融合类模型,B表示图像情感分析模型. 括号内D_5表示使用D_5评估,D_5 + ASR表示联合D_5和ASR增强后的样本完成评估,其余模型采用完整数据集进行评估. 如表2所示,M^{2}模型在两个数据集上均表现优异,下面从4个角度进行深入分析.

1) 与微调后的深度学习模型相比,M^{2}的性能在两个数据集上均有较大幅度提升. 在TW数据集上,M^{2}相对于VGG16(D_5)提升了4.41%. 微调后的深度学习模型的深层特征判别性一般,制约了分类精度. 此外,深度学习模型需要大量样本来拟合网络. M^{2}模型能有效应对上述问题:它使用SENet作为情感分析模型的基础网络,基于DML架构实现异构网络间的双向知识传递并挖掘SENet中关键判别信息;M^{2}模型还运用AR策略积极探寻图像中情感语义强烈的局部区域,为模型训练提供了更丰富且鲜明的情感语义. 此外,M^{2}模型设计多头数据增强策略,为训练精度优良的图像情感分析模型奠定了坚实的数据基础.

2) 与特征融合类模型相比,在TW数据集上,M^{2}比GS-SVM(D_5)提高了2.25%. CCA融合后特征维度偏高,且执行CCA较为耗时,不利于模型部署;GS类模型需提取多组图像特征且特征融合操作复杂,复现难度大;执行DCA融合后特征维数更低,可能会丢失一些关键判别信息;gradKCCA模型需计算多个核矩阵,模型空间复杂度相对较高.

3) 与主流图像情感分析模型相比,M^{2}模型表现优异. 在TW数据集上,它比WSCNet (D_5+ASR)提升了1.57%;在FI数据集上,相较于SR-w-DCA模型提升了5.42%. 对比需要人工进行局部区域标注的WILDCAT,M^{2}模型虽也挖掘情感局部区域,但其使用的AR策略不需额外标注,可极大节省人力成本. 其次,SPN等模型基于深度学习网络,需更多计算资源;而M^{2}模型在多粒度语义挖掘中引入DML框架,在相同情况下M^{2}模型能够取得更高的分类精度. 再者,M^{2}模型优于MSGAN模型,不同于基于数据生成策略的方法,M^{2}中的FAA自适应选取数据增强策略,其训练过程更简单,易于快速复现和部署.

4) 与基于样本精选的图像情感分析模型相比,在TW与FI数据集上,M^{2}相比于ASRF^{2}分别提升了0.91%与5.37%,相比于CCM分别提升了1.07%与0.82%. M^{2}是有效且鲁棒的,主要原因是:M^{2}设计递进式数据增强模型,将主动样本精选策略与主流数据增强模型FAA结合,在生成多样性数据基础上实现动态样本精选,从全局视角完成数据集“质”和“量”的提升. 同时,M^{2}模型设计情感区域数据增强方法,以检测情感语义强烈的局部区域,从局部视角完成数据集“质”和“量”的飞越;其次,M^{2}采用DML框架构建局部区域情感分析模型,通过双向知识传递深入挖掘异构SENet网络间互补的情感语义,以更好地刻画图像视觉内容. 不同于ASRF^2,M^{2}仅使用D_5数据,对原始数据集的依赖性更低. 此外,不同于基于聚类相关挖掘的CCM,M^{2}模型构建一种更简便且有效的自适应特征融合模块,以融合DML框架中异构SENet的特征层,获取判别性更强的新特征,最终提高图像情感分析性能.

2.2 定性实验结果分析

除定量分析之外,本节通过定性实例来验证M^{2}模型的有效性. 如图3第2行所示,每一张图像都有形态上的变化. 例如M^{2}模型对第5张图像进行翻转、裁剪以及增加对比度等操作,可以突出图像中细节并展示关键图像局部区域. 此外,ASR策略挑选出的增强图像拥有较高质量,且未破坏图像中原情感极性;相反,它增强了能凸显情感极性的局部区域. 例如第6张图像整体颜色较暗沉,主要表达负面情感,执行PDA后图像整体颜色未发生变化,不会影响最终情感预测. 因此,PDA模型是有效的,它既能增加样本数量,还可以保留图像中原情感语义.

图3 原图像与执行PDA生成的图像对比

AR策略能生成不同类别的情感语义强烈的局部区域. 在图像情感分析研究中,若要判断图像中人的情感,其表情是一个非常重要的参考因素. 如图4

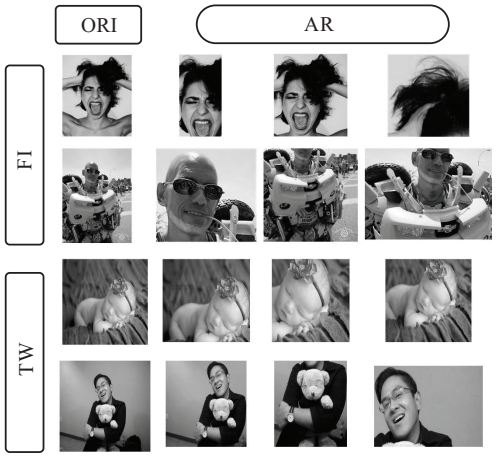


图4 执行AR策略后生成的图像(ORI表示原图像,AR表示执行AR策略后随机挑选的图像)

所示,这4张图像的核心描述对象都是人,故人物的表情对于最终情感预测至关重要. 本文提出的AR策略能准确聚焦完整图像中情感语义强烈的局部区域,即识别这些人的面部表情,这些表情都蕴含非常鲜明的情感极性,为改善情感分析精度奠定了坚实基础.

综上,定性实验结果表明:递进式数据增强模型能生成高质量图像,从全局视角完成数据集“质”与“量”的提升,进而增强模型的鲁棒性. 情感区域数据增强模型能准确捕捉图像中情感语义强烈的局部区域,从局部视角完成数据集“质”与“量”的双重飞越. 递进式数据增强模型与情感区域增强模型形成互补,因此多头数据增强策略是有效的,它从多维视角提升数据集质量,为训练模型夯实数据基础.

2.3　模型参数

2.3.1　网络实时曲线展示

为更好地体现模型的实时运行状态,本节展示在TW与FI数据集上M^2的准确率与loss曲线. 图5和图6分别为loss曲线和准确率曲线.

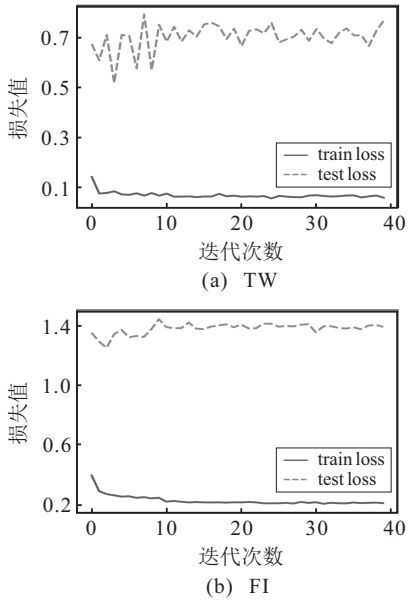


图5　M^2模型的loss曲线

如图5所示:两个数据集中的训练损失曲线都较平滑,模型在第5次迭代就已达到局部最优,这说明在训练集上,M^2具有较快的收敛速度,主要原因是,SENetXT50和SENetXT101这两个网络通过互学习相互传递有价值的监督信息,共同推动模型向局部最优逼近;而测试损失曲线变化较剧烈,尤其是TW数据集,在1～10次迭代期间曲线变化幅度较大,主要原因是,不同类别数据之间的差异性较大,同时数据集在增强后包含多粒度图像,导致模型出现“数据颠簸”现象,而伴随迭代次数增加,测试损失曲线都逐渐趋于平缓,说明M^2在不断适应不同尺度的多粒度图

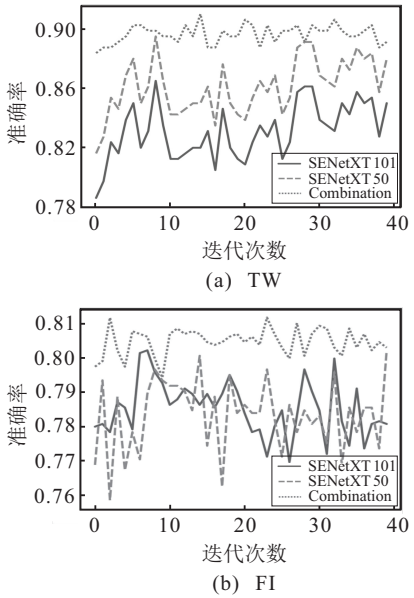


图6　M^2模型中单个网络及特征融合后的分类精度

像,并学习到充足且多样性的判别信息.

如图6所示,“Combination”表示M^2准确率. 由图6可知,执行自适应特征融合后M^2的性能明显优于单个网络,这表明自适应特征融合方法是有效的,它能充分挖掘异构网络之间的互补性监督信息,促使分类性能提升. 图6还表明:在TW数据集中,SENetXT50优于SENetXT101,即在粗粒度情感图像中,浅层次神经网络能捕获到更多有价值的情感语义,进而促进精度提升;而在FI数据集中,两个SENetXT网络性能差异不大,即在细粒度情感图像中,网络的深度不是关键因素,而应充分挖掘异构网络间的互补性来提升模型分类精度. 总之,M^2融合了来自SENetXT50和SENetXT101判别信息的新特征,因此,具有更强的图像情感预测能力.

2.3.2　消融分析实验

为更好地验证M^2模型每个模块在图像情感分析中的重要作用,本节对M^2模型进行消融分析,得到如下4个模型变种:$M^2(D_5)$、M^2(FAA)、M^2(PDA)、M^2,它们的含义分别是仅使用D_5数据的M^2、仅执行FAA策略的M^2、执行PDA后的M^2以及同时执行多头数据增强与多粒度语义挖掘的M^2(本文模型). 表3给出了消融分析结果,其中最优值用粗体表示.

表3　消融分析结果　　%

数据集	模型	准确率	模型	准确率
TW	$M^2(D_5)$	87.21	M^2(PDA)	88.72
	M^2(FAA)	90.23	M^2	**90.97**
FI	$M^2(D_5)$	77.82	M^2(PDA)	78.00
	M^2(FAA)	77.66	M^2	**81.84**

如表3所示,依次使用FAA、ASR等策略后,M^2

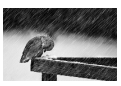
的性能逐步提升. 在TW数据集上,执行FAA策略后M^2性能提升近3%,但在FI数据集上出现微弱衰减,这说明:在粗粒度数据集上,FAA数据增强策略可发挥更大作用;而在细粒度数据集上,执行FAA操作会产生一些噪声图像,这些图像影响了特征的判别性及最终图像情感分析精度. 继续执行ASR策略后,在TW数据集上,相对于M^2(FAA),M^2性能出现衰减,而在FI数据集上提升了0.34%,表明ASR策略对于识别细粒度情感更有效,挑选出了优质的增强图像,降低了噪声图像干扰. 当执行AR策略生成情感区域后,M^2模型在TW和FI数据集上分别提高了2.25%和3.84%. 这充分说明:M^2能有效挖掘图像中包含的多粒度情感语义;同时,在引入包含全局图像和局部情感区域的多粒度信息后,M^2能获取更充分的互补性判别信息,这些判别信息有助于改善M^2模型分类性能. 因此,本文所提出的多粒度语义挖掘思路是有效且鲁棒的,而且它仅需在数据层面对不同尺度的图像进行混合,操作非常简单.

综上,M^2模型各组成部分在图像情感分析中都扮演了重要角色,它们形成合力以共同促进模型预测精度提升.

2.3.3 实证分析结果

为进一步验证M^2的实用性,选取8张微博图像对M^2进行客观评估,8张图像分别对应FI中的8个细粒度情感类别,以完成实证分析,如表4所示,其中,错误预测结果用粗体表示.

表4 采用微博上的真实图像测试模型性能

图像				
标签	amusement	anger	awe	**contentment**
预测	amusement	anger	awe	**sadness**
图像				
标签	disgust	**excitement**	fear	**sadness**
预测	disgust	**amusement**	fear	**awe**

在表4中,M^2在处理真实场景图像时有5张图像预测正确,3张图像预测错误,实证效果较好. 其中,M^2在预测情感标签更明确的图像时表现优异,例如第2张图像是一个愤怒的男人,符合“anger”标签. M^2在预测一些语义模糊图像时表现一般,例如M^2将第4张图像预测为“sadness”,此时图像原标签语义较模糊,存在一定歧义. 此外,M^2在识别标签为“sadness”的图像时将其错误预测为“awe”,因为这两种情感存在潜在语义交集,故它们在视觉内容上具有相似性,使得模型出现识别误差. 未来,拟考虑引入多模态分析方法[11,30],以获取更明确的情感语义描述,进一步改善实证分析效果.

3 结 论

图像中蕴含复杂的情感信息,对图像进行情感分析具有重要价值. 对此,本文提出了联合多头数据增强与多粒度语义挖掘的图像情感分析模型M^2:1)构建了PDA模型,生成并精选出更多高质量图像;2)设计了情感区域增强模型,动态检测图像中情感语义强烈的局部区域,联合整幅图像和局部区域构建多粒度图像数据;3)设计了局部区域情感分析模型,充分挖掘来自异构SENet的图像局部区域中的情感语义,并将其迁移到多粒度图像上,获取更优的分类精度. 实验表明:M^2优于主流基线,M^2模型能对图像情感数据集进行有效增强,为模型训练提供情感语义明确的多粒度数据;能充分利用情感语义强烈的局部区域挖掘图像中隐含的情感信息,为多粒度语义挖掘及提升识别精度做出了重要贡献. M^2对社交平台图像进行情感分析时预测效果良好,具备实用价值.

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]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7584-7592.

[3] Ge W F, Lin X R, Yu Y Z. Weakly supervised complementary parts models for fine-grained image classification from the bottom up[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 3029-3038.

[4] 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.

[5] Zhang H B, Shi H W, Xiong Q P, et al. Image sentiment analysis via active sample refinement and cross-modal semantics mining[J]. Control and Decision, 2022, 37(11): 2949-2958.

[6] Wu L F, Liu S, Jian M, et al. Reducing noisy labels in weakly labeled data for visual sentiment analysis[C]. Proceedings of the IEEE International Conference on Image Processing. Beijing, 2017: 1322-1326.

[7] Sun M, Yang J F, Wang K, et al. Discovering affective regions in deep convolutional neural networks for

visual sentiment prediction[C]. Proceedings of the IEEE International Conference on Multimedia and Expo. Seattle, 2016: 1-6.

[8] 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.

[9] Zhu Y, Zhou Y Z, Ye Q X, et al. Soft proposal networks for weakly supervised object localization[C]. Proceedings of the IEEE International Conference on Computer Vision. Venice, 2017: 1859-1868.

[10] Rao T R, Li X X, Xu M. Learning multi-level deep representations for image emotion classification[J]. Neural Processing Letters, 2020, 51(3): 2043-2061.

[11] Zhang F, Li X C, Dong C R, et al. Deep emotional arousal network for multimodal sentiment analysis and emotion recognition[J]. Control and Decision, 2022, 37(11): 2984-2992.

[12] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]. Proceedings of the IEEE International Conference on Computer Vision. Venice, 2017: 2242-2251.

[13] Lin C, Zhao S C, Meng L, et al. Multi-source domain adaptation for visual sentiment classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence. New York, 2020: 2661-2668.

[14] Zhang H B, Shi H W, Hou J Y, et al. Image sentiment analysis via active sample refinement and cluster correlation mining[J]. Computational Intelligence and Neuroscience, 2022: 2477605.

[15] Zhang Y, Xiang T, Hospedales T M, et al. Deep mutual learning[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4320-4328.

[16] Lim S, Kim I, Kim T, et al. Fast autoaugment[C]. Proceedings of Conference on Neural Information Processing Systems. Vancouver, 2019: 1-11.

[17] Yang J F, She D Y, Sun M, et al. Visual sentiment prediction based on automatic discovery of affective regions[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2513-2525.

[18] Zitnick C L, Dollár P. Edge boxes: Locating object proposals from edges[C]. Proceedings of the European Conference on Computer Vision. Zurich, 2014: 1-15.

[19] You Q Z, Luo J B, Jin H L, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]. Proceedings of the AAAI Conference on Artificial Intelligence. Austin, 2015: 381-388.

[20] 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.

[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. 2014, arXiv: 1409.1556.

[22] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.

[23] Zhou B L, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 2921-2929.

[24] 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.

[25] Hotelling H. Relations between two sets of variates[C]. Breakthroughs in Statistics. New York: Springer, 1992: 162-190.

[26] Haghighat 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.

[27] 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.

[28] Durand T, Mordan T, Thome N, et al. WILDCAT: Weakly supervised learning of deep ConvNets for image classification, pointwise localization and segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 5957-5966.

[29] Al-Halah Z, Aitken A, Shi W Z, et al. Smile, be happy:) emoji embedding for visual sentiment analysis[C]. Proceedings of the IEEE International Conference on Computer Vision Workshop. Seoul, 2019: 4491-4500.

[30] Yadav A, Vishwakarma D K. A deep multi-level attentive network for multimodal sentiment analysis[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2022, 19(1): 1-19.

作者简介

张红斌(1979–), 男, 教授, 博士, 硕士生导师, 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;

侯婧怡(1998–), 女, 硕士生, 从事计算机视觉、图像情感分析等研究, E-mail: 914602773@qq.com;

石晾炜(1996–), 男, 硕士生, 从事计算机视觉、图像情感分析等研究, E-mail: 18007001885@163.com;

吕敬钦(1984–), 男, 副教授, 博士, 从事行人检测、机器视觉等研究, E-mail: jingqinlv@ecjtu.edu.cn;

李雄(1987–), 男, 副教授, 博士, 从事生物信息学、机器学习等研究, E-mail: 4185350@qq.com;

李广丽(1977–), 女, 教授, 硕士, 从事医学影像处理、跨媒体检索、推荐系统等研究, E-mail: 1333@ecjtu.edu.cn.