

基于多层次注意力与图模型的图像多标签分类算法

朱旭东^{1,2}, 熊 贇^{1,2}

(1. 复旦大学 计算机科学与技术学院, 上海 200433; 2. 上海市数据科学重点实验室, 上海 200433)

摘 要: 图像多标签分类作为计算机视觉领域的重要研究方向, 在图像识别、检测等场景下得到广泛应用。现有图像多标签分类方法无法有效利用标签相关性信息以及标签语义与图像特征的对应关系, 导致分类能力较差。提出一种图像多标签分类的新算法, 通过利用标签共现信息和标签先验知识构建图模型, 使用多尺度注意力学习图像特征中目标, 并利用标签引导注意力融合标签语义特征和图像特征信息, 从而将标签相关性和标签语义信息融入到模型学习中。在此基础上, 基于图注意力机制构建动态图模型, 并对标签信息图模型进行动态更新学习, 以充分融合图像信息和标签信息。在图像多标签分类任务上的实验结果表明, 相比于现有最优算法 MLGCN, 该算法在 VOC-2007 数据集及 COCO-2012 数据集上的 mAP 值分别提高了 0.6、1.2 个百分点, 性能有明显提升。

关键词: 多标签; 标签语义; 图像特征; 注意力机制; 动态图; 多尺度

开放科学(资源服务)标志码(OSID):



中文引用格式: 朱旭东, 熊贇. 基于多层次注意力与图模型的图像多标签分类算法[J]. 计算机工程, 2022, 48(4): 173-178, 190.

英文引用格式: ZHU X D, XIONG Y. Multi-label image classification algorithm based on multi-scale attention and graph model[J]. Computer Engineering, 2022, 48(4): 173-178, 190.

Multi-label Image Classification Algorithm Based on Multi-scale Attention and Graph Model

ZHU Xudong^{1,2}, XIONG Yun^{1,2}

(1. School of Computer Science, Fudan University, Shanghai 200433, China;

2. Shanghai Key Laboratory of Data Science, Shanghai 200433, China)

[Abstract] As an important research direction in the field of computer vision, multi-label image classification is widely used in recognition, detection, and other applications. Existing multi-label image classification methods cannot effectively use label correlation information and the corresponding relationship between label semantics and image features, resulting in poor classification ability. A new algorithm for multi-label image classification is proposed. By using tag co-occurrence information and tag prior knowledge to build a graph model, multi-scale attention is used to learn the target in image features, and tag guided attention is used to fuse tag semantic features and image feature information to integrate tag correlation and tag semantic information into model learning. On this basis, a dynamic graph model is constructed based on the graph attention mechanism, and the label information graph model is dynamically updated and learned to integrate the image and label information fully. The experimental results on a multi-label image classification task show that, compared with the existing optimal algorithm, Multi-Label Graph Convolutional Network (MLGCN), the mean Average Precision (mAP) values of the algorithm on the Visual Object Classes-2007 (VOC-2007) and Common Object in COntext-2012 (COCO-2012) datasets are improved by 0.6 and 1.2 percentage points, respectively, improving the performance significantly.

[Key words] multi label; label semantic; image feature; attention mechanism; dynamic graph; multi scale

DOI: 10.19678/j.issn.1000-3428.0061072

0 概述

图像多标签分类技术作为计算机视觉领域的经典技术, 在图像识别、检测等场景下得到广泛应用, 具有重要研究意义。如何通过有效学习标签之间相关性的信息, 提升多标签的分类效果, 已受到越来越

多研究人员的关注。目前, 多标签场景下还存在数据分布不均衡、正负样本学习等问题。现有对多标签关系的建模多基于标签信息统计, 一定程度上捕捉了标签相关信息, 但仅使用标签统计信息的标签泛化性较差, 容易产生不同样本分布类别间的偏差, 不能有效利用图像信息学习标签关系。

基金项目: 国家自然科学基金(U1636207); 上海市科委项目(19511121204)。

作者简介: 朱旭东(1995—), 男, 硕士研究生, 主研方向为计算机视觉、图神经网络; 熊贇, 教授、博士。

收稿日期: 2021-03-10 **修回日期:** 2021-04-30 **E-mail:** 18212010052@fudan.edu.cn

本文提出一种基于图结构与多层次注意力的图像多标签分类算法,通过建立不同尺度大小的特征层次空间,对特征层次空间注意力和通道注意力的权重进行分析与学习。此外,将不同特征层次和不同特征通道的注意力权重与图像信息进行叠加,促进局部信息和整体信息的融合,从而获取标签语义空间和对应图像特征空间之间的关系信息。另外,通过分析标签相关性建立标签语义空间和图像特征空间的联系,并对标签语义特征和图像特征进行融合,引入图像特征信息和图注意力机制进行动态图模型更新。

1 相关工作

多标签分类的主要任务是建立标签信息空间和数据特征空间的对应关系,与普通单标签分类任务最大的区别在于多标签分类中,同一实体的数据特征空间和标签信息空间存在多个映射关系。

在图像多标签分类的相关研究中,文献[1]在图像特征空间进行划分学习,文献[2-4]注重对图像中不同标签对应目标的图像特征进行学习,充分利用图像特征与分类相关的信息,但这些文献中所提算法忽略了标签相关性信息和标签语义信息,其中MRAR算法^[2]通过循环注意力模块提取图像特征不同区域信息,ORVA算法^[3]通过引入无序循环网络对图像特征进行加权学习。文献[5]通过概率分布计算辅助标签分类,文献[6-8]利用标签语义信息融合图像特征辅助图像分类,但 these 方法对标签相关性信息利用不充分。其中,文献[6]所提到的CMA算法通过融合标签信息和图像进行多

模态学习,而文献[8]所提的LGS算法利用相似性网络捕捉标签与图像对应关系。文献[9-11]对标签间关系进行建模,其中CNN-RNN算法^[11]利用序列神经网络构建标签关系,结合卷积神经网络的图像特征进行学习,MLGCN算法^[12]和FGCN算法^[13]等通过构建多标签图模型,并利用图模型中的边定义标签共现关系。这些方法建模的标签关系利用了标签相关信息进行辅助分类,但是泛化性和类别分类能力不平衡,有待改进。

针对现有方法存在的一些问题,如MLGCN算法^[12]在构建多标签关系时,只考虑到标签统计信息,容易造成标签关系过拟合问题;FGCN算法^[13]在对标签关系模型进行更新时,只使用标签信息,没有利用到图像特征信息,使得构建的标签关系无法有效辅助图像特征分类。本文提出基于动态图的多层级注意力(Multi Level Attention based Dynamic Graph, MLADG)算法,通过结合基于知识先验和标签共现相关性建立的图模型进行标签信息和图像特征融合,并采用图注意力动态更新机制进行模型学习。

2 MLADG 模型

本节主要介绍图像多标签分类算法MLADG,算法的主要框架包括通过多尺度多通道图像注意力机制学习图像特征,标签语义引导注意力机制融合标签语义,通过在共现关系和先验知识上的标签图模型结构,动态更新捕捉图像特征和标签之间关系。图1为MLADG模型整体架构示意图(彩色效果见《计算机工程》官网HTML版)。

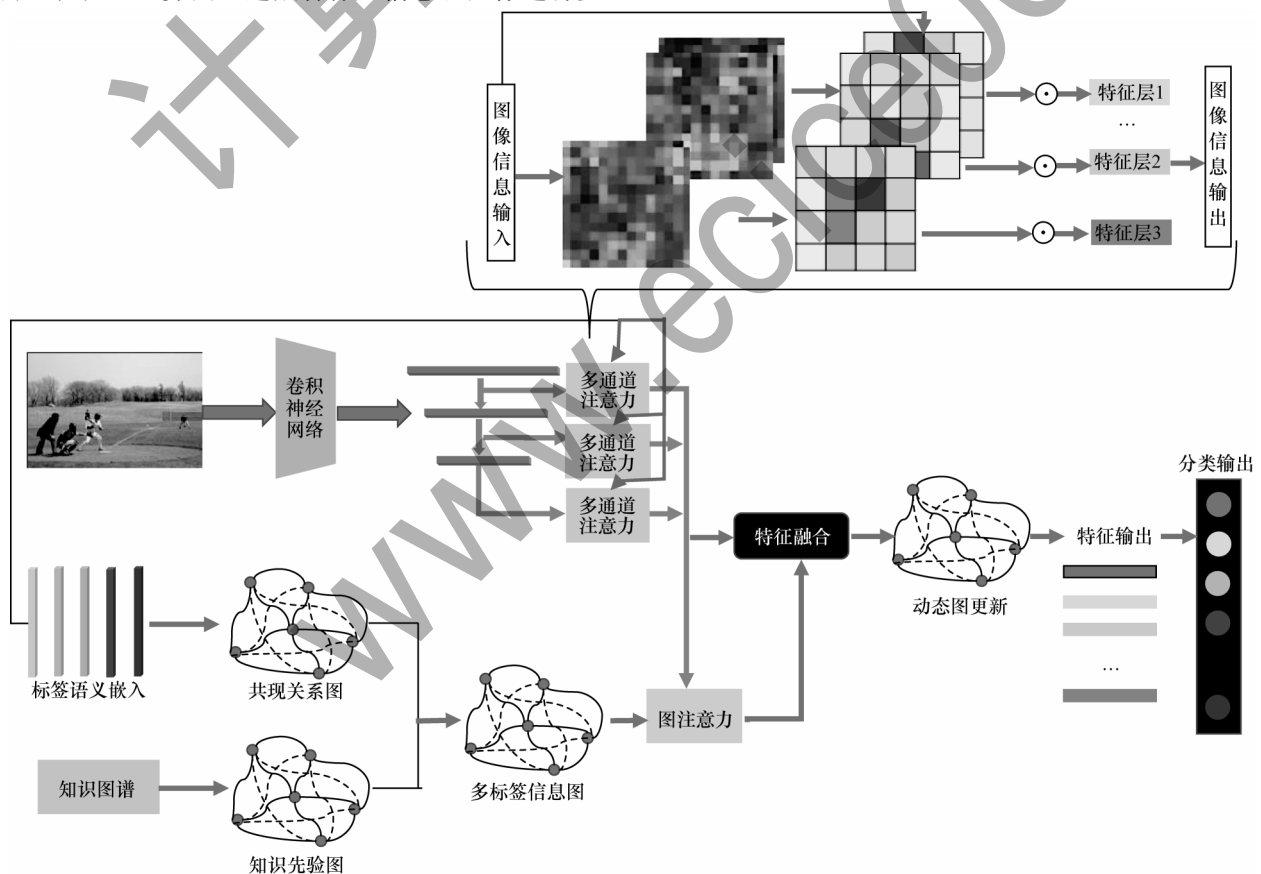


图1 MLADG模型结构

Fig.1 Structure of MLADG model

2.1 多标签图模型构建

多标签分类问题中标签间的相互关系包含多标签分类的有效辅助信息。对于给定图像 I 的标签序列 $L=[l_1, l_2, \dots, l_n]$, 标签序列中任一标签 l_i 通过词嵌入层的权重矩阵 W 获得对应的低维表征向量 e_i 。标签词语通常具有多种词性和词义, 标签语义嵌入重点关注标签在词嵌入模型中所具有的名词词性代表的语义。通过 GloVe 模型获得图像标签词语义信息。定义图像多标签共现关系图模型 $G_s=\{V_s, E_s\}$, 其中节点关系集合 $V_s \in \{v_{s1}, v_{s2}, \dots, v_{sc}\}$ 表示嵌入的标签特征信息; 边集合为 $E_s \in \{e_{s1}, e_{s2}, \dots, e_{sc}\}$ 为图像多标签间共现关系, 包含与每条边相关非负权值。对于其中一条边 e_{ij} 的值由 $p_{ij}=P(v_i|v_j)$ 确定, 即当标签 l_j 标记情况下标签 l_i 同时进行标记。同理, p_{ji} 表示当标签 l_i 标记的情况下标签 l_j 同时标记的概率。

标签共现图通过统计标签共现次数, 可以有效学习到标签的分布情况。但标签共现图受训练集数据分布影响很大, 可能同时存在共现次数较多的标签关系, 造成频率偏差, 使得模型分类准确性受到数据分布不均衡的影响。为此, 除共现关系图外, 借助知识图谱 ConceptNet^[14] 信息构建知识先验关系图 $G_k=\{V_k, E_k\}$, 其中节点集合 $V_k \in \{v_{k1}, v_{k2}, \dots, v_{kc}\}$ 为嵌入的标签语义特征, 边集合 $E_k \in \{e_{k1}, e_{k2}, \dots, e_{kc}\}$ 为知识图谱中标签相关性。ConceptNet 中各个标签知识实体通过多元组形式构成, 将各个标签在知识图谱中同一元组的数量占该实体全部元组比重作为边权重。

对于构建的标签共现图 G_s 和标签知识图 G_k , 分别对其边权重矩阵进行标准化后得到 G'_s 和 G'_k , 设定图融合参数 $\lambda \in (0, 1)$, 最终得到的多标签信息融合图的计算式如式(1)所示:

$$G_{sk} = \lambda \times G'_s + (1 - \lambda) \times G'_k \quad (1)$$

2.2 多层次注意力模块

多标签图像数据集 $I=\{i_1, i_2, \dots, i_n\}$ 中图片 i 所包含的图像特征空间信息与标签信息空间存在映射关系。但在图像中, 真正对确定其标签有帮助的特征信息只是图像局部的关键特征信息, 除此之外还存在对确定图像标签类别存在干扰的噪声信息。通过多尺度多通道图像特征注意力模块在不同特征尺度和维度上对每个标签语义对应的重点信息进行提取, 标签映射注意力模块负责捕捉标签语义信息和图像特征信息相关性。

2.2.1 多通道注意力模块

给定的多标签图像中不同标签所对应的图像区域目标大小不同, 占据更大图像区域的标签可以提供更多的图像特征信息。通过进行卷积神经网络学习, 多标签对应的图像特征也分布在特征空间的各个通道。选取残差网络 ResNet 作为图像特征提取的主干神经网络, 将残差网络不同层次卷积层 Conv1、Conv2、Conv3 的输出特征 V_1 、 V_2 、 V_3 作为多尺度注意

力的输入, 其中 Conv1 为 ResNet 中第 1 个 BottleNeck 模块最后一层卷积层, Conv2 为 ResNet 中第 2 个 BottleNeck 最后一层卷积层, Conv3 为 ResNet 中第 3 个 BottleNeck 最后一层卷积层。相应的 V_1 表示较浅层特征信息, V_2 和 V_3 分别表示中间层和深层特征信息。此外, 对于每一层图像特征, 通过多通道分组注意力模块得到图像通道特征注意力 A_1 、 A_2 、 A_3 , 再通过 softmax 函数获得归一化后的注意力信息 A'_1 、 A'_2 、 A'_3 , 最后图像 I 对应的特征表示如式(2)所示:

$$V'_i = [A'_1]_1' \times [V_1]_1' + [A'_2]_2' \times [V_2]_2' + [A'_3]_3' \times [V_3]_3' \quad (2)$$

图像 I 经过卷积神经网络特征提取后得到的特征空间为 $X \in \mathbb{R}^{H \times W \times C}$, 其中: H 和 W 分别表示特征尺度的高和宽; D 表示特征空间的通道数。对于某个特定的类别 l_c 通过多通道注意力机制获得其相应注意力权重信息 a_c , 将权重信息和特征空间进行点积运算可以获得加权后的特征信息。对于全部类别的注意力信息空间 $A \in H \times W$, 通过注意力学习后得到的图像信息表征如式(3)所示:

$$x_c = a_c^T \times X' = \sum_{i=1}^H \sum_{j=1}^W [a_c]_{ij} \times X'_{ij} \quad (3)$$

通过 C 个 1×1 卷积对特征空间 $X \in \mathbb{R}^{H \times W \times D}$ 进行降维至 $X' \in \mathbb{R}^{H \times W \times C}$, 对 X' 中第 c_i 个通道内特征进行注意力加权, \hat{X}'_{ci} 为加权处理后的 c_i 通道特征, μ_x 和 σ_x 分别为 X 特征空间中特征信息的均值和方差。 \hat{X}'_{ci} 的表达式如式(4)所示:

$$\hat{X}'_{ci} = \frac{X_{ci} - \mu_{ci}}{\sigma_{ci} + \varepsilon} \quad (4)$$

随后通过设置特征恢复参数 γ 和 β 生成注意力权重, 表达式如式(5)所示:

$$a_{ci} = \gamma \hat{X}'_{ci} + \beta \quad (5)$$

其中: a_{ci} 为特征通道 c_i 中学习到的注意力权重信息。

2.2.2 标签引导模块

为促进标签语义特征和图像特征的融合, 更好地学习图像特征的标签意义, 对于嵌入学习到的标签 l_c 的标签词语义 e_c , 通过特征向量余弦相似度的计算捕捉标签语义和图像特征对应关系, 可以得出标签 k 在特征 v 的位置 i 处的注意力 a_{zk}^i 。

$$a_{zk}^i = \text{ReLU}\left(\frac{[v^i]^T e_c}{\|v^i\| \times \|e_c\|}\right) \quad (6)$$

对于图像特征中的某一区域, 较高的 a_{zk}^i 值可以说明该处特征信息和标签语义的相关度较高, 利用标签映射权重信息对不同类别图像进行加权平均得到标签 k 对应的图像特征最终表征向量 h_k :

$$h_k = \sum_{i=1}^M a_{zk}^i \times v^i \quad (7)$$

2.2.3 动态特征图融合更新

标签共现关系图和知识先验图注重于标签相关性和标签语义知识的挖掘, 图模型的构建依赖于静态的标签统计信息和知识信息。为利用图像信息对

图模型进行更新,本文采取基于图注意力^[15]模式的动态特征融合图更新模块。具体实现方式:标签融合图最后一层输出 G_f 和多尺度注意力模块输出特征 V_f 进行Concat融合,得到动态图模型的输入信息 h_i 。

$$h_i = \left[\begin{bmatrix} G_i \\ A_i \end{bmatrix} \right]_f \quad (8)$$

之后,通过如式(9)所示的图注意力公式计算得出动态特征融合图注意力神经网络中注意力系数 α_{ij} :

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}_d^T [W_d h_i \| W_d h_j]))}{\sum_{k \in n_i} \exp(\text{LeakyReLU}(\vec{a}_d^T [W_d h_i \| W_d h_k]))} \quad (9)$$

其中: \vec{a}_d^T 为动态特征图前馈神经网络参数; W_d 为设定的待学习的参数矩阵;设置动态特征图注意力网络层数为 k ;单层图神经网络的输出 \vec{h}_i^T 可由注意力权重推导得到:

$$\vec{h}_i^T = \sigma(\sum_{j \in N} \alpha_{ij} W_d \vec{h}_j^T) \quad (10)$$

在模型训练的过程中,不同批次数据的输入提供动态的图像信息,通过对多尺度模块输出特征的动态变化更新图注意力层的输入,使得模型在捕捉多标签相关性的同时保持对图像特征信息的敏感性。

3 实验与结果分析

3.1 实验参数和环境设置

为验证本文算法的有效性,选取2个数据集进行实验,并和相关算法进行对比。实验采用ResNet101^[16]模型作为图像特征提取器,在8×1 080 ti显卡服务器训练,使用的优化器为SGD,动量参数设置为0.9,初始学习率为0.01,训练周期为100,特征融合参数 λ 设置为0.5。

3.2 实验评价指标

图像多标签分类评价指标最常用的指标为平均均值精度(mean Average Precision, mAP),同时也会采用平均每类精度(Class Precision, CP)、平均每类召回(Class Recall, CR)、平均每类(Class F1, CF1)以及整体平均精度(Overall Precision, OP)、整体平均召回(Overall Recall, OR)、整体平均(Overall F1, OF1)。相关表达式分别如式(11)~式(16)所示。

$$C_{CP} = \frac{1}{C} \sum_{i \in C} \frac{N_i^c}{N_i^p} \quad (11)$$

$$C_{CR} = \frac{1}{C} \sum_{i \in C} \frac{N_i^c}{N_i^g} \quad (12)$$

$$O_{OP} = \frac{1}{N} \sum_{s \in N} \frac{N_s^c}{N_s^p} \quad (13)$$

$$O_{OR} = \frac{1}{N} \sum_{s \in N} \frac{N_s^c}{N_s^g} \quad (14)$$

$$C_{CP} = \frac{1}{C} \sum_{i \in C} \frac{N_i^c}{N_i^p} \quad (15)$$

$$M_{mAP} = \frac{1}{N} \sum_{n \in N} \int_0^1 p(r) dr \quad (16)$$

其中: N_i^c 代表预测类别是 i 并且实际类别也是 i 的样本总数; N_i^p 表示预测类别为 i 的全部样本数量; N_i^g 表示第 i 类中样本总数; N_s^c 代表具有分类预测结果的样本 s ; N_s^p 表示分类正确的样本 s ; N_s^g 表示样本 s 的实际标签; $p(r)$ 为某一类别分类准确率。

3.3 数据集

使用VOC-2007^[17]和COCO-2014^[18]数据集。其中,VOC-2007数据集包括Person、Animal、Vehicle等20个标签类别,分为train、val、test这3部分,共有9 963张图片;COCO-2014数据集是使用最为广泛的图像数据集之一,包含共80现实场景常见类别,82 783张训练图片,40 775张测试图片。

3.4 对比算法

为评估本文算法的有效性,选取ResNet101^[16]、CNN-RNN^[11]、RMIC^[19]、MLGCN^[12]、F-GCN^[20]5种算法进行对比实验分析。

ResNet101算法通过设置残差模块解决多层神经网络随网络深度增加带来的梯度弥散问题,ResNet101是其中神经网络共101层的版本。CNN-RNN算法通过卷积神经网络特征提取器和序列神经网络的结合,建立多标签序列捕捉多标签相关性,为图像分类提供辅助信息。RMIC算法基于强化学习中原始多标签问题框架,通过组合图像特征和前序标签信息作为下一轮标签预测的新状态。MLGCN算法通过以标签为节点,标签相关性为边构建图模型,边权重为标签共现概率,使用图卷积神经网络进行图模型更新。F-GCN算法提出一种基于图卷积神经网络的快速模型,利用多模态分解双线性池化融合图像信息和标签嵌入信息,以端到端方式进行训练。

3.5 结果分析

不同算法在各个数据集的实验结果如表1、表2所示。其中表1为VOC-2007数据集下的实验结果,表2为COCO-2014数据集下的实验结果,表中加粗数字表示该组数据最大值。

表1 不同算法在VOC-2007数据集下的实验结果对比

Table1 Comparison of experimental results of different algorithms on VOC-2007 data set

| 算法 | mAP | CP | CR | CF1 | OP | OR | OF1 |
|-----------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| ResNet101 | 89.9 | 89.1 | 90.3 | 89.6 | 90.7 | 90.2 | 90.3 |
| CNN-RNN | 84.0 | 85.1 | 83.2 | 84.3 | 84.7 | 84.1 | 84.6 |
| RMIC | 93.7 | 92.4 | 93.0 | 92.9 | 92.2 | 91.5 | 91.7 |
| MLGCN | 94.0 | 93.6 | 94.4 | 94.3 | 93.2 | 94.2 | 93.9 |
| F-GCN | 94.1 | 93.9 | 94.6 | 94.5 | 94.3 | 94.1 | 94.2 |
| MLADG | 94.6 | 94.7 | 94.9 | 94.8 | 94.6 | 94.3 | 94.5 |

表2 不同算法在COCO-2014数据集下的实验结果对比
Table2 Comparison of experimental results of different algorithms on COCO-2014 data set %

| 算法 | mAP | CP | CR | CF1 | OP | OR | OF1 |
|-----------|------|------|------|------|------|------|------|
| ResNet101 | 79.7 | 82.7 | 67.4 | 74.3 | 86.4 | 71.8 | 78.4 |
| CNN-RNN | 72.1 | 71.7 | 72.9 | 71.8 | 71.1 | 70.3 | 70.8 |
| RMIC | 81.3 | 82.2 | 81.9 | 82.6 | 83.5 | 83.1 | 83.2 |
| MLGCN | 83.0 | 85.1 | 72.0 | 78.0 | 85.8 | 75.4 | 80.3 |
| F-GCN | 83.2 | 85.4 | 72.4 | 78.3 | 86.0 | 75.7 | 80.5 |
| MLADG | 83.4 | 88.6 | 73.8 | 79.6 | 91.0 | 77.3 | 81.2 |

从表1、表2中可得以下结论:

1) 本文所提 MLADG 算法在 VOC-2007 和 COCO-2014 两个数据集中,在 mAP、CP、OP 等指标上超过另外 5 个对比算法。MLADG 算法在图像多标签分类的任务上有较好的表现。

2) 由于 MLGCN 算法相对于 Resnet101 算法增加了对于标签共现数据信息的学习,因此指标有明显提升。但使用标签共现关系,可能导致数量较少的标签种类被正确分类的概率降低,而数量较多的标签种类更易被分类。本文算法能够增强对图像特征和标签相关性的关系学习。

3) F-GCN 算法相对于 MLGCN 算法增加标签语义信息和图像特征信息的融合,有助于标签语义指导图像特征的学习,而 MLADG 算法通过动态学习,能够更好地利用图像特征更新图模型。

3.6 模型有效性分析与可视化关系

为验证 MLADG 算法相关模型设置的有效性,将原有模型中的各结构通过以下设置:

1) 多尺度多通道注意力验证。移除多尺度多通道注意力模块,直接使用 ResNet101 算法提取的特征进行后续运算,其他保持不变。

2) 标签引导注意力验证。移除标签引导注意力模块,其他保持不变。

3) 动态图更新验证。移除动态特征图更新机制,只使用图注意力网络,其他保持不变。

图2为 MLADG 算法中各结构有效性消融实验的结果。由图2可知,有效性实验中各结构的移除均导致图像多标签分类各指标的下降,说明 MLADG 中各结构对于提高图像多标签分类效果均有正向作用,且各结构能有效协同。其中多尺度多通道注意力模块对 mAP、CP、OP 的影响最大,说明该模块对图像中重点信息注意力的学习有效。此外,移除标签引导注意力模块指标下降,说明该模块对标签信息和图像特征的融合有效引导了图像信息的学习。相比于 MLGCN^[12]等算法,MLADG 算法将标签关系图和图像特征信息分开学习,动态特

征图能够将图像特征学习和标签关系建模结合,在 mAP 指标上相对于静态更新的标签关系图提升了 1.2%。

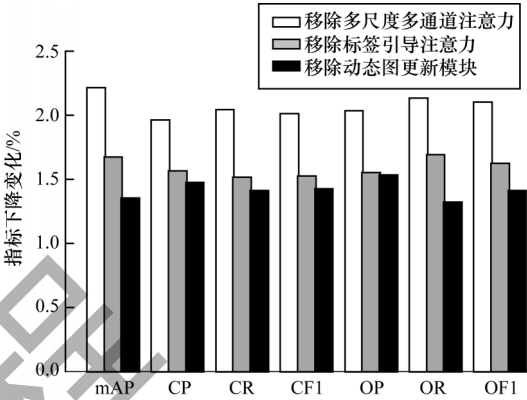


图2 MLADG 模块消融实验

Fig.2 Ablation experiment of MLADG module

对 MLADG 建模的多标签信息关系图中标签关系进行可视化,如图3所示。对于两个标签 l_i 和 l_j ,第 i 行第 j 列的示意图方格表示为 $p_{i,j} = P(l_i|l_j)$,颜色越深表示标签间相关性越强。

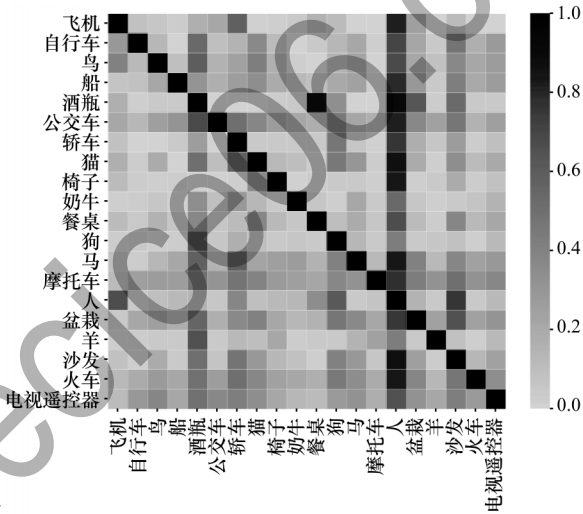


图3 多标签相关性可视化示意图

Fig.3 Visualization diagrammatic sketch of multi label correlation

由图3中可知,MLADG 算法能够对图像多标签中标签相关性进行清晰有效地建模。

3.7 参数分析

为证明 MLADG 算法中相关参数的有效性,本文在 VOC-2007 数据集下进行不同参数验证实验。

在 MLADG 算法中,参数 λ 为标签共现图和知识先验图平衡参数,参数 k 为图注意力神经网络层数。在参数敏感性实验中,将 λ 在 0~1 间进行遍历,将图注意力层数由 1~6 进行遍历,实验其他参数设置和硬件环境配置与对比实验相同。

图4和图5分别为参数 k 和参数 λ 验证实验的结果,由图4可知,当 $k=2$ 时,本文所提MLADG算法达到最好的效果。这可能是当 $k=1$ 时图注意力层过浅,无法有效建模多标签关系,而当 k 过大时可能存在标签特征嵌入过度平滑的问题,导致模型测试指标的下降。由图5可知,随着参数 λ 增加,融合多标签图模型中标签共现图模型所占权重不断增大,mAP指标逐渐上升,直到 λ 达到0.7后开始下降。这可能是由于标签共现图模型中包含更丰富的标签相关性信息,但由于训练数据和测试数据的分布差异性以及数据集中不同类别图像样本数量不均衡分布对模型训练的影响,导致分类器在不同类别图像样本中的分类能力偏差较大,整体指标降低。因此在参数 λ 过大时,融合多标签图模型中标签相关性信息容易出现不平衡的现象。

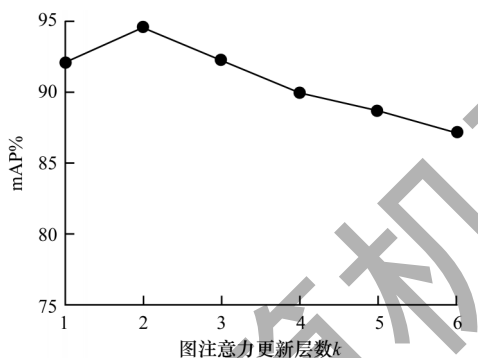


图4 k 参数在 VOC-2007 数据集上的效果验证

Fig.4 Effect of parameter k in VOC-2007 data set

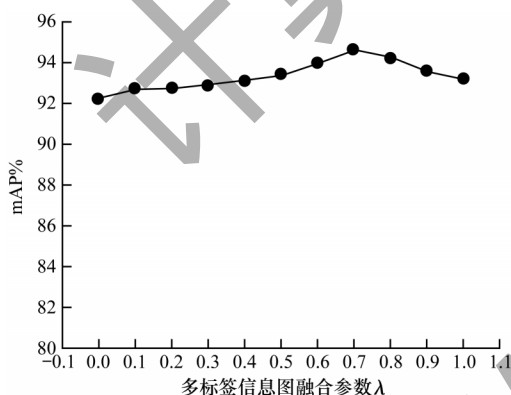


图5 λ 参数在 VOC-2007 数据集上的效果验证

Fig.5 Effect of parameter λ in VOC-2007 dataset

4 结束语

在图像多标签分类中,根据标签关系进行建模存在标签与图像间学习关联性弱的问题,且容易导致过拟合。本文提出一种基于多层次注意力和动态图注意力机制的图像多标签分类算法 MLADG。通过构建多标签信息图模型建立标签关系,将使用多层次注意力获取的多层次图像特征信息与标签语义信息进行融合,构建标签语义和图像特征相关性,并将动态图模型引入图像信息中以更新模型。实验结果表明,在相同测试环境下,与 MLGCN、F-GCN 等图像多标签分类算法相比,本文算法在 mAP、CP、OP 等指标上均取得了较

好效果。下一步将进行更高阶维度的标签相关性学习,如多个标签间连续相关性的学习,并从标签互斥角度对标签关系的定义进行深度挖掘。

参考文献

- [1] ZHU F, LI H S, OUYANG W L, et al. Learning spatial regularization with image-level supervisions for multi-label image classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA: IEEE Press, 2017: 5513-5522.
- [2] WANG Z X, CHEN T S, LI G B, et al. Multi-label image recognition by recurrently discovering attentional regions [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2017: 464-472.
- [3] CHEN S F, CHEN Y C, et al. Order-free RNN with visual attention for multi-label classification[C]//Proceedings of AAAI Conference on Artificial Intelligence. Los Angeles, USA: AAAI Press, 2018: 765-777.
- [4] 生龙, 马建飞, 杨瑞欣, 等. 基于特征交换的 CNN 图像分类算法研究[J]. 计算机工程, 2020, 46(9): 268-273. SHENG L, MA J F, YANG R X, et al. Research on CNN image classification algorithm based on feature exchange[J]. Computer Engineering, 2020, 46(9): 268-273. (in Chinese)
- [5] 王一宾, 郑伟杰, 陈玉胜, 等. 基于 PLSA 学习概率分布语义信息的多标签分类算法[J]. 南京大学学报(自然科学版), 2021, 57(1): 75-89. WANG Y B, ZHENG W J, CHENG Y S, et al. Multi label classification algorithm based on PLSA learning probability distribution semantic information[J]. Journal of Nanjing University (Natural Science), 2021, 57(1): 75-89. (in Chinese)
- [6] YOU R C, GUO Z Y, CUI L, et al. Cross-modality attention with semantic graph embedding for multi-label classification [C]//Proceedings of AAAI Conference on Artificial Intelligence. Menlo Park, USA: AAAI Press, 2020: 12709-12716.
- [7] 于玉海, 林鸿飞, 孟佳娜, 等. 跨模态多标签生物医学图像分类建模识别[J]. 中国图象图形学报, 2018, 57(1): 917-927. YU Y H, LIN H F, MENG J N, et al. Classification modeling and recognition for cross modal and multi-label biomedical image [J]. Journal of Image and Graphics, 2018, 23(6): 917-927. (in Chinese)
- [8] YA W, HE D L, LI F, et al. Multi-label classification with label graph superimposing [C]//Proceedings of AAAI Conference on Artificial Intelligence. Menlo Park, USA: AAAI Press, 2020: 12265-12272.
- [9] CHEN T S, XU M X, HUI X L, et al. Learning semantic-specific graph representation for multi-label image recognition[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C. , USA: IEEE Press, 2019: 522-531.
- [10] 顾广华, 曹宇尧, 李刚, 等. 基于语义标签生成和偏序结构的图像层级分类[J]. 软件学报, 2020, 31(2): 531-543. GU G H, CAO Y Y, LI G, et al. Image hierarchical classification based on semantic label generation and partial order structure[J]. Journal of Software, 2020, 31(2): 531-543. (in Chinese)

(下转第 190 页)

(上接第 178 页)

- [11] JIANG W, YANG Y, MAO J H, et al. CNN-RNN: a unified framework for multi-label image classification [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA : IEEE Press, 2016; 2285-2294.
- [12] CHEN Z M, XIU S, WEI X S, et al. Multi-label image recognition with graph convolutional networks [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA : IEEE Press, 2019; 5177-5186.
- [13] HOU D B, ZHAO Z J, et al. Multi-label learning with visual-semantic embedded knowledge graph for diagnosis of radiology imaging [J]. IEEE Access, 2021, 11: 15720-15730.
- [14] LIU H, PUSH S. ConceptNet—a practical commonsense reasoning tool-kit[J]. BT Technology Journal, 2004, 22(4): 211-226.
- [15] VELIEKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. [2021-02-04]. <https://arxiv.org/abs/1710.10903>.
- [16] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C. , USA : IEEE Press, 2016; 770-778.
- [17] EVERINGHAM M, LUC V G, CHRISTOPHER K W, et al. The pascal visual object classes challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [18] LIN T Y, MICHAEL M, SERGE B, et al. Microsoft coco: common objects in context[J]. European Conference on Computer Vision, 2014, 2(4): 740-744
- [19] HE S Y, XU C, GUO T Y, et al. Reinforced multi-label image classification by exploring curriculum [C]// Proceedings of AAAI Conference on Artificial Intelligence. Menlo Park, USA : AAAI Press, 2018; 376-388.
- [20] WANG Y T, XIE Y Z, LIU Y, et al. Fast graph convolution network based multi-label image recognition via cross-modal fusion [C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York, USA : ACM Press, 2020; 1575-1584.

编辑 赖玉玲