

doi:10.16055/j.issn.1672-058X.2022.0001.005

基于图注意力网络的多标签图像分类模型

张辉宜, 张进, 黄俊

(安徽工业大学 计算机科学与技术学院, 安徽 马鞍山 243000)

摘要:针对 ML-GCN 中标签共现嵌入维度过高影响模型分类性能和 ML-GCN 中没有充分发掘标签之间不对称关系的问题,提出一种基于图注意力网络的多标签图像分类模型 ML-GAT;ML-GAT 模型首先对高维标签语义嵌入矩阵进行降维;然后通过降维后的低维标签语义嵌入表示和标签类别共现图得到标签共现嵌入;与此同时 ML-GAT 将多标签原始图像输入卷积神经网络进行图像通用特征提取,将卷积神经网络提取出的多标签图像通用特征按照图注意力网络计算得到的标签共现嵌入的维度进行维度统一;最后 ML-GAT 融合标签共现嵌入和图像通用特征得到每一张多标签图像的标签预测评分;在 VOC 2007 与 MS-COCO 2014 上的实验结果表明:在训练样本充分且标签类别数足够多的情况下,ML-GAT 取得了较好的实验结果,通过和其他模型比较分析,ML-GAT 模型所采取的策略可以一定程度上提升模型的多标签图像分类性能。

关键词:多标签分类;图注意力网络;卷积神经网络;深度学习

中图分类号:TP181

文献标志码:A

文章编号:1672-058X(2022)01-0034-08

0 引言

传统监督学习中每个样本只含有一个语义信息,但是现实世界的数据往往含有多个类别的语义信息,即单个样本关联着多个语义标签。例如,一幅天空的图像可以同时标注“蓝天”、“白云”等语义标签;一段新闻文档可以同时属于“时事”、“政策”等多个类别。针对这些含有多个语义标签的多标签数据,如果只考虑单一语义标签对其进行学习,就很难获得很好的分类效果。多标签学习的应用领域十分广泛,包含了图像分类^[1]、文本分类^[2]、音乐分类^[3]以及生物学分类^[4]等多个领域。随着现实生活中

多标签图像数量越来越多、种类越来越复杂,多标签学习在图像分类上的应用也显得更加重要。利用标签之间的相关性可以提升多标签模型的性能^[5]。根据标签相关性挖掘的程度,可以将多标签分类模型分为 3 类:没有利用到标签相关性的一阶方法^[6-7];挖掘标签对之间关系的二阶方法^[8-9];和利用所有标签或类别标签子集中标签关系的高阶方法^[10-11]。初期采用浅层分类模型^[12-13]对多标签图像进行分类,在人工干预提取数据特征的情况下,浅层模型一般都能取得较好的分类结果。近十年来应用深度学习理论^[14]尤其是卷积神经网络(Convolutional Neural Network, CNN)构建了一批经典深度卷积神经网络模型,例如 AlexNet^[15]、

收稿日期:2021-01-05;修回日期:2021-03-03.

基金项目:国家自然科学基金项目(61806005);安徽省高校自然科学研究重点项目(KJ2018A0050);安徽省教育厅教学研究重点项目(2018JYXM1050).

作者简介:张辉宜(1963—),男,四川富顺人,教授,从事机器学习研究.

通讯作者:张进(1996—),男,江苏南京人,硕士研究生,从事机器学习研究. Email:jzhang_ahut@foxmail.com.

VGG^[16]、ResNet^[17],这些模型可以对大量多标签图像样本进行有效的深层特征学习,但单独利用卷积神经网络对多标签图像进行分类缺乏对标签相关性的利用,这会影响模型的性能,因此多标签分类的现有工作往往会利用标签相关性以提高性能。在标签相关性中标签的共现关系可以通过概率图模型很好地表述,在以往的研究工作中,有很多基于这种数学理论的方法可以对标签关系进行建模^[18,19],但是概率图模型的计算成本过高,为了解决这个问题,使用递归网络将标签编码为嵌入向量,以实现标签间相关性建模的方法被提出^[20],该方法也存在着递归神经网络模型依赖于预定义或学习的标签顺序的不足,且无法很好地获得标签全局依赖性。2019年Chen等提出了ML-GCN模型^[21],ML-GCN利用训练数据集中所有标签类别的标签共现关系建立了整体的标签相关性,在最终分类阶段使用图卷积网络(Graph Convolutional Network, GCN)^[22]传播标签共现嵌入并将标签共现嵌入与CNN特征合并,但是ML-GCN学习到的标签共现嵌入维度远远高于需要分类的标签类别数,这会严重影响模型的性能。提出基于图注意力网络(Graph Attention Network, GAT)^[23]的多标签图像分类模型ML-GAT,ML-GAT

采用降维^[24]的方法对ML-GCN中标签共现嵌入维度过高的问题进行改进,同时采用图注意力网络对标签之间的关系进行更加精确的建模。

1 多标签图注意力网络模型架构

针对通过图卷积神经网络得到标签共现嵌入维度过高的问题,ML-GAT采用词嵌入降维模块对高维双向Transformer的表征编码器(Bidirectional Encoder Representation from Transformers, BERT)^[25]标签语义嵌入表示矩阵进行降维,得到低维标签语义嵌入表示矩阵。为了学习标签之间非对称的关系特征,将低维标签语义嵌入表示矩阵和标签类别共现图输入GAT,获取标签共现嵌入模块,得到维度合适的低维标签共现嵌入。同时ML-GAT采用图像特征提取模块提取图像特征。为了匹配低维标签共现嵌入维度,图像特征需要经过图像特征降维模块进行降维,在降维的同时也减少了图像特征中的冗余部分。最后,将标签共现嵌入与降维后的图像特征通过图像特征与标签共现嵌入融合模块进行融合,得到多标签预测评分。多标签图注意力网络模型结构如图1所示。

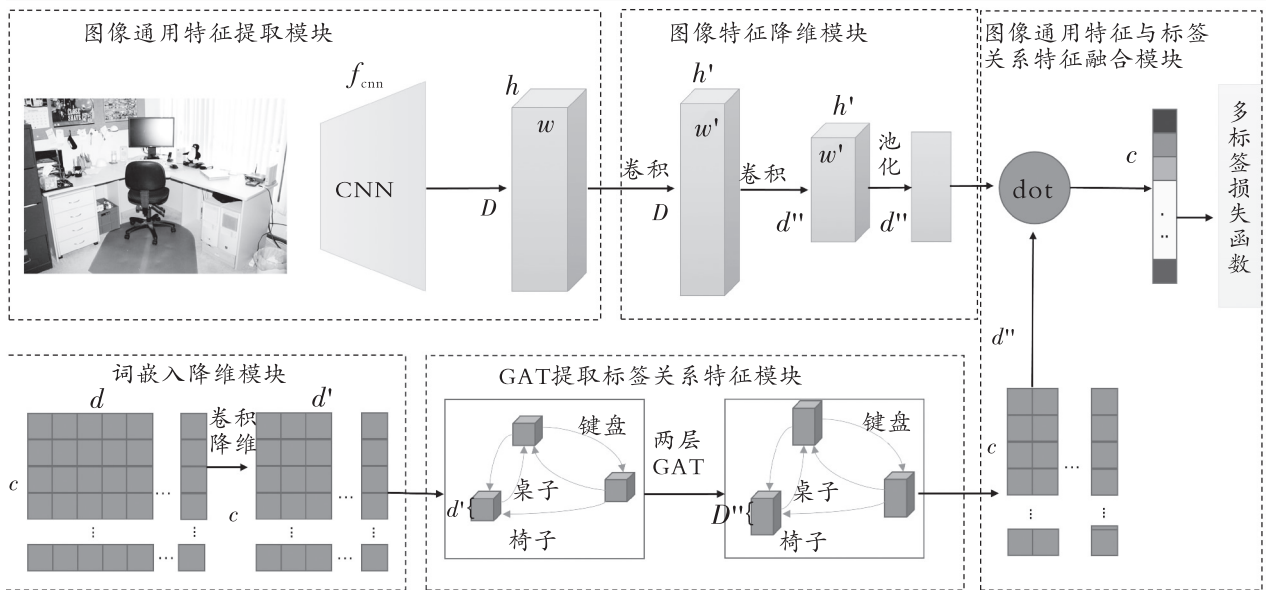


图 1 多标签图注意力网络模型结构

Fig. 1 Model structure of multi-label graph attention network

1.1 图像特征提取模块

ML-GAT 中图像通用特征提取模块使用 101 层 ResNet, 即 ResNet-101 模型。ResNet-101 是目前主流的 CNN 之一, 其优点是易于调整, 可以比较方便地利用在多标签图像分类任务上, 并且有较强的特征提取能力。因为 ML-GAT 采用的是在 ImageNet 上预训练的 ResNet-101, 所以需要去除用来对 ImageNet 进行分类的全连接层, 为了控制图像维度, 需要同时去除 ResNet-101 的自适应池化层, 这样可以得到多标签图像特征提取器。将解析度为 448×448 的多标签图像样本 I 输入多标签图像特征提取器, 可提取多标签图像的特征图 F :

$$F = f_{\text{ResNet}}(I; \theta_{\text{ResNet}}) \in \mathbf{R}^{W \times H \times D}$$

其中, 特征图 F 的长宽为 W, H , 通道数为 D , f_{ResNet} 表示图像通用特征提取模块, θ_{ResNet} 是 ResNet-101 模型参数。

1.2 图像特征降维模块

因为在图像通用特征与标签共现嵌入融合模块中需要将图像特征与标签共现嵌入维度进行匹配, 同时对图像特征进行降维, 可以一定程度上提高图像特征的判别力, 故在 ML-GAT 中采取以下步骤对特征图 F 的长宽 W, H 以及通道数 D 进行降维, F 首先通过卷积层 conv1 下采样, 得到 $F' \in \mathbf{R}^{W' \times H' \times D}$, W' 和 H' 代表降维后特征图 F' 的长与宽, 再通过一层卷积层 conv2 对特征图 F' 的通道数 D 进行降维, 得到 $F'' \in \mathbf{R}^{W' \times H' \times d''}$, d'' 为降维后 F'' 的通道数, 最后经过全局最大值池化层 GMP, 提取多标签图像的特征纹理, 去除无用特征。这样可以为每一张图像提取一个维度为 $\mathbf{R}^{d''}$ 的图像特征向量 x :

$$x = f_{\text{GMP}}(f_{\text{conv2}}(f_{\text{conv1}}(F); \theta_{\text{conv1}}); \theta_{\text{conv2}}) \in \mathbf{R}^{d''}$$

其中, f_{GMP} 为全局最大值池化运算, f_{conv1} 和 f_{conv2} 分别为卷积层 conv₁ 与 conv₂ 进行的卷积运算, θ_{conv1} , θ_{conv2} 分别为卷积层 conv₁ 与卷积层 conv₂ 的模型参数。

1.3 词嵌入降维模块

高维 BERT 标签语义嵌入表示虽然可以很好地表达标签自身的含义, 但是其维度较高, 所以在数据总的标签数较少时, 应该采用合适的降维方式对 GAT 学习得到的标签共现嵌入维度进行控制。为

了对 GAT 得到的标签共现嵌入进行降维, ML-GAT 在词嵌入降维模块采用一层卷积核长度为 f 宽度为 1 的卷积层, 对维度为 $\mathbf{R}^{C \times L}$ 的高维 BERT 标签语义嵌入矩阵 Z^0 表示进行下采样, C 表示标签类别数, L 表示预训练 BERT 标签次嵌入矩阵维度, 进行卷积计算时水平步长为 f , 垂直步长为 1。这样可以得到维度为 $\mathbf{R}^{C \times \frac{L}{f}}$ 的低维标签语义嵌入表示矩阵 Z^1 , 其中 $\frac{L}{f} = d'$ 。

1.4 GAT 获取标签共现嵌入模块

为了提取含有标签与标签之间的非对称关系的标签共现嵌入, ML-GAT 采用了 GAT 获取标签共现嵌入模块, 输入标签类别共现关系图与低维标签语义嵌入表示矩阵 Z^1 , Z^1 每一行都是一个类别标签的低维标签语义嵌入表示向量, $Z^1 = [\vec{z}_1, \vec{z}_2; \dots, \vec{z}_c]; \vec{z}_i \in \mathbf{R}^{d'}$, 经过两层 GAT 后输出标签共现嵌入 $Z^{1+2} \in \mathbf{R}^{C \times d''}$ 。

GAT 首先针对每一个标签节点 i , 计算标签节点 i 与包括自身节点自身在内的所有邻居节点 j 之间的相关系数 e_{ij} :

$$e_{ij} = a([\mathbf{W} \vec{z}_i \parallel \mathbf{W} \vec{z}_j]), j \in N_i$$

其中, $\mathbf{W} \in \mathbf{R}^{M \times d'}$ 是共享参数, 对标签节点 i 的特征 z_i 和标签节点 j 的特征 z_j 进行增维, 增维后的维度为 M , 在 ML-GAT 中最后一层 $M = d''$, $[\cdot \parallel \cdot]$ 表示对标签节点 i, j 的特征进行拼接可以将两个维度为 \mathbf{R}^M 的向量拼接为 \mathbf{R}^{2M} 的向量, $j \in N_i$ 是与标签节点 i 存在共现关系的一跳邻居节点。 $a \in \mathbf{R}^{2M}$ 运算将拼接特征映射到一个实数上。

对相关系数采用归一化运算得到注意力系数:

$$a_{ij} = \frac{\exp(\text{LeakyReLU}(e_{ij}))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(e_{ik}))}$$

其中, LeakyReLU 是激活函数。用注意力系数 a_{ij} 用来计算每个节点的最终输出特征:

$$z'_i = \sigma\left(\sum_{j \in N_i} a_{ij} \mathbf{W} z_j\right)$$

其中, σ 为非线性激活函数。因为通过计算得到的标签 i 对标签 j 的注意力系数与标签 j 对标签 i 的注意力系数不同, 所以 GAT 得到的标签节点特征可以一定程度上表示多标签学习中标签与标签之间的非

对称关系,例如“飞机”和“天空”这一对标签,有“天空”这一标签时“飞机”有着很小的概率同时出现,而“飞机”标签出现时则会大概率伴随着“天空”标签的出现。ML-GAT 采用 GAT 可以单独为每一对标签计算注意力系数,得到能更加准确表达标签之间关系的标签共现嵌入。

在 GAT 获取标签共现嵌入模块,经过两层 GAT 的计算,可以得到一个带有类别标签间非对称关系,维度为 $\mathbf{R}^{C \times d'}$ 的标签共现嵌入 \mathbf{Z}^{l+2} 。每一层的 GAT 计算为

$$\mathbf{Z}^{l+1} = f_{\text{GAT}}(\mathbf{Z}^l, \mathbf{U}) + \mathbf{Z}^l$$

其中, f_{GAT} 表示一层 GAT 计算, $\mathbf{U} \in \mathbf{R}^{C \times C}$ 表示标签节点从标签类别共现图中获得的相关矩阵建立方式与 ML-GCN 中相同, \mathbf{U} 中元素 u_{ij} 取值取决于类别标签 i 与类别标签 j 之间的共现次数,为了能更好地将上一层的信息传递到下一层,因此在计算时将加上之前一层 GAT 的计算结果 \mathbf{Z}^l 。

1.5 图像通用特征与标签共现嵌入融合模块

将维度为 \mathbf{R}^d 的图像特征向量 \mathbf{x} 与多标签节点特征矩阵 \mathbf{Z}^{l+2} 进行矩阵乘法,得到在第 i 张多标签图像上的标签预测评分 \hat{y}_i :

$$\hat{y}_i = \mathbf{Z}^{l+2} \mathbf{x}$$

对于一张多标签图像样本,本模型使用的多标签分类损失函数(Multi-label Soft Margin Loss):

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

其中,第 i 张图像的真实标签为 $\mathbf{y}_i \in \mathbf{R}^C$, \hat{y}_{ij} 和 y_{ij} 分别表示 \hat{y}_i 和 y_i 的第 j 个元素。

2 模型参数设置与评价指标设计

2.1 实验平台及模型参数设置

实验所采用的软硬件环境为 Intel Pentium G4560 @ 3.50 GHz, NVIDIA GeForce GTX 1080Ti 11 GB 显卡, 12 GB 内存, 操作系统为 Ubuntu 16.04, 编程语言为 Python, 深度学习框架为 Pytorch 1.5。

ML-GAT 在两种常用多标签图像数据集上进行对比实验,分别是: Microsoft COCO 2014 (MS-COCO 2014)^[26] 和 PASCAL Visual Object Classes Challenge (VOC 2007)^[27]。MS-COCO 2014 拥有 80 个类别的多标签图像,包含 82 081 张图像组成的训练集和 40 504 张图像组成的验证集,平均每张图像都拥有 2.9 个类别标签。VOC 2007 数据集包含 9 963 张图像组成的训练集、验证集和测试集,包含 20 个常见物体类别标签。

在 ML-GAT 中,将维度为 $\mathbf{R}^{C \times L}$ 的高维 BERT 标签语义嵌入矩阵 \mathbf{Z}^0 输入到词嵌入降维模块,预训练 BERT 标签次嵌入矩阵维度 L 取值为 1 024,经过一层卷积核长度 4 宽度为 1 的卷积层进行下采样,水平步长为 4 垂直步长为 1,得到低维标签语义嵌入表示矩阵 $\mathbf{Z}^1 \in \mathbf{R}^{C \times d'}$, d' 此时为 256,将 \mathbf{Z}^1 输入 GAT 获取标签共现嵌入模块,经过两层 GAT 计算得到标签共现嵌入 $\mathbf{Z}^{l+2} \in \mathbf{R}^{C \times d''}$ 。为了将标签共现嵌入应用在图像特征上,将多标签图像解析度设置为 448×448 ,将其输入图像通用特征提取模块,得到多标签图像特征图 $\mathbf{F} \in \mathbf{R}^{W \times H \times D}$, D 为 2 048, W, H 均为 14,针对 VOC 2007 数据集模型,采用的卷积层 conv_1 不改变其 W, H ,使 W', H' 与 W, H 相等。MS-COCO 2014 数据集中 W, H 通过长宽为 5 卷积核的 conv_1 计算得到值均为 10 的 W', H' ,在两种数据集上均经过长宽为 1 的卷积核的卷积层 conv_2 ,对特征图通道数 D 进行降维,得到 $\mathbf{F}'' \in \mathbf{R}^{W'' \times H'' \times d''}$,最后采用池化核大小为 $W' \times H'$ 的全局最大值池化层 GMP 得到维度为 \mathbf{R}^d 的图像特征向量 \mathbf{x} , d'' 是图像特征向量 \mathbf{x} 的维度,同时也是标签共现嵌入的列维度,在 VOC 2007 数据集上的取值分别为 {300, 512, 768},而在 MS-COCO 2014 数据集上设置 d'' 为 {1 024, 1 280, 1 536}, d'' 参数设置由参数搜索和数据集中的标签类别标签个数共同决定,参数搜索策略为试错法,由于 MS-COCO 2014 所含有的类别标签数是 VOC 2007 中所含类别标签数的 4 倍,因此 d'' 也同步增加。设置初始学习率为 0.005,采用随机梯度下降作为优化器,权重衰减设置为 10^{-4} ,动量设置为 0.9,总共训练 100 轮。

2.2 评价指标

测试采用的评价指标有:平均每类精度 (CP)、

平均每类召回率(CR)和平均每类 $F_1(CF_1)$ 值。另外针对整体分类结果使用平均整体精度(OP),平均整体召回率(OR),平均整体 $F_1(OF_1)$ 进行评价。针对每个类别的分类准确度,取平均值得到平均精度均值(mAP)^[28],评价指标定义如下:

$$OP = \frac{\sum_i N_i^C}{\sum_i N_i^P} \quad OR = \frac{\sum_i N_i^C}{\sum_i N_i^g}$$

$$OF_1 = \frac{2 \times OP \times OR}{OP + OR} \quad CF_1 = \frac{2 \times CP \times CR}{CP + CR}$$

$$CP = \frac{1 N_i^C}{C N_i^P} \quad CR = \frac{1 N_i^C}{C N_i^g}$$

其中, N_i^C 表示对第 i 个类别标签正确分类的图像数, N_i^P 表示对第 i 个类别标签有预测的图像数, N_i^g 表示对第 i 个类别标签存在真实值的标签数。所有评价指标均是数值越高越好。

3 实验结果及分析

在 MS-COCO 2014 数据集的实验中,因为实验设备条件有限,且数据集中样本相对较多,故进行实验时,采用随机抽取部分训练样本用作训练模型,再

将训练出的模型在全部测试样本上进行测试的方法。对于 ML-GCN 和 ResNet-101 进行同样的采样、训练、测试方法进行实验,在 MS-COCO 2014 训练样本列表中采用 Python 的 Random 模块,从 82 081 张训练样本中随机抽取 4 000 个样本,采样 3 次,训练出 3 个模型分别测试,对所有测试产生的评价指标,取 3 次测试的均值作为实验结果。VOC 2007 数据集采用全部训练样本和测试样本进行实验。实验结果中各评价指标中最佳值均已加粗。

ML-GAT 在 VOC2007 上的测试结果如表 1 所示,经过与近年来的主流深度多标签图像分类模型进行对比(实验数据来源中除 ResNet-101、ML-GAT,其他方法数据均来自各论文中给出的测试结果),在 d'' 设置为 512 的情况下,ML-GAT 在 mAP 这一指标上达到了 94.3,在 14 个类别的分类上为最佳值。在 MS-COCO 2014 数据集上 ML-GAT 的测试结果如表 2 所示,此时 d'' 设置为 1 280,在所有标签上的预测与前 3 个标签上的预测结果中,有 7 个主要分类指标超过或持平 ML-GCN,说明 ML-GAT 模型可以在多个常用数据集上取得较好的分类结果。

表 1 在 VOC 2007 上的实验结果

Table 1 Experimental results on VOC 2007

方法 类别	飞机	单车	鸟类	船只	瓶子	巴士	汽车	猫类	椅子	奶牛	桌子	狗类	马匹	摩托	人类	植物	羊类	沙发	火车	电视	mAP
CNN-SVM	88.5	81.0	83.5	82.0	42.0	72.5	85.3	81.6	59.9	58.5	66.5	77.8	81.8	78.8	90.2	54.8	71.1	62.6	87.4	71.8	73.9
CNN-RNN	96.7	83.1	94.2	92.8	61.2	82.1	89.1	94.2	64.2	83.6	70.0	92.4	91.7	84.2	93.7	59.8	93.2	75.3	99.7	78.6	84.0
ResNet-101	99.6	97.7	97.7	96.7	67.5	92.1	96.0	97.4	75.4	78.8	85.0	98.2	96.4	95.8	98.3	73.8	79.7	80.4	98.7	90.7	89.8
ML-GCN	99.5	98.5	98.6	98.1	80.8	94.6	97.2	98.2	82.3	95.7	86.4	98.2	98.4	96.7	99.0	84.7	96.7	84.3	98.9	93.7	94.0
ML-GAT	99.9	98.6	98.6	98.3	80.6	94.5	97.4	97.8	83.3	96.5	86.0	98.3	98.4	97.1	98.9	85.4	96.5	86.3	99.3	94.5	94.3

表 2 在 MS-COCO 2014 上的实验结果

Table 2 Experimental results on MS-COCO 2014

方法 指标	ALL							TOP3						
	mAP	CP	CR	CF1	OP	OR	OF1	CP	CR	CF1	OP	OR	OF1	
ResNet-101	35.4	28.0	6.4	10.4	80.2	24.2	37.2	26.5	6.0	9.8	80.9	23.7	36.6	
ML-GCN	68.6	65.6	65.6	65.6	69.6	69.7	69.6	72.6	56.7	63.7	78.9	59.8	68.0	
ML-GAT	69.2	75.1	58.6	65.8	78.2	64.0	70.3	78.1	53.0	63.1	83.3	57.5	68.0	

为了比较不同数据集上标签共现嵌入列维度 d'' 对分类性能的影响,分别对两个数据集设置不同的 d'' 进行对比实验,如图 2 所示,在 VOC 2007 数据集中, d'' 取值为 512 时,ML-GAT 在 mAP 评价指标上达到最佳,在 MS-COCO 2014 数据集上进行一次采样测试, d'' 大小为 1 280 时得到最佳 mAP,这说明 MS-COCO 2014 数据集中的标签类别更多,标签共现嵌入中冗余部分较少。而 VOC 2007 因为标签类别较少,因此标签共现嵌入冗余部分较多。通过在这两种数据集上进行对比实验,验证了 ML-GAT 在标签共现嵌入降维,与对标签之间非对称关系的提取上采取的策略是有效的。

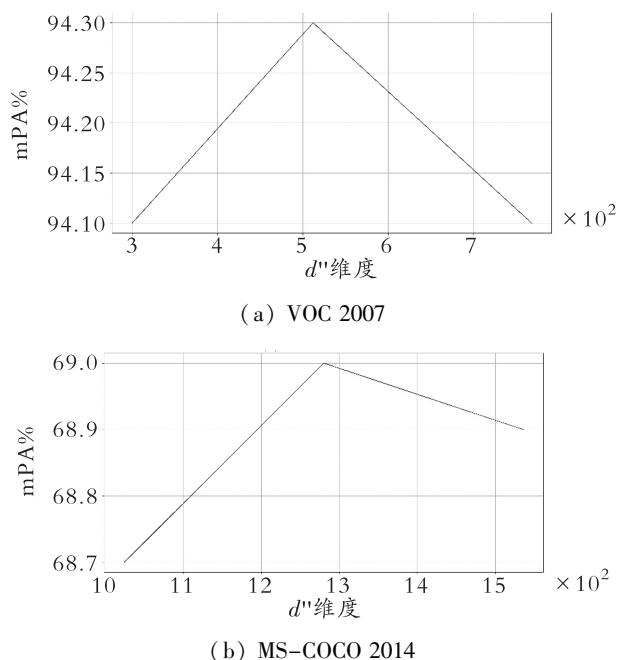


图 2 不同 d'' 值对 mAP 的影响

Fig. 2 The influence of different d'' values on mAP

4 结束语

图卷积神经网络与 CNN 结合的深度多标签图像分类模型 ML-GCN 在多标签图像的分类上取得了很好的效果,但是 ML-GCN 中通过 GCN 获取到的标签共现嵌入维度过高,标签共现嵌入没有很好的反应标签之间非对称关系,针对 ML-GCN 存在的这两点不足,提出一种基于图注意力网络的多标签图像分类模型 ML-GAT。ML-GAT 通过对输入 GAT 的高维标签语义嵌入表示矩阵进行降维,解决了 ML-

GCN 利用 GCN 获取标签共现嵌入时,冗余部分降低模型分类准确度问题,同时 GAT 可以对标签邻居之间计算不同注意力系数,学习标签之间非对称关系特征,促进模型分类。通过在主流数据集上与多标签深度学习经典模型进行对比实验,ML-GAT 模型在多标签图像分类主要评价指标上,相较经典深度多标签图像分类模型有一定的改进,实验证明了 ML-GAT 模型的有效性。

参考文献(References):

- [1] WEI Y C, XIA W, LIN M, et al. HCP: a flexible CNN framework for multi-label image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(9): 1901—1907.
- [2] LIU J, CHANG W C, WU Y, et al. Deep learning for extreme multi-label text classification[C] //Proc of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2017.
- [3] SANDEN C, ZHANG J Z. Enhancing multi-label music genre classification through ensemble techniques[C] //Proc of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2011.
- [4] ZHANG J, ZHANG Z, WANG Z, et al. Ontological function annotation of long non-coding RNAs through hierarchical multi-label classification[J]. Bioinformatics, 2018, 34(10): 1750—1757.
- [5] 张永,刘浩科,张洁.基于类属特征和实例相关性的多标签分类算法[J].模式识别与人工智能,2020,33(5): 439—448.
ZHANG Yong, LIU Hao-ke, ZHANG Jie. Multi-label classification algorithm based on generic features and instance correlation [J]. Pattern Recognition and Artificial Intelligence, 2020, 33(5): 439—448.
- [6] AMORIM W P, FAICÃO A X, PAPA J P. Multi-label semi-supervised classification through optimum-path forest [J]. Information Sciences, 2018, 465(10): 86—104.
- [7] BOUTELL M R, LUO J, SHEN X, et al. Learning multi-label scene classification[J]. Pattern Recognition, 2004, 37(9): 1757—1771.
- [8] HUANG J, LI G R, HUANG Q M, et al. Learning label-specific features and class-dependent labels for multi-

- label classification[J]. *IEEE Transactions on Knowledge & Data Engineering*, 2016, 28(12): 3309—3323.
- [9] HUANG J, LI G R, WANG S H, et al. Multi-label classification by exploiting local positive and negative pairwise label correlation [J]. *Neurocomputing*, 2017, 257(27): 164—174.
- [10] HUANG J, XU L C, WANG J, et al. Discovering latent class labels for multi-label learning[C]//Proc of the 29th International Joint Conferences on Artificial Intelligence. Springer, 2020.
- [11] HUANG J, QIN F, ZHENG X, et al. Improving multi-label classification with missing labels by learning label-specific features [J]. *Information Sciences*, 2019, 492(8): 124—146.
- [12] CLARE A, KING R D. Knowledge discovery in multi-label phenotype data [C]//Proc of the 5th European Conference on Principles of Data Mining and Knowledge Discovery. Springer, 2001.
- [13] ZHANG M L, ZHOU Z H. ML-KNN: A lazy learning approach to multi-label learning [J]. *Pattern Recognition*, 2007, 40(7): 2038—2048.
- [14] YAN L C, YOSHUA B, GEOFFREY H. Deep learning [J]. *Nature*, 2015, 521(7553): 436—444.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E, et al. ImageNet classification with deep convolutional neural networks[C]//Proc of the 26th International Conference on Advances in Neural Information Processing Systems. Morgan Kaufmann Publishers, 2012.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large scale image recognition [EB/OL]. <https://arxiv.org/pdf/1409.1556.pdf>.
- [17] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016.
- [18] LI Q, QIAO M, BIAN W, et al. Conditional graphical lasso for multi-label image classification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016.
- [19] LI X, ZHAO F, GUO Y. Multi-label image classification with a probabilistic label enhancement model [C]//Thirtieth Conference on Uncertainty in Artificial Intelligence. AUAI, 2014.
- [20] WANG J, YANG Y, MAO J H, et al. CNN-RNN: a unified framework for multi-label image classification [C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016.
- [21] CHEN Z M, WEI X S, WANG P, et al. Multi-label image recognition with graph convolutional networks [C]//Proc of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019.
- [22] KIPF T N, WELING M. Semi supervised classification with graph convolutional networks [EB/OL]. <https://arxiv.org/pdf/1609.02907.pdf>.
- [23] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. <https://arxiv.org/pdf/1710.10903.pdf>.
- [24] 张平照, 张辉宜. 多标记数据的特征及标记降维方法综述[J]. *重庆工商大学学报(自然科学版)*, 2020, 37(5): 23—29.
ZHANG Ping-zhao, ZHANG Hui-yi. A review of features and labels dimensionality reduction methods of multi label data[J]. *Journal of Chongqing Technology and Business University(Natural Science Edition)*, 2020, 37(5): 23—29.
- [25] DEVLIN J, CHANG M W, LEE K T, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proc of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long and Short Papers). ACL, 2019.
- [26] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proc of the 13th European Conference on Computer Vision. Springer, 2014.
- [27] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. *International Journal of Computer Vision*, 2010, 88(2): 303—338.
- [28] WU X Z, ZHOU Z H. A unified view of multi-label performance measures[C]//Proc of the 34th International Conference on Machine Learning. ACM, 2017.

Multi-label Image Classification Model Based on Graph Attention Network

ZHANG Hui-yi, ZHANG Jin, HUANG Jun

(School of Computer Science and Technology, Anhui University of Technology,
Anhui Maanshan 243000, China)

Abstract: In order to solve the problem that the high co-occurrence dimension of labels in ML-GCN reduces the model classification performance and the asymmetrical relationship between labels is not fully explored in ML-GCN, a multi-label image classification model of ML-GAT based on graph attention network is proposed. Firstly, the ML-GAT model reduces the dimensionality of the semantic embedding matrix of high dimensional labels. Then the label co-occurrence embedding is obtained by the low dimensional label semantic embedding representation and the label category co-occurrence graph after dimensionality reduction. At the same time, ML-GAT inputs the original multi-label image into the convolutional neural network to extract the general features of the image, and the general features of the multi-label image extracted by the convolutional neural network are unified in dimension according to the embedded dimensions of the labels calculated by the graph attention network. Finally, ML-GAT fusion of the image features after co-occurrence and dimensionality reduction of labels is used to obtain the label prediction score of each multi-label image. Experimental results on VOC 2007 and MS-COCO 2014 show that ML-GAT achieves good experimental results under the condition of sufficient training samples and sufficient number of label categories. By comparing with other models, the strategy adopted by ML-GAT model can improve the multi-label image classification performance of the model to a certain extent.

Key words: multi-label classification; graph attention network; convolutional neural network; deep learning

责任编辑:罗珊珊

引用本文/Cite this paper:

张辉宜, 张进, 黄俊. 基于图注意力网络的多标签图像分类模型[J]. 重庆工商大学学报(自然科学版), 2022, 39(1): 34—41.
ZHANG Hui-yi, ZHANG Jin, HUANG Jun. Multi-label Image Classification Model Based on Graph Attention Network[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2022, 39(1): 34—41.