

赖金水,万中英,曾雪强.基于情感轮和多任务卷积神经网络的图像情感分布学习[J].江西师范大学学报(自然科学版),2022,46(4):363-371.

LAI Jinshui, WAN Zhongying, ZENG Xueqiang. The image emotion distribution learning based on emotion wheel and multi-task convolutional neural network [J]. Journal of Jiangxi Normal University(Natural Science) 2022, 46(4) : 363-371.

文章编号: 1000-5862(2022) 04-0363-09

基于情感轮和多任务卷积神经网络的 图像情感分布学习

赖金水,万中英,曾雪强*

(江西师范大学计算机信息工程学院,江西 南昌 330022)

摘要: 图像情感分布学习可以对多种情绪同时进行建模,但现有的模型缺乏有效的方法直接考虑情绪之间的相关性。针对这一问题,该文提出一种基于情感轮和多任务卷积神经网络(EW-MTCNN)的图像情感分布学习模型,通过先验知识模块将心理学情感知识直接引入到深度神经网络中。基于 Mikel's 情感轮定义成对情绪之间的相关性,EW-MTCNN 模型采用多任务卷积神经网络学习情绪之间的相关性信息,同时优化情感分布预测和情绪分类任务。EW-MTCNN 模型由 3 个模块组成,3 个模块分别是图像特征提取层、情感轮先验知识层和多任务损失层。在情感分布数据集(Emotion6)和单标签数据集(Artphoto)上进行对比实验的结果表明:EW-MTCNN 模型在情感分布预测与情绪分类任务上比其他情感分布学习方法具有更优的性能。

关键词: Mikel's 情感轮;多任务卷积神经网络;情感分布学习;情绪分类;标记分布学习

中图分类号: TP 391 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2022.04.06

0 引言

图像情感分析任务的目标是挖掘出图像中隐含的人们的情感倾向,是一项涉及计算机视觉、心理学和美学的多学科交叉任务^[1]。近年来,图像情感分析因其在人工智能领域中的广泛应用而已成为机器学习领域的一个研究热点并受到大量关注^[2-3]。

与传统的计算机视觉任务相比,图像情感分析任务更具有挑战性,主要体现在 2 个方面:(i) 图像很少表达单一的情绪,而往往是多种不同情绪的混合^[4];(ii) 人类各种情绪之间存在较高的相关性^[5]。图 1 展示了来自 Emotion6 数据集的 6 个样本及其对应的情绪标注,由图 1(a)和图 1(e)可见,高兴情绪与惊讶情绪表现出较高正相关性;由图 1(d)可见,该图像的主导情绪是愤怒,但同时也蕴含了大量的恐惧与惊讶的情绪。因此,多情绪分析和情绪间的

相关性是情绪分析模型需要考虑的重要因素。

目前,图像情感分析方法主要包括 2 个大类。一类是基于单标签学习或多标签学习^[6]的方法,其主要特点是为每幅图像分配 1 个或多个情绪标签。基于单标签学习或多标签学习的方法可以借鉴经典的分类模型来实现情绪标签的预测,但不能对图像在各种情绪上的表达程度进行定量分析。另一类是基于情感分布学习的方法。Zhou Ying 等^[7]提出情感分布学习(emotion distribution learning,EDL),其核心思想是将图像在各个基本情绪上的表达程度作为图像的情感分布,学习图像中的特征与情感分布之间的相关信息。EDL 模型对图像中蕴含的多种情绪同时建模,从而可以更好地处理多情绪标签的相关性和模糊性。

近年来,国内外学者在国际顶级会议和期刊上发表了多个 EDL 相关的研究工作,其中在 EDL 模型中考虑情绪类别之间的相关性是一个重要的研究

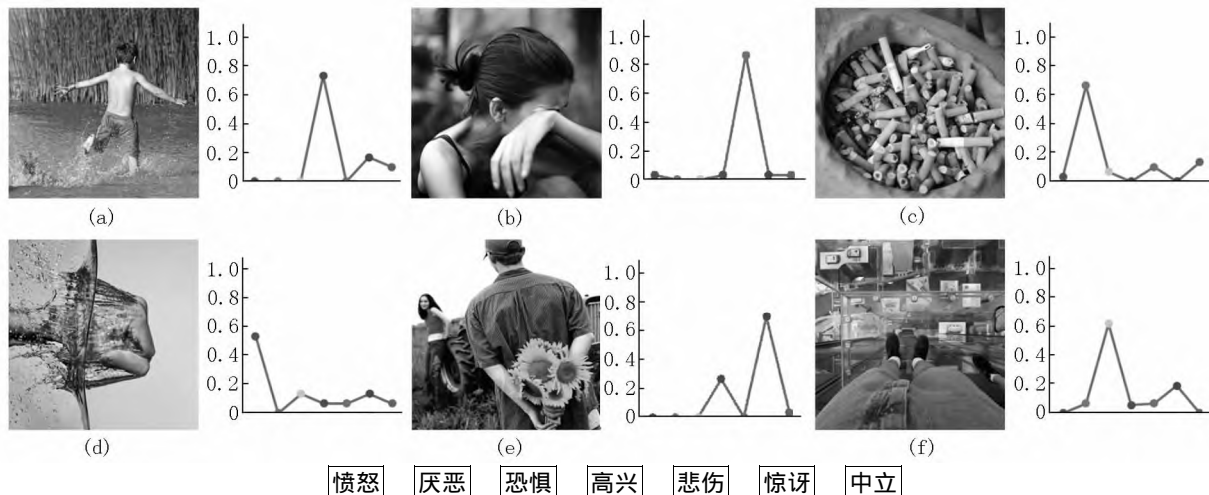
收稿日期:2022-01-13

基金项目:国家自然科学基金(61866017)资助项目。

通信作者:曾雪强(1978—),男,江西南昌人,教授,博士,博士生导师,主要从事标记分布学习、自然语言处理和数据降维的研究。E-mail: xqzeng@jxnu.edu.cn

思路。如 Jia Xiuyi 等^[8]提出了一种利用局部低秩结构捕捉标签局部相关性的 EDL 方法,Zhou Deyu 等^[9]提出了一种基于情绪之间的约束关系的 EDL 方法,Xiong Haitao 等^[10]提出了利用情感极性和情绪标签稀疏性的 EDL 方法。这些研究工作都表明,

考虑在训练数据中的情绪相关性可明显提升 EDL 模型的性能。情绪间的相关性既可从训练数据中抽取,也可基于心理学先验知识获得。但是,目前较少有基于图像的 EDL 研究工作,特别是在预测模型中直接运用心理学先验知识考虑情绪之间的相关性。



注:横轴是情绪 纵轴是表达程度。

图1 Emotion6 数据集的 6 幅图像及其情绪标注

Mikel's 情感轮(Mikel's emotion wheel)是心理学家 J. A. Mikel's 等^[11]提出的心理学情绪模型。Mikel's 情感轮描述了人类的 8 种基本情绪之间的相关性,分别是愤怒(anger)、厌恶(disgust)、悲伤(sadness)、恐惧(fear)、兴奋(excitement)、惊讶(awe)、满意(contentment)、高兴(amusement)。这 8 种情绪按一定顺序组成一个情感轮,在情感轮上的位置关系反映了情绪类别之间的相关性,处于相邻位置的 2 种情绪呈正相关,处于相对位置的 2 个情绪呈负相关。

本文将 Mikel's 情感轮心理学先验知识引入情感分布学习中,提出了一种基于情感轮和多任务卷积神经网络的图像情感分布学习(emotion wheel enhanced multi-task convolutional neural network for image emotion distribution learning,EW-MTCNN)模型。EW-MTCNN 的深度神经网络模型包括 3 个模块:图像特征提取层、情感轮先验知识层和多任务损失层。图像特征提取层通过多层卷积神经网络提取图像中的特征信息;情感轮先验知识层基于 Mikel's 情感轮计算成对情绪间的相关程度,将图像特征提取层的输出转换为情感分布;多任务损失层采用交叉熵损失和 Kullback-Leibler(KL)损失分别度量情绪分类任务和情感分布预测任务,通过综合这 2 种损失训练目标,同时优化情绪分类和情感分布预测任务。EW-MTCNN 模型以端到端的方式进行训练,将在输出的情感分布中表达程度最大的情绪作为情绪分类

任务的预测值。在图像情感分布数据集(Emotion6)和情绪单标签数据集(Artphoto)上的对比实验结果表明:EW-MTCNN 模型在情感分布预测和情绪分类任务上的性能优于对比的其他情感分布学习方法。

1 相关工作

图像情感分析的相关工作主要包括图像特征提取和情感分布学习。

1.1 图像特征提取

传统的图像特征提取方法主要基于特征工程,人工提取图像中的底层特征(如颜色^[12]、纹理^[13]和构图^[14]等)。C. Colombo 等^[15]将图像分割成若干个区域,将每个区域和其他区域在颜色、饱和度、色调等特征上进行比对,将图像分为 4 种不同情绪。J. Machajdik 等^[16]基于艺术原理和心理学理论定义了多种人工处理特征的组合,包括构图、颜色变化和图像语义。研究表明,这些数据大多来源于抽象绘画与艺术照片,通过人工处理的图像特征在一些特定的数据集上是有效的。因为基于特征工程的方法太过于经验化,且需要耗费大量的人力,因此不适用于大规模数据集。此外,对于来自不同领域的图像,人工提取图像中情感信息的方法往往存在一定差异。因此,传统的图像情感分析方法不具有泛化性能。

近年来,随着卷积神经网络(convolutional neural network, CNN)在许多计算机视觉任务上取得成功, CNN 也被运用到图像情感分析中。基于 CNN 的图像特征提取方法主要通过深层卷积神经网络提取在图像中的深层特征,不再需要人工从图像中提取特征,实现了从图像中学习特征和参数,并且在各种数据集上具有良好的泛化性能。如 Rao Tianrong 等^[17]提出了一个多层深度网络(MldrNet),它将图像的底层特征与深层特征进行统一。Yang Jufeng 等^[18]将人工提取的图像底层特征与深层卷积神经网络提取的深层特征进行对比实验,结果表明深层特征在情感分析模型上表现更优。

1.2 情感分布学习

传统的图像情感分析致力于图像的情感极性分类,即判别在一幅图像中的情感是正面还是负面。然而,这种情感极性识别只能适用于简单的情感分析任务(如情感二分类问题),无法实现对多个情绪类别的识别。因此,单标签学习(single label learning)将简单的正面、负面情绪细分为若干个基本情绪,通过人工标注,从细分的情绪类别中为每幅图像分配一个情绪标签。在此基础上,多标签学习(multi label learning)为每幅图像分配多个情绪标签进行区分,这种标注方法适用于细粒度的图像情感分析。

虽然基于单标签或多标签的方法可以实现细粒度情绪上的情绪分类,但是无法对图像在每种情绪上的表达程度进行定量分析。因此,Zhou Ying 等^[7]提出了情感分布学习(emotion distribution learning, EDL),情感分布学习能够对多种情绪进行建模,学习图像特征与图像真实情感分布之间的相关信息,适用于处理具有情绪模糊性的任务。此外,Geng Xin^[19]提出了标签分布学习(label distribution learning, LDL),其中包括 3 种策略,即问题转化(PT)、算法改造(AA)和专用算法(SA);基于这 3 种策略提出了 6 种 LDL 模型,它们分别是 PT-Bayes、PT-SVM、AA-kNN、AA-BP、SA-HIS 和 SA-BFGA 模型,本文将在后续实验中展示这 6 种标签分布学习模型的性能。

目前,许多学者在国内外顶级会议与期刊上发表了关于 EDL 的相关工作。Zhang Yuxiang 等^[20-21]提出了基于多任务卷积神经网络(multi-task convolutional neural network, MT-CNN)模型用于文本情感分布学习。由于未考虑情绪标签之间的相关性, MT-CNN 模型的性能有限。Jia Xiuyi 等^[8]提出一种利用局部低秩结构捕捉标签局部相关性的情感分布学习方法并将它用于面部表情识别任务,该方法旨在从

训练数据中学习情绪标签之间的相关性,在不同数据集上不具有泛化性能。He Tao 等^[22]通过引入外部先验知识来表示情绪之间的相关性,提出一种基于情感轮的情绪图卷积网络的方法(EmotionGCN),但由于 EmotionGCN 模型结构较为简单,不能有效地学习情绪之间的相关性,模型性能有限。与 MT-CNN 和 EmotionGCN 模型不同,本文提出的 EW-MTCNN 模型将 Mikel's 情感轮心理学先验知识引入情感分布学习中,并采用多任务卷积神经网络,以端到端的方式进行训练,学习情绪之间的相关性,共同优化情感分布预测与情绪分类任务。

2 基于情感轮和多任务卷积神经网络的图像情感分布学习

2.1 Mikel's 情感轮

有心理学研究表明:人类情绪之间存在高度相关性^[5],一些正相关的情绪经常同时出现,一些负相关的情绪则通常互斥出现。基于心理学理论,J. A. Mikel's 等^[11]提出了 Mikel's 情感轮(Mikel's emotion wheel),用于描述 8 种基本情绪之间的相关性,它们分别是愤怒(anger)、厌恶(disgust)、悲伤(sadness)、恐惧(fear)、兴奋(excitement)、惊讶(awe)、满意(contentment)、高兴(amusement),这 8 种基本情绪在情感轮的相对位置如图 2 所示。根据 Mikel's 情感轮,将 2 个情绪之间的距离定义为在 Mikel's 情感轮上从一个情绪到另一个情绪所经过的步数。如兴奋与惊讶之间的距离为 1,高兴与恐惧之间的距离为 4。总体而言,距离近的 2 类情绪之间相关程度高,距离远的 2 类情绪相关程度低。



图2 Mikel's 情感轮

目前已有一些研究工作将情感轮运用于情感分布学习,如 He Tao 等^[22]提出一种基于情感轮的图卷积网络,但由于未能很好地学习情绪标签之间的相关性信息,所以该模型性能有限。总体而言,基于情感轮的情感分布学习方法仍比较少,本文基于

Mikel's 情感轮定义成对情绪之间的相关程度,并采用基于情感轮和多任务卷积神经网络的图像情感分布学习模型,同时优化情感分布预测与情绪分类任务。

2.2 基于 Mikel's 情感轮和多任务卷积神经网络的图像情感分布学习

本文提出的基于情感轮和多任务卷积神经网络的图像情感分布学习(emotion wheel enhanced multi-task convolutional neural network for image emotion distribution learning, EW-MTCNN)模型主要由3个模块组成,它们分别为图像特征提取层、情感轮先验知识层和多任务损失层。具体的模型架构如图3所示。

1) 图像特征提取层。鉴于 VGGNet^[23] 在许多计算机视觉任务的特征提取上表现出优秀的性能,本文基于 VGGNet 构建图像特征提取层,取 VGGNet 倒数第2层的全连接层作为图像信息的表示,即采用了 VGGNet 的5组卷积层与池化层和2层全连接层用于提取图像中隐含的深层图像特征。对于给定彩色图像 $x_i \in \mathbf{R}^w \times \mathbf{R}^h \times \mathbf{R}^3$,在 VGGNet 之后增加一个全连接层,用于对应各个目标情绪。最后,图像特征提取层提取出图像信息表征向量 $v = (v_1, v_2, \dots, v_C)$, C 为目标情绪的数量。

2) 情感轮先验知识层。定义转换矩阵 $W = (f_1, f_2, \dots, f_C)$, 其中 C 为情绪类别数量。根据各种情绪在 Mikel's 情感轮上的位置关系,为每个情绪 μ 生成一个先验情感分布 $f_\mu = \{f_\mu^j\}_{j=1}^C$, 其中 $\mu \in \{1, 2, \dots, C\}$, f_μ^j 表示情绪 j 在先验情感分布 f_μ 上的表达程度,情绪 μ 在先验情感分布中的表达程度应该最大,其余情绪的表达程度随着 Mikel's 情感轮上与情绪 μ 之间的距离增加而逐步递减。因此,总体上先验情感分布 f_μ 是以情绪 μ 为最大值,两边逐渐递减的对称分布。参照 Geng Xin 等^[24-27] 基于 LDL 的人脸年龄预

测工作的方法,假设先验情感分布服从高斯分布,先验情感分布 f_μ^j 的计算公式为

$$g_\mu^j = (\exp(-|j - \mu|^2 / (2\sigma^2))) / (\sqrt{2\pi}\sigma), \quad (1)$$

$$f_\mu^j = \{g_\mu^j / \sum_{k=1}^C g_\mu^k\}_{j=1}^C, \quad (2)$$

其中 σ 为先验情感分布标准差,用于描述先验情感分布 f_μ 的离散程度, $|j - \mu|$ 为情绪 j 与情绪 μ 在 Mikel's 情感轮上的距离。

采用基于 Mikel's 情感轮的转换矩阵 W 对图像特征提取层的输出 $v = (v_1, v_2, \dots, v_C)$ 进行加权计算,将其转换为包含心理学先验知识的情感分布,其计算公式为

$$p = v_1 f_1 + v_2 f_2 + \dots + v_C f_C.$$

3) 多任务损失层。采用情绪分类和情感分布预测2个任务,同时对神经网络进行训练,用基于情感轮先验知识层预测情感分布。多任务损失层采用交叉熵损失和 Kullback-Leibler(KL) 损失分别度量情绪分类任务和情感分布预测任务,通过加权的方式,将这2种损失函数综合起来,同时优化情绪分类和情感分布预测任务,具体的损失函数为

$$L = (1 - \lambda) L_{cls}(x, y) + \lambda L_{sdl}(x, d), \quad (3)$$

其中 L_{cls} 和 L_{sdl} 分别表示情绪分类损失和情感分布预测损失,参数 λ 用于控制在训练过程中这2种损失的相对权重。

在训练过程中,由于交叉熵损失用于最大化正确类别的概率,因此采用交叉熵损失用于情绪分类任务。交叉熵损失为

$$L_{cls}(x, \mu) = -\frac{1}{N} \left(\sum_{i=1}^N \sum_{j=1}^C 1(y_i = j) \ln p_i^j \right),$$

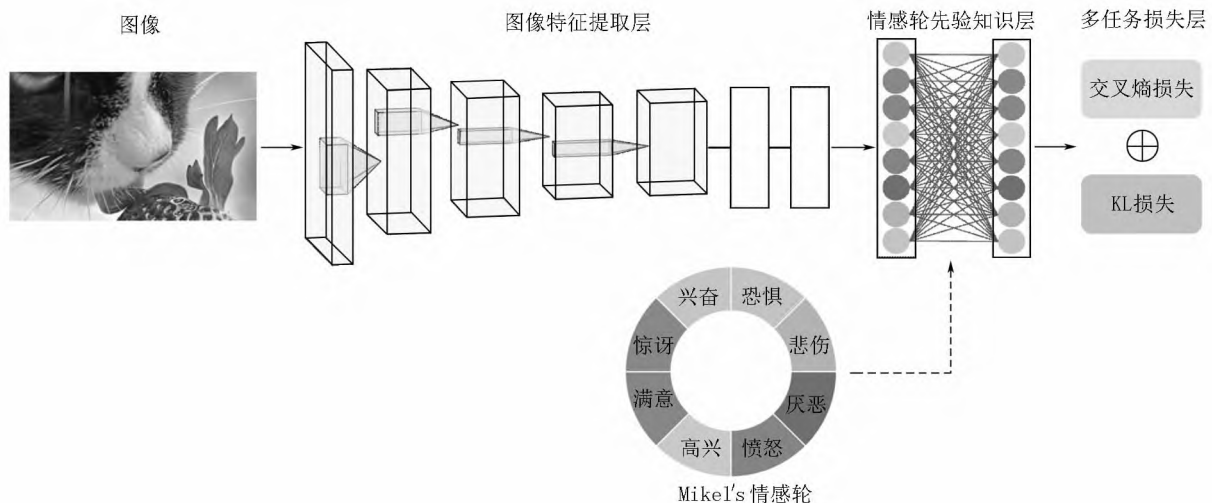


图3 基于情感轮和多任务卷积神经网络的图像情感分布学习模型总体框架

其中函数 $1(\delta) = \begin{cases} 1 & \delta = \text{真实的} \\ 0 & \delta = \text{不真实的} \end{cases}$, p_i^j 表示第 i 幅图像在第 j 类情绪上的表达程度。

KL 损失是度量预测情感分布与先验情感分布或真实情感分布之间的相似性的指标之一,因此,将 KL 损失用于情感分布预测任务, KL 损失为

$$L_{sdl}(x, d) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C d_i^j \ln p_i^j.$$

对于式 (3), 采用随机梯度下降 (SGD) 来最小化损失函数。根据链式法则, 梯度的计算公式为

$$\begin{aligned} \partial L / \partial v_i^j &= (1 - \lambda) \sum_k \partial L_{cls} \partial p_i^k / (\partial p_i^k \partial v_i^j) + \\ &\lambda \sum_k \partial L_{sdl} \partial p_i^k / (\partial p_i^k \partial v_i^j) = (1 - \lambda) (p_i^j \sum_k y_i^k - y_i^j) + \\ &\lambda (p_i^j \sum_k d_i^k - d_i^j) = p_j^{(i)} - (1 - \lambda) y_j^{(i)} - \lambda d_j^{(i)}, \quad (4) \end{aligned}$$

多任务损失层根据式 (4) 进行梯度的反向传播。

3 实验结果与分析

为了评估 EW-MTCNN 模型的性能, 本文在 2 个图像情感数据集上设置了 3 组实验: 对比 EW-MTCNN 模型与多种情感分布学习模型在情感分布预测与情绪分类任务上的性能、分析先验情感分布参数 σ 对 EW-MTCNN 模型性能的影响以及对比 EW-MTCNN 模型与基于深度网络的情绪分析模型在传统情绪分类任务上的性能。

3.1 数据集

本文采用图像情感分布数据集 (Emotion6) 和情绪单标签数据集 (Artphoto) 进行 EDL 模型的对比实验。Emotion6 数据集是从 Flickr 中集成而来, 常用于情感预测基准数据集, 其中标注了 7 个情绪类别: 愤怒、厌恶、高兴、恐惧、悲伤、惊讶、中性, 总共包含 1 980 幅图像。在 Emotion6 数据集中每幅图像都对应一个真实情感分布, 即图像在每个情绪上的表达程度。Artphoto 数据集来源于一个艺术图像分享网站, 每幅图像都分配至 8 个情绪标签中的一个, 8 个情绪分别是愤怒、厌恶、悲伤、恐惧、兴奋、惊讶、满意和高兴, 该数据集总共包含 806 幅艺术图像。Emotion6 数据集和 Artphoto 数据集各个情绪的图像数量如表 1 所示。

3.2 实验设置

本文实验设置采用分层十折交叉验证。具体实验步骤如下: 在保持各种情绪的图像数量比例的前提下, 将数据集平均分成 10 份, 每份数据作为测试集使用 1 次, 剩余 9 份数据作为训练集, 如此重复 10 次, 取各评价指标在 10 次实验中的平均值作为

模型的最终性能。为了合理对比 EW-MTCNN 模型与其他图像情感分布学习模型, 参与对比的模型均采用相同的数据划分。

表 1 实验数据集的各个情绪的图片数量

情绪	Emotion6	Artphoto
愤怒	330	77
厌恶	330	70
悲伤	330	166
惊讶	330	102
恐惧	330	115
高兴	330	101
兴奋		105
满意		70
总计	1 980	806

将 EW-MTCNN 模型的图像特征提取层的权重初始化为 VGGNet^[23] 在 ImageNet 数据集上预训练的权重, 并在 Emotion6 或 Artphoto 数据集上进行了微调。这种训练方法能够有效提高模型的收敛速率, 适用于大规模数据集上的训练。

参照 Yang Jingyuan 等^[28] 的方法, 将多任务损失函数权重系数 λ 设为 0.7。此外, EW-MTCNN 模型的其余参数设置如下: 图像特征提取层和多任务损失层的学习率分别设为 0.001 和 0.010; 为防止过拟合, 在全连接层采用 Dropout 比率为 0.5, 设置 Batch Size 为 32, 并采用随机梯度下降 (SGD) 对整个网络的所有层的权重进行微调, 具体的模型参数设置如表 2 所示。

表 2 EW-MTCNN 模型参数设置

参数	设定值
优化器	SGD
λ	0.7
图像特征提取层学习率	0.001
多任务损失层学习率	0.010
Dropout 率	0.5
Batch Size	32
Epochs	20

对于情感分布预测任务, 采用 6 种常用于评价真实情感分布和情感分布预测之间距离或相似度的指标, 它们分别是 Chebyshev distance、Clark distance、Canberra metric、KL divergence、Cosine 和 Intersection。其中, 前 4 个指标为距离指标, 值越小表明模型性能越优; 后 2 个指标为相似度指标, 值越大表明模型越优。由于当分母为 0 时, KL divergence 无定义, 因此设 $\varepsilon = 1 \times 10^{-10}$ (近似 0 值)。对于情绪分类任务, 采用准确率评估模型的性能, 将在情感分布预测中表达程度最大的情绪作为图像的情绪标签用于情绪分类。

本文实验的硬件配置为 Intel 酷睿 i9-40900X 3.70 GHz 10 核 CPU, 128 G 内存, NVIDIA GeForce

RTX 3060 显卡. 操作系统为 Ubuntu 18.04 ,深度学习框架采用 Tensorflow2.4 和 Keras2.4.3.

3.3 多种图像情感分析模型的情感分布预测和情绪分类性能对比

为了评估 EW-MTCNN 模型在图像情感分布数据集上的情感分布预测和情绪分类性能,将 EW-MTCNN 模型与 7 种 LDL 模型(分别是 PT-Bayes、PT-SVM、AA-kNN、AA-BP、SA-IIS、SA-BFGA 和 SA-CPNN^[24]模型)进行对比. 参照 Yang Jufeng 等^[29]的

方法,将 VGGNet 中倒数第 2 层全连接层的输出作为图像信息表征,并采用主成分分析(PCA)将图像信息表征的维度降至 280 维,将降维后的图像信息表征作为输入,采用上述 LDL 模型作为分类器进行训练,输出情感分布预测. 此外,将 EW-MTCNN 模型与图像情感分布学习模型 DLDL 模型^[30]进行对比, DLDL 模型通过多层卷积神经网络提取图像中的深层特征,并采用 KL 损失函数进行训练,在情感分布预测上具有良好性能,对比实验的具体结果如表 3 所示.

表 3 在 Emotion6 数据集上 9 种情感分布学习模型的性能对比结果

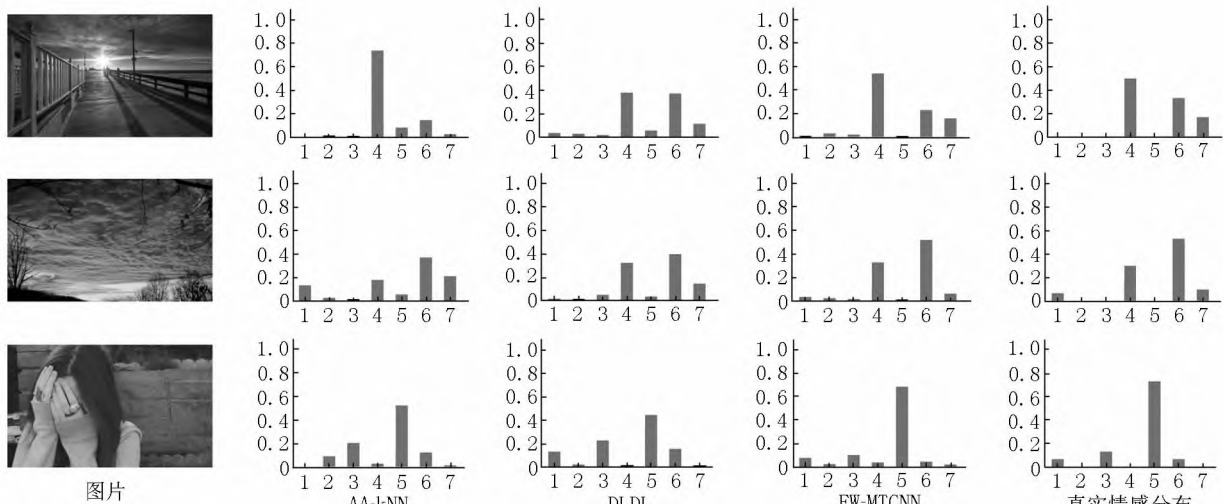
模型	情感分布预测						情绪分类 准确率/%
	Cheb(↓)	Clark(↓)	Canber(↓)	KLdiv(↓)	Cos(↑)	Int(↑)	
PT-Bayes	0.352 5(7)	1.971 8(9)	4.475 2(8)	2.380 4(9)	0.672 7(7)	0.504 6(8)	38.27(7)
PT-SVM	0.361 3(9)	1.896 0(8)	4.582 3(9)	1.068 9(8)	0.479 7(9)	0.401 8(9)	36.45(8)
AA-kNN	0.282 3(3)	1.637 9(2)	3.617 6(3)	0.864 0(6)	0.786 4(2)	0.629 7(3)	44.90(4)
AA-BP	0.287 1(4)	1.781 9(7)	3.816 0(6)	0.662 7(5)	0.728 4(4)	0.592 8(7)	47.16(3)
SA-IIS	0.321 8(6)	1.683 2(4)	3.787 0(4)	0.622 3(4)	0.691 2(5)	0.615 2(4)	41.10(6)
SA-BFGA	0.354 1(8)	1.811 1(6)	4.125 5(7)	0.998 5(7)	0.645 4(8)	0.601 2(6)	36.39(9)
SA-CPNN	0.301 6(5)	1.695 6(5)	3.796 0(5)	0.574 2(3)	0.683 1(6)	0.607 2(5)	42.03(5)
DLDL	0.269 7(2)	1.642 8(3)	3.608 4(1)	0.469 6(2)	0.781 8(3)	0.632 4(2)	48.96(2)
EW-MTCNN	0.250 4(1)	1.627 4(1)	3.612 9(2)	0.456 2(1)	0.790 2(1)	0.671 5(1)	53.12(1)

注: ↑表示该指标越大越好, ↓表示该指标越小越好.

表 3 展示了 EW-MTCNN 模型与 8 种情感分布学习模型在 Emotion6 数据集上的情感分布预测和情绪分类性能,括号中展示了在单项指标上各模型性能的排序结果. 实验结果表明:在情感分布预测和情绪分类任务的 7 个指标上,EW-MTCNN 模型均优于 PT-Bayes、PT-SVM、AA-kNN、AA-BP、SA-IIS、SA-BFGA 和 SA-CPNN 模型. 在情感分布预测任务上,EW-MTCNN 模型除了在 Canberra 指标上略低于 DLDL 模型外,在其余 5 项情感分布预测指标上均优于对比的情感分布学习模型. 在情绪分类任务上,EW-MTCNN 模型在准确率上表现出优秀的性能,比次优的 DLDL 模型高 4.16%. 由此可见,EW-MTC-

NN 模型比其他情感分布学习模型在情感分布预测和情绪分类任务上均取得了优越的性能. 实验结果表明:EW-MTCNN 模型能够有效利用情绪之间的相关性信息,综合度量情感分布预测损失与情绪分类损失,同时优化情感分布预测和情绪分类任务.

图 4 展示了在 Emotion6 数据集上 AA-kNN、DLDL 和 EW-MTCNN 模型的情感分布预测结果. 由图 4 可见:EW-MTCNN 模型在情绪极性、情绪类型和情感分布上都与真实情感分布相似度最高. 此外,与 DLDL 模型仅采用 KL 损失函数相比,EW-MTCNN 模型结合交叉熵损失与 KL 损失,在情感分布预测上具有优越的性能.



情绪标签 1: 愤怒、2: 厌恶、3: 恐惧、4: 高兴、5: 悲伤、6: 惊讶、7: 中立; 横轴是情绪 纵轴是表达程度.

图 4 在 Emotion6 数据集上 3 种情感分布学习模型的情感分布预测结果

3.4 先验情感分布参数 σ 对模型性能的影响

先验情感分布参数 σ 是先验情感分布的标准差,用于描述由情绪标签生成先验情感分布的离散程度,是影响 EW-MTCNN 模型的重要参数。 σ 越小,生成的先验情感分布越集中,曲线越瘦高。 σ 越大,生成的先验情感分布越分散,曲线越扁平。为了分析先验情感分布参数 σ 对 EW-MTCNN 模型性能的影响,在 Artphoto 数据集上进行实验。首先根据图像情绪标签,由式(1)~(2)为每幅图像生成对应先验情感分布,其次将参数 σ 从 0~1 取值,每隔 0.1 取值 1 次,采用情感分布预测指标 KL divergence 和情绪分类指标的准确率来评估先验情感分布参数 σ 对 EW-MTCNN 模型性能的影响。由式(1)可知,当 $\sigma = 0$ 时无定义,因此采用 $\varepsilon = 1 \times 10^{-10}$ 近似 0 值。先验情感分布参数 σ 对 EW-MTCNN 模型性能的影响如图 5 所示。

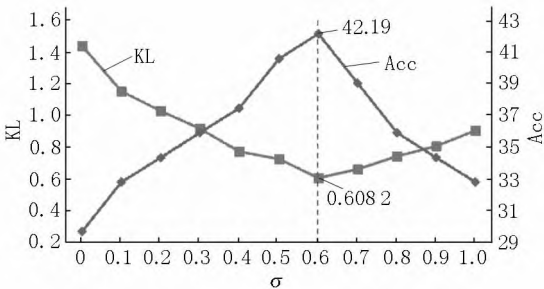


图 5 在 Artphoto 数据集上参数 σ 对模型性能的影响

由图 5 可见:随着 σ 的增加,EW-MTCNN 模型在情绪分类指标的准确率上的性能逐步上升,并在 $\sigma = 0.6$ 时达到最高(42.19%),随后逐步下降,在情感分布预测指标 KL divergence 上的性能逐步下降,在 $\sigma = 0.6$ 时达到最低(0.6082),随后逐步上升。实验结果表明:相比于传统的基于单标签的情绪分类模型($\sigma = 0$),将情绪标签扩展为情感分布能有效提高模型在情感分布预测和情绪分类任务上的性能,并且适当增加先验情感分布的离散程度能有效促进 EW-MTCNN 模型的性能。当 $\sigma = 0.6$ 时,先验情感分布的离散程度最佳。以恐惧情绪为例,当 $\sigma = 0.6$ 时先验情感分布为 $[1.48 \times 10^{-10}, 2.47 \times 10^{-6}, 2.57 \times 10^{-3}, 1.66 \times 10^{-1}, 6.64 \times 10^{-1}, 1.66 \times 10^{-1}, 2.57 \times 10^{-3}, 2.47 \times 10^{-6}]$ 。当 σ 取值为 0.6~1.0 时,EW-MTCNN 模型在准确率指标上的性能逐步下降,这说明先验情感分布离散程度过大会混淆主导情绪标签,影响 EW-MTCNN 模型情绪分类性能。由于 EW-MTCNN 模型在 $\sigma = 0.6$ 时性能最优,因此,在实验中设 $\sigma = 0.6$ 。

3.5 基于深度网络的情绪识别模型的情绪分类性能对比

为了评估 EW-MTCNN 模型在传统情绪分类任

务上的性能,本文在 Emotion6 和 Artphoto 数据集上进行对比实验。对于 Emotion6 数据集,采用在真实情感分布中表达程度最高的情绪作为图像的情绪标签。鉴于 AlexNet^[31]、VGGNet^[23] 和 ResNet^[32] 3 种深度网络模型在图像特征提取上已经表现出优越的性能,本文展示了这 3 种模型在传统情绪分类任务上的性能,以及在对对应数据集上对这 3 种模型进行微调后的模型性能,具体的 9 种基于深度网络的情绪识别模型在情绪分类任务上的性能对比结果如表 4 所示。9 种对比模型简要介绍如下:

1) AlexNet、VGGNet 和 ResNet。模型权重初始化为 3 种模型在 ImageNet 数据集上预训练的权重,将最后一层用于分类的全连接层神经元数量修改为数据集的情绪类别数量,采用交叉熵损失,将其余层权重冻住,只对最后一层全连接层权重进行更新。

2) Fine-tuned AlexNet、Fine-tuned VGGNet 和 Fine-tuned ResNet。基于 AlexNet、VGGNet 和 ResNet 模型,在目标数据集上对 3 个模型的权重进行微调。

3) EW-MTCNN (AlexNet)、EW-MTCNN (VGGNet) 和 EW-MTCNN (ResNet)。为了分析不同图像特征提取层对 EW-MTCNN 模型的性能的影响,基于 3 种深度网络模型构建图像特征提取层,得到对应的模型。

表 4 9 种基于深度网络的情绪识别模型在 Emotion6 数据集和 Artphoto 数据集上的情绪分类性能对比

数据集	模型	准确率/%
Emotion6	AlexNet	38.43
	VGGNet	43.62
	ResNet	45.83
	Fine-tuned AlexNet	40.68
	Fine-tuned VGGNet	46.88
	Fine-tuned ResNet	47.92
	EW-MTCNN (AlexNet)	45.96
	EW-MTCNN (VGGNet)	53.12
	EW-MTCNN (ResNet)	53.60
Artphoto	AlexNet	29.38
	VGGNet	33.72
	ResNet	34.64
	Fine-tuned AlexNet	33.53
	Fine-tuned VGGNet	36.85
	Fine-tuned ResNet	37.04
	EW-MTCNN (AlexNet)	37.56
	EW-MTCNN (VGGNet)	42.19
	EW-MTCNN (ResNet)	42.08

表 4 的实验结果表明:相比于基于 AlexNet 的情绪识别模型,基于 ResNet 或 VGGNet 模型的表现更优。具体而言,在 Emotion6 数据集上,EW-MTCNN (ResNet) 模型性能最优(53.60%),在 Artphoto 数据集上,EW-MTCNN (VGGNet) 模型性能最优

(42.19%)。同时,在 Emotion6 和 Artphoto 数据集上,基于 ResNet 或 VGGNet 的模型之间的性能差异都比较小。即 ResNet 和 VGGNet 均可以较好地提取图像中的图像特征信息。

从预训练模型是否进行微调的角度观察,在 Emotion6 和 Artphoto 数据集上进行了微调的情绪识别模型的情绪分类性能均优于只在 ImageNet 数据集上预训练的模型的情绪分类性能。这说明对预训练模型进行微调可以提高情绪识别模型的性能。

另外,EW-MTCNN 模型总体上表现出最优的性能,其准确率比次优的模型在 Emotion6 数据集上高出 5.68%,在 Artphoto 数据集上高出 5.34%。实验结果表明:相比于只使用交叉熵损失进行训练的情绪分类模型,多任务同时进行训练,可以提高情绪识别模型在情绪分类任务上的性能。

4 总结与展望

针对多情绪分析任务,本文提出了一种基于情感轮和多任务卷积神经网络的图像情感分布学习模型。EW-MTCNN 模型将 Mikel's 情感轮心理学先验知识引入情感分布学习中,基于 Mikel's 情感轮计算成对情绪之间的相关程度,采用多任务卷积神经网络以端到端的方式进行训练,有效利用情绪之间的相关性,共同优化情感分布预测与情绪分类任务。在图像情感分布数据集和情绪单标签数据集上的对比实验表明:EW-MTCNN 模型在情感分布预测与情绪分类任务上比现有的 EDL 模型具有更优越的性能。

在下一步的工作中,将对提高图像特征提取层与多任务学习的性能上加以研究,尝试将 EW-MTCNN 的图像特征提取层更换为更加有效的预训练图像特征提取模型,并提出能够更有效的学习情绪间相关性的多任务损失函数,进一步提高图像情感分布学习模型的性能。

5 参考文献

- [1] 王仁武,孟现茹.图片情感分析研究综述[J].图书情报知识,2020(3):119-127.
- [2] JIA Jia,WU Sen,WANG Xiaohui,et al. Can we understand van gogh's mood? learning to infer affects from images in social networks [EB/OL]. [2021-12-12]. <https://doi.org/10.1145/2393347.2396330>.
- [3] BORTH D,JI Rongrong,CHEN Tao,et al. Large-scale visual sentiment ontology and detectors using adjective noun pairs [EB/OL]. [2021-12-11]. <https://dl.acm.org/doi/pdf/10.1145/2502081.2502282>.
- [4] ZHAO Sicheng,GAO Yue,JIANG Xiaolei,et al. Exploring principles-of-art features for image emotion recognition [EB/OL]. [2021-12-11]. <https://dl.acm.org/doi/epdf/10.1145/2647868.2654930>.
- [5] PLUTCHIK R. A general psycho evolutionary theory of emotion [M]. Pittsburgh: Academic Press,1980:3-33.
- [6] ZHANG Minling,ZHOU Zhihua. A review on multi-label learning algorithms [J]. IEEE Transactions on Knowledge and Data Engineering,2013,26(8):1819-1837.
- [7] ZHOU Ying,XUE Hui,GENG Xin. Emotion distribution recognition from facial expressions [EB/OL]. [2021-12-10]. <https://dl.acm.org/doi/10.1145/2733373.2806328>.
- [8] JIA Xiuyi,ZHENG Xiang,LI Weiwei,et al. Facial emotion distribution learning by exploiting low-rank label correlations locally [EB/OL]. [2021-12-04]. <https://ieeexplore.ieee.org/document/8953790>.
- [9] ZHOU Deyu,ZHANG Xuan,ZHOU Yin,et al. Emotion distribution learning from texts [EB/OL]. [2021-12-12]. https://www.researchgate.net/publication/311990212_Emotion_Distribution_Learning_from_Texts.
- [10] XIONG Haitao,LIU Hongfu,ZHONG Bineng,et al. Structured and sparse annotations for image emotion distribution learning [J]. Proceedings of the AAAI Conference on Artificial Intelligence,2019,33(1):363-370.
- [11] MIKELS J A,FREDRICKSON B L,LARKIN G R,et al. Emotional category data on images from the International Affective Picture Sysem [J]. Behavior Research Methods,2005,37(4):626-630.
- [12] GONG Rui,WANG Qing,HAI Yan,et al. Investigation on factors to influence color emotion and color preference responses [J]. Optik,2017,136:71-78.
- [13] MATTHEWS T,NIXON M S,NIRANJAN M. Enriching texture analysis with semantic data [EB/OL]. [2021-12-15]. <https://dl.acm.org/doi/10.1109/CVPR.2013.165>.
- [14] YAO Lei,SURYANARAYAN P,QIAO Mu,et al. Oscar: on-site composition and aesthetics feedback through exemplars for photographers [J]. International Journal of Computer Vision,2012,96(3):353-383.
- [15] COLOMBO C,DEL BIMBO A,PALA P. Semantics in visual information retrieval [J]. Ieee Multimedia,1999,6(3):38-53.
- [16] MACHAJDIK J,HANBURY A. Affective image classification using features inspired by psychology and art theory [EB/OL]. [2021-12-19]. <https://dl.acm.org/doi/10.1145/1873951.1873965>.
- [17] RAO Tianrong,XU Min,XU Dong. Learning multi-level deep representations for image emotion classification [J]. Neural Processing Letters,2020,51(3):2043-2061.
- [18] YANG Jufeng,SHE Dongyu,SUN Ming. Joint image emotion classification and distribution learning via deep convolutional neural network [EB/OL]. [2021-12-19]. <https://dl.acm.org/doi/10.5555/3172077.3172345>.

- [19] GENG Xin. Label distribution learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(7): 1734-1748.
- [20] 曾雪强,罗明珠,陈素芬,等. 基于自适应多重多元回归的人脸年龄估计 [J]. 江西师范大学学报(自然科学版), 2019, 43(1): 68-75.
- [21] ZHANG Yuxiang, FU Jiamei, SHE Dongyu, et al. Text emotion distribution learning via multi-task convolutional neural network [EB/OL]. [2021-12-12]. <https://dl.acm.org/doi/abs/10.5555/3304222.3304409>.
- [22] HE Tao, JIN Xiaoming. Image emotion distribution learning with graph convolutional networks [EB/OL]. [2021-12-12]. <https://dl.acm.org/doi/pdf/10.1145/3323873.3326593>.
- [23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2021-12-18]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [24] GENG Xin, YIN Chao, ZHOU Zhihua. Facial age estimation by learning from label distributions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(10): 2401-2412.
- [25] PENG Kuanchuan, CHEN Tsuhan, SADOVNIK A, et al. A mixed bag of emotions: model, predict, and transfer emotion distributions [EB/OL]. [2021-12-12]. <https://ieeexplore.ieee.org/document/7298687/>.
- [26] MACHAJDIK J, HANBURY A. Affective image classification using features inspired by psychology and art theory [EB/OL]. [2021-12-12]. <https://dl.acm.org/doi/10.1145/1873951.1873965>.
- [27] CAMPOS V, SALVADOR A, GIRÓ-I-NIETO X, et al. Diving deep into sentiment: understanding fine-tuned CNNs for visual sentiment prediction [EB/OL]. [2021-12-12]. <https://arxiv.org/pdf/1508.05056.pdf>.
- [28] YANG Jingyuan, LI Jie, LI Leida, et al. A circular-structured representation for visual emotion distribution learning [EB/OL]. [2021-12-12]. <https://arxiv.org/abs/2106.12450>.
- [29] YANG Jufeng, SUN Ming, SUN Xiaoxiao. Learning visual sentiment distributions via augmented conditional probability neural network [EB/OL]. [2021-12-12]. <https://dl.acm.org/doi/10.5555/3298239.3298273>.
- [30] GAO Binbin, XING Chao, XIE Chenwei, et al. Deep label distribution learning with label ambiguity [J]. IEEE Transactions on Image Processing, 2017, 26(6): 2825-2838.
- [31] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [32] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [EB/OL]. [2021-12-12]. <https://ieeexplore.ieee.org/iel7/7776647/7780329/07780459.pdf>.

The Image Emotion Distribution Learning Based on Emotion Wheel and Multi-Task Convolutional Neural Network

LAI Jinshui, WAN Zhongying, ZENG Xueqiang*

(School of Computer & Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: Image emotion distribution learning can model multiple emotions simultaneously, but existing models lack effective methods to directly consider the correlation between emotions. In response to this problem, the emotion wheel enhanced multi-task convolutional neural network for image emotion distribution learning (EW-MTCNN) model is proposed. Psychological emotional knowledge is directly introduced into the deep learning network. Based on Mikel's emotion wheel to define the correlation between paired emotions, the EW-MTCNN model adopts a multi-task convolutional neural network to learn the correlation information between emotions to jointly optimize the emotion distribution prediction and emotion classification tasks. The EW-MTCNN model consists of 3 modules, namely the emotional wheel prior knowledge layer, the visual feature extraction layer and the multi-task loss layer. Comparative experiments on emotion distribution dataset (Emotion6) and single-label dataset (Artphoto) show that the EW-MTCNN model has better performance than other emotion distribution learning methods on emotion distribution prediction and emotion classification tasks.

Key words: Mikel's emotion wheel; multi-task convolutional neural network; emotion distribution learning; emotion classification; label distribution learning

(责任编辑: 冉小晓)