基于图神经网络的东巴画小样本分类

黎克, 钱文华*, 王成学, 徐丹

(云南大学信息学院计算机科学与工程系 昆明 650504)
(whqian@ynu.edu.cn)

摘 要: 针对纳西族东巴画艺术形象线条突出、色彩分明、样本较少的特点, 提出一种端到端的基于图神经网络的东巴画小样本分类方法. 首先, 设计多分辨率多尺度的图像特征提取网络, 图像特征与边缘特征融合后嵌入图神经网络中作为节点, 构建分类图神经网络; 其次, 以边标记作为架构并采用二维边标记特征作为图像分类依据, 保留节点分类时所需的类内相似性和类间相异性; 最后, 提出自注意力机制与特征显著性注意力机制相结合的方法更新节点特征, 增强了节点之间的特征关联性. 模型使用 Python 实现并用 RTX 2080Ti 在自建东巴画数据集上进行实验, 结果表明, 所提方法较好地提取了东巴图像特征, 保留了东巴画分类所需的图像局部细节和节点间相似性特征, 与对比算法相比, 提高了分类准确度, 并有更低的分类精度标准差.

关键词: 东巴画; 小样本分类; 边标记; 图神经网络; 特征显著性注意力

中图法分类号: TP391.41 **DOI:** 10.3724/SPJ.1089.2021.18618

Dongba Painting Few-Shot Classification Based on Graph Neural Network

Li Ke, Qian Wenhua*, Wang Chengxue, and Xu Dan

(Department of Computer Science and Engineering, School of Information Science and Engineering, Yunnan University, Kunming 650504)

Abstract: In view of the outstanding lines and distinguishing colors of the Naxi Dongba paintings, an end-to-end few-shot classification method based on graph neural network is proposed. Firstly, a multi-resolution and multi-scale image feature extraction network is designed, the image feature and edge feature are fused and embedded into the graph as nodes to construct classification graph neural network. Secondly, with edge-labeling framework, the 2-dimensional edge feature is used as the image classification basis, and the intra-class similarity and interclass dissimilarity required for node classification are preserved. Finally, a method combining self-attention and feature saliency attention mechanism is proposed to update the node features and enhance the feature correlation between nodes. The model is implemented by Python, and RTX 2080Ti is used to conduct classification experiments on the self-built Dongba painting few-shot dataset. The experimental results show that the method can better extract the features of Dongba painting image and more sufficiently retain local details and structural features required for image classification of Dongba painting. The accuracy and standard deviation of the classification are better than others on the few-shot dataset of Dongba painting image.

收稿日期: 2020-08-31; 修回日期: 2020-12-21. **基金项目:** 国家自然科学基金(61662087, 62061049, 61761046); 云南省中青年学术技术带头人后备人才项目(2019HB121); 云南省重大科技专项(202002AD080001, 202001BB050043); 云南省科技厅应用基础研究计划(2019FA044). 黎克(1994—), 男, 硕士研究生, 主要研究方向为图像处理、图像分类; 钱文华(1980—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 论文通讯作者, 主要研究方向为图像处理、非真实感绘制; 王成学(1994—), 男, 硕士研究生, 主要研究方向为图像处理、图像融合; 徐丹(1968—), 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为图形绘制技术、图像融合、虚拟现实、视觉计算.

Key words: Dongba painting; few-shot classification; edge-labeling; graph neural network; feature saliency attention

纳西族东巴画是我国少数民族的非物质文化遗产,是我国西南地区形成的独具特色的文化瑰宝,许多画面亦字亦画,保留了浓郁的象形文字书写特征,是研究人类原始绘画艺术的“活化石”。根据不同的绘画对象,纳西族东巴画中的对象可分为自然景观、鬼神、动物、植物和日常生活^[1-3]等不同类型。这些形象生动的东巴画艺术形象象征着纳西族人民生活的一部分,也表现了东巴文化对各种自然现象的崇拜。

目前对于纳西族东巴画的研究大多集中于研究其艺术价值^[2-3]与绘画风格^[3-5]。纳西族是位于偏远地区的少数民族,经济发展相对落后。东巴画缺乏完善保护措施,对东巴画进行数字化处理并分类归档,建立数据库进行数据化存储,有利于东巴画的保存。通过将画作中的形象进行分类,为进一步研究东巴画的语义分割、画作风格迁移等研究打下基础,为东巴画传统文化与现代文化进行融合创新,以及扩大东巴画的传播与影响力添砖加瓦。

传统的神经网络学习方法进行图像分类研究时需要有巨大的样本量作为支持。通过大量数据训练才能产生一个可以对数据分布进行预测的分类器,即大样本学习方法。但现存东巴画作由于经过长时间的保存,存世量较少,导致可用作构成数据集的样本较少,每类仅有少量的已标记图像,不足以训练出一个可以完整反映类间区别的模型,可能导致分类效果不理想,故难以采用大样本学习方法。小样本学习方法通过表征学习,借助最近邻分类的思想产生关系聚类,最后产生分类预测。因此,仅通过少量数据样本就可以建立一个反映类间区别的模型。

小样本学习方法的一大关键是需要充分利用支持集和查询集间的关系进行学习训练。近年来,图神经网络(graph neural networks, GNN)^[6-9]在处理拥有丰富关系结构但数据量较少的数据中颇有成效。GNN 通过相邻节点间信息传递聚合邻域节点特征,这种方式可以将数据间复杂的关系有效表示出来。因此,将 GNN 应用于小样本学习满足了小样本学习过程中对于处理数据间结构信息的要求。Garcia 等^[6]提出了一种支持集和查询集紧密相连的图结构,图中每个输入节点由嵌入特征(如卷积神经网络的输出)和给定的标签信息(如独热编码标签)表示,然后通过迭代更新节点特征并聚

合邻域节点特征对未标记的查询进行分类。Liu 等^[7]提出了一种推理传导网络(transductive propagation networks, TPN),将从神经网络获得的节点特征和已标记的标签信息嵌入无向图中,再将已标记的信息通过相似性计算的方式,传导到未标记的节点上预测节点分类。

已有的基于 GNN 的小样本分类方法,大多是将图中节点作为分类标记的,即节点标记型的模型,这种模型隐式地将类内相似性和类间相异性做了建模。但 Gong 等^[8]探索了如何更大限度地利用一个图中的边特征,即边标记的方法,通过将多种类型的边结构引入 GNN,使边特征能够表达出更多信息。Kim 等^[9]同样采用了一种边标记图神经网络(edge-labeling graph neural networks, EGNN),能够显式地表示图在嵌入空间中的聚类,从而通过边的标记表明关联的 2 个节点是否属于同一个类(簇)。不同于节点标记的结构,边标记的结构可以直接表达类间关系,并可以利用不同类型的边表示更多的图结构信息。此外,不同于节点标记的方法需要明确规定出图中类别的数量,边标记的方法通过判断与边相连的 2 个节点是否属于同一类进行聚类,而不用事先规定类别数量。因此,边标记架构很适合作为关系聚类的方法,也应用在 GNN 标注网络或动态系统^[10-11],但是很少应用于小样本学习,尤其是图像分类中。

从东巴画的数量来看,其经过长时间保存和较为封闭的传承,存世量稀少且有不同程度的噪声和褪色,可用作数据集的东巴画作也较少,若采用大样本方法,由于数据量较少,会导致模型过拟合,使模型的泛化性能下降,导致分类准确度不高。从东巴画风格来看,其虽然形象生动,但与现实场景中的人物、动物等仍有较大差别,存在一定程度的抽象和简化,若直接使用现实场景形象对模型进行训练,会使模型对东巴画的特点(如独特的边缘和色彩结构)缺乏关注和估计,导致分类性能下降。针对东巴画的上述特点,本文提出一种基于边标记的小样本 GNN 的东巴画分类方法。(1) 根据东巴画艺术形象线条突出、颜色结构分明的特点,设计了一个多分辨率、多尺度的空间金字塔特征提取模块,对图像主要细节部分和边缘信息分别提取,补充了原图特征,改善了分类准确度。(2) 在节点特征更新的过程中,采用了自注意力机

制与特征显著性注意力机制相结合的方式,使节点更新过程对于特征显著性有了筛选的能力,保留了更多显著的特征,弱化了不显著特征,提高了节点信息传递的有效性。(3) 本文小样本网络采用多层 GNN,以层次聚类的思想自下而上地将样本类别合并聚集,每层都由二维边特征分别记录节点间相似性特征和相异性特征,并以此作为更新依据结合注意力机制对网络进行更新,它们对应的参数由幕式训练的方法获得。

1 相关工作

1.1 GNN

GNN 是通过神经网络直接处理图结构数据的一种网络结构。Beck 等^[12]在基础的 GNN 架构上加入了门控循环单元(gated recurrent unit, GRU)和一些其他现代优化方法,进一步扩展增强了 GNN 的效果。GNN 通过将邻居节点进行迭代聚合,将邻居节点的特征通过边进行传递和转换,以达成表征学习。Ghorbani 等^[13]通过将一般化的基于卷积的传递方法应用到图上,尤其是将其加入了半监督图学习中,使处理图结构数据的方法具有了可扩展性。还有一些基于节点标记架构的方法^[14-18]也研究了小样本的 GNN 学习。然而,节点标记架构的 GNN 分类依据是节点本身或节点所处的某一个类,无法对于不确定类别数的样本进行分类,而且由于其分类过程的局限性,无法最大限度地将分类所需的信息在图中进行表达。

1.2 边标记 GNN

边标记 GNN 的核心是将图结构中的边类型丰富化,使其可以表达更多的图结构信息或节点间的关系信息。相关性聚类(correlation clustering, CC)^[19]是第 1 个将边而不是节点作为聚类重心的 GNN 结构,其通过考虑同时最大化类内相似性和类间相异性,并把其表示在边的标签上进行聚类。Finley 等^[20]进一步将支持向量机模型应用到 CC 中,用以进行非词组聚类和新文章分类。Taskar 等^[21]在 CC 中加入了最大差公式计算边标签的权值,以此对同一幅图像产生 2 个不同的分类标签。Kim 等^[22]探索了将 CC 应用于超图中的图像分类,超图是指一种由多个小节点共同组成一个传统图中节点的图。

最近,随着图注意力网络中图注意力机制的发展,将可以同时表征局部内容特征和全局神经网络层特征的边标记加入图分类网络中^[10]。Kipf 等^[11]进一步将可解释学习的原理加入边标记 GNN

架构中,用以解决动态交互网络的学习。Johnson^[23]提出门控图转换网络(gated graph transformer neural networks, GGT-NN)解决自然语言处理任务,其在节点状态更新、信息传播和边更新中加入了多类别的边标签和几种不同的图变换操作解决这个问题。边标记 GNN 的出现,有效地改善了传统节点标记型的 GNN 在聚类时的性能表现,并通过将关系聚类的方法引入 GNN,使 GNN 在处理分类任务时效果更好。

1.3 小样本分类

小样本分类是元学习的一个分支,旨在通过少量样本训练获得一个可以完整表达样本数据间关系的模型。目前一种主流的方法是表征学习,将图像特征表征为节点特征,再通过相似度最近邻聚类的方法对节点特征进行聚类产生分类结果。相似度的表达可以是一个简单的距离度量,如余弦距离或欧几里得度量。孪生网络(Siamese networks, SIAMESE)^[24]在研究对偶文字分类中使用了一个可训练权值的 L_1 距离。在匹配网络(matching networks, MATCH)^[25]中进一步使用了来源于图注意力机制中的最近邻分类器作为相似度的度量。原生网络(prototypical networks, PROTO)^[26]则更进一步将最近邻分类器的基准定义为支持集中每个类嵌入后的均值。深度元学习(deep meta-learning, DEML)^[27]提出了一种概念学习器,将大量数据中的高维概念提取为特征,提高了分类的准确度,即通过更好的数据表征提高小样本分类方法的性能。

小样本学习优化模型参数的方法也脱胎于元学习。元长短期记忆网络(meta long short-term memory, meta-LSTM)^[28]中使用长短期记忆网络作为模型的更新器,并将模型参数作为模型的隐藏状态,通过输入小样本图像将模型参数的初始化和参数更新当作普通神经网络中的隐藏层更新来对待。模型无关元学习(model-agnostic meta-learning, MAML)^[29]只是将模型的参数做随机初始化,简单地通过梯度下降方法更新模型。这虽然是一个黑盒模型的方法,但可以应用于监督学习和增强学习的任务。爬行网络(reptile networks, RPTL)^[30]与 MAML 类似,不过它只是用一阶梯度进行模型的更新。简单注意力元学习器^[31]则是采用了时序卷积和软性注意力机制。

1.4 艺术风格分类

曹建芳等^[32]探索了一种基于特征融合的古壁画形象分类方法,采用一种双通道融合的特征

提取模块的深度卷积神经网络,对古代壁画的特征信息进行额外的提取,提高了分类的准确性.但其古代壁画分类任务中的数据集规模远大于东巴画,对于小样本的分类对象并不适用.高峰等^[33]提出了一种基于检测关键点和笔画特征的国画风格分类方法,通过将国画风格中不同手法独有点和线的特征进行稀疏表示建模,以此作为分类依据.但这种分类方法缺少对具体对象的分类,且只对具有特定笔画区别的对象适用.盛家川^[34]提出一种基于小波变换的国画特征提取分类,通过对不同时期的国画灰度图像进行小波分解,并通过分解后的小波系数进行分类.但其只对有明显风格区别的画作进行了分类,对于相似、相近风格的画作分类效果不佳,且不能对其中具体对象进行分类.

2 本文方法

2.1 东巴画的小样本分类

本节将从东巴画分类、小样本分类、幕式训练方法 3 方面综合描述本文的东巴画小样本分类方法.

2.1.1 东巴画的分类

东巴画中有丰富的艺术形象^[2-4].不同类型的东巴画象征着纳西族人民生活的一部分,表现了东巴文化对各种自然、鬼神、动物、天堂、人间等事物的崇拜.图 1 中的 4 种形象是东巴画中

图 1 东巴画艺术形象

最为典型和常用的艺术形象,本文采用它们作为东巴画的分类对象.同样地,东巴画与世界其他知名画作中的类似形象(如水墨画中的人物动物、抽象画中扭曲的人物动物、油画中的人物动物)进行区分分类也是分类任务之一.

2.1.2 小样本分类任务

本文采用小样本分类任务的形式作为东巴画小样本分类任务的定义.给定一个分类任务 T ,其中包含支持集 S ,即标签与输入图像成对的一个数据集,以及一个查询集 Q ,包含没有标签的待分类的图像.如果支持集 S 中包含 N 个类别,每个类别中有 K 个标记过的图像,那么这个分类任务就定义为 N -way K -shot 的分类任务.

2.1.3 幕式训练

东巴画创作对象来源于纳西族人民的现实生活场景,且其表现形式非常形象生动,本文将现实场景中的人物、动物等加入模型训练中,添补东巴画分类中所需要的信息,组成幕式训练作为本文的训练方法.其旨在给网络一个相对较大的已标记训练集,从中挑选出少量样本组成训练任务即一幕,近似模拟成一个小样本训练任务.因为选取出的训练任务与原数据集具有近似的分布,所以训练出的模型对于原数据集会有更好的表现.这是一种在元学习中很有效且被广泛应用于各种类似领域研究中的方法^[25,29,35].此外,通过加入特征显著性注意力机制,将东巴画风格明显的特征进行着重强化,实现了幕式训练在东巴画分类上的适用.

对于一个 N -way K -shot 的训练及测试任务,其幕式训练可定义为如下形式.对于任务 T ,有 $T = S \cup Q$,其中支持集 $S = \{(x_i, y_i)\}_{i=1}^{N \times K}$,查询集 $Q = \{x_i\}_{i=N \times K + 1}^{N \times K + \tau}$, τ 表示查询样本的数量, $x_i, y_i \in \{C_1, \dots, C_N\} = C_T \subset C_{all}$ 为第 i 个带有标签的输入图像, C_T 为任务 T 中包含的分类集合, C_{all} 为包含训练集和测试集中所有分类的集合.训练及测试任务中的图像是从原数据集中抽样选取的,且保证 $C_{train} \cap C_{test} = \emptyset$.那么在每一幕中,通过将支持集中已标记的图像数据输入模型进行训练去获得模型在查询集上的损失最小化.最终在若干次迭代后模型收敛获得最终的分类模型.

2.2 本文模型

本节将详细描述本文提出的小样本分类网络的具体结构.分类过程如图 2 所示.

图 2 本文图神经网络分类过程

2.2.1 特征提取网络

特征提取网络采用卷积神经网络作为基础架构, 并在其中加入了多分辨率与边缘特征融合的结构. 1/2 分辨率图像由原图通过下采样获得, 边缘图像由扩展高斯差分的方法获得. 每个卷积块包含 3×3 的卷积层, 批标准化层, Leaky-ReLU 激活层和池化层. 通道整合模块为将原图特征、1/2 图特征和边缘特征按通道连接后通过一个线性层, 进行批标准化, 然后通过 Leaky-ReLU 激活层输出嵌入特征 V_i . 该网络的具体构造如图 3 所示, 其中, N 表示该卷积块输出的特征通道数.

图 3 特征提取模块构造

2.2.2 特征显著性注意力模块

特征显著性注意力模块将原特征通过一个卷积网络产生显著性注意力指标, 给予原特征中显著性强的特征高的权值, 使整个特征更新的过程更加有针对性. 加入特征显著性注意力模块后, 在节点更新的过程中不再是只由边特征和节点特征指导更新, 还将特征显著性也作为参数加入更新过程, 可以使特征主体更加突出, 使相似特征聚合更新更加具有针对性, 如

$$\beta_{v,l-1}, \gamma_{v,l-1} = I(v_i^{(l)}; \theta_l^{(l-1)}) \quad (1)$$

其中, I 表示参数集为 θ_l 的显著性网络; v_i 表示结点 V_i 的特征; β 和 γ 为网络输出的特征显著性指标. 对于给定的一个节点特征 f_1 和其邻居节点特征 f_2 , 将 f_1 通过一个显著性表达网络 $I(v; \theta_l)$ 获得其特征显著性注意力表达, 然后以元素相乘的形式融合到 f_2 上, 将原本常用于 GNN 中的以链接聚合节点的更新方式 $W_1 f_1 \parallel W_2 f_2$, 提升为以特征显著性作为权值, 并围绕着 $W_1 f_1 \odot W_2 f_2$ 更新节点, 使节点聚合时信息的传递集中在显著特征上. 在幕式训练的过程中, 首先将东巴画之一作为分类任务中固定的一类, 其余类别随机产生, 通过一定次数的迭代后再将这一限制解除, 以此获得显著性模块对于东巴画显著特征的强化表达.

2.2.3 边标记 GNN 模块

对于给定的分类任务 T , 首先将输入数据通过一个多尺度、多分辨率的嵌入网络, 将输入的原图像转换为输入特征, 在这一步中对输入图像的形状特征、纹理特征和边缘特征进行提取. 然后初始化一个图结构, 图中的每个节点储存一个输入的图像特征, 图中的每 2 个节点之间的边表示其连接的 2 个节点之间的关系(如是否为同一类). 具体而言, $G = (V_G, E_G; T)$ 表示这个用来表达任务 T 的图结构. 其中, $V_G = \{V_i\}_{i=1, \dots, |T|}$ 表示图中的节点; $E_G = \{E_{ij}\}_{i,j=1, \dots, |T|}$ 表示图中节点间的边集; 用 v_i 和 e_{ij} 表示 V_i 和 E_{ij} 的特征, 且对于一个总样本数量为 $|T|$ 的分类任务 T , 有 $|T| = N \times K + \tau$.

(1) 节点特征的初始化

节点特征的初始化采用特征提取网络作为嵌入网络, 将原输入图像转换为节点特征 $v_i^{(0)} = f_{\text{emb}}(x_i; \theta_{\text{emb}})$, 其中, θ_{emb} 表示特征提取网络对应

的参数集. 通过多分辨率提取以及融合边缘特征后可以将原图像中的信息尽可能全面地转换成特征向量.

(2) 边特征的定义与初始化

边特征定义为 $\mathbf{e}_{ij} = \{e_{ijn}\}_{n=1,2}^2 \in [0,1]^2$, 其中, i 和 j 为这条边所连的 2 个节点序号, n 为维度. 边特征的 2 个维度分别用来表示其连接的 2 个节点之间的类内关系和类间关系, 即将 2 个节点之间的相似度和相异度分开进行表示和利用. 接着将节点对之间的相似性情况记录为 L_{ij} , 并根据节点对的标签对 L_{ij} 进行赋值, 如

$$L_{ij} = \begin{cases} 1, & L_i = L_j \\ 0, & \text{其他} \end{cases} \quad (2)$$

其中, L_i 表示数据 i 所带有的标签. 然后将边特征进行初始化, 如

$$\mathbf{e}_{ij}^{(0)} = \begin{cases} [1,0], & L_{ij} = 1, i, j \leq N \times K \\ [0,1], & L_{ij} = 0, i, j \leq N \times K \\ [0.5,0.5], & \text{其他} \end{cases} \quad (3)$$

其中, L_{ij} 由式(2)求得. 将边特征第 1 维作为节点相似性表达, 第 2 维作为节点相异性表达, 这样的设置便于接下来在对图中节点进行信息传递时将通过边类型的不同将节点间的特征相似性进行单独记录并传播, 其相异性也同理.

(3) 图计算与更新

本文的 GNN 包含 L 层的图计算, 在其中进行节点特征更新、信息传递、邻域节点聚合等操作. 给定处于 $l-1$ 层的节点特征和边特征 $\mathbf{v}_i^{(l-1)}$, $\mathbf{e}_{ij}^{(l-1)}$, 节点特征的更新首先进行邻域节点信息聚合. 对于处在 l 层的节点特征 $\mathbf{v}_i^{(l)}$ 的更新, 首先以其与邻居节点的边特征值作为将相似特征进行传递聚合的依据, 然后将节点特征 $\mathbf{v}_i^{(l-1)}$ 通过特征显著性注意力模块提取为显著性参数 $\beta_{v,l-1}$ 和 $\gamma_{v,l-1}$, 并将其同样作为依据加入. 可以将其表现为如下带有自注意力与特征显著性注意力相结合的方式, 即

$$\mathbf{f}_n = \beta_{v,l-1} \mathbf{v}_i^{(l-1)} + \gamma_{v,l-1} \quad (4)$$

$$\mathbf{v}_i^{(l)} = f_v^{(l)} \left(\left(\sum_j a_{ij1} \mathbf{f}_1 \mathbf{v}_j^{(l-1)} \right) \parallel \left(\sum_j a_{ij2} \mathbf{f}_2 \mathbf{v}_j^{(l-1)} \right); \theta_v^{(l)} \right) \quad (5)$$

其中, $a_{ijn} = \frac{e_{ijn}}{\sum_k e_{ikn}}$; f_v 为参数集为 θ_v 的节点更新

网络. 采用边特征指导节点特征的更新, 为了避免输出的节点特征尺度发生变化, 将节点特征进行

了正则化. 将式(1)(3)(4)(5)结合即构成了节点更新模块, 在 $f_v^{(l)}$ 这个网络中, 采用 3 层多层感知网络, 更新过程同时考虑了邻域节点中相似的节点的信息传递以及相异节点的信息传递, 即通过特征向量连接的方式将相似特征及相异特征同时传入节点更新网络中, 更新获得 l 层的节点特征.

获得了 l 层的所有新的节点特征后, 进行 l 层边特征的更新, 使用新的节点特征, 结合 $l-1$ 层的边特征重新计算边特征, 表示为

$$\bar{e}_{ij1}^{(l)} = \frac{f_s^{(l)}(\mathbf{v}_i^{(l)}, \mathbf{v}_j^{(l)}; \theta_s^{(l)}) \sum_k e_{ik1}^{(l-1)}}{\sum_k f_s^{(l)}(\mathbf{v}_i^{(l)}, \mathbf{v}_k^{(l)}; \theta_s^{(l)}) e_{ik1}^{(l-1)}} e_{ij1}^{(l-1)} \quad (6)$$

$$\bar{e}_{ij2}^{(l)} = \frac{(1 - f_s^{(l)}(\mathbf{v}_i^{(l)}, \mathbf{v}_j^{(l)}; \theta_s^{(l)})) \sum_k e_{ik2}^{(l-1)}}{\sum_k (1 - f_s^{(l)}(\mathbf{v}_i^{(l)}, \mathbf{v}_k^{(l)}; \theta_s^{(l)})) e_{ik2}^{(l-1)}} e_{ij2}^{(l-1)} \quad (7)$$

$$\mathbf{e}_{ij} = \frac{\bar{\mathbf{e}}_{ij}}{\|\bar{\mathbf{e}}_{ij}\|} \quad (8)$$

其中, $f_s^{(l)}$ 为参数集为 $\theta_s^{(l)}$ 的相似性度量网络, 用于计算 2 个节点的相似程度, 网络构成为一个简单的多层感知网络. 即以 2 个节点特征之差的绝对值作为网络的输入, 通过 3 层线性感知网络获得 2 个节点的相似性度量值, 相异性度量同理. 然后将它们与 $l-1$ 层的边特征向量中代表相似性与相异性的维度结合产生 l 层的边特征向量.

上述边特征的更新过程中, 节点特征的信息在显著性引导下先流向边特征, 然后边特征向量再由节点间相似性度量进行更新. 这样边特征中就同时包含了待更新节点与其他邻居节点间的相似性度量与相异性度量, 使相似性与相异性在边特征中记录与传播, 使每层节点特征更新与边特征更新更有针对性, 也保证了尽可能多的相似性信息在边特征中传递.

在经过共 L 层的全部节点特征和边特征的更新后, 此时的边特征(边标记)就代表了待分类节点与所有其他节点间的相似性相异性度量. 具体而言, 若此时待预测节点为 $\mathbf{z}_i$, 它的一个边特征为 $z_{ij1}^{(L)} = e_{ij1}^L$ 且 $z_{ij1}^{(L)} \in [0,1]$, 其边特征可以视为待分类节点 $\mathbf{z}_i$ 与已知类别节点 $\mathbf{v}_j$ 间的相似性度量, 2 个节点越相似, 那么这个边特征的值就越接近 1, 所以用待分类节点 $\mathbf{z}_i$ 与它的相邻节点的边特征权值判断 $\mathbf{z}_i$ 属于哪个类, 即 $\mathbf{z}_i$ 会被归类为边权值最大的边所连的另一个节点的类别. 特别地, 待分类节点

z_i 在任务 T 中类别为 C_k 的概率为

$$P(z_i = C_k | T) = P_i^{(k)} = \text{softmax}(z_{ij} \delta(z_j = C_k)) \quad (9)$$

其中, δ 为克罗内克函数, 其在函数内表达式相等时值为 1, 其他情况下值为 0; z_{ij} 表示待预测节点与各个支持集数据产生的节点之间的边特征值. 在常见的传统图分类方法中, 通常使用图聚类的方法, 即将图中的节点先分为各个聚类, 然后使用边的权值优化调整聚类, 最后通过待分类节点所在聚类具有的支持集标签对待分类节点进行分类预测. 而本文方法将边的权值直接作为待预测节点的分类依据, 使用式(9)进行分类预测. 本文方法的伪代码如下.

输入. 图像分类任务 $T = S \cup Q$, 其中支持集 $S = \{(x_i, y_i)\}_{i=1}^{N \times K}$, 查询集 $Q = \{x_i\}_{i=N \times K+1}^{N \times K+\tau}$.

输出. 查询集分类预测 $\{\hat{z}_i\}_{i=N \times K+1}^{N \times K+\tau}$.

Step1. 采用特征提取网络获得节点特征 $v_i = f_{\text{emb}}(x_i; \theta_{\text{emb}})$.

Step2. 初始化图 $G = (V_G, E_G; T)$, 将节点特征嵌入为 $v_i^{(0)}$, 采用式(3)初始化边特征为 $e_{ij}^{(0)}$.

Step3. GNN 模块:

```
for  $l = 1, \dots, L$  do //按层更新图
  for  $i = 1, \dots, V$  do //节点特征更新模块
     $I(v_i^{(l-1)}; \theta_i^{(l-1)}) \rightarrow \beta_{v_i, l-1}, \gamma_{v_i, l-1}$  //获取显著性注意力
     $f_v(e_{ij}^{(l-1)}, \beta_{v_i, l-1}, \gamma_{v_i, l-1}; \theta_v^l) \rightarrow v_i^{(l)}$  //更新节点特征
  end
  for  $(i, j) = 1, \dots, E$  do //边特征更新模块
     $f_e(v_i^{(l-1)}, e_{ij}^{(l-1)}, \gamma_{v_i, l-1}, \beta_{v_i, l-1}; \theta_e^l) \rightarrow e_{ij}^{(l)}$  //更新边特征
  end
end
```

Step4. 分类预测, 由式(9)获得待预测节点的分类.

东巴画动物、人物、鬼、神等形象共 14 个类别的 mini-东巴数据集.

除此之外, 还额外搜集了水墨画、油画、抽象画中的艺术形象, 共 9 类, 每类 20 幅图像, 加上东巴数据集中的 4 类图像, 构成多风格形象数据集. 其中所有图像为 84×84 像素分辨率的 RGB 图像, 共有 13 个类别. 不同于 mini-东巴数据集为现实形象与东巴画的数据集, 多风格形象数据集是加入了其他画作的艺术形象后, 包含了 4 种不同艺术风格中 13 类艺术形象的数据集. 使用本数据集进行艺术画作领域的模型训练与验证. 图 4 展示了上述数据集中的部分代表图像, 图 4a 从左至右分别为东巴画中人、神、动物、鬼怪; 图 4b 从左至右分别为水墨画卷人物、白底水墨画人物、油画人物、抽象画人物. 东巴画形象相较于其他艺术风格形象最明显的区别是其形象具有清晰明显的边缘结构, 且形象立体度比较扁平, 形象中色彩对比度较高, 较少使用渐变. 通过实验验证本文方法是否对东巴画艺术形象分类更加有效.

a. 部分东巴画形象

b. 部分其他风格形象

图 4 数据集部分示例

3 实验结果与分析

3.1 数据集构成

本文搜集真实东巴画图像并进行剪裁、翻转等处理后建立了东巴数据集, 其中东巴画图像为人工搜集与标注的图像, 包含 4 个分类分别为动物、人物、鬼、神, 每个分类包含 15~20 幅图像, 采用 84×84 像素分辨率以及 RGB 颜色. mini-imagenet 是一个常用的小样本分类数据集, 是由 ILSVRC-12 中分离出来的子数据集. 其中所有的图像为 84×84 像素分辨率的 RGB 图像, 包含 100 个不同的分类, 每个分类有 600 幅图像, 从中选出 10 个分类, 每类 100 幅, 用于构建本文的数据集. 通过将上述 2 个数据集合并到一起, 构成包含真实动物、人物以及

3.2 实验设置

本文的模型训练采用初始学习率为 5×10^{-4} 的 Adam 优化器, 权重衰减为 10^{-6} . 对于 mini-东巴数据集, 5-way 的训练采用大小为 16 的 meta-set, 每 15 000 幕迭代后将学习率减半. 对于多风格形象数据集, 5-way 的训练采用大小为 10 的 meta-set, 每 5 000 幕迭代后将学习率减半. 使用 Python 进行编程并在 NVIDIA RTX 2080Ti 进行训练.

本文的模型评估采用小样本学习中的标准方法 5-way 5-shot 的设置, 即 5 个类别, 每个类别 5 幅图像作为训练数据; 然后为这 5 个类别每类再随机抽取 3 幅图像作为测试数据. 模型性能评估使用 500 次从测试集中随机产生的幕式训练任务结果

的均值作为最终结果。

3.3 小样本分类结果

本文提出的模型在 2 个包含东巴画的数据集上的小样本分类性能如表 1 所示, 所有的结果都是 500 次测试结果的均值。表 1 中同时对比了几个经典小样本分类方法。在传导列中, “否”表示没有使用传导学习方法, 即每个查询样本的分类仅由支持集中已标注样本数据进行预测; “是”表示该模型应用了传导学习方法, 即分类过程中查询样本也作为分类的部分依据; “BN”表示该模型在分类过程中使用了查询集的一些统计特征而不是只使用全局批标准化, 所以也可在一定程度上看做传导学习。

表 1 不同模型是否使用传导方式准确率对比

模型	传导	准确率/%	
		mini-东巴数据集	多风格形象数据集
MATCH ^[25]	否	56.97	56.31
RPTL ^[30]	否	64.56	63.39
PROTO ^[26]	否	67.61	66.79
GNN ^[6]	否	68.45	67.62
EGNN ^[9]	否	68.73	67.35
本文	否	69.82	68.70
MAML ^[29]	BN	64.81	63.57
RPTL+BN	BN	67.90	67.08
Relation Net ^[35]	BN	68.95	67.78
MAML+传导	是	71.26	69.34
TPN ^[7]	是	71.51	69.58
TPN(高 K 值)	是	71.94	69.92
EGNN+传导	是	76.32	74.03
本文+传导	是	78.51	76.15

本文在 5-way 5-shot 的分类设置上分别进行了非传导和传导学习方法的实验。在非传导学习方法的实验中, 本文模型在 2 个东巴画数据集上都达到了最高的准确率, 并且其比同样是边标记结构的 EGNN 效果好。EGNN 并没有显著特征检测的机制, 也没有多分辨率特征提取机制, 而仅是对输入图像做了粗略的提取, 本文方法考虑了东巴画风格独特鲜明的特点, 加入了更加丰富的特征提取以及显著特征的强化, 并以显著特征为引导进行边特征的更新, 使更新更有效率的同时也更有针对性, 所以在东巴画的分类效果上更佳。

在传导学习的实验中, 本文模型比 TPN 和 EGNN 的分类准确率都高。在节点更新的过程中, EGNN 只考虑了边可用作节点特征更新, 而本文方法将边与特征显著性注意力权值共同作为更新

依据, 使节点更新的效率更高, 更有针对性。TPN 中使用一个拉普拉斯矩阵判断支持集与查询集的相似性, 然后将支持集的标签按照节点间的相似度传播到查询集上, 这意味着在 TPN 中图像间的信息传递只限于它们嵌入后的特征之间的相似性。不同于 TPN, 本文模型在节点间的信息传递不仅有节点间的相似特征, 同时还包含节点本身显著特征的表达, 以及他们之间边的特征, 都进行了传递。而且 TPN 中的节点特征是固定不变的, 这样模型就没法在迭代更新的过程中适应不同的分类需求。而本文方法无论是节点特征还是特征显著性注意力参数以及边特征都是动态的, 在迭代的过程中可以不断适应任务的需求, 所以灵活性更好。

3.4 东巴画小样本分类结果

为了对比本文模型在纯东巴画分类的效果, 采用第 3.2 节中描述的 5-way 5-shot 的方法, 设置了东巴画的类别在 5 个类别中占不同的比例。例如, 20% 表示 5 个类中有 1 个类别为东巴画相关的类别。表 2 和表 3 展示了东巴画占比为 20%~80% 的实验结果; “传导”表示使用了传导学习方法, 并与 EGNN 进行了对比。在东巴画比例为 80% 的实验对比中, 同样是非传导学习方法, 本文模型比 EGNN 的 58.98% 准确率高出约 6 个百分点, 并且在使用传导学习的方法时更是高出了约 10 个百分点。可以看出, 在各种东巴画类别的占比中, 本文模型比 EGNN 和 GNN 有更好的分类准确率和更低的标准差。即本文方法在东巴画分类上更加有效。

表 2 各模型对东巴画不同占比下分类准确率对比

模型	占比/%			
	20	40	60	80
GNN ^[6]	76.37	67.96	60.49	54.38
EGNN ^[9]	78.65	70.75	65.12	58.98
本文	88.93	80.92	72.99	65.16
本文+传导	90.97	82.78	75.01	68.63

表 3 各模型对东巴画不同占比下分类标准差对比

模型	占比/%			
	20	40	60	80
GNN ^[6]	6.69	6.73	6.72	6.71
EGNN ^[9]	6.70	6.78	6.82	6.70
本文	6.65	6.77	6.80	6.68
本文+传导	6.32	6.37	6.45	6.34

3.5 本文其他实验分析

为了验证本文提出的多尺度多特征融合的特征提取器对模型性能的影响, 对不同的特征提取

器在模型分类准确率上的影响做了实验, 实验采用第 3.2 节中相同的实验设置, 在 mini-东巴数据集上进行测试. 表 4 展示了本文采用多特征融合特征提取器的有效性, 其中“本文+CNN”指只采用主特征提取器, 去掉了边缘特征和1/2分辨率特征提取器; “本文+边缘特征”指增加了边缘特征提取器; “本文+多特征融合”指使用了全部 3 种特征提取并做融合. 在增加了边缘特征和低分辨率图特征后分类准确率有一定程度的提高. 边缘特征包含了形象的轮廓特征, 低分辨率图像则淡化了形象的具体细节纹理, 重点突出了形象的全局结构特征, 扩大了感受野, 使模型能获得更多的空间信息, 改善了分类准确性.

表 4 不同特征提取器分类准确率对比 %

模型	准确率
本文+CNN	77.13
本文+边缘特征	77.62
本文+多特征融合	78.51

为了直观地展示本文的分类过程和效果, 通过 t -SNE 对本文模型产生的边特征值做了可视化, 用以表明整个分类过程中, 如何从嵌入图像的初始化边特征值到最终得到分类结果的边特征值. 图 5 所示为一个 3-way 3-shot 的东巴画小样本分类的过程, 其中对角线上的色块表示样本本身, 其他位置的色块为其对应行列的样本间的边特征值, 其颜色越接近红色, 代表边特征值越接近 1, 蓝色则为 0. 其中, 图 5a 表示所有的边刚初始化完, 图中前 9 个为支持集的边特征值, 后 9 个为查询集的边标签. 初始化时产生的边特征值, 所有查询集节点的边特征值都被设为 0.5, 然后通过迭代传播, 边特征值逐渐被真实的边特征值所改变并在 GNN 的层之间传播. 可以看到, 在初始化时在同一个类之间的边特征值为 1(红色), 查询集与支持集之间的标签被设置为 0.5(灰色), 然后随着相似特征在边上的传播, 由第 1 层(图 5b)传播到第 2 层(图 5c), 直至最后 1 层(图 5d), 查询集与相应分类的支持集间的边特征逐渐接近 1, 即产生了正确的分类. 将图 5d 中第 1 个类别的 3 个支持集样本与其同类别的一个查询样本以及一个不同类别的查询样本进行详细表示, 如图 6 所示. 前 3 行的图像为“东巴画人物”的支持集, 后 2 行的图像为查询样本 1 和查询样本 2, 其中查询样本 1 是来自“东巴画人物”类别的查询图像, 查询样本 2 为来自“东巴画鬼神”类

别的查询图像. 最终每个图像间的相似性信息如图 6 所示, 人物类别的查询样本与支持集之间的边特征值很接近 1(橙红色), 而鬼神类别的查询样本与支持集之间的边特征值则接近 0(浅蓝色), 说明了本文模型的分类正确性.

图 5 t -SNE 边特征可视化

图 6 t -SNE 边特征可视化的一部分放大

4 结 语

本文针对东巴画图像小样本分类的问题, 提

出将多特征融合以及特征显著性注意力加入边标记 GNN 中,并基于特征线性模块设计并实现了一种特征显著性注意力模块,同时以节点特征、边特征以及显著特征作为引导进行聚合,通过相似性计算再更新边特征,最终由边特征产生分类结果.多特征融合的特征提取方法使从原图像中获得的信息更加全面,特征显著性注意力机制使特征更新的过程更有针对性,更符合东巴画主体突出的特点.以边特征作为分类信息的载体的方法使东巴画分类过程中信息传递更加丰富有效.与 GNN 和 EGNN 相比,本文采用了多分辨率多尺度的特征提取器,提高了图像特征的采集和表达.采用复合注意力机制,使聚类过程对东巴画更有针对性.并在东巴画的数据集上获得了更好的分类准确率,更好地解决了将东巴画进行分类的问题.但还存在对部分与神佛穿戴特征相似性很高的人物分类准确率不高,没有对人物男女进行区分分类的问题,可能是由于模型缺少更进一步的对细粒度特征的注意力机制,这也是下一步需要研究的.

小样本分类是目前计算机辅助分类技术的一个重要方向.然而,当数据量较大时小样本分类的方法相比传统神经网络计算量较大,处理数据的效率不如传统的神经网络.因此,将小样本分类的方法扩展到大数据领域是将来进一步需要研究的方向.

参考文献(References):

- [1] Liu Liwei. Interpretation and reconstruction of Dongba characters in modern Dongba paintings[J]. Yi Hai, 2018(5): 58-60 (in Chinese)
(刘丽伟. 现代东巴画中东巴文的解读与重构[J]. 艺海, 2018(5): 58-60)
- [2] Wu Shusheng. Cultural and artistic precious of Dongba of Naxi nationality: Dongba painting[J]. Local Culture Research, 2019 (6): 2(in Chinese)
(吴树生. 纳西族东巴文化艺术结晶: 东巴画[J]. 地方文化研究, 2019(6): 2)
- [3] Zhang Lei. A brief analysis of the characteristics of Naxi Dongba painting[J]. Ability and Wisdom, 2016(36): 200(in Chinese)
(张垒. 纳西族东巴绘画的特征浅析[J]. 才智, 2016(36): 200)
- [4] He Yong. On the characteristics of Dongba painting and its appreciation methods[J]. Art of the Public, 2015(14): 107(in Chinese)
(和勇. 试论东巴画的特征及欣赏方法[J]. 大众文艺, 2015(14): 107)
- [5] Qian Wenhua, Xu Dan, Xu Jin, *et al.* Simulation of Dongba art style painting[J]. Journal of System Simulation, 2020, 32(7): 1349-1359(in Chinese)
(钱文华, 徐丹, 徐瑾, 等. 东巴画艺术风格绘制[J]. 系统仿真学报, 2020, 32(7): 1349-1359)
- [6] Garcia V, Bruna J. Few-shot learning with graph neural networks[OL]. [2020-08-31]. <https://arxiv.org/pdf/1711.04043.pdf>
- [7] Liu Y B, Lee J, Park M, *et al.* Learning to propagate labels: transductive propagation network for few-shot learning[OL]. [2020-08-31]. <https://arxiv.org/pdf/1805.10002v5.pdf>
- [8] Gong L Y, Cheng Q. Exploiting edge features for graph neural networks[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 9203-9211
- [9] Kim J, Kim T, Kim S, *et al.* Edge-labeling graph neural network for few-shot learning[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 11-20
- [10] Al-Shedivat M, Bansal T, Burda Y, *et al.* Continuous adaptation via meta-learning in nonstationary and competitive environments[OL]. [2020-08-31]. <https://arxiv.org/pdf/1710.03641.pdf>
- [11] Kipf T, Fetaya E, Wang K C, *et al.* Neural relational inference for interacting systems[C] //Proceedings of the 35th International Conference on Machine Learning. New York: ACM Press, 2018: 2688-2697
- [12] Beck D, Haffari G, Cohn T. Graph-to-sequence learning using gated graph neural networks[C] //Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL Press, 2018: 273-283
- [13] Ghorbani M, Baghshah M S, Rabiee H R. MGCN: semi-supervised classification in multi-layer graphs with graph convolutional networks[C] //Proceedings of International Conference on Advances in Social Networks Analysis and Mining. Los Alamitos: IEEE Computer Society Press, 2019: 208-211
- [14] Battaglia P W, Hamrick J B, Bapst V, *et al.* Relational inductive biases, deep learning, and graph networks[OL]. [2020-08-31]. <https://arxiv.org/pdf/1806.01261.pdf>
- [15] Bronstein M M, Bruna J, LeCun Y, *et al.* Geometric deep learning: going beyond Euclidean data[J]. IEEE Signal Processing Magazine, 2017, 34(4): 18-42
- [16] Xu K, Hu W H, Leskovec J, *et al.* How powerful are graph neural networks?[OL]. [2020-08-31]. <https://arxiv.org/pdf/1810.00826.pdf>
- [17] Gilmer J, Schoenholz S S, Riley P F, *et al.* Neural message passing for quantum chemistry[C] //Proceedings of the 34th International Conference on Machine Learning. New York: ACM Press, 2017: 1263-1272
- [18] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains[C] //Proceedings of the IEEE International Joint Conference on Neural Networks. Los Alamitos: IEEE Computer Society Press, 2005: 729-734
- [19] Bansal N, Blum A, Chawla S. Correlation clustering[J]. Machine Learning, 2004, 56: 89-113
- [20] Finley T, Joachims T. Supervised clustering with support vector machines[C] //Proceedings of the 22nd International Conference on Machine Learning. New York: ACM Press, 2005: 217-224
- [21] Taskar B, Chatalbashev V, Koller D, *et al.* Learning structured

- prediction models: a large margin approach[C] //Proceedings of the 22nd International Conference on Machine Learning. New York: ACM Press, 2005: 896-903
- [22] Kim S, Yoo C D, Nowozin S, *et al.* Image segmentation using higher-order correlation clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(9): 1761-1774
- [23] Johnson D D. Learning graphical state transitions[OL]. [2020-08-31]. <https://openreview.net/forum?id=HJONvFzxl>
- [24] Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition[OL]. [2020-08-31]. <http://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf>
- [25] Vinyals O, Blundell C, Lillicrap T, *et al.* Matching networks for one shot learning[C] //Proceedings of the Annual Conference on Neural Information Processing System. Cambridge: MIT Press, 2016: 3630-3638
- [26] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[C] //Proceedings of Annual Conference on Neural Information Processing System. Cambridge: MIT Press, 2017: 4077-4087
- [27] Zhou F W, Wu B, Li Z G. Deep meta-learning: learning to learn in the concept space[OL]. [2020-08-31]. <https://arxiv.org/pdf/1802.03596.pdf>
- [28] Ravi S, Larochelle H. Optimization as a model for few-shot learning[OL]. [2020-08-31]. <https://openreview.net/forum?id=rJY0-KcII>
- [29] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C] //Proceedings of the 34th International Conference on Machine Learning. New York: ACM Press, 2017: 1126-1135
- [30] Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms[OL]. [2020-08-31]. <https://arxiv.org/pdf/1803.02999.pdf>
- [31] Mishra N, Rohaninejad M, Chen X, *et al.* A simple neural attentive meta-learner[OL]. [2020-08-31]. <https://openreview.net/forum?id=B1DmUzWAW>
- [32] Cao Jianfang, Cui Hongyan, Zhang Qi. Ancient mural classification with AlexNet model using feature fusion[J]. Journal of Image and Graphics, 2020, 25(1): 92-101(in Chinese)
(曹建芳, 崔红艳, 张琦. 特征融合 AlexNet 模型的古代壁画分类[J]. 中国图象图形学报, 2020, 25(1): 92-101)
- [33] Gao Feng, Nie Jie, Huang Lei, *et al.* Traditional Chinese painting classification based on painting techniques[J]. Chinese Journal of Computers, 2017, 40(12): 2871-2882(in Chinese)
(高峰, 聂婕, 黄磊, 等. 基于表现手法的国画分类方法研究[J]. 计算机学报, 2017, 40(12): 2871-2882)
- [34] Sheng Jiachuan. Automatic categorization of traditional Chinese paintings based on wavelet transform[J]. Computer Science, 2014, 41(2): 317-318(in Chinese)
(盛家川. 基于小波变换的国画特征提取及分类[J]. 计算机科学, 2014, 41(2): 317-318)
- [35] Sung F, Yang Y X, Zhang L, *et al.* Learning to compare: relation network for few-shot learning[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 1199-1208