

半监督场景下多视角信息交互的图卷积神经网络*

王悦天¹, 傅司超¹, 彭勤牧¹, 邹斌², 荆晓远³, 尤新革¹



¹(华中科技大学 电子信息与通信学院, 湖北 武汉 430074)

²(湖北大学 数学与统计学学院, 湖北 武汉 430062)

³(武汉大学 计算机学院, 湖北 武汉 430072)

通信作者: 傅司超, E-mail: fuschao_hust@hust.edu.cn; 尤新革, E-mail: youxg@hust.edu.cn

摘要: 在当前数据来源多样化且人工标记难度大的现实生活中, 半监督场景下多视角数据的分类算法在各个领域中都具有重要的研究意义. 近年来, 基于图神经网络的半监督多视角分类算法研究已经取得了很大的进展. 但是现有的图神经网络算法大多是在分类阶段进行多视角互补信息的融合, 反而忽略了训练阶段同一样本不同视角间互补信息的交互. 针对上述问题, 提出半监督场景下多视角信息交互的图卷积神经网络算法 MIGCN (multi-view interaction graph convolutional network). 该方法通过在不同视角上训练的图卷积层之间引入 Transformer Encoder 模块, 使得同一样本在训练阶段都可以通过注意力机制自适应的在不同视角间获取互补性信息, 进而加强自身的训练; 除此之外, 还通过引入一致性约束损失让不同视角最终特征表达的相似关系尽可能一样, 促使图卷积神经网络在分类阶段更加合理的利用多视角特征之间的一致性和互补性信息, 进一步提升多视角融合特征的鲁棒性. 最后, 在多个真实世界多视角数据集上的实验表明, 相比于基于图的半监督多视角分类模型, MIGCN 可以更好地学习到多视角数据的本质特征, 进而提升半监督多视角分类的准确性.

关键词: 半监督学习; 多视角学习; 图神经网络; 信息交互; 节点分类

中图法分类号: TP18

中文引用格式: 王悦天, 傅司超, 彭勤牧, 邹斌, 荆晓远, 尤新革. 半监督场景下多视角信息交互的图卷积神经网络. 软件学报, 2024, 35(11): 5098–5115. <http://www.jos.org.cn/1000-9825/7007.htm>

英文引用格式: Wang YT, Fu SC, Peng QM, Zou B, Jing XY, You XG. Multi-view Interaction Graph Convolutional Network for Semi-supervised Classification. Ruan Jian Xue Bao/Journal of Software, 2024, 35(11): 5098–5115 (in Chinese). <http://www.jos.org.cn/1000-9825/7007.htm>

Multi-view Interaction Graph Convolutional Network for Semi-supervised Classification

WANG Yue-Tian¹, FU Si-Chao¹, PENG Qin-Mu¹, ZOU Bin², JING Xiao-Yuan³, YOU Xin-Ge¹

¹(School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China)

²(Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China)

³(School of Computer Science, Wuhan University, Wuhan 430072, China)

Abstract: In current real life where data sources are diverse, and manual labeling is difficult, semi-supervised multi-view classification algorithms have important research significance in various fields. In recent years, graph neural networks-based semi-supervised multi-view classification algorithms have achieved great progress. However, most of the existing graph neural networks carry out multi-view information fusion only in the classification stage, while neglecting the multi-view information interaction between the same sample during the training stage. To solve the above issue, this study proposes a model for semi-supervised classification, named multi-view interaction graph convolutional network (MIGCN). The Transformer Encoder module is introduced to the graph convolution layer trained on different

* 基金项目: 国家重点研发计划 (2022YFC3301004); 国家自然科学基金 (62172177); 中央高校基本科研业务费专项资金 (2022JYCXJ034)
收稿时间: 2022-08-03; 修改时间: 2022-11-16; 采用时间: 2023-08-16; jos 在线出版时间: 2023-11-29
CNKI 网络首发时间: 2023-12-01

views, which aims to adaptively acquire complementary information between different views for the same sample during the training stage. More importantly, the study introduces the consistency constraint loss to make the similar relationship of the final feature expressions of different views as similar as possible. This operation can make graph convolutional neural networks during the classification stage better utilize the consistency and complementarity information between different views reasonably, and then it can further improve the robust performance of the multi-view fusion feature. Extensive experiments on several real-world multi-view datasets demonstrate that compared with the graph-based semi-supervised multi-view classification model, MIGCN can better learn the essential features of multi-view data, thereby improving the accuracy of semi-supervised multi-view classification.

Key words: semi-supervised learning; multi-view learning; graph neural network (GNN); information interaction; node classification

随着互联网、物联网、云计算和大数据等信息化技术的快速发展,人类所产生的数据正以几何式增长,每时每刻都有海量的数据生成.在这样的大数据时代里,由于各个领域研究的不断深入与网络应用的不断丰富,现实生活中数据的生产渠道、表示模态和信息化形式都变得更加多样化、复杂化;数据采集技术和数据处理技术的快速发展,也让不同来源、不同类型和不同特征描述下的数据获取更加方便和快捷.如图1所示,一个现实对象可以通过图片、文本、视频和音频等形式来描述;一个新闻事件通常会被不同国家的多个新闻媒体进行报道;一个数字图像可以通过不同的视觉特征来描述;一个特定物体可以从不同的视角进行拍摄与观察.这种针对同一个样本具有多个信息源(异源)或者多种不同数据表示结构(异构)的数据被称为多视角或者多视图数据(multi-view data)^[1,2].

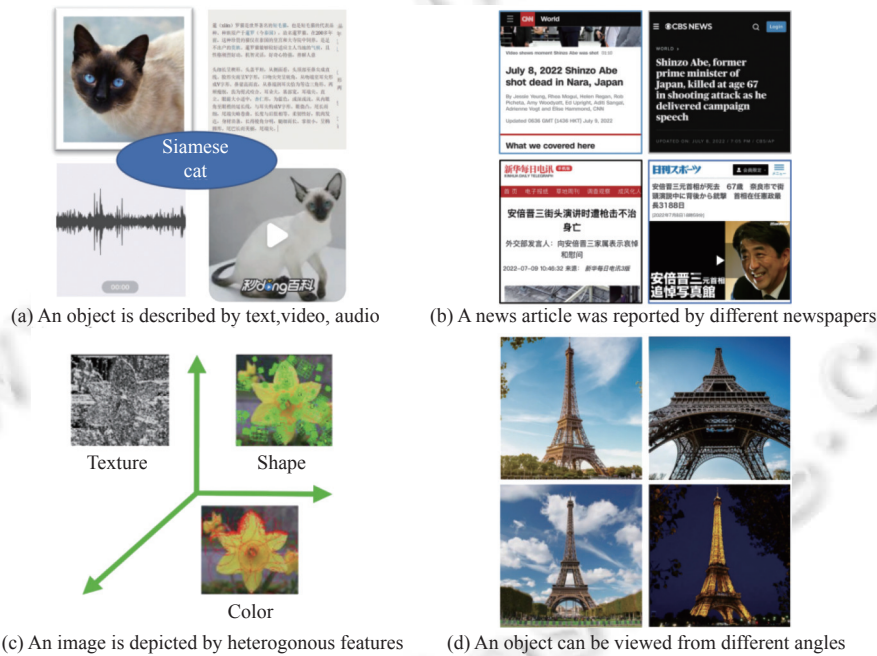


图1 多视角数据(multi-view data)示例

与此同时,用于机器学习模型训练的数据标签通常需要利用相关领域专家的先验知识进行人工标注,这往往会耗费巨大的人力、财力与时间.例如,在医疗图像分析与工业异常检测任务中,标记数据往往需要主任医生与高级工程师等专家参与,获取标注的成本高、难度大^[3].因此,在处理现实生活场景的任务时,数据中往往存在着大量未标记的节点,即所得到的数据一般只带有少许的标记信息.

综上,随着信息技术的快速发展,数据的形式与来源越来越多样化,同一对象从不同方式或不同层次获取的多视角数据目前已十分普遍;同时,样本标注的高昂成本也使得真实世界数据中往往只含有少量的标记信息,而半监督学习通过综合利用有限的标记数据与丰富的无标记数据来对数据分布进行建模,能够解决标记获取困难、耗时的现实矛盾,进一步拓展了机器学习的适用性.因此,半监督多视角学习(semi-supervised multi-view learning)近年来

引起了众多机器学习与深度学习研究人员的广泛关注^[4].

现有的半监督多视角分类算法大致包括基于子空间、基于核、基于分歧和基于图的 4 种主流范式^[5]. 基于子空间的方法旨在通过子空间学习为所有的视角学习一个共同的潜在子空间, 在低维子空间中保留特定的统计属性并进行标签传播^[6-10]. 基于核的模型旨在通过核方法找出样本的结构关联, 对不同视角的特征空间采取不同的核, 然后将多个核来学习结果进行融合从而提升整体的学习性能^[11-13]. 基于分歧和基于图的方法因为和深度学习的结合研究而迅速发展, 涌现出众多优秀的算法模型^[1], 本文的工作也与之相关.

基于分歧的方法源于协同训练 (co-training)^[14], 而协同训练最初是为只有两个不同视角的数据设计的. 它首先利用标记的数据来训练特定视角的分类器, 然后在每个视角上对未标记的数据分别进行分类, 将每个分类器中置信度最高的预测样本添加到另一分类器的训练集中, 过程一直持续到所有训练好的分类器达成一致. 但是, 在面对多视角数据时, 协同训练具有大量的超参数需要调整, 很难满足现实中的标签噪声假设, 且最终分类精度很容易受到误导性伪标签的影响, 因此在实践中的适用性非常有限. 这从而衍生出一些协同正则化与基于分歧的工作, 如协同期望最大化 (co-trained expectation maximization)^[15]、协同正则化 (co-regularization)^[16]和贝叶斯协同训练 (Bayesian co-training)^[17]等. 与此同时, 文献 [18] 中指出, 协同训练可以视为在一个汇聚了各个视图结构信息的联合图上进行信息交互的过程, 这也启发了 Co-GCN^[19]与我们的工作.

源于图神经网络强大的信息提取能力, 近年来也有许多工作尝试将其应用在一些多源或具有多种异构关系的数据上^[20], 相关的代表性工作如下: Li 等人提出了能够利用混合拉普拉斯矩阵交换视角间结构信息的多视角分类方法 Co-GCN^[19], 其将协同训练和图卷积网络统一到一个框架中, 在各视角下分别建立图卷积神经网络, 通过交替优化分别更新各视角的拉普拉斯矩阵与图卷积层参数的方法, 从而利用多视角数据的信息; MGAT^[21]是一种基于图注意力网络 (graph attention network, GAT) 的多视角图神经网络算法, 其通过不同视角下网络参数受到一个正则化项约束的 GAT 模型, 进而从每个视角中学习节点的特征, 最后, 利用一种基于注意力的方法来整合来自多个视角的节点信息.

通过上述分析可以发现, 现有的图卷积神经网络在解决半监督场景下多视角分类问题时, 大多数都是在分类阶段进行多视角互补信息的融合, 即在送入到分类器之前, 通过简单的拼接/求和/加权操作进行多视角数据的特征融合. 然而, 在训练过程中, 基于不同视角数据训练的相互独立的单视角图神经网络却无法利用同一样本不同视角的互补信息加强自身的训练. 因此, 现在的方法很难保证相互独立的单视角图神经网络在不同视角上最终特征提取上的鲁棒性, 进而削弱了在分类阶段基于多视角互补信息融合特征的鲁棒性.

针对上述问题, 本文提出了多视角信息交互的图卷积神经网络模型 MIGCN (multi-view interaction graph convolutional network), 以更加有效地解决多视角数据的半监督分类任务. MIGCN 通过在模型的每层 block 中先后应用各视角间相互独立的 GCN layer 及不同视角间进行信息交互的 Transformer encoder layer, 促使每个样本都可以通过图卷积操作与注意力机制自适应地从视角内邻居节点与不同视角间的互补信息中提取有价值的信息, 进而加强自身的训练, 进而提升单视角图神经网络的鲁棒性; 除此之外, 为了促使模型在分类阶段更加合理的利用多视角数据间的一致性信息, 引入了一种新的视角间的一致性约束损失来让不同视角最终特征表达的余弦相似度矩阵的差值尽可能最小; 最后, 通过使用基于注意力的全局特征融合方法让样本自适应的聚合各视角的互补信息得到其最终的全局特征表达, 进一步提升了多视角融合特征的鲁棒性.

在多个公开的多视角数据集上的实验证明, 相比现有的基于图的半监督多视角分类方法, 我们的方法可以更加合理的利用多视角数据间的互补性信息, 最大化多视角数据之间的一致性信息, 进而提升模型分类性能.

本文工作的主要贡献总结如下.

(1) 提出了一种新颖的半监督场景下多视角信息交互的图卷积神经网络算法. 该方法通过综合利用图卷积神经网络与注意力机制实现训练阶段相互独立的单视角图神经网络在多视角互补信息上的交互, 进而加强自身的训练. 并进一步结合视角间的一致性约束与基于注意力的全局特征融合, 可以为每个样本自适应的学习到其更好的全局特征表达, 有效地提升了半监督多视角分类的准确性.

(2) 设计了一种新的视角间的一致性约束函数. 通过约束不同视角间最终特征表达余弦相似度矩阵的差异, 促

使模型更加合理的利用多视角间的一致性信息,从而进一步提升模型的性能。

(3) 通过实验验证了本文所提出方法的有效性,并设计消融实验,进一步验证了基于自注意力机制的视角间互补信息交互在提取多视角样本中更加全面的特征中的关键作用,与基于注意力的特征融合与视角间一致性约束带来的分类精度提升。

本文第1节介绍模型主要应用到的图卷积神经网络与自注意力机制的相关工作与知识.第2节详细介绍了本文提出的多视角信息交互的图卷积神经网络模型(MIGCN).第3节为实验部分,在真实数据集上进行了大量的实验,测试本文提出MIGCN模型的性能.最后在第4节总结全文,并展望未来研究及后续改进的方向。

1 相关工作

本文所提出的方法主要基于图卷积神经网络、基于图的多视角半监督分类和自注意力机制,下面将对相关工作 and 知识予以介绍。

1.1 图卷积神经网络

深度学习在图像、视频、文本和语音等应用上取得的成功,给众多技术领域带来了巨大进步与深刻变革,但其中的算法模型多用于处理欧氏空间数据(例如,文本和语音为序列结构,图像为矩阵结构),相关技术无法直接应用于非欧结构数据.与此同时,在数理表达与逻辑抽象方面表示能力强大的非欧结构(尤其是图结构)数据又在现实生活中广泛地存在与应用.因此,为了解决这一矛盾冲突,关于在深度学习中如何合理高效利用图结构数据的研究越来越受到重视,图神经网络(graph neural network, GNN)^[22,23]应运而生并由此兴起,并受到了研究者的广泛关注。

根据文献[23],目前主流的图神经网络可分为4大类别:卷积图神经网络(convolutional graph neural networks, ConvGNNs)、循环图神经网络(recurrent graph neural networks, RecGNNs)、图自编码器(graph autoencoders, GAEs)和图时空网络(spatial-temporal graph neural networks, STGNNs).其中,卷积图神经网络(ConvGNNs)受启发于卷积神经网络(CNN)使用卷积运算提取图像信息的方式,并将其从二维结构化数据推广到图领域.其主要思想是通过对节点自身特征与其邻居特征的收集聚合,进而生成节点更新后的表示.下面,我们将选择其中最为经典同时也是应用最为广泛的图卷积神经网络(graph convolutional network, GCN)做简要介绍。

最初的基于谱域的卷积图神经网络^[24]的想法是希望从信号的角度在图域上做谱卷积(spectral convolution)操作,对于图域信号 $x \in \mathbb{R}_N$ 与参数为 $\theta \in \mathbb{R}_N$ 的滤波算子 $g_\theta = \text{diag}(\theta)$,其在谱域的图滤波操作如公式(1)所示:

$$g_\theta \star x = U g_\theta U^T x \quad (1)$$

其中, U 是由归一化图拉普拉斯矩阵 $L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^T$ 的特征向量组成的矩阵, Λ 是由对应的特征值组成的对角矩阵, $U^T x$ 是图信号 x 的图傅里叶变换。

这里 g_θ 可视是为对应特征频率 Λ 的函数,即 $g_\theta(\Lambda)$.然而,此情况下模型需要学习的参数量很大,因为特征向量矩阵 U 与所有的节点都相关.此外,对于稠密图或者大规模图来说,对 L 矩阵进行特征值分解时的计算成本也相当高.为解决这些问题,有学者提出可以通过切比雪夫多项式 $T_k(x)$ 到 K^{th} 阶的截断展开来近似 $g_\theta(\Lambda)$ 的操作,如公式(2)所示:

$$g_\theta(\Lambda) \approx \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}) \quad (2)$$

其中, $\tilde{\Lambda} = \frac{2}{\lambda_{\max}} \Lambda - I_N$, λ_{\max} 表示拉普拉斯矩阵的最大特征值. $\theta' \in \mathbb{R}_K$ 是对应项的系数向量.这里,切比雪夫多项式由以下的递推关系生成 $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$,式中有 $T_0(x) = 1$, $T_1(x) = x$.

对图信号 x 应用切比雪夫滤波器,可以得到:

$$g_\theta \star x = \sum_{k=0}^K \theta'_k T_k(\tilde{L})x \quad (3)$$

其中, $\tilde{L} = \frac{2}{\lambda_{\max}} L - I_N$.

在 GCN^[25]中, 则进一步将最大特征值近似为 2, 即 $\lambda_{\max} \approx 2$. 在此情况下, 公式 (3) 简化为公式 (4):

$$g_{\theta} \star x \approx \theta_0 x + \theta_1 (L - I_N) x = \theta_0 x - \theta_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \quad (4)$$

其中, θ_0 与 θ_1 是可以自由优化的参数, 在实践中, Kipf 发现可以令 $\theta = \theta_0 = -\theta_1$, 从而对公式进一步简化, 如公式 (5) 所示:

$$g_{\theta} \star x \approx \theta (I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) x \quad (5)$$

再利用如公式 (6) 所示的重整归一化技巧, 其中 $\tilde{A} = A + I_N$, $\tilde{D}_i = \sum_j \tilde{A}_{ij}$, 可以得到:

$$I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \rightarrow \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \quad (6)$$

上文是单通道的图滤波, 实践中的图信号通常是多通道的. 对具有 C 维的多通道和 F 个频域算子的图信号 $X \in R^{N \times C}$ 来说, 上述操作可扩展为公式 (7):

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta \quad (7)$$

其中, $\Theta \in R^{C \times F}$ 是对应的多通道图滤波器的算子参数, $Z \in R^{N \times F}$ 是进行图滤波后的输出节点特征.

进一步, 可以定义图卷积神经网络的前向传播规则如公式 (8) 所示:

$$H^{(k+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(k)} W^{(k)}) \quad (8)$$

其中, $H^{(k+1)}$ 和 $H^{(k)}$ 分别为该层输入与输出的特征矩阵, 且有 $H^{(0)} = X$; $W^{(k)}$ 是每一层中可学习的参数矩阵, $\sigma(\cdot)$ 表示激活函数.

最后, 整个 GCN 网络的损失函数通过在所有标记样本上的交叉熵计算, 如公式 (9) 所示:

$$L = - \sum_{i=1}^l \sum_{j=1}^c y_{ij} \ln H_{ij}^{(k)} \quad (9)$$

GCN 的出现带来了近几年神经网络的火热, 后续涌现出大量的工作对其进行改进、适配与研究, 如 GAT^[26]、GraphSAGE^[27]和 FastGCN^[28]等. 此外, 也出现了许多神经网络的通用框架, 试图为众多神经网络算法寻求统一解释与实现方法, 其中影响力最大的是信息传递神经网络 (message passing neural networks, MPNNs)^[29]框架, 它将现有的神经网络算法抽象为一个包括信息传播与读出的两阶段学习过程, 许多神经网络工具框架, 如 PyG^[30]与 DGL^[31]都是基于 MPNN 实现.

总的来说, 图卷积神经网络 (GCN) 填补了基于谱域的方法与基于空间的方法之间的差距, 具有较好的表现与可解释性, 并且能够直接应用于具有图结构的半监督数据, 是一种使用最广泛的图分析方法.

1.2 基于图的多视角半监督分类

基于图的方法是多视图半监督分类的一类典型方法. 基于图的半监督分类自出现以来就展现出良好的性能表现, 它将有标签和无标签的数据视为一个共同图的顶点, 然后通过图结构中的边关系传播标签信息. 目前, 已经有许多有效的研究将其应用于多视角数据^[32], 其中具有代表性的有: 文献 [33] 提出了一种基于图的自适应融合不同视角的半监督学习, 学习共享类信息矩阵用于图像分类; 文献 [34] 中提出了利用稀疏权值学习最优的融合图用于标签传播; AMGL (autoweighted multiple graph learning)^[35]是一种基于谱聚类的无参数多视角学习方法, 其通过建立一个凸优化模型, 使得该模型可以自动学习每个数据视角的最优权重, 可以推广到半监督分类任务中并转变为一个直推式的半监督学习模型; DMCL (deep multi-view concept learning)^[36]是一种深度多视角概念学习方法, 其通过标签信息指导下的图嵌入来捕捉数据的语义结构, 使用 NMF 并结合分组稀疏性约束来平衡一致性和互补性信息, 从而为多视角数据寻求共同的高层表示; MLAN (multi-view learning with adaptive neighbours)^[37]是一种自适应邻居的多视角学习算法, 其通过构建具有自适应邻居的相似图来学习各个视角的权重, 并在模型训练中不断地对最优的图进行调整, 进而得到分类结果.

近年来, 随着图卷积神经网络在各个领域中的巨大成功, 越来越多的研究者利用现有图卷积神经网络的优越性去解决多视角半监督分类的问题, 例如: Multi-GCN^[38]是一种结合流形学习的多视角模型, 主要包括 3 个步骤: 首先使用子空间分析方法有效地合并同一图上的多个视角信息, 其次使用排序来识别图中信息量最大的子成分, 并对

图进行剪枝,最后应用图卷积神经网络来进行半监督场景下的节点分类;mGCN^[39]是一种多视角的图卷积神经网络,其通过使用特定视角表示与全局表示来分别捕捉各视角下的独立信息与全局的统一信息,并在对每个视角的特定表示进行建模时分别从视角内和共同特征中获取信息,最后使用全连接层将不同视角的表示合并为全局表示。

1.3 自注意力机制

最近,自注意力机制(self-attention mechanism)在捕获全局信息上的优秀表现,使其成为目前深度学习的主流方法与讨论热点,并在各大研究领域中大放异彩,被广泛应用于解决机器翻译、图像分类、目标检测与文本生成等实际任务^[40,41]。

自注意力机制有多种表现形式,我们选取其中应用最广的缩放点积自注意力(scaled dot-product attention),对其原理如下介绍。

对于一组输入变量(例如,在本文的任务中,多个视角下的中间层特征就被视为一组输入变量): $X = [x_1, \dots, x_n] \in \mathbb{R}^{D_k \times N}$,其中 $n \in [1, N]$ 。首先,我们通过利用3个不同参数的可学习权值矩阵 $W^Q \in \mathbb{R}^{D_k \times D_k}$, $W^K \in \mathbb{R}^{D_k \times D_k}$ 和 $W^V \in \mathbb{R}^{D_v \times D_k}$ 将其映射到3个不同的向量空间中,如公式(10)所示:

$$\begin{cases} Q = W^Q X \in \mathbb{R}^{D_k \times N} \\ K = W^K X \in \mathbb{R}^{D_k \times N} \\ V = W^V X \in \mathbb{R}^{D_v \times N} \end{cases} \quad (10)$$

这样就可以分别得到查询向量(query): $Q \in \mathbb{R}^{D_k \times N}$ 、键向量(key): $K \in \mathbb{R}^{D_k \times N}$ 与值向量(value): $V \in \mathbb{R}^{D_v \times N}$ 。其中, D_k 为 K 的维度, D_v 为 V 的维度。

进一步使用缩放点积(scaled dot-product)作为注意力权值打分函数,得到注意力输出向量 $H = [h_1, \dots, h_n]$,如公式(11)所示:

$$H = V \times \text{Softmax}\left(\frac{K^T Q}{\sqrt{D_k}}\right) \quad (11)$$

此外,为了增强模型的表达能力,在实际应用中多使用多头注意力机制(multi-head attention)。即通过利用多个查询 $Q = [q_1, \dots, q_M]$ 来并行地从输入中独立提取多种特征表示,促使每个注意力头(head)关注输入变量中的不同模式,如公式(12)所示,其中, \oplus 表示拼接操作:

$$\text{att}((K, V), Q) = \text{att}((K, V), q_1) \oplus \dots \oplus \text{att}((K, V), q_M) \quad (12)$$

需要注意,多头注意力的输出通常都需要经过另一个线性转换,它对应着 h 个头连结后的结果,如公式(13)所示,其中 $W_o \in \mathbb{R}^{d_o \times h \cdot d_o}$,上标中 \cdot 运算体现多头注意力的相关维度。

$$W_o \begin{bmatrix} h_1 \\ \vdots \\ h_h \end{bmatrix} \in \mathbb{R}^{d_o} \quad (13)$$

总的来说,自注意力机制可以为每个输入自适应地从所有输入中提取相关信息,从而得到更好的特征表达。尽管其最初是被Transformer应用于在文本数据上的序列到序列学习,但现在已经被推广到各种前沿的深度学习研究中,在自然语言、计算机视觉、语音处理、多模态和强化学习领域中占据着领先地位。

2 多视角信息交互的图卷积神经网络

本节将对提出的多视角信息交互的图卷积神经网络模型MIGCN(multi-view interaction graph convolutional network)进行详细介绍。我们将首先声明主要使用的符号,然后简述MIGCN的总体框架,最后再对模型的各个部分进行了详细介绍。

2.1 符号声明

考虑到本文所提出算法使用的变量较多,在介绍各个部分的细节之前,我们首先在表1中对主要使用的符号与变量进行必要的描述与声明。

表 1 符号声明

符号	声明	符号	声明
V	多视角数据集的视角数量	L	模型总单元层数
C	数据集的类别数量	$E_v^{(l)}$	第 l 层单元第 v 个视角GCN输出的特征矩阵
N	数据集的样本数量	$H_v^{(l)}$	第 l 层单元第 v 个视角输入的特征矩阵
Y_L	有标记信息训练样本的集合	$H_v^{(l+1)}$	第 l 层单元第 v 个视角输出的特征矩阵
Y	有标记信息训练样本的真实标签	Z_v	第 v 个视角下的最终节点特征
\hat{Y}	样本通过模型得到的分类输出	Z_{v-nor}	标准化后的 Z_v
k	KNN中选取最近邻居的数量	α_v	全局特征融合时第 v 个视角的注意力权重
X_v	第 v 个视角下所有样本的特征集合	Z	全局节点特征
G_v	第 v 个视角通过KNN建立的图结构	S_v	第 v 个视角的余弦相似度矩阵
A_v	第 v 个视角所建图结构的邻接矩阵	$\ \cdot\ _2^2$	矩阵的范数的平方
D	模型中间层的特征维度	M^T	矩阵 M 的转置操作

2.2 模型总览

MIGCN 的模型结构如图 2 所示, 其算法流程如下: 首先, 针对多视角数据的各个视角, 通过 KNN (K-nearest neighbor) 构建各视角下的相似图结构作为输入. 然后, 在模型的每一个单元中, 先通过在不同视角上训练的单层图卷积神经网络进行视角内的邻居信息传递, 学习特定视角的节点特征表示; 再针对同一节点的不同视角特征通过单层的 Transformer encoder layer 实现基于自注意力的不同视角间的互补信息交互, 得到各视角下新的节点特征表示. 同时, 通过基于余弦相似度矩阵的视角间一致性约束损失 \mathcal{L}_{com} , 来约束不同视角下最终特征表达的一致性. 之后, 通过注意力机制得到每个节点不同视角表示的注意力权重, 将各视角下最终的节点特征表示进一步融合为全局节点特征表示, 使节点将注意力集中在信息更丰富的视角上. 最后, 将全局节点表示通过全连接层输出得到各节点分类结果.

由图 1 可知, MIGCN 主要由 4 部分组成: (1) 基于图卷积神经网络的视角内邻居信息传递; (2) 基于自注意力机制的不同视角间的互补信息交互; (3) 基于注意力机制的全局特征融合; (4) 基于余弦相似度矩阵的视角间一致性约束. 下面, 我们将在各个小节内对上述 4 个部分的细节进行详细介绍.

2.3 基于图卷积神经网络的视角内邻居信息传递

对于多视角数据集的输入数据 (X_1, X_2, \dots, X_V) , 其中 V 为数据集的视角数量, X_v 是第 v 个视角下所有样本的特征集合 (即 $X_v = [\mathbf{x}_v^1, \mathbf{x}_v^2, \dots, \mathbf{x}_v^N]$, 其中 N 为数据集的样本数量), 首先, 我们利用 K-最近邻 (K-nearest neighbor, KNN) 的方式先获取所有节点的 k 个近邻节点, 并以此构建各视角下的相似图结构作为输入, 如公式 (14) 所示:

$$G_v = (A_v, X_v) \quad (14)$$

其中, $A_v \in \mathbb{R}^{N \times N}$ 是第 v 个视角所建图结构的邻接矩阵.

接下来, 在模型第 l 层单元 (block) 中, 我们首先基于图神经网络 (如图卷积神经网络 GCN^[25]、图注意力网络 GAT^[26]等) 的图卷积层, 进行各视角间相互独立的视角内邻居信息传递过程. 这里, 我们使用图卷积神经网络层 GCN layer, 如公式 (15) 所示:

$$E_v^{(l)} = \sigma \left(\tilde{D}_v^{-\frac{1}{2}} \tilde{A}_v \tilde{D}_v^{-\frac{1}{2}} H_v^{(l)} W_v^{(l)} \right) \quad (15)$$

其中, $H_v^{(l)} \in \mathbb{R}^{N \times D}$ 是第 l 层单元内第 v 个视角 GCN layer 的输入特征矩阵, 在整个模型输入端有 $H_v^{(1)} = X_v$; $E_v^{(l)}$ 是第 l 层单元内第 v 个视角 GCN layer 的输出特征矩阵; $\tilde{A}_v = A_v + I_N$ 是第 v 个视角带有自环的无向图的邻接矩阵, 这里 I_N 是单位矩阵; D_v 是对应的度矩阵, $\tilde{D}_v = \sum_j \tilde{A}_{v,j}$; $W_v^{(l)}$ 是第 l 层单元内第 v 个视角的 GCN 对应的可学习参数矩阵; $\sigma(\cdot)$ 是非线性的激活函数, 在这里, 我们选择使用 ReLU.

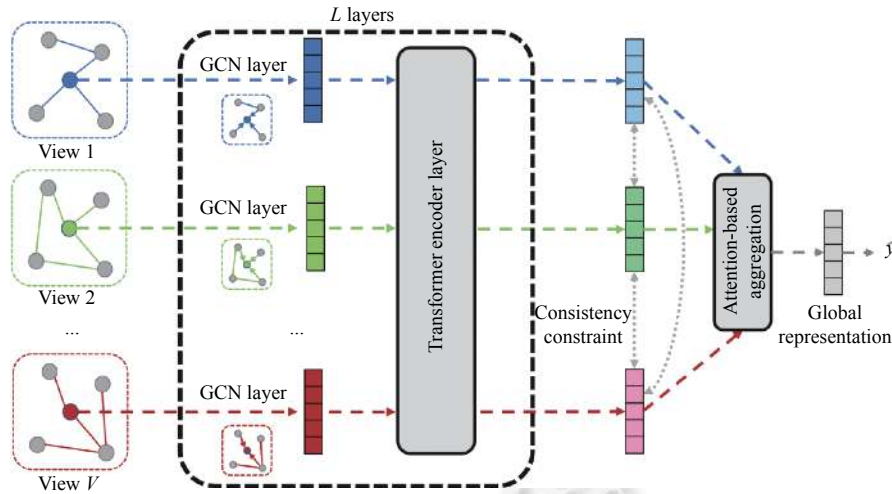


图2 MIGCN 模型结构图

这样,我们就通过各视角独立的单层 GCN layer,实现了第 l 层单元中的视角内邻居信息传递,完成了视角内的信息聚合.

2.4 基于自注意力机制的不同视角间的互补信息交互

在模型第 l 层单元中完成基于图卷积神经网络的视角内邻居信息传递之后,再进行基于自注意力的不同视角间的互补信息交互过程.

这部分,我们参考基于自注意力的经典模型 Transformer,利用其中的 Transformer encoder layer 实现相关操作,如公式 (16) 所示:

$$(H_1^{(l+1)}, H_2^{(l+1)}, \dots, H_V^{(l+1)}) = \text{TransformerEncoderLayer}(E_1^{(l)}, E_2^{(l)}, \dots, E_V^{(l)}) \quad (16)$$

其中, $(E_1^{(l)}, E_2^{(l)}, \dots, E_V^{(l)})$ 表示每个样本的多视角特征, Transformer encoder layer 的结构如图 3 所示,其主要包括多头自注意力 (multi-head attention)、残差连接 (add)、层归一化 (layer normalization, LN) 与前馈网络 (feed-forward networks, FFN).

简单来说,在 Transformer encoder layer 中,我们首先将 $(E_1^{(l)}, E_2^{(l)}, \dots, E_V^{(l)})$ 送入多头自注意力层 (MSA),使每个样本利用视角间的互补性,自适应地在视角间交互信息,更新视角内的节点表示为 $(E_1^{(l')}, E_2^{(l')}, \dots, E_V^{(l')})$,如公式 (17) 所示:

$$(E_1^{(l')}, E_2^{(l')}, \dots, E_V^{(l')}) = \text{MSA}(E_1^{(l)}, E_2^{(l)}, \dots, E_V^{(l)}) \quad (17)$$

接下来,对特征表示使用残差连接与层归一化,如公式 (18) 所示:

$$E_v^{(l')} = \text{LayerNorm}(E_v^{(l')} + E_v^{(l)}) \quad (18)$$

然后,通过前馈网络更新其特征表达.该网络依次应用于每个视角上.其由两个线性变换组成,两者间有一个非线性激活函数 ReLU,具体如公式 (19) 和公式 (20) 所示:

$$\text{FFN}(x) = \max(0, W_1 x + b_1) W_2 + b_2 \quad (19)$$

$$E_v^{(l'')} = \text{FFN}(E_v^{(l')}) \quad (20)$$

最后,再次作用残差连接与 LayerNorm 得到 Transformer encoder layer 输出:

$$H_v^{(l+1)} = \text{LayerNorm}(E_v^{(l')} + E_v^{(l'')}) \quad (21)$$

在这里,我们取最后一层单元中 Transformer encoder layer 的输出作为各视角下的最终的节点特征表示,如公式 (22) 所示:

$$Z_v = H_v^{(L+1)} \quad (22)$$

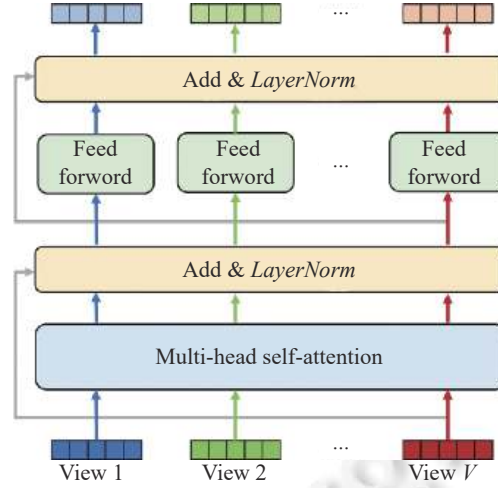


图3 模型使用的 Transformer encoder layer 结构图

2.5 基于注意力机制的全局特征融合

在得到各视角下的最终的节点特征表示后,我们通过基于注意力的融合方法来学习不同视角的特征表达的注意力权重,将多个视角的节点特征融合为全局节点特征,得到各视角特征的潜在共享信息;并将全局节点特征通过全连接层加 *Softmax* 输出,得到最终各个节点的分类结果.具体细节如下.

通过最后一个单元中 Transformer encoder layer 的输出,根据公式 (22) 可以得到所有节点各视角下的最终节点特征 (Z_1, Z_2, \dots, Z_V) ,为了将各视角下的最终节点特征融合为全局节点特征,我们使用注意力机制 $att(Z_1, Z_2, \dots, Z_V)$ 来学习它们对应的注意力权重 $(\alpha_1, \alpha_2, \dots, \alpha_V)$:

$$(\alpha_1, \alpha_2, \dots, \alpha_V) = att(Z_1, Z_2, \dots, Z_V) \quad (23)$$

具体来说,对于样本节点 i ,其对应第 v 个视角下的最终节点表示为 z_v^i ,首先利用一个使用 \tanh 激活函数的全连接层对该表示进行投影,然后使用一个可学习的统一 Query 向量 $q^T \in \mathbb{R}^{h \times 1}$ 来获得各视角的注意力评分 ω_v^i ,如公式 (24) 所示:

$$\omega_v^i = q^T \times \tanh(W \times (z_v^i)^T + b) \quad (24)$$

其中, $W \in \mathbb{R}^{h' \times h}$ 是可学习的权重参数矩阵, $b \in \mathbb{R}^{h' \times 1}$ 是对应的偏置向量.类似地,我们计算得到样本节点 i 在其他视角下注意力评分 $(\omega_1^i, \omega_2^i, \dots, \omega_V^i)$.接下来,通过使用 *Softmax* 函数对得到的注意力评分进行归一化,以获得最终的注意力权重 $(\alpha_1, \alpha_2, \dots, \alpha_V)$.具体计算如公式 (25) 所示:

$$\alpha_v^i = Softmax(\omega_v^i) = \frac{\exp(\omega_v^i)}{\sum_{j=1}^V \exp(\omega_j^i)} \quad (25)$$

通过这样的方法,我们可以得到所有样本节点各视角下的注意力权重:

$$\alpha_v = \text{diag}(\alpha_v^i) \quad (26)$$

然后,基于此注意力权重,将各视角下的最终节点特征融合为全局节点特征,如公式 (27) 所示:

$$Z = \sum_{v=1}^V \alpha_v \times Z_v \quad (27)$$

得到融合后的全局节点表示 Z 后,我们通过接在模型最后的全连接层与 *Softmax* 函数得到用于分类任务的输出,如公式 (28) 所示,其中 $*$ 表示卷积操作.

$$\hat{Y} = Softmax(W * Z + b) \quad (28)$$

2.6 基于余弦相似度矩阵的视角间一致性约束

考虑到不同视角下的最终特征表示都描述着同一个样本节点的某些特征属性,他们之间应该存在着一致性.因此,为了约束在各视角下的最终的节点特征表示空间的一致性,进一步增强模型的表达能力,我们提出了基于余弦相似度矩阵的视角间一致性约束损失 \mathcal{L}_{com} ,其具体描述如下.

针对第 v 个视角下的最终节点特征 Z_v ,我们首先通过 L2-normalization 得到标准化后的表示 $Z_{v-\text{nor}}$,然后,可以使用两个标准化节点表示矩阵来生成第 v 个视角的余弦相似度矩阵,如公式 (29) 所示:

$$S_v = Z_{v-\text{nor}} \times Z_{v-\text{nor}}^T \quad (29)$$

一致性意味着不同视角间的余弦相似度矩阵应该是相似的.因此,我们可以将各视角相似度矩阵两两间差值矩阵的范数的平方作为一致性约束损失 \mathcal{L}_{com} ,具体计算如公式 (30) 所示:

$$\mathcal{L}_{\text{com}} = \sum_{i,j \in V} \|S_i - S_j\|_2^2 \quad (30)$$

2.7 目标函数

整个模型目标函数由两部分组成:标记数据上的监督损失函数 \mathcal{L}_{sup} 与视角间的一致性约束损失函数 \mathcal{L}_{com} .

视角间的一致性约束损失 \mathcal{L}_{com} 已经在第 2.6 节中做过介绍,监督损失函数 \mathcal{L}_{sup} 使用在所有的有标记训练样本上的交叉熵损失,具体如公式 (31) 所示.其中, Y_{li} 表示有标记训练样本的真实标签, y_L 为所有含标记训练样本的集合, \hat{Y}_{li} 为对应样本通过公式 (28) 得到的分类输出, C 是该数据集的类别数量.

$$\mathcal{L}_{\text{sup}} = - \sum_{l \in y_L} \sum_{i=1}^C Y_{li} \ln \hat{Y}_{li} \quad (31)$$

结合节点分类任务的监督损失函数与一致性约束损失函数,总的损失函数如公式 (32) 所示,其中 λ 是调节监督损失与一致性约束损失间比例的权重参数.

$$L = \mathcal{L}_{\text{sup}} + \lambda \mathcal{L}_{\text{com}} \quad (32)$$

这样,我们就可以在标签信息与一致性约束的引导下,通过反向传播对模型进行优化,并利用模型对所有节点的进行分类.

3 实验分析

本节将展示本文所提出的 MIGCN 在多个真实多视角基准数据集上的性能表现,并与基于图的半监督多视角分类模型进行对比;同时,对模型的超参数与收敛性进行了实验分析;此外,我们还将通过消融实验,进一步验证模型中基于自注意力的不同视角间互补信息交互、基于注意力的特征融合与视角间一致性约束等模块的有效性.

3.1 数据集

为了证明模型的有效性,我们的实验在几个真实世界的多视角数据集上进行,这些数据集涵盖多种不同的应用,包括通用图像物体分类、手写数字分类和新闻文章分类等.在这些数据集上的实验可以充分检验本文提出算法在实际应用时的性能表现与泛化能力.这些数据集的简要信息如表 2 中所示.

表 2 实验数据集

数据集	样本数 (N)	视角数 (V)	类别数 (C)	各视角特征维度
HW	2000	6	10	240/76/216/47/64/6
Caltech7	1474	6	7	48/40/254/1984/512/928
Caltech20	2386	6	20	48/40/254/1984/512/928
Caltech101	9144	6	101	48/40/254/1984/512/928
3sources	169	3	6	3560/3631/3068
BBC	685	4	5	4659/4633/4665/4684
BBCSports	544	2	5	3283/3183

以上数据集的详细信息如下。

(1) HW (handwritten numerals): 是一个关于手写数字 (0-9) 的图像分类数据集, 来自 UCI 存储库: <http://archive.ics.uci.edu/ml/datasets/Multiple+Features>. 其中每一个数字包含 200 张图片, 总共有 2000 个样本, 每个样本有 6 个视角的图像特征可用, 分别是: 216 维轮廓描述 (profile-correlation, FAC) 特征, 76 维频域系数 (Fourier-coefficient, FOU) 特征, 64 维卡尔胡宁-洛夫系 (Karhunen-Loeve-coefficient, KAR) 特征, 240 维图像平均值 (intensity-averaged, PIX) 特征, 47 维 Zernike 矩 (Zernike moment, ZER) 特征和 6 维形态 (morphological, MOR) 特征。

(2) Caltech101: 是由加州理工学院整理收集的物体分类图像数据集, 来自 <http://www.vision.caltech.edu/archive.html>. 其中包括来自 101 个类别的 9144 个对象图像, 每幅图像有 6 个视角的特征可用, 分别是: 48 维滤波器 (Gabor) 特征, 254 维直方图 (CENTRIST) 特征, 40 维 wavelet 特征, 1984 维霍格变换 (HOG) 特征, 512 维空间包络 (GIST) 特征和 928 维平局灰度 (LBP) 特征. 参照以前的工作^[19], 我们还选取了该数据集的两个经典子集进行实验, 分别是包括 1474 张来自 7 个主要子类别数字图像的 Caltech7 与 2386 张来自 20 个主要子类别数字图像的 Caltech20.

(3) 3sources: 3sources 数据集来自 <http://mlg.ucd.ie/datasets/3sources.html>. 其中包括从 3 个著名的在线新闻来源 (BBC、路透社、卫报) 收集的 169 篇新闻文章, 分 6 个主题. 每篇新闻都在 3 个来源中进行了报道, 每个来源分别具有 3068 维、3560 维和 3631 维的文本特征。

(4) BBC: 是一个来自 <http://mlg.ucd.ie/datasets/segment.html> 的新闻文章分类数据集. 其中包括来自 BBC 新闻网站的 685 条新闻报道, 涉及 6 个热点话题, 分别是商业、娱乐、政治、体育和技术. 对于每篇新闻报道, 分别从不同视角提取了 4633、4659、4665 和 4684 维的语义特征。

(5) BBCSports: 是一个来自 <http://mlg.ucd.ie/datasets/segment.html> 体育新闻文章分类数据集. 具体来说, 它包含 544 篇体育新闻文章, 对应 5 个类别, 即田径、板球、足球、橄榄球和网球. 每个样本都具有 3283 维和 3183 维的两个视角特征。

3.2 实验设定

对于本文所提出的方法 MIGCN: 在使用 KNN 建立图结构时, 可以选择欧氏距离或余弦距离, 本实验中所有数据集都使用欧氏距离, 近邻数量从 $k \in [3, 5, 7, 9, 11]$ 中选取; 模型单元 (block) 总层数 L 设置为 2; 所有的图卷积网络层 (CCN layer), 其输出特征维度都被设置为 256, 并使用比率为 0.3 的 Dropout; Transformer encoder layer 中多头注意力的 head 数目被设置为 4, FFN 的中间层维度设置为 512, 并使用比率为 0.1 的 Dropout; 调节监督损失与一致性约束损失间比例的权重参数 λ 设置为 0.01. 在训练时, 使用 AdamW^[42] 作为优化器, 初始学习率设置为 0.001, 并采用 CosineAnnealing^[43] 学习率衰减策略。

在所有数据集上, 我们首先随机采样 10% 的数据作为验证集, 然后在剩下的数据中随机采样比例等于监督率 γ 的数据作为含标记信息的训练数据. 每次都选用不同的随机种子进行 10 次重复实验, 最后汇报这 10 次实验的平均分类精度作为评价指标. 所有未固定的超参数, 根据在第 1 次实验中的分类精度进行选取, 然后在后面的 9 次重复实验中都保持不变。

3.3 对比方法

为了评估本文提出模型的性能, 我们与近年来许多经典的基于图的半监督多视角分类方法 AMGL^[35]、MVAR^[44]、MLAN^[37]、Co-GCN^[19]、GCN fusion、ERL-MVSC^[45] 和 MGLNN^[46] 进行了比较. 每种方法的简要介绍如下。

AMGL (autoweighted multiple graph learning): 是一种无参数的基于谱聚类的多视角学习方法, 可以推广到半监督分类任务中转变为一个直推式半监督学习模型. 其通过建立一个凸优化模型, 使得该模型可以自动学习每个数据视角的最优权重。

MVAR (multi-view semi-supervised classification via adaptive regression): 是一种基于自适应回归的半监督多视角模型. 其对每个视角采用一个基于回归的损失函数, 将损失度量线性组合为最终的目标函数, 最后通过非光滑 (2, 1) 范数极小化直接预测类标签。

MLAN (multi-view learning with adaptive neighbours): 是一个利用图结构信息, 通过在模型训练中对不断对最优图进行学习的半监督多视角学习方法. 其通过构建了具有自适应邻域的相似图, 迭代学习多个视角的自动权重。

Co-GCN: 是一种基于协同训练、谱图理论与图卷积神经网络的半监督多视角学习方法. 其在各视角下分别建

立图卷积网络,通过交替优化分别更新各视角的拉普拉斯矩阵与图卷积层参数的方法,从而利用多视角的信息。

GCN fusion: 参考 Co-GCN 的实验设置,我们设计了同样的 GCN fusion. GCN fusion 是使用了 GCN 的集成学习方法,每一个视图的 GCN 网络仅使用了该视图自身的结构信息,然后对输出结果进行加权投票得到模型的分类结果。

ERL-MVSC: 是一种用于多视图半监督分类的嵌入正则化学习方案. 其通过在融合的嵌入正则化器上使用 L2 与 L1 范数,利用样本的稀疏局部结构传递有价值的分类信息,将半监督多视角分类问题归结为一个嵌入正则化器和预测标签矩阵的联合优化问题。

MGLNN: 是一种多图学习 (multiple graph learning) 神经网络算法,可用于多视角数据的半监督分类. MGLNN 的目标是从多个图结构中学习一个最优的图结构,并将多图学习和图神经网络相结合,以此最大限度地提升 GNN 在多视角数据上的性能表现。

3.4 实验结果与分析

我们在多个监督率 γ 下进行实验,并根据平均分类精度 Acc (即正确分类样本占有所有样本的比例) 来评估各个算法的性能. 在监督率分别为 $\gamma = 5\%$ 、 $\gamma = 10\%$ 与 $\gamma = 20\%$ 时,各数据集上所有比较方法的分类精度如表 3、表 4 与表 5 所示,其中最好的结果以加粗突出显示,而次优的结果以下划线突出显示,一表示该算法未进行相应实验。

表 3 各数据集所有算法分类精度表 (%) (监督率 $\gamma = 5\%$)

算法	HW	Caltech7	Caltech20	Caltech101	3sources	BBC	BBCSports
AMGL	85.29	55.47	39.36	19.02	31.17	46.11	46.97
MVAR	76.02	84.03	61.89	<u>42.96</u>	45.48	64.37	67.09
MLAN	94.97	89.77	77.64	24.56	36.44	68.52	74.43
GCN fusion	89.37	87.78	79.37	38.21	47.16	67.25	84.09
Co-GCN	91.40	88.58	79.71	40.25	<u>55.28</u>	<u>70.14</u>	<u>85.48</u>
MGLNN	97.36	93.73	85.69	—	—	—	—
MIGCN	<u>95.59</u>	<u>91.93</u>	<u>81.25</u>	44.62	66.90	74.91	87.26

表 4 各数据集所有算法分类精度表 (%) (监督率 $\gamma = 10\%$)

算法	HW	Caltech7	Caltech20	Caltech101	3sources	BBC	BBCSports
AMGL	88.57	59.19	45.05	24.42	39.87	53.72	55.65
MVAR	78.34	84.07	68.90	43.49	46.81	75.30	83.71
MLAN	95.72	89.95	77.64	31.81	45.50	74.06	76.84
GCN fusion	91.04	91.04	82.73	47.15	59.64	77.27	84.60
Co-GCN	94.93	92.05	83.63	46.23	60.13	<u>81.90</u>	84.80
ERL-MVSC	94.31	—	—	<u>46.26</u>	<u>69.34</u>	77.85	<u>88.52</u>
MGLNN	97.79	95.77	<u>85.87</u>	—	—	—	—
MIGCN	<u>97.12</u>	<u>92.54</u>	86.01	48.59	70.73	83.24	89.47

表 5 各数据集所有算法分类精度表 (%) (监督率 $\gamma = 20\%$)

算法	HW	Caltech7	Caltech20	Caltech101	3sources	BBC	BBCSports
AMGL	92.13	62.43	49.56	29.58	41.93	58.73	59.97
MVAR	55.64	89.86	70.34	47.74	66.29	85.67	89.54
MLAN	<u>96.52</u>	91.90	81.46	37.19	55.04	73.74	62.82
Co-GCN	91.60	<u>92.46</u>	<u>87.29</u>	52.67	75.96	86.51	<u>91.33</u>
ERL-MVSC	96.83	—	—	<u>56.04</u>	<u>78.72</u>	83.45	90.24
MIGCN	97.85	94.32	88.14	56.58	79.09	<u>85.94</u>	91.95

根据上述结果,我们可以做出以下分析。

(1) 首先,通过实验结果中 MIGCN 与现有最新的基于图的半监督多视角分类方法比较,本文提出的 MIGCN

模型在大多数数据集上要比其他主流方法的分类性能精度更高,这证明了我们方法的优越性与有效性.同时,在 3sources 数据集上,当监督率 $\gamma = 5\%$ 时,我们的算法获得的大幅领先其他算法的表现,这可能是由于该数据集的样本个数太少,使得通过 KNN 建立的图结构中的存在较多噪声,同时受到该数据集背后的一些潜在因素的影响,使得其他算法利用该图结构难以学习到好的分类结果,而 MIGCN 模型则可以通过样本级的自注意力,在多个视角的联合图上自适应的提取到有用的互补信息,从而达到更高的精度.

(2) 其次,我们可以观察到,在几乎所有的数据集上,利用图卷积神经网络的方法 MIGCN 与 Co-GCN、GCN fusion 和 MGLNN 都取得了更好的结果,这进一步证明了图卷积神经网络利用图结构解决半监督节点分类任务时的高效性能.同时,可以看到 AMGL 在除了 HW 所有数据集上都取得了较差的表现,即无参数方法相比可自学习并优化参数的方法具有一定差距.

(3) 在部分数据集上, MGLNN 取得了相比 MIGCN 更高的分类精度,因为相比 MIGCN 简单地使用 KNN 构建各视角下的相似图结构作为输入, MGLNN 则通过模型训练中的迭代学习从多个图中寻找最优的图结构从而提升 GNN 的性能表现,相对的,我们的 MIGCN 则在训练中保持输入的图结构不变.因此,关于 MIGCN 如何在训练中不断自适应地对现有图结构进行优化,是一个值得我们后续研究的方向.

(4) 从 AMGL、MVAR、MLAN、Co-GCN 和 GCN fusion 的简要介绍来看,这些模型都是通过引入某种方法自适应的学习每个视角提取特征对于当前任务的贡献度,进而通过加权组合的方式生成其最终的多视角融合特征.然而,上述多视角融合特征的鲁棒性除了依赖于每个视角提取特征的合理贡献度,很大程度上还依赖于每一视角提取特征的鲁棒性.为了有效地提升相互独立的每一视角提取特征的鲁棒性,我们在不同视角上训练的图卷积层之间引入 Transformer encoder 模块,使得同一样本在训练阶段都可以通过注意力机制合理的利用多个视角数据之间的互补性信息,加强自身的训练,进而提升了每一视角提取特征和多视角融合特征的鲁棒性,在表 3、表 4、表 5 中的大量试验结果证明了我们方法的有效性.为了进一步促使模型在分类阶段更加合理地在多视角特征之间的一致性和互补性信息间取得平衡,让注意力机制更加合理的生成每个视角特征提取特征的贡献度,保证多视角融合特征对全局节点表示的全面性,除了经常使用的在标记数据上的监督损失函数之外,我们还引入了基于余弦相似度矩阵的视角间一致性约束函数来约束同一样本不同视角下输出表征的特征一致性信息,在消融实验部分的大量实验依次验证了我们模块的有效性.

(5) 对几乎所有算法,随着监督率的提升,都可以取得更高的分类精度,即模型分类正确率与标签数据的比例呈正相关,这证明了标记信息的重要作用.因此,在实际应用中可以通过简单地增加标记数据的方法来提高算法的性能.

3.5 超参数分析实验

为了研究实验中的各类超参数对于模型精度的影响,我们在监督率 $\gamma = 10\%$ 下的 Caltech20 数据集上针对 MIGCN 使用 KNN 算法的近邻数量 k , Transformer encoder layer 中多头注意力的 head 数目和用于调节监督损失与一致性约束损失间比例的权重参数 λ 进行了大量的实验分析.具体来说,本实验中的近邻数量 k 从 [3, 5, 7, 9, 11] 中选取, head 的数目从 [2, 4, 8, 16] 中选取, λ 的变化范围为 0.005-1.

图 4 显示了在不同超参数设置下的 MIGCN 在所选数据集上的分类精度.从中我们可以看到, MIGCN 在某些参数范围内是稳定的,也就是说,当我们在某些范围内改变这些超参数的值时, MIGCN 的性能仅会发生轻微的变化.我们通过选取适合该数据集的最优超参,促使 MIGCN 取得更好的模型性能.

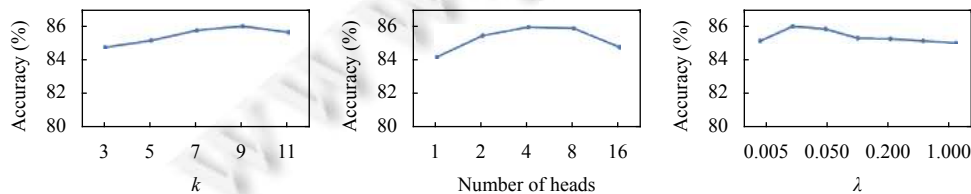


图 4 在 Caltech20 数据集上的超参数分析实验结果

3.6 收敛性分析

同时,为了验证 MIGCN 算法的收敛性,我们选取了监督率 $\gamma = 10\%$ 下的 3 个数据集进行了实验分析. 具体来说,图 5 给出了 $\gamma = 10\%$ 下 HW, Caltech7 与 Caltech20 数据集上 MIGCN 总的损失函数值与迭代 epoch 轮次的关系曲线,同时,在表 6 中展示了对应数据集上 MIGCN 的参数数量与其在实验设备 (NVIDIA RTX 1080Ti 12 GB) 上训练 500 个 epoch 的所用时长. 从表 6 中可以看出,越复杂的数据集意味着更长的训练时间与更大的收敛轮次,但总的来说, MIGCN 算法仅需较短的训练时长与较少的迭代 epoch 数.

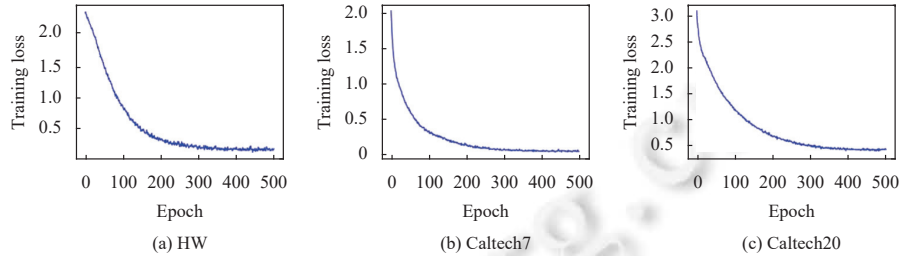


图 5 部分数据集上的损失函数下降曲线

表 6 部分数据集上 500 个 epoch 的训练时长

数据集	HW	Caltech7	Caltech20
参数量	451 242	849 831	851 508
训练时长 (s)	323.6	291.8	406.2

3.7 消融实验

为了进一步验证本文提出的基于自注意力的不同视角间的互补信息交互、视角间一致性约束与基于注意力的特征融合方法在模型中的作用,我们选择利用 HW、Caltech7 和 Caltech20 数据集,在监督率 $\gamma = 5\%$ 的情况下进行消融实验 (ablation experiment),实验结果如表 7 所示.

表 7 消融实验分类精度 (%)

算法/数据集	HW	Caltech7	Caltech20
MIGCN (w/o-TE)	92.71	88.45	69.56
MIGCN (w/o-Lc)	93.88	90.97	79.28
MIGCN-mean	93.65	91.09	79.48
MIGCN (Lc-emb)	93.37	90.42	78.60
MIGAT	94.97	91.89	81.07
MIGCN	95.59	91.93	81.25

表 7 中使用的各模型的介绍如下.

- (1) MIGCN: 即本文中提出的模型 MIGCN, 其中各部分与第 3 节中的介绍保持一致.
- (2) MIGCN (w/o TE): 保持其余结构不变, 在每层单元中仅进行基于图卷积神经网络的视角内邻居信息传递, 而不使用 Transformer encoder layer 进行基于自注意力的不同视角间的互补信息交互.
- (3) MIGCN (w/o Lc): 保持其余结构不变, 在训练中不使用视角间一致性约束损失 \mathcal{L}_{com} , 即总的损失函数只包括标记数据上的监督损失函数 \mathcal{L}_{sup} .
- (4) MIGCN-mean: 保持其余结构不变, 仅在全局特征融合时使用各视角特征的平均池化替代基于注意力机制的方法进行全局特征融合.
- (5) MIGCN (Lc-emb): 保持其余结构不变, 在训练中将视角间一致性约束损失 \mathcal{L}_{com} 修改为直接约束各视角下的最终的节点特征表示差异的 $\mathcal{L}_{com-emb}$, 如公式 (33) 所示:

$$\mathcal{L}_{\text{com-emb}} = \sum_{i,j \in V} \|Z_i - Z_j\|_2^2 \quad (33)$$

(6) MIGAT: 为了验证所提方法的通用性与扩展性, 保持其余结构不变, 我们仅将 MIGCN 中的图神经网络层由 GCN layer 替换为 GAT layer, 将修改后的模型命名为 MIGAT.

通过实验结果可以得出以下结论.

(1) 通过 Transformer encoder layer 进行基于自注意力的不同视角间的互补信息交互, 对模型充分利用多个视角数据之间的互补性信息进行算法优化, 从而对提取多视角样本中更加全面的特征具有重要作用.

(2) 通过增加基于余弦相似度矩阵的视角间一致性约束损失 \mathcal{L}_{com} , 约束在各视角下的最终的节点表示空间的一致性, 可以进一步增强模型的表达能力.

(3) 通过基于注意力的方法来学习不同视角最终表示的注意力权重, 进而进行多视角的节点表示融合, 相比于直接的平均池化, 可以得到更好的全局节点表示.

(4) MIGCN (Lc-emb) 的分类精度低于 MIGCN, 说明相比使用 L2 Loss 来直接强硬的约束不同视角下输出表征一致的方法来说, MIGCN 这种通过对不同视角下的相似关系进行约束, 来促使节点特征表示空间层面的一致性约束, 可以促使模型更加合理地在一致性和互补性间取得平衡.

(5) MIGAT 与 MIGCN 取得了几乎相同的分类精度, 证明了所提方法良好的扩展性能, 即更换所用图神经网络层的具体结构, 引入 Transformer encoder 模块的操作同样有效.

综上所述, 本文提出的基于自注意力的不同视角间的互补信息交互、视角间一致性约束与基于注意力的特征融合方法均在模型中起到了作用, 可以带来分类精度的提升.

3.8 t-SNE 可视化

最后, 为了验证 MIGCN 可以利用多视角间的互补性并进行可视化分析, 我们进行了如下实验.

在监督率 $\gamma = 5\%$ 的 Caltech7 数据集上, 我们在每个视角上训练一个神经元数目设置与 MIGCN 中相同的单视角 GCN, 其分类精度如表 8 所示.

表 8 各视角下独立训练 GCN 的分类精度 (%)

视角序号	1	2	3	4	5	6	MIGCN
分类精度	85.99	75.98	83.63	86.71	83.56	80.20	91.93

同时, 为了更直观地进行分析, 我们利用 t-SNE^[47]对每个视角下 GCN 的输出层前的特征嵌入 (embedding) 进行了可视化, 结果如图 6 所示.

可见, MIGCN 可以通过更好地利用多视角间的互补性信息, 从而提取到比单一视角情况下更有区分度的特征.

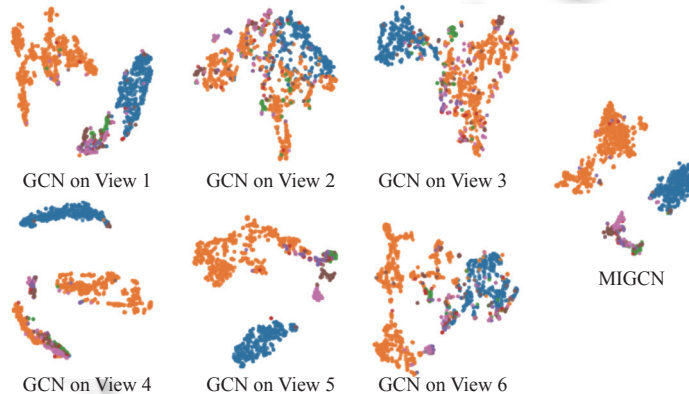


图 6 各视角下 GCN 特征嵌入可视化

4 总结与讨论

面对世界真实数据特征表示多样化且获取标记信息难度较大的需求,本文对半监督多视角分类问题进行了研究,提出了一种更加合理的利用图卷积神经网络与自注意力机制的半监督多视角分类算法 MIGCN. 该方法通过视角间独立的 GCN Layer 与不同视角互补信息交互的 Transformer encoder layer 模块,并结合视角间的一致性约束与基于注意力的全局特征融合,可以为每个样本自适应的学习到更加全面的全局特征,从而提高多视角数据的分类准确性. 在多个真实的多视角基准数据集上良好性能表现,证明了 MIGCN 相对于现有的基于图的半监督多视角分类方法的优越性与有效性. 此外,针对各部分结构所设计的消融实验,进一步验证了它们在模型中的关键作用.

在解决现实应用问题时,本文提出的 MIGCN 能够灵活地适应于不同的多视角数据场景,表现出较强的模型通用性和扩展性. 具体来说:(1)在第 3.4 节中给出的在多种类型数据集上的大量实验结果表明,本文提出的 MIGCN 并不依赖于特定的数据类型或视角数量,只需要根据数据的特点构建其合适的图结构,就可以实现图卷积神经网络中多视角信息的交互,进而加强多视角融合特征的鲁棒性;(2)在第 3.7 节中给出的图神经网络层替换实验结果(MIGCN 到 MIGAT)表明, MIGCN 可以与不同类型的图神经网络模型结合,并仍然可以有效地提升 MIGCN 模型在不同多视角数据场景下的任务性能.

最后,基于本文的研究内容,在未来的仍有许多值得进一步探索的方向,例如:(1)本文所提出的 MIGCN 是一种直推式的半监督学习,如何利用归纳式(inductive)的图神经网络算法,如 FastGCN 与 GraphSAGE 等,通过某种方法把新的节点特征加入动态的多视角图结构学习过程中,并直接应用于非直推的半监督多视角分类,进一步提升所提出模型在不同多视角数据场景下的任务性能,仍是个需要进一步研究的问题.(2)MIGCN 在整个多视角信息交互的过程始终保持着各视角初始的图结构,所以,能否借鉴 Co-GCN 等方法^[19]的思想,在多视角融合的过程中不断自适应地对各个图结构进行调整与优化,也是一个值得继续研究的方向.(3)本文所提出的 MIGCN 通过 Transformer encoder layer 进行基于自注意力的视角间协同信息交互,但最近的一些研究表明^[48],似乎有着更加高效的交互方式. 因此,如何将最新的自适应信息交互方法应用于本文提出的模型,也需要进一步的深入研究.

References:

- [1] Yan XQ, Hu SZ, Mao YQ, Ye YD, Yu H. Deep multi-view learning methods: A review. *Neurocomputing*, 2021, 448: 106–129. [doi: [10.1016/j.neucom.2021.03.090](https://doi.org/10.1016/j.neucom.2021.03.090)]
- [2] Sun SL. A survey of multi-view machine learning. *Neural Computing and Applications*, 2013, 23(7–8): 2031–2038. [doi: [10.1007/s00521-013-1362-6](https://doi.org/10.1007/s00521-013-1362-6)]
- [3] Yang XL, Song ZX, King I, Xu ZL. A survey on deep semi-supervised learning. *IEEE Trans. on Knowledge and Data Engineering*, 2023, 35(9): 8934–8954. [doi: [10.1109/TKDE.2022.3220219](https://doi.org/10.1109/TKDE.2022.3220219)]
- [4] Jiang LK. Multi-view semi-supervised classification overview. In: *Proc. of the 2nd Int'l Conf. on Artificial Intelligence and Information Systems*. Chongqing: ACM, 2021. 1–7. [doi: [10.1145/3469213.3470387](https://doi.org/10.1145/3469213.3470387)]
- [5] Liu WF, Ma XQ, Zhou YC, Tao DP, Cheng J. p -Laplacian regularization for scene recognition. *IEEE Trans. on Cybernetics*, 2018, 49(8): 2927–2940. [doi: [10.1109/TCYB.2018.2833843](https://doi.org/10.1109/TCYB.2018.2833843)]
- [6] Xie Y, Zhang WS, Qu YY, Dai LQ, Tao DC. Hyper-Laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning. *IEEE Trans. on Cybernetics*, 2020, 50(2): 572–586. [doi: [10.1109/TCYB.2018.2869789](https://doi.org/10.1109/TCYB.2018.2869789)]
- [7] Zhang CQ, Fu HZ, Wang J, Li W, Cao XC, Hu QH. Tensorized multi-view subspace representation learning. *Int'l Journal of Computer Vision*, 2020, 128(8–9): 2344–2361. [doi: [10.1007/s11263-020-01307-0](https://doi.org/10.1007/s11263-020-01307-0)]
- [8] Zhang CQ, Hu QH, Fu HZ, Zhu PF, Cao XC. Latent multi-view subspace clustering. In: *Proc. of the 2017 IEEE Conf. on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 4333–4341. [doi: [10.1109/CVPR.2017.461](https://doi.org/10.1109/CVPR.2017.461)]
- [9] Li LK, Kang Z, Long B. Riemannian manifold based multi-view spectral clustering algorithm. *Computer Engineering*, 2023, 49(1): 113–120, 129 (in Chinese with English abstract). [doi: [10.19678/j.issn.1000-3428.0062723](https://doi.org/10.19678/j.issn.1000-3428.0062723)]
- [10] Liu J, Jiang Y, Li ZC, Zhou ZH, Lu HQ. Partially shared latent factor learning with multiview data. *IEEE Trans. on Neural Networks and Learning Systems*, 2015, 26(6): 1233–1246. [doi: [10.1109/TNNLS.2014.2335234](https://doi.org/10.1109/TNNLS.2014.2335234)]
- [11] Chao GQ, Sun SL. Multi-kernel maximum entropy discrimination for multi-view learning. *Intelligent Data Analysis*, 2016, 20(3): 481–493. [doi: [10.3233/IDA-160816](https://doi.org/10.3233/IDA-160816)]
- [12] Lanckriet GRG, Cristianini N, Bartlett P, Jordan MI. Learning the kernel matrix with semidefinite programming. *The Journal of Machine*

- Learning Research, 2004, 5: 27–72.
- [13] Wang Z, Chen SC. Multi-view kernel machine on single-view data. *Neurocomputing*, 2009, 72(10–12): 2444–2449. [doi: [10.1016/j.neucom.2008.11.017](https://doi.org/10.1016/j.neucom.2008.11.017)]
- [14] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Proc. of the 11th Annual Conf. on Computational Learning Theory. Madison: ACM, 1998. 92–100. [doi: [10.1145/279943.279962](https://doi.org/10.1145/279943.279962)]
- [15] Nigam K, Ghani R. Analyzing the effectiveness and applicability of co-training. In: Proc. of the 9th Int'l Conf. on Information and Knowledge Management. McLean: ACM, 2000. 86–93. [doi: [10.1145/354756.354805](https://doi.org/10.1145/354756.354805)]
- [16] Niu XS, Han H, Shan SG, Chen XL. Multi-label co-regularization for semi-supervised facial action unit recognition. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2019. 82.
- [17] Yu SP, Krishnapuram B, Rosales R, *et al.* Bayesian co-training. In: Proc. of the 21st Annual Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2007. 1665–1672.
- [18] Wang W, Zhou ZH. A new analysis of Co-training. In: Proc. of the 27th Int'l Conf. on Int'l Conf. on Machine Learning. Haifa: Omnipress, 2010. 1135–1142.
- [19] Li S, Li WT, Wang W. Co-GCN for multi-view semi-supervised learning. In: Proc. of the 37th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2020. 4691–4698. [doi: [10.1609/aaai.v34i04.5901](https://doi.org/10.1609/aaai.v34i04.5901)]
- [20] Cheng JF, Wang QQ, Tao ZQ, Xie DY, Gao QX. Multi-view attribute graph convolution networks for clustering. In: Proc. of the 29th Int'l Joint Conf. on Artificial Intelligence. Yokohama: IJCAI.org, 2021. 411.
- [21] Xie Y, Zhang YQ, Gong MG, Tang ZD, Han C. MGAT: Multi-view graph attention networks. *Neural Networks*, 2020, 132: 180–189. [doi: [10.1016/j.neunet.2020.08.021](https://doi.org/10.1016/j.neunet.2020.08.021)]
- [22] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. *IEEE Trans. on Neural Networks*, 2009, 20(1): 61–80. [doi: [10.1109/TNN.2008.2005605](https://doi.org/10.1109/TNN.2008.2005605)]
- [23] Wu ZH, Pan SR, Chen FW, Long GD, Zhang CQ, Yu PS. A comprehensive survey on graph neural networks. *IEEE Trans. on Neural Networks and Learning Systems*, 2021, 32(1): 4–24. [doi: [10.1109/TNNLS.2020.2978386](https://doi.org/10.1109/TNNLS.2020.2978386)]
- [24] Chen YS, Zhao X, Jia XP. Spectral-spatial classification of hyperspectral data based on deep belief network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2381–2392. [doi: [10.1109/JSTARS.2015.2388577](https://doi.org/10.1109/JSTARS.2015.2388577)]
- [25] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc. of the 5th Int'l Conf. on Learning Representations. Toulon: OpenReview.net, 2017.
- [26] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018.
- [27] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.
- [28] Chen J, Ma TF, Xiao C. FastGCN: Fast learning with graph convolutional networks via importance sampling. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018.
- [29] Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: JMLR.org, 2017. 1263–1272.
- [30] Fey M, Lenssen JE. Fast graph representation learning with PyTorch geometric. arXiv:1903.02428, 2019.
- [31] Zheng D, Wang MJ, Gan Q, Song X, Zhang Z, Karypis G. Scalable graph neural networks with deep graph library. In: Proc. of the 14th ACM Int'l Conf. on Web Search and Data Mining. ACM, 2021. 1141–1142. [doi: [10.1145/3437963.3441663](https://doi.org/10.1145/3437963.3441663)]
- [32] Zhao K, Liu L, Han M. Semi-supervised classification based on transformed learning. *Journal of Computer Research and Development*, 2023, 60(1): 103–111 (in Chinese with English abstract). [doi: [10.7544/jssn1000-1239.202110811](https://doi.org/10.7544/jssn1000-1239.202110811)]
- [33] Cai X, Nie FP, Cai WD, *et al.* Heterogeneous image features integration via multi-modal semi-supervised learning model. In: Proc. of the 2013 IEEE Int'l Conf. on Computer Vision. Sydney: IEEE, 2013. 1737–1744. [doi: [10.1109/ICCV.2013.218](https://doi.org/10.1109/ICCV.2013.218)]
- [34] Karasuyama M, Mamitsuka H. Multiple graph label propagation by sparse integration. *IEEE Trans. on Neural Networks and Learning Systems*, 2013, 24(12): 1999–2012. [doi: [10.1109/TNNLS.2013.2271327](https://doi.org/10.1109/TNNLS.2013.2271327)]
- [35] Nie FP, Li J, Li XL. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. New York: AAAI Press, 2016. 1881–1887.
- [36] Xu C, Guan ZY, Zhao W, Niu YF, Wang Q, Wang ZH. Deep multi-view concept learning. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: IJCAI.org, 2018. 2898–2904.
- [37] Nie FP, Cai GH, Li J, Li XL. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Trans. on Image Processing*, 2018, 27(3): 1501–1511. [doi: [10.1109/TIP.2017.2754939](https://doi.org/10.1109/TIP.2017.2754939)]
- [38] Khan MR, Blumenstock JE. Multi-GCN: Graph convolutional networks for multi-view networks, with applications to global poverty. In:

- Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI Press, 2019. 75. [doi: [10.1609/aaai.v33i01.3301606](https://doi.org/10.1609/aaai.v33i01.3301606)]
- [39] Ma Y, Wang SH, Aggarwal CC, Yin D, Tang J. Multi-dimensional graph convolutional networks. In: Proc. of the 2019 SIAM Int'l Conf. on Data Mining. Calgary: SIAM, 2019. 657–665.
- [40] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai XH, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. In: Proc. of the 9th Int'l Conf. on Learning Representations. Vienna: OpenReview.net, 2021.
- [41] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- [42] Loshchilov I, Hutter F. Decoupled weight decay regularization. In: Proc. of the 7th Int'l Conf. on Learning Representations. New Orleans: OpenReview.net, 2019.
- [43] Loshchilov I, Hutter F. SGDR: Stochastic gradient descent with warm restarts. In: Proc. of the 5th Int'l Conf. on Learning Representations. Toulon: OpenReview.net, 2017.
- [44] Tao H, Hou CP, Nie FP, Zhu JB, Yi DY. Scalable multi-view semi-supervised classification via adaptive regression. IEEE Trans. on Image Processing, 2017, 26(9): 4283–4296. [doi: [10.1109/TIP.2017.2717191](https://doi.org/10.1109/TIP.2017.2717191)]
- [45] Huang AP, Wang Z, Zheng YN, Zhao TS, Lin CW. Embedding regularizer learning for multi-view semi-supervised classification. IEEE Trans. on Image Processing, 2021, 30: 6997–7011. [doi: [10.1109/TIP.2021.3101917](https://doi.org/10.1109/TIP.2021.3101917)]
- [46] Jiang B, Chen S, Wang BB, Luo B. MGLNN: Semi-supervised learning via multiple graph cooperative learning neural networks. Neural Networks, 2022, 153: 204–214. [doi: [10.1016/j.neunet.2022.05.024](https://doi.org/10.1016/j.neunet.2022.05.024)]
- [47] van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research, 2008, 9(86): 2579–2605.
- [48] Tang CX, Zhao YC, Wang GT, Luo C, Xie WX, Zeng WJ. Sparse MLP for image recognition: Is self-attention really necessary? In: Proc. of the 36th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI Press, 2022. 2344–2351. [doi: [10.1609/aaai.v36i2.20133](https://doi.org/10.1609/aaai.v36i2.20133)]

附中文参考文献:

- [9] 李林珂, 康昭, 龙波. 基于黎曼流形的多视角谱聚类算法. 计算机工程, 2023, 49(1): 113–120, 129. [doi: [10.19678/j.issn.1000-3428.0062723](https://doi.org/10.19678/j.issn.1000-3428.0062723)]
- [32] 康昭, 刘亮, 韩蒙. 基于转换学习的半监督分类. 计算机研究与发展, 2023, 60(1): 103–111. [doi: [10.7544/issn1000-1239.202110811](https://doi.org/10.7544/issn1000-1239.202110811)]



王悦天(2000—), 男, 硕士生, 主要研究领域为深度学习, 图表示学习, 计算机视觉.



邹斌(1969—), 男, 博士, 教授, 博士生导师, 主要研究领域为统计学习理论, 机器学习.



傅司超(1995—), 男, 博士生, 主要研究领域为流形学习, 图表示学习.



荆晓远(1971—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为机器学习, 人工智能, 软件工程.



彭勤牧(1985—), 男, 博士, 副教授, 主要研究领域为视觉计算, 机器学习, 医学图像分析.



尤新革(1969—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为模式识别, 图像与信号处理, 计算机视觉, 生物特征识别与智能防伪.