计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(6):286−292 [doi: 10.15888/j.cnki.csa.009148]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于多头图注意力网络与图模型的多标签图像分类①

石琇赟, 李顺勇, 韩　翔

(山西大学 数学科学学院, 太原 030006)
通信作者: 李顺勇, E-mail: lisy75@sxu.edu.cn

摘　要: 多标签图像分类是多标签数据分类问题中的研究热点. 针对目前多标签图像分类方法只学习图像的视觉表示特征, 忽略了图像标签之间的相关信息以及标签语义与图像特征的对应关系等问题, 提出了一种基于多头图注意力网络与图模型的多标签图像分类模型 (ML-M-GAT). 该模型利用标签共现关系与标签属性信息构建图模型, 使用多头注意力机制学习标签的注意力权重, 并利用标签权重将标签语义特征与图像特征进行融合, 从而将标签相关性与标签语义信息融入到多标签图像分类模型中. 为验证本文所提模型的有效性, 在公开数据集 VOC-2007 和 COCO-2014 上进行实验, 实验结果表明, ML-M-GAT 模型在两个数据集上的平均均值精度 (mAP) 分别为 94% 和 82.2%, 均优于 CNN-RNN、ResNet101、MLIR、MIC-FLC 模型, 比 ResNet101 模型分别提高了 4.2% 和 3.9%. 因此, 本文所提的 ML-M-GAT 模型能够利用图像标签信息提高多标签图像分类性能.

关键词: 图像分类; 残差神经网络; 多头注意力; 图模型

引用格式: 石琇赟,李顺勇,韩翔.基于多头图注意力网络与图模型的多标签图像分类.计算机系统应用,2023,32(6):286–292. http://www.c-s-a.org.cn/1003-3254/9148.html

Multi-label Image Classification Based on Multi-head Graph Attention Network and Graph Model

SHI Xiu-Yun, LI Shun-Yong, HAN Xiang

(School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China)

Abstract: Multi-label image classification is a research hotspot in multi-label data classification. The existing multi-label image classification methods only learn the visual representation features of images and ignore the relevant information between image labels and the correspondence between label semantics and image features. In order to solve these problems, a multi-label image classification model based on a multi-head graph attention network and graph model (ML-M-GAT) is proposed. By using label co-occurrence and attribute information, the model builds a graph model, and it employs the multi-head attention mechanism to learn the attention weight of the label. In addition, the model utilizes label weights to fuse label semantic features and image features, so as to integrate label correlation and label semantic information into the multi-label image classification model. In order to verify the effectiveness of the proposed model, experiments are carried out on the public datasets VOC-2007 and COCO-2014, and the experimental results show that the average mean accuracy (mAP) of the ML-M-GAT model on the two datasets is 94% and 82.2%, respectively, which are better than that of CNN-RNN, ResNet101, MLIR, and MIC-FLC models and are 4.2% and 3.9% higher than that of ResNet101 models, respectively. Therefore, the proposed model can improve the performance of multi-label image classification by using image label information.

Key words: image classification; residual neural network (RNN); multi-head attention; graph model

① 基金项目: 国家自然科学基金 (82274360, 61976128); 2022 年度山西省研究生教育教学改革课题 (2022YJJG010); 山西省高等学校教学改革创新项目 (J2021059); 高等学校大学数学教学研究与发展中心项目 (CMC20210315)
收稿时间: 2022-12-06; 修改时间: 2023-01-17; 采用时间: 2023-02-03; csa 在线出版时间: 2023-04-25
CNKI 网络首发时间: 2023-04-27

传统的分类器大多是依据单标签数据设计的, 但是随着数据资源的迅猛增长, 多标签数据分类问题成为研究热点, 多标签图像分类是其中一个重要的研究方向. 传统的机器学习算法, 例如 K 近邻算法 (K-nearest neighbors, KNN)[1]、支持向量机 (support vector machine, SVM)[2] 等主要用于单标签图像分类, 而深度学习算法, 例如卷积神经网络 (convolutional neural networks, CNN)、图卷积网络 (graph convolutional networks, GCN) 等能够很好地提取图像特征用于多标签图像分类, 但大部分模型都只学习图像的视觉表示特征, 忽略了图像标签的语义特征信息. 本文提出一种基于多头图注意力机制的图像多标签分类模型, 通过建立多头图注意力机制, 对图像标签的注意力权重进行学习, 获得标签之间的不对称相关关系. 并将标签的注意力权重与图像视觉特征进行融合.

现有的多标签图像分类模型大致可以分为 3 类: 不基于标签间的相关关系进行图像分类的模型、基于标签对之间的关系信息进行图像分类的模型以及基于所有标签之间的关系信息进行图像分类的模型. 根据模型利用标签相关性信息的程度高低可以将多标签图像分类算法分为一阶、二阶和高阶的算法. Amorim 等[3] 提出了一种基于单标签分类最优连通性的半监督学习方法, 并将其扩展到多标签分类. Huang 等[4] 提出了包含多个二元分类器的多标签分类算法 LLSF, 该方法将每个类别标签视为二元分类问题, 通过学习每个类标签的特定数据表示特征进行类别单标签分类. 这些方法虽然有一定效果, 但是无法有效利用图像标签之间相互依赖的相关性信息. Huang 等[5] 又提出了 LPLC 贝叶斯模型, 该模型通过学习局部正负成对标签相关性进行多标签分类. Wang 等[6] 提出 CNN-RNN 来学习标签语义特征依赖性以及标签相关性, 并将两个信息集成到一个统一的框架中. Chen 等[7] 提出了基于 GCN 的多标签图像分类模型, 该方法通过 ResNet101 模型[8] 提取图像特征, 通过图卷积神经网络 (GCN)[9] 学习标签特征. 最近研究[10,11] 通过构建不同的卷积神经网络 (CNN) 模型框架, 可以同时识别多标签图像的标签语义信息和标签相关性, 并通过 CNN 中不同的优化器自适应学习多标签图像分类器. 虽然, 深度网络模型在多标签图像分类中能够很好地学习图像特征以及图像标签关系特征, 但随着网络的加深也会伴随模型过拟合、梯度爆炸的问题. 最近, 越来越多的学者将注意力机制融入深度学习模型[12,13] 进行图像分类任务, 并取得了良好的分类效果.

随着深度学习的进一步发展, 注意力机制被广泛应用于计算机视觉的各项任务中, 注意力机制通过对重要信息分配更高的权重提升模型分类性能. Transformer 的初次提出是为了解决机器翻译问题, 因为它能捕捉到全局的上下文信息, Transformer 的全局属性主要体现在它的编码方式和多头注意力机制 (multiple head attention, MHA)[14]. 王延召[15] 提出了基于多头自注意力机制的三维点云分类方法, 使用多个并行的自注意力模块分别从不同的特征维度提取各个特征向量之间的关联信息并将结果融合, 以此来提升模型的分类性能. 李金星等[14] 利用融合多头注意力机制关注全局特征, 通过交叉注意力综合提取 X 线胸片图像的浅层直观特征和深层抽象特征, 使所提模型具有优异的肺炎诊断分类性能. 张健飞等[16] 提出了一种以结构振动加速度信号为输入的基于多头自注意力的 CNN 模型, 利用多头自注意力机制学习输入数据的全局特征, 提高了模型的识别精度与辨识能力.

基于以上, 本文提出一种基于多头图注意力机制与图模型的多标签图像分类模型 (multi-label image classification algorithm based on multi-head graph attention network and graph model, ML-M-GAT), 该模型利用标签共现关系与标签属性信息构建图模型, 使用多头注意力机制学习标签的注意力权重, 并利用标签权重将标签语义特征与图像特征进行融合, 从而将标签相关性与标签语义信息融入到多标签图像分类模型中. 最后, 在两大公开数据集上与多种模型进行对比实验, 验证了本文所提模型的有效性.

1 本文算法

本文提出 ML-M-GAT, 该算法主要包括 4 个部分: 基于 ResNet101 模型的图像特征提取模块、基于词嵌入和图结构的标签向量转换模块、基于多头图注意力机制的标签注意力权重学习模块以及基于融合特征的分类器模块, 模型结构如图 1 所示.

由图 1 可知, ML-M-GAT 利用 ResNet101 模型提取每一张输入图像的特征, 采用词嵌入模型获得标签的词嵌入矩阵, 并结合标签类别共现矩阵转换为图结构, 将标签信息图输入多头图注意力模型获得标签间不对称相关关系权重矩阵. 为匹配该权重矩阵, 在图像特征提取模块后添加特征降维模块, 最后将降维后的图像特征与标签注意力权重进行融合, 输入多标签分类器进行分类预测.

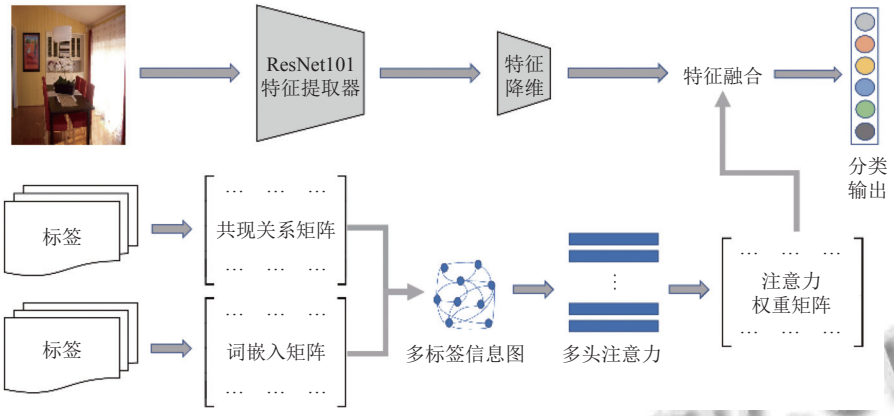

图 1 ML-M-GAT 模型结构

1.1 图像特征提取-降维模块

残差网络 (residual network, ResNet) 在 2015 年由 He 等[17] 提出, 解决了 CNN 模型深度加深出现的梯度爆炸、消失问题, 最具代表性的是: ResNet50、ResNet101 等. 如图 2 所示, 残差网络从输入 X 引出一条快速连接, 与经过两层卷积层处理后的特征相加, 最后通过 ReLU 函数得到 $H(X)$.

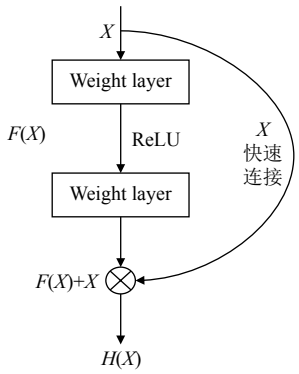

图 2 Residual block 模块

ML-M-GAT 使用在训练数据集上训练好的 ResNet101 模型, 并修改 ResNet101 模型最后的全连接层参数, 该层原始参数设定输入维度为 2 048, 输出维度为图像标签种类数, 保持输入参数不变, 修改输出参数为 2 048 后得到多标签图像特征提取器. ML-M-GAT 将图像 I_i 尺寸裁剪为 224×224 后输入 ResNet101 图像特征提取器, 获得多标签图像的特征张量 f_i, 计算过程如式 (1):

$$f_i = f_{\text{ResNet}}(I_i;\theta_{\text{ResNet}}) \in \mathbb{R}^{W\times H\times D} \tag{1}$$

其中, f_{ResNet} 表示 ResNet101 图像特征提取器, I_i 表示第 i 张图像, θ_{ResNet} 表示 ResNet101 图像特征提取器的参数, W、H、D 分别是特征张量 f_i 的长、宽与通道数.

为了最终将图像特征与标签注意力权重进行融合, 匹配图像特征与标签注意力权重矩阵的维度, 需要对多标签图像的特征张量 f_i 进行降维. 同时, 为了最大程度的保留图像原始信息并简化模型, ML-M-GAT 在图像特征提取模块设置特征张量 f_i 的长、宽均为 1, 因此只需对通道数 D 进行降维. ML-M-GAT 通过一层卷积层 conv1 得到降维后的特征向量 x_i, 计算过程如式 (2):

$$x_i = f_{\text{conv1}}(f_i;\theta_{\text{conv1}}) \in \mathbb{R}^d \tag{2}$$

其中, f_{conv1} 表示卷积层 conv1, θ_{conv1} 表示卷积层 conv1 的参数, d 为降维后特征张量 x_i 的维数.

1.2 标签向量转换图结构模块

对于给定图像 I_i 的标签序列 $L_i = [l_1, l_2, \cdots, l_n]$, ML-M-GAT 通过预训练 Word2Vec 模型[18] 获得对应的标签高维表征向量 $a_i \in \mathbb{R}^{d'}$. 由于多标签图像数据中标签的种类数远小于标签高维表征向量的维度 d', 这不利于 ML-M-GAT 中 M-GAT 标签注意力权重学习模块获得多标签图像数据集中标签间的不对称相关关系, 同时也会造成后续图像特征与标签注意力权重融合过程中的维度不匹配问题. 因此 ML-M-GAT 采用一层卷积层 conv2 对标签高维表征向量进行降维, 计算过程如式 (3):

$$a_i' = f_{\text{conv2}}(a_i;\theta_{\text{conv2}}) \in \mathbb{R}^{d''} \tag{3}$$

其中, f_{conv2} 表示卷积层 conv2, θ_{conv2} 表示卷积层 conv2 的参数, d'' 为降维后表征向量 a_i' 的维数. 通过统计所有图像的标签信息获得全局标签类别共现关系矩阵 $P \in \mathbb{R}^{C\times C}$, C 为标签的总种类数. ML-M-GAT 通过统计多标签图像数据集中各标签的共现次数, 计算标签

l_i 与标签 l_j 同时标记的概率 $p_{ij}=P(l_j|l_i)$. p_{ij} 表示当标签 l_i 标记情况下标签 l_j 同时标记的概率, 计算过程如式 (4):

$$p_{ij}=P\left(l_j|l_i\right)=\frac{Q_{ij}}{Q_i} \tag{4}$$

其中, Q_i 表示标签 l_i 在多标签数据集中出现的总次数, Q_{ij} 表示标签 l_i 和 l_j 在多标签数据集中同时出现的总次数. 由式 (4) 可知共现概率 $p_{ij}\neq p_{ji}$, 即全局标签类别共现关系矩阵 P 为非对称矩阵. 如图 3 所示, 为多标签图像数据集 VOC-2007 中标签贡献概率矩阵.

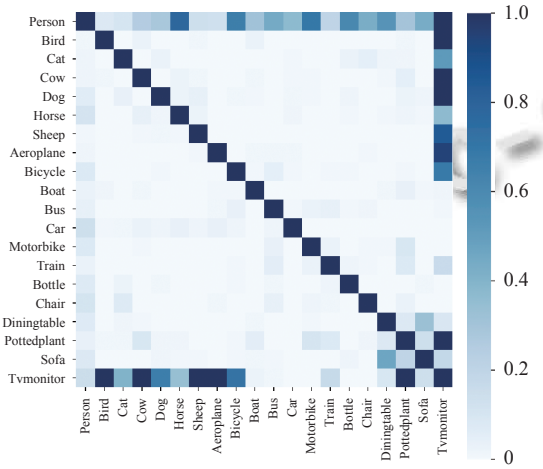

图 3 VOC-2007 数据集标签共现概率可视化

由图 3 可知, VOC-2007 数据集中“person”标签与其他标签间存在强烈的共现关系, 即当“horse”“tvmonitor”等标签出现时通常都伴随着“person”标签的出现, 但由标签贡献概率的非对称性可知, 当“person”标签出现时, “horse”“tvmonitor”等标签不一定伴随出现, 即图像标签之间存在不对称共现关系.

ML-M-GAT 定义图像多标签信息图模型为 $G_l=\{V_l,E_l\}$, 其中节点集合 $V_l=\{v_1,v_2,\cdots,v_C\}$ 代表多标签数据集中各标签节点, 边集合 $E_l=\{e_1,e_2,\cdots,e_C\}$ 代表各标签节点间的共现关系, 并将降维后表征向量 a_i' 作为节点特征赋给每个节点. 当且仅当 $p_{ij}>0$ 时, 给节点 v_i 到节点 v_j 之间连接一条有向边, 并将全局标签类别共现关系矩阵 P 中的元素 p_{ij} 作为权重赋给此有向边.

1.3 M-GAT 标签注意力权重学习模块

ML-M-GAT 采用多头图注意力模型 (multi-head graph attention network, M-GAT)[19] 学习多标签图像数据集中标签间的不对称相关关系, 输入图像多标签信息图 $G_l=\{V_l,E_l\}$, 经过 M-GAT 后输出标签注意力权重矩阵 $Z\in\mathbb{R}^{C\times C}$.

在 M-GAT 的每一个 GAT 中, 首先将线性变换得到的节点 i 和节点 j 的嵌入 $z_i^{(l)}$ 和 $z_j^{(l)}$ 进行拼接, 经过式 (5) 的计算就得到了第 l 层上节点 i 和节点 j 之间未经加工的注意力系数 $e_{ij}^{(l)}$.

$$e_{ij}^{(l)}=LeakyReLU(\vec{a}^{(l)^{\mathrm{T}}}(z_i^{(l)}\|z_j^{(l)})) \tag{5}$$

其中, $z_i^{(l)}=W^l h_i^l$, $\vec{a}^{(l)^{\mathrm{T}}}\in\mathbb{R}^{2M}$ 表示注意力核, $W\in\mathbb{R}^{M\times d^a}$ 为线性变换权重矩阵, $LeakyReLU$ 为激活函数, $\|$ 代表拼接运算.

对节点 i 的所有邻居节点进行归一化, 得到第 l 层中归一化后的注意力系数 $\alpha_{ij}^{(l)}$, 归一化的公式如式 (6).

$$\alpha_{ij}^{(l)}=\frac{\exp(e_{ij}^{(l)})}{\sum_{k\in N(i)}\exp(e_{ik}^{(l)})} \tag{6}$$

其中, $N(i)$ 表示包括节点 i 自身在内的节点 i 的一阶邻居节点集合, $e_{ij}^{(l)}$ 为由式 (5) 得到的注意力系数.

将邻居节点的特征聚合起来, 并且根据注意力系数 $\alpha_{ij}^{(l)}$ 进行缩放, 由式 (7) 得到节点 i 聚合了邻居节点特征得到的新特征 $h_i^{(l+1)}$.

$$h_i^{(l+1)}=\sigma\left(\sum_{j\in N(i)}\alpha_{ij}^{(l)}z_j^{(l)}\right) \tag{7}$$

其中, $\sigma(\cdot)$ 表示非线性激活函数, $\alpha_{ij}^{(l)}$ 为由式 (6) 得到的归一化后的注意力系数, $z_j^{(l)}$ 为节点 j 的嵌入向量 .

最后, ML-M-GAT 利用式 (8) 整合多个注意力机制的输出结果作为对应节点的特征输出.

$$h_i^{(l+1)}=\sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in N(i)}\alpha_{ij}^{k}W^k h_j^{(l)}\right) \tag{8}$$

其中, K 为多头注意力的头数, $\sigma(\cdot)$ 表示非线性激活函数, $N(i)$ 表示包括节点 i 自身在内的节点 i 的一阶邻居节点集合, $\alpha_{ij}^{(l)}$ 为由式 (6) 得到的归一化后的注意力系数, W^k 为线性变换权重矩阵, $h_i^{(l)}$ 为聚合了邻居节点的新特征.

1.4 分类器模块

ML-M-GAT 将 M-GAT 所提取的多标签注意力权重矩阵 Z 与降维后的图像特征向量 x_i 相乘, 得到多标签图像融合特征, 再经过一层全连接层 fc 后, 得到每一张图像的多标签分类预测结果 $\hat{y}_i$, 具体计算过程如式 (10).

$$\hat{y}_i=f_{fc}\left(Zx_i;\theta_{fc}\right)\in\mathbb{R}^C \tag{9}$$

其中, f_{fc}表示全连接层 fc, θ_{fc}表示全连接层 fc 的参数, Z 为多标签注意力权重矩阵, x_i 为降维后的图像特征向量.

针对每一张多标签图像的分类预测结果, ML-M-GAT 使用 multi label soft margin loss 作为模型的损失函数, 具体计算公式如式 (11).

$$\mathcal{L}(\hat{y}_i,y_i)=\frac{1}{C}\sum_{j=1}^{C}y_{ij}\log\left(\left(1+\exp\left(-\hat{y}_{ij}\right)\right)^{-1}\right)+\left(1-y_{ij}\right)\log\left(\frac{\exp\left(-y_{ij}\right)}{1+\exp\left(-y_{ij}\right)}\right) \tag{10}$$

其中, C 为标签的总种类数, $y_i\in\mathbb{R}^C$为第i张多标签图像的真实标签, y_{ij}为y_i的第 j 个元素, $\hat{y}_{ij}$为$\hat{y}_i$的第 j 个元素. ML-M-GAT 模型算法过程如算法 1 所示.

算法 1. ML-M-GAT 模型算法过程

输入: 图像集合$I=\{I_1,I_2,\cdots,I_N\}$, 图像I_i标签集合$L_i=\{l_1,l_2,\cdots,l_n\}$, 通道数 D, 注意力头数 K, 标签种类数 C.
输出: 预测标签集合$Y=[\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_N]$.

1) for (i=1 to N) do
2) $f_i=f_{\text{ResNet101}}(I_i,\theta_{\text{ResNet101}})$
3) $x_i=f_{\text{conv1}}(f_i;\theta_{\text{conv1}})$
4) $a_i=\text{Word2Vec}(L_i)$
5) $a_i'=f_{\text{conv2}}(a_i;\theta_{\text{conv2}})$
6) $p_{ij}=P(l_j|l_i)$
7) $G_l=\{V_l,E_l\}, V_l=\{v_1,v_2,\cdots,v_C\},E_l=\{e_1,e_2,\cdots,e_C\}$
8) $Z=MultiHeadGAT(G_l)$
9) $\hat{y}_i=f_{fc}(Zx_i;\theta_{fc})$
10) return $Y=[\hat{y}_1,\hat{y}_2,\cdots,\hat{y}_N]$.

2 实验结果与分析

2.1 实验环境配置与参数设置

为验证本文所提多标签图像分类模型的有效性, 选取 2 个多标签图像数据集进行实验, 并在多个指标层面与经典的多标签图像分类算法进行对比. ML-M-GAT 使用 Python 进行编程, 软件环境为: Python 3.9、PyTorch 1.12.1, 使用 SGD 作为模型优化器.

ML-M-GAT 中 ResNet101 图像特征提取器获得多标签图像的特征张量f_i的通道数 D=2048, 通过卷积层 conv1 得到降维后的特征向量x_i的维数 d=C, C 为多标签图像数据集中标签的种类数, 预训练 Word2Vec 模型获得对应的标签高维表征向量 a_i 的维数d' = 1024, 通过卷积层 conv2 得到降维表征向量a_i'的维数d'' = 256, M-GAT 中多头注意力模型的头数 K=8. 设置初始学习率为 0.01, 训练周期为 100.

2.2 实验数据集介绍

实验使用 PASCAL visual object classes challenge 2007 (VOC-2007)[20] 和 Microsoft COCO 2014 (COCO-2014)[21] 数据集. VOC-2007 数据集中 train、validation、test 共有 9 963 张图像, 标签总类别数为 20; COCO-2014 数据集中 train、test 共有 123 558 张图像, 标签总类别数为 80. 采用 VOC-2007 完整数据集用于训练测试, 由于 COCO-2014 数据集图片数量庞大, 故从 82 783 张训练图像样本中随机抽取 20 000 张图像样本进行训练. 图 4 为 VOC-2007 数据集的部分示例, 图 5 为 COCO-2014 数据集的部分示例.

图 4 VOC-2007 数据集部分示例

图 5 COCO-2014 数据集部分示例

2.3 实验评价指标

本文使用平均均值精度 (mean average precision, mAP)、平均每类精度 (class precision, CP)、平均每类召回 (class recall, CR)、整体平均精度 (overall precision, OP)、整体平均召回 (overall recall, OR) 作为多标签图像分类模型的评价指标[22–24].

由于 COCO-2014 数据集测试集包含 40 000 多张图像, 故本文在进行测试时从所有测试图像中随机抽取 400 张图像作为实验测试集, 随机抽取 3 次, 取 3 次测试实验评价指标得分的平均值作为最后的指标得分.

2.4 实验结果分析

ML-M-GAT 在 VOC-2007 数据集上训练 30 次后 mAP 已经达到 90% 并趋于饱和, loss 也已经降至 0.1 以下并趋于稳定. ML-M-GAT 在 COCO-2014 数据集上训练 50 次后 mAP 已经达到 80% 并趋于饱和, loss

也已经降至 0.1 以下并趋于稳定.

选取 CNN-RNN、ResNet101、MLIR、MIC-FLC 共 4 种多标签图像分类算法与本文所提 ML-M-GAT 模型进行对比实验分析. CNN-RNN[6] 通过卷积神经网络与序列神经网络学习标签语义特征依赖性以及标签相关性, 最终实现多标签图像分类. 基于 ResNet101[8] 构建的多标签图像分类模型通过引入残差学习, 解决了传统卷积网络在信息传递过程中造成的信息丢失、损耗问题. MLIR[25] 在利用 ResNet101 对图像进行特征提取的过程中引入了注意力机制, 将标签和图像特征投影到公共潜在向量空间完成多标签图像分类. MIC-FLC[26] 通过交替学习多标签分类器和新类检测器实现多标签图像分类问题. 表 1 为 5 种算法在 VOC-2007 数据集上的实验结果, 表 2 为 5 种算法在 COCO-2014 数据集上的实验结果, 加粗字体表示各指标的最优表现.

表 1 VOC-2007 数据集上的实验结果 (%)

算法	mAP	CP	OP	CF1	CR
CNN-RNN	84.0	—	—	—	—
ResNet101	89.9	80.7	82.0	82.0	**83.2**
MLIR	91.9	85.3	85.9	81.9	78.7
MIC-FLC	90.4	84.7	85.5	82.0	79.5
ML-M-GAT	**94.0**	**95.5**	**96.4**	**82.0**	82.1

表 2 COCO-2014 数据集上的实验结果 (%)

算法	mAP	CP	OP	CF1	CR
CNN-RNN	62.0	—	—	—	—
ResNet101	78.3	80.2	70.8	**72.8**	66.7
MLIR	80.5	79.1	84.0	67.4	58.7
MIC-FLC	80.0	78.8	84.0	66.2	57.2
ML-M-GAT	**82.2**	**85.8**	**86.9**	68.2	**67.9**

由表 1 可知, 本文算法在 VOC-2007 数据集上 mAP 达到了 94.0%, 相较于 CNN-RNN 、ResNet101、MLIR、MIC-FLC 模型, mAP 分别提升了 10%、4.1%、2.1%、3.6%; CP 和 OP 都达到了 95.5% 和 96.4%, 为所有对比算法中的最优; CF1 达到 82%, 与 ResNet101 和 MIC-FLC 模型持平; CR 达到 82.1%, 仅次于 ResNet101 模型.

由表 2 可知, 本文算法在 COCO-2014 数据集上 mAP 达到了 82.2%, 相较于 CNN-RNN 、ResNet101、MLIR、MIC-FLC 模型, mAP 分别提升了 20.2%、3.9%、1.7%、2.2%; CP 和 OP 都达到了 85.8% 和 86.9%, 为所有对比算法中的最优; CF1 达到了 68.2%, 仅次于 ResNet101 模型; CR 达到了 67.9%, 为所有对比算法中的最优.

由实验结果可以证明, 本文所提出的一种基于多头图注意力机制与图模型的多标签图像分类模型 (ML-M-GAT), 具有较好的多标签图像分类效果.

3 结束语

充分挖掘图像标签之间的相关关系, 是提升多标签图像分类模型精度的一大研究热点, 本文提出一种基于多头图注意力机制与图模型的多标签图像分类模型 (ML-M-GAT), 该模型在利用 ResNet101 模型提取图像特征的基础上, 利用标签共现关系与标签属性信息构建图模型, 使用多头注意力机制学习标签的注意力权重, 并利用标签权重将标签语义特征与图像特征进行融合, 从而将标签相关性与标签语义信息融入到多标签图像分类模型中. 通过在 VOC-2007 和 COCO-2014 数据集上与 4 种多标签图像分类算法进行对比实验分析, ML-M-GAT 在多个多标签图像分类指标上取得较好结果, 验证了模型的有效性. 下一步将关注多标签数据集中样本分布不均衡问题, 从平衡样本分布角度继续深入研究.

参考文献

1 Zhou NR, Liu XX, Chen YL, *et al*. Quantum K-nearest-neighbor image classification algorithm based on K-L transform. International Journal of Theoretical Physics, 2021, 60(4): 1209–1224.

2 Yousefi S, Mirzaee S, Almohamad H, *et al*. Image classification and land cover mapping using Sentinel-2 imagery: Optimization of SVM parameters. Land, 2022, 11(7): 993. [doi: 10.3390/land11070993]

3 Amorim WP, Falcão AX, Papa JP. Multi-label semi-supervised classification through optimum-path forest. Information Sciences, 2018, 465: 86–104. [doi: 10.1016/j.ins.2018.06.067]

4 Huang J, Li GR, Huang QM, *et al*. Learning label-specific features and class-dependent labels for multi-label classification. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(12): 3309–3323. [doi: 10.1109/TKDE.2016.2608339]

5 Huang J, Li GR, Wang SH, *et al*. Multi-label classification by exploiting local positive and negative pairwise label correlation. Neurocomputing, 2017, 257: 164–174. [doi: 10.1016/j.neucom.2016.12.073]

6 Wang J, Yang Y, Mao JH, *et al*. CNN-RNN: A unified framework for multi-label image classification. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 2285–2294.
7 Chen ZM, Wei XS, Wang P, *et al*. Multi-label image recognition with graph convolutional networks. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5172–5181.
8 Tan MX, Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 6015–6114.
9 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2016.
10 Song LY, Liu J, Qian BY, *et al*. A deep multi-modal CNN for multi-instance multi-label image classification. IEEE Transactions on Image Processing, 2018, 27(12): 6025–6038. [doi: 10.1109/TIP.2018.2864920]
11 Kareem RSA, Ramanjineyulu AG, Rajan R, *et al*. Multilabel land cover aerial image classification using convolutional neural networks. Arabian Journal of Geosciences, 2021, 14(17): 1681. [doi: 10.1007/s12517-021-07791-z]
12 宋一格, 王宁, 李宏昌, 等. 基于分组卷积与双注意力机制的河流水面污染图像分类. 计算机系统应用, 2022, 31(9): 250–256. [doi: 10.15888/j.cnki.csa.008688]
13 李文书, 王志骁, 李绅皓, 等. 基于注意力机制的弱监督细粒度图像分类. 计算机系统应用, 2021, 30(10): 232–239. [doi: 10.15888/j.cnki.csa.008141]
14 李金星, 孙俊, 李超, 等. 融合多头注意力机制的新冠肺炎联合诊断与分割. 中国图象图形学报, 2022, 27(12): 3651–3662.
15 王延召. 基于多头自注意力机制的三维点云分类分割方法研究 [硕士学位论文]. 哈尔滨: 哈尔滨理工大学, 2022.
16 张健飞, 黄朝东, 王子凡. 基于多头自注意力机制和卷积神经网络的结构损伤识别研究. 振动与冲击, 2022, 41(24): 60–71.
17 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016. 770–778.
18 席笑文, 郭颖, 宋欣娜, 等. 基于 Word2Vec 与 LDA 主题模型的技术相似性可视化研究. 情报学报, 2021, 40(9): 974–983. [doi: 10.3772/j.issn.1000-0135.2021.09.007]
19 Zhang BY, Ling HF, Li P, *et al*. Multi-head attention graph network for few shot learning. Computers, Materials & Continua, 2021, 68(2): 1505–1517.
20 Everingham M, van Gool L, Williams CKI, *et al*. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision, 2010, 88(2): 303–338. [doi: 10.1007/s11263-009-0275-4]
21 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.
22 朱旭东, 熊赟. 基于多层次注意力与图模型的图像多标签分类算法. 计算机工程, 2022, 48(4): 173–178, 190.
23 陈琳琳, 朱惠娟, 朱俊, 等. 基于卷积神经网络的多尺度注意力图像分类模型. 南京理工大学学报, 2020, 44(6): 669–675. [doi: 10.14177/j.cnki.32-1397n.2020.44.06.005]
24 吴东东. 基于图神经网络的多标签图像分类研究 [硕士学位论文]. 西安: 电子科技大学, 2021.
25 Wen SP, Liu WW, Yang Y, *et al*. Multilabel image classification via feature/label co-projection. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 51(11): 7250–7259. [doi: 10.1109/TSMC.2020.2967071]
26 Zhang Y, Wang Y, Liu XY, *et al*. Large-scale multi-label classification using unknown streaming images. Pattern Recognition, 2020, 99: 107100. [doi: 10.1016/j.patcog.2019.107100]

(校对责编: 孙君艳)