

# 基于图卷积神经网络的自注意力的融合节点分类框架<sup>①</sup>



姜发健<sup>1</sup>, 王金凤<sup>1,2</sup>, 招奕钧<sup>1</sup>, 郑志燊<sup>1</sup>

<sup>1</sup>(华南农业大学 数学与信息学院, 广州 510642)

<sup>2</sup>(广州市智慧农业重点实验室, 广州 510642)

通信作者: 王金凤, E-mail: [wangjinfeng@scau.edu.cn](mailto:wangjinfeng@scau.edu.cn)

**摘要:** 图神经网络因其强大的建模能力引起广泛关注, 常常被用来解决图上的节点分类任务. 现阶段常用的以图卷积神经网络 (graph convolutional network, GCN) 为内核的模型解决此类问题, 但往往因为出现过拟合与过平滑而导致深层的节点嵌入表示效果并不好. 因此, 本文提出了一种基于 GCN 内核的结合残差连接与自注意力方法——GCN RN 模型, 以提升 GCN 的泛化能力. 同时, 为了整合更深入的信息, 本文引入融合机制, 采用模糊积分融合多个分类器, 最终提高模型测试精度. 为了验证所提出方法的优越性, 本文采用 ogbn-arxiv 与常用的引文数据集进行了对比实验. GCN RN 模型与多个以 GCN 为内核的现有模型相比, 节点分类准确率平均提高了 2%, 且避免了传统的过拟合和过平滑现象. 此外, 实验结果表明, 增加了基于模糊积分的融合模块的多分类器模型比传统融合方法具有更好的分类效果.

**关键词:** 图卷积神经网络; 残差连接; 自注意力; 模糊积分; 模糊融合

引用格式: 姜发健, 王金凤, 招奕钧, 郑志燊. 基于图卷积神经网络的自注意力的融合节点分类框架. 计算机系统应用, 2023, 32(7): 251-260. <http://www.c-s-a.org.cn/1003-3254/9159.html>

## Fusion Node Classification Framework Based on Self-attention in Graph Convolutional Networks

JIANG Fa-Jian<sup>1</sup>, WANG Jin-Feng<sup>1,2</sup>, ZHAO Yi-Jun<sup>1</sup>, ZHENG Zhi-Shen<sup>1</sup>

<sup>1</sup>(College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China)

<sup>2</sup>(Guangzhou Key Laboratory of Intelligent Agriculture, Guangzhou 510642, China)

**Abstract:** Graph neural networks (GNNs) have attracted widespread attention due to their powerful modeling capabilities, and they are often used to solve node classification tasks on graphs. At this stage, the commonly used model with the graph convolutional network (GCN) as the core solves such problems. However, due to over-fitting and over-smoothing, the deep node embedding representation effect is not positive. Therefore, this study proposes a graph convolutional neural residual networks (GCN RN) model that combines residual connection and self-attention based on GCN kernel to improve the generalization ability of GCN. At the same time, in order to integrate more in-depth information, this study introduces a fusion mechanism, uses fuzzy integral to fuse multiple classifiers, and finally improves the model testing accuracy. In order to verify the superiority of the proposed method, this study uses ogbn-arxiv and commonly used citation datasets to conduct comparative experiments. Compared with many existing models with GCN as the core, the GCN RN model has an average improvement of node classification accuracy by 2% and avoids the traditional over-fitting and over-smoothing phenomena. In addition, the experimental results show that the multi-classifier model with the fusion module based on fuzzy integral has a better classification effect than the traditional fusion method.

**Key words:** graph convolutional networks (GCN); residual connection; self-attention; fuzzy integral; fuzzy fusion

<sup>①</sup> 基金项目: 智能农业广州重点实验室 (201902010081); 广东省科技规划项目 (2017A040406023); 广州市科技规划项目 (201804010353)

收稿时间: 2023-01-02; 修改时间: 2023-01-19; 采用时间: 2023-02-09; csa 在线出版时间: 2023-05-24

CNKI 网络首发时间: 2023-05-25

## 1 引言

近年来,将深度学习应用于图结构数据受到广泛关注,并广泛应用于自然语言处理、计算机视觉和推荐系统.图神经网络可以看作是节点之间的连接和消息传递之间的一个过程.自从引入图卷积神经网络(graph convolutional networks, GCN)<sup>[1]</sup>以来,许多高性能的设计已经被提出来解决节点分类问题.例如,图注意力网络(graph attention networks, GAT)<sup>[2]</sup>, GraphSAGE<sup>[3]</sup>和 JKNet<sup>[4]</sup>等.随着越来越多的网络结构的提出,模型的精度也稳步提高.

图卷积神经网络的目标是将节点特征转换为能够感知图的结构特征,可以通过聚合多层邻居信息来增加感知域,并提高对图结构的认识.但是,随着GCN层数的加深,将会导致梯度消失、过平滑、过拟合等一系列问题.随着图卷积神经网络的发展,有许多方法可以减少过平滑.例如GCNII<sup>[5]</sup>通过初始残差连接和恒等式映射,缓解了过平滑问题. DeeperGCN<sup>[6]</sup>提出了两个广义的均值-最大聚合函数,分别是 *Softmax* 聚合函数与 *PowerMean* 聚合函数. JKNet 需要保存每一层的中间层信息,最后进行层聚合.虽然已有部分文章提出缓解图卷积神经网络过平滑问题,但是它们提出的方法并不是都适用于图卷积神经网络节点分类任务,如果随意使用常见的缓解过平滑的技巧,往往会使得模型效果适得其反.

然而,单个分类器的性能是有局限性的,通过融合多个模型的技术可以进一步提升分类的精度.这些信息融合的概念产生于20世纪70年代末,最开始起源于军事领域,但随着信息融合技术本身、计算机和网络通信以及控制技术的发展,信息融合已成为信息处理的一种通用工具和常用的思维模式.

信息融合技术应用于分类问题也称为多分类融合技术,旨在利用各分类器检测的信息和不同的处理方法以获得对分类对象的全面检测信息,从而提高分类精度和可靠性.目前,常见的融合方法有投票法<sup>[7]</sup>、最大值法<sup>[8]</sup>和权重平均法<sup>[9]</sup>.投票法是一种遵循多数票原则的集成学习模型.最大值法是一种选择样本并选择分类效果概率最高的方法.加权平均法根据不同的模型对最终结果的贡献不同,这可以代表不同模型的重要性.虽然这些上述传统的融合方法是常用且高效的,但是他们并不能具有对分类器的准确性、描述的质量和可区分性具有可解释性.

然而,采用模糊积分融合技术替代传统融合技术进行多分类器模型融合.与传统的融合技术相比,恰恰能够弥补它们的不足.在计算模糊测度值中,通过计算模糊测度公式描述分类器的准确性、质量以及可区分性.这使得使用模糊积分融合技术进行多分类器融合相比于传统融合方法具有更优的可解释性与效果.

综合以上分析,本文主要围绕图卷积神经网络用于节点分类性能不佳的问题,为了减轻GCN的深层影响和提高泛化性,提出了一种结合层注意力的GCNRN模型,利用残差连接结构,使得节点特征获得更好的表达,从而有效提升其泛化能力.最后,为基于模糊积分的多分类器融合机制构建融合分类框架,以整合更深入的节点信息,提高分类精度.

本文的主要工作如下.

(1) 提出了一种基于图卷积神经网络的结合残差网络与层注意力的分类算法,从而更好地表达节点的表征.

(2) 基于模糊积分构建融合分类框架,以获得更深层次的信息,从而提升整体分类效果.

(3) 在4个公共基准数据集上进行实验,验证了本文的分类算法的分类指标高于同类算法.

本文架构如下:第2节将简要介绍相关工作;第3节提出本文的创新方法,分别从核心分类模型、融合算法和整体框架展开介绍;第4节分别在多个数据集上展开对比验证,并分析和讨论了本文算法的特点;最后一节给出简要总结和未来研究方向.

## 2 相关工作

### 2.1 基于图神经网络的分类算法

图神经网络GNN自2009年模型被提出<sup>[10]</sup>,直到2017年GCN模型的出现使得图神经网络得到快速发展. Bruna 等人<sup>[11]</sup>首先利用谱域内的图的傅里叶变换和卷积神经网络相结合,产生频谱图卷积神经网络. Kipf 等人<sup>[1]</sup>提出了使用切比雪夫多项式简化图卷积.图卷积神经网络中的消息聚合和消息更新是消息传递机制中的两个重要过程.首先聚合每个邻居节点的信息,并将其传输到目标节点,然后通过消息更新将目标节点的信息和邻居节点传递来的信息进行更新,最终得到目标节点的表示.图卷积神经网络层是由频率响应函数的线性近似推导出来的,但是和深度学习中深度网络给模型带来的强大拟合能力一样,通过堆叠多

层图卷积神经网络,在某种程度上,可以达到高阶多项式形式的频率相应函数的滤波能力。

图卷积神经网络架构被广泛使用,它依赖于层级的传播规则。我们定义 $G = \{V, E, A\}$ 表示无向图,其中, $V$ 表示节点集合, $|V| = n$ 表示图上共有 $n$ 个节点, $E$ 表示边集合, $A$ 表示邻接矩阵,定义节点之间的相互连接,且在无向图中 $A_{i,j} = A_{j,i}$ , $L = D - A$ 表示图上的拉普拉斯矩阵,其中, $D$ 表示一个对角阵, $D_{i,i}$ 表示第 $i$ 个节点的度, $X^l$ 表示第 $l$ 层的表征当前节点的表征。

$$X^{(l+1)} = \sigma \left( D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X^{(l)} W^{(l)} \right) \quad (1)$$

目前以GCN内核为基础的模型中,例如GraphSAGE、DeeperGCN、JKNet和GCNII模型,在面对多层图卷积过平滑问题,大多对初始残差连接块、聚合函数和中间层进行填充聚合器进行改进以适应深层模型。

但是上述基于GCN内核的分类算法仍无法解决深层过平滑问题和融合模型的问题。因此,本文首先结合注意力机制与残差网络结构,再利用模糊积分融合多分类器的方式提取更深层次的信息,最终提升模型性能。

## 2.2 分类器融合机制

信息融合技术应用于分类问题称为多分类融合技术,从而提高分类精度和可靠性,对于分类问题,主要分为3个层次上的融合,数据融合、特征级融合和决策级融合。常见的融合工具有投票法、最大值法、权重平均法等。

传统的融合技术是常用且高效的,但仍各自存在缺陷。最大值法只是对样本使用分类结果概率最高的方法,但该方法不能忽略分类器对样本描述的质量和区别。投票模型融合方法采用多数票规则进行投票,但该方法依赖于模型的区分度。如果相似的模型融合,投票融合效果会更差。权重平均模型融合方法根据准确率为不同的模型分配不同的权重,但该方法考虑的分类器指标太少,无法影响最终的融合效果。

为了避免上述传统方法存在的问题,能够充分考虑多个分类器的联合贡献率,本文采用模糊积分<sup>[12]</sup>做多分类器融合算法。

在基于模糊积分的多分类融合系统中,模糊测度的取值决定着模糊积分的值,从而对分类结果产生较大影响。依靠模糊测度能够对分类器间交互作用的特性,实现将多个分类器融合后性能的正向提升。在应用

模糊积分进行融合时,需要学习出最优的模糊测度值,以更好地表现各分类器自身在融合系统中具有的重要程度和分类器之间的交互作用,以达到有效提高融合的分类性能的目的。

## 3 模型设计

本文提出了一种基于初始残差连接和层注意力方法的GCNRN模型。该方法基于ResNet首次提出了跳跃连接,可以缓解深度GCN的过平滑问题,提高了模型的泛化性。另外,GCNRN模型添加了一个层注意力机制建模每一层嵌入的重要程度,最后根据每一层的嵌入结果完成分类任务。

### 3.1 GCNRN模型框架

GCNRN模型框架如图1所示,其中,包含线性变化层、图卷积层、残差网络层、自注意力层和分类层。

线性变换层为了避免高维输入特征 $e$ 造成的计算负载,该模型需要将输入 $e$ 的特征通过线性变换层得到 $e_0$ 的输入特征向量。这个 $e_0$ 也就是图卷积神经网络中的第0跳表达,然后通过残差连接结构与自注意力机制得到最终商品的最终表达 $e_i$ 。

图卷积神经网络层包含 $Norm^{[13]}$ 、 $ReLU^{[14]}$ 和 $Dropout$ 模块<sup>[15]</sup>, $Norm$ 模块主要用于加速网络的收敛速度,具有一定程度的正则化效果。经过归一化操作后,不同维度的特征具有数值可比性。 $ReLU$ 是一个分段线性函数,相较于线性函数更具表现力。由于非负区间的梯度是常数,因此 $ReLU$ 函数不存在梯度消失的问题,可以使模型的收敛速度处于稳定状态。 $Dropout$ 模块可以防止过拟合,在训练过程中减少了特征之间的交互作用,以减少过拟合,在该模型中用于提高模型的鲁棒性和测试的准确性。

残差连接结构需要将模型的输入 $e_0$ 按比例连接到每一层。 $\alpha$ 、 $\beta$ 和 $\gamma$ 是根据每层的尺度比例来计算每层的特征矩阵, $\alpha$ 代表初始输入 $e_0$ 的比例, $\beta$ 代表上一层的特征矩阵的比例, $\gamma$ 代表当前层的特征矩阵的比例,使用残差网络解决不仅可以加速收敛,而且可以缓解深层的过度平滑,使用式(2)计算当前 $i$ 层的节点表征 $e_i$ 。

$$e_i = \alpha \times e_0 + \beta \times e_{i-1} + \gamma \times e_i \quad (2)$$

为更好地得到图卷积神经网络的最终表达,采用自注意力机制来建模每一层嵌入的重要程度,使用 $att_z$ 权重系数代表每层的重要度, $att_z$ 包含了第0层到

第  $z$  层的嵌入表达的权重,  $Softmax$  函数用于将  $z$  层嵌入的权重归一化.

$$att_z = Softmax(ReLU(e_0 \times e_z)) \quad (3)$$

利用注意力向量将每层的嵌入向量加权求和可以得到最终的节点表征  $e$ , 如式 (4) 所示.

$$e = \sum_{z=0}^k att_z \times e_z \quad (4)$$

### 3.2 基于模糊积分的多分类器融合方法

基于模糊积分的多分类融合方法是一种扩展的加权平均法, 可以根据分类器的输出与模糊测度动态为各分类器赋予权重, 从而使得融合后的分类性能比单个分类器效果好.

#### 3.2.1 模糊积分

模糊积分原理是建立在模糊测度上的概念. 设  $X$  是任意集合,  $R$  是  $X$  的子集构成的  $\sigma$  代数, 集函数  $\mu: R \rightarrow [0, 1]$  当且仅当满足以下条件时, 称为定义在  $R$  的模糊测度.

- (1) 平凡性. 若  $\emptyset \in R$ , 那么  $\mu(\emptyset) = 0$ .
- (2) 单调性. 若  $A \in R, B \in R, A \subset B$ , 则有  $\mu(A) \leq \mu(B)$ .

由此定义模糊积分.  $\mu$  是集合  $X$  上的模糊测度,  $x_i$  是集合  $X$  上的元素, Choquet 积分是关于  $\mu$  从  $X$  到  $R^+$  的函数, 定义如下<sup>[16]</sup>.

$$\int f d\mu = \sum_{i=1} [f(x_i) - f(x_{i-1})] \mu(A_i) \quad (5)$$

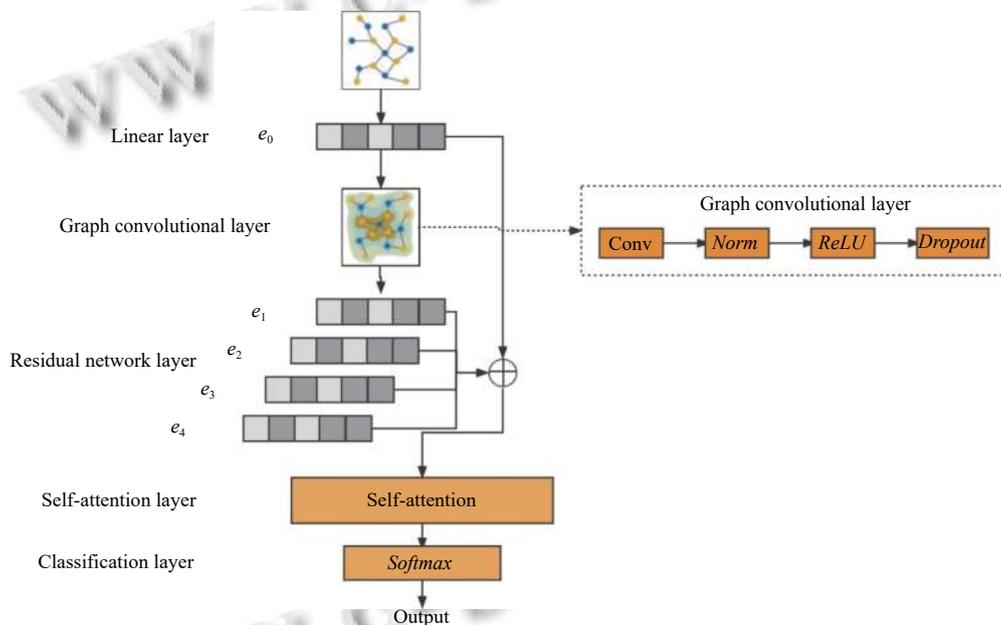


图1 GCNRN 模型

#### 3.2.2 模糊积分的模糊测度值算法

对于给定训练样本集  $Z = \{z_1, z_2, \dots, z_L\}$ ,  $C = \{C_1, C_2, \dots, C_n\}$  是  $n$  个目标类别组成的集合分类器,  $X = \{x_1, x_2, \dots, x_m\}$  是  $m$  个分类器组成的集合. 在  $X$  中取一分类器  $x_i$ , 则  $x_i$  对这些样本的隶属程度输出用  $D_L^i$  表示.

为了进一步给出模糊测度的计算式, 我们需要引入分类器识别过程中的 3 个不确定性表征量<sup>[17]</sup>.

(1) 随机性或概率不确定性. 分类正确率是对这一不确定性的反映, 我们使用  $R_i$  表示  $x_i$  对所有样本的分类准确率.

(2) 清晰度. 分类器对某类别的正确认知能力以及认知的清晰度大小称为分类器对这一类的清晰度, 定义为  $F_j^i$ , 同时也表示分类器  $x_i$  对所有样本中正确归类为  $C_j$  类的清晰度. 推广到全体样本, 对  $s$  个被分类成功的样本, 有:

$$F_j^i = \frac{\sum_{k=1}^s f_j^k}{s} \quad (6)$$

(3) 描述品质. 当样本被正确分类的时候, 输出向量中最大分量的值越大, 其他分量的值越小, 则该输出

向量的歧义性越小,分类器的分辨能力越强;输出向量的歧义性反映出的分类器的分辨能力称为分类器的描述品质,用符号 $D_j^i$ 如式(7)表示。

$$D_j^i = \left[ \frac{\sum_{k=1}^s (2 \times a_{kj}^i - 1)}{s} \right] + \left[ \frac{\sum_{k=1}^{L-s} (1 - 2 \times a_{kj}^i)}{L-s} \right] \quad (7)$$

其中,对分类器 $x_i$ ,所有样本在 $C_j$ 类上的分量为 $Z_j = \{z_{j1}, z_{j2}, \dots, z_{jL}\}$ ,相应的预测状态为 $(a_{1j}^i, a_{2j}^i, a_{3j}^i, \dots, a_{Lj}^i)$ . $s$ 为 $L$ 个样本中事实属于 $C_j$ 类的个数, $a_{kj}^i$ 为分类器给出的隶属程度, $L-s$ 为余下的不属于 $C_j$ 类的个数, $a_{kj}^i$ 为其隶属程度.输出向量的二元分布越明显,分类器的描述品质越高;输出向量的二元分布越模糊,分类器的可分辨度越低。

最终对分类器 $x_i$ ,对目标类别 $C_j$ ,模糊测度的计算如式(8)<sup>[18]</sup>。

$$V_j^i = R_i \times F_j^i \times e^{-\frac{(2-D_j^i)}{4}} \quad (8)$$

### 3.2.3 模糊积分的融合模型

基于模糊积分的多分类融合模型,要求分类器的输出为非负值向量,对于一个共有 $C$ 个类别的分类问题中,第 $i$ 个分类器的输出形式为 $[d_{i,1}, d_{i,2}, \dots, d_{i,c}]$ ,其中, $d_{i,1} \in [0, 1]$ , $j = 1, 2, \dots, C$ .把所有的分类器作为一个集合,记作 $X = [X_1, X_2, \dots, X_L]$ ,为每一类 $C_i$ 在集合 $X$ 的幂集上定义模糊测度,在第3.2.2节中进行模糊测度的计算,得到模糊测度 $V^i$ 后使用模糊积分把各个分类器对样本的输出进行综合.系统把最大可能性对应的类作为对样本的分类结果。

### 3.3 模型算法的具体实现

如果学习节点嵌入能够考虑更深入的信息同时保持模型的泛化性,那么就可以缓解模型过平滑问题.本文提出的GCNRN模型采用了残差结构与层注意力方案,恰恰能够解决这一问题。

为了避免高维输入引起的操作复杂性,我们需要通过全连接层来降低特征矩阵的维数.每一层都需要 $Norm$ 、 $ReLU$ 和 $Dropout$ 层.在每一层中,残差网络结构需要将输入 $e$ 按比例连接到每一层. $\alpha$ 、 $\beta$ 和 $\gamma$ 是根据每层的尺度比例来计算每层的特征矩阵,这可以缓解深层的过度平滑.为了更好地得到图卷积神经网络的最终表达,采用自注意力机制来建模每一层嵌入的重要程度,使用自注意力机制能够很好地学习到每一层的

重要度并且能提升模型效率,对最终的节点分类准确率作用是很大的.GCNRN模型的伪代码如下。

#### 算法1. GCNRN模型框架

输入: 图 $G\{V, E\}$ ; 输入特征 $\{e_v, \forall v \in V\}$ ; 图卷积深度 $K$ ; 残差连接系数 $\alpha, \beta, \gamma$ .  
输出: 表征向量 $Z_v, v \in V$ .

```

1.  $e^{(0)} \leftarrow e_v, \forall v \in V$ ;
2. for  $k=1, \dots, K$  do
3.    $e^{(k)} \leftarrow GCNs(e^{(k-1)}, G)$ ;
4.    $e^{(k)} \leftarrow Norm(e^{(k)})$ ;
5.    $e^{(k)} \leftarrow ReLU(e^{(k)})$ ;
6.    $e^{(k)} \leftarrow Dropout(e^{(k)})$ ;
7.    $e^{(k)} \leftarrow \alpha \times e^{(0)} + \beta \times e^{(k-1)} + \gamma \times e^{(k)}$ ;
8. end for
9. for  $k=1, \dots, K$  do
10.   $att_k \leftarrow Softmax(ReLU(e_0 \times e_k))$ 
11.   $e_k \leftarrow att_k \times e_k$ 
12. end for
13.  $Z_v \leftarrow Sum(e^{(1)}, \dots, e^{(k)}), \forall v \in V$ ;
14.  $Z_v \leftarrow Linear(Z)$ ,  $\forall v \in V$ ;
15.  $Z_v \leftarrow LogSoftmax(Z)$ ,  $\forall v \in V$ ;

```

### 3.4 模型损失函数

如图1所示,分类层将最终的节点表征 $e$ 作为输入,经过 $Softmax$ 分类器后对节点进行分类预测 $q_{i,c}$ 。

模型训练时,本文使用交叉熵 $L$ 作为损失函数进行最小化,如式(9)所示.其中, $N$ 表示样本个数, $M$ 代表类别个数, $p_{i,c}$ 和 $q_{i,c}$ 分别代表节点 $i$ 真实的类别与节点 $i$ 预测的类别.具体的收敛性分析将在第4.4节中详细给出。

$$L = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{c=0}^{M-1} p_{i,c} \log(q_{i,c}) \quad (9)$$

## 4 实验

本节将验证所提出的GCNRN模型的有效性,在多个数据集上进行图节点分类实验,并与多个基于经典模型展开对比.同时,针对融合模块,与传统的模型融合方法展开对比研究,证实了模糊积分融合机制的有效性。

### 4.1 基于模糊积分的多分类器融合方法

本文采用的4个数据集分别是ogbn-arxiv<sup>[19]</sup>、Cora、Citeseer和PubMed数据集,详情如表1所示。

表1 数据集描述

数据集	节点数	边	特征	标签
ogbn-arxiv	169343	1166243	128	40
Cora	2708	5429	1433	7
Citeseer	3327	4732	3703	6
PubMed	19717	44338	500	3

ogbn-arxiv 是一个真实的、大规模、多样的论文引文网络数据集, 每一个节点代表一篇 arXiv 论文, 每条有向边代表节点之间的引用关系, 每篇论文都有 128 维的特征向量, 它是通过对标题和摘要中单词的嵌入进行平均得到的, 常被用于 40 类的多分类问题. 另外 3 个数据集选取了被广泛使用的 Cora、Citeseer 和 PubMed 数据集, 相比于 ogbn-arxiv, 它们的数据量是在 2000 到 20000 节点内, 属于轻量级数据. 本文需要在大型和小型数据集的应用中, 验证 GCN 模型的泛化性能以及分类性能.

本次实验模型是在 PyTorch 框架下实现的, CPU 是锐龙 R7 2700X 3.8 GHz. 内存是 16 GB, 显卡是 GeForce GTX 3060.

### 4.2 实验超参数设定

为了保证实验的可靠性与公平性, 本文使用交叉验证方法将数据集划分训练集、验证集和测试集, 比例为 8:1:1. 本文 5 个对比模型的最终嵌入向量维度与本文提出的 GCN 模型维度相同.

本模型中的参数大多取多个模型的经验数据设定如表 2 所示, 只有其中的图卷积迭代次数  $la$  会对计算复杂度和分类性能产生一定影响. 本文对  $la$  做了取值变化测试, 在 4 个数据集上分别从 1 到 6 展开实验, 性能变化趋势如图 2 所示, 对每个数据库都可以看出,  $la$  取值为 5 时为性能较优的设置. 尤其对于 ogbn-arxiv 数据集, 卷积次数从  $la=1$  到  $la=2$  变化时, 分类准确率波动很大, 由于其数据量巨大, 导致了图卷积次数过少就会造成欠拟合, 而图卷积次数的数量对另外 3 个数据集影响波动较小, 经过实验结果的细微对比和宏观验证, 确定模型中的图卷积次数为  $la=5$ .

表 2 模型参数表

参数名称	参数含义	参数值
$d$	嵌入向量维度	128
$t$	自注意力机制的维度	128
$\alpha$	残差结构中初始嵌入向量比例	0.1
$\beta$	残差结构中上一层嵌入向量比	0.2
$\gamma$	残差结构中本层嵌入向量比	0.8
$lr$	学习率	1E-3
$la$	图卷积迭代次数	5
$dp$	Dropout 中保留比例	0.5

### 4.3 评价指标

本文的模型分类指标采用准确率 ( $accuracy$ )、验证集准确率 ( $valid$ )、宏平均 ( $macro\_avg$ ) 和权重平均

( $weighted\_avg$ ) 4 种来评价模型性能. 准确率如式 (10) 所示, 是针对测试数据集模型正确分类的样本数与总样本数的比值, 反映了模型对整个样本的判定能力. 其中,  $TP$  代表把正样本成功预测为正,  $TN$  代表把负样本成功预测为负,  $FP$  代表把负样本错误预测为正,  $FN$  代表把正样本错误预测为负.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$precision = \frac{TP}{TP + FP} \quad (11)$$

$$recall = \frac{TP}{TP + FN} \quad (12)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (13)$$

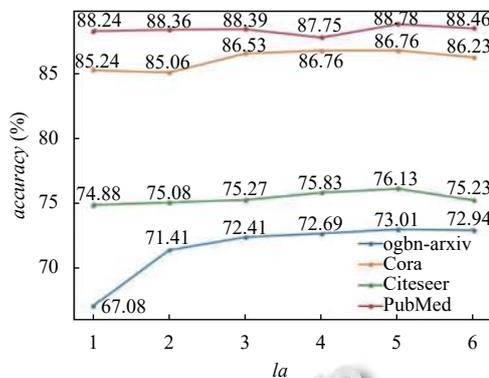


图 2 图卷积迭代次数调整对比实验

验证集准确率是对给定的验证数据集, 分类器正确分类的样本数与总样本数的比值.

宏平均如式 (14) 所示, 是指所有类别的每一个指标值的算数平均值, 也就是宏精确率  $precision$ 、宏召回率  $recall$  和宏  $F1$  值的算数平均值, 并且给他们赋予相等的权重, 这个方法平等看待每个类别.

$$macro\_avg = \frac{precision + recall + F1}{3} \quad (14)$$

权重平均如式 (15) 所示, 是用每一个类别样本数量在所有类别的样本总数的占比作为  $W$  权重, 再做加权平均计算. 该方法考虑了类别不平衡的情况, 相比于宏平均方法更具有可解释性.

$$weighted\_avg = \frac{W_p \times precision + W_r \times recall + W_F \times F1}{3} \quad (15)$$

### 4.4 对比与分析

在完成数据准备、参数设定和评价指标确定之后,

本文所提出的方法将与以下几个经典模型进行对比,已验证模型性能.相关模型包括多层感知机 MLP<sup>[20]</sup>、GCN、DeeperGCN、JKNet 和 GraphSAGE,具体介绍如下.

(1) 多层感知机 MLP. 这是一种前向结构的人工神经网络,映射一组输入向量到一组输出向量,MLP 是由多个节点层组成,每一层都全连接到下一层.将 MLP 应用于图数据,不需考虑边的结构信息,实现简单.

(2) 图卷积神经网络 GCN. 是能够直接作用于图并且利用其结构信息的图卷积神经网络,对于每个节点都需要考虑所有邻居节点与自身的特征信息,从而得到一个可以传入神经网络的结果向量.

(3) 深度图卷积神经网络 DeeperGCN. 此方法在堆叠多层图卷积神经网络的基础上,对于通过加深图卷积层以获得更大的信息传播范围,采用恒等映射与残差结构进行缓解过平滑问题.

(4) JKNet. 在堆叠多层图卷积神经网络的基础上加入了聚合层,在加深传播范围的同时做整体的聚合,

并且在多层聚合时提出 concatenation 聚合方式,将所有节点表针做串联后进行线性转换与最大池化操作,以选择信息最丰富的表征层方法.

(5) GraphSAGE. 此算法的核心思想是如何聚合节点邻居特征信息,它是通过采用随机采样的方法对邻居节点进行聚合以此降低计算复杂度,得到各节点的表征以供下游任务使用.

(6) GCNRRN. 本文采用了残差连接结构与层自注意力,在残差结构中对嵌入向量以确定最优的超参数,采用层自注意力机制计算多层图卷积神经网络的各层表征权重系数,以期解决过平滑现象.

以上多个模型分别在 ogbn-arxiv、Cora、Citseer 和 PubMed 四个数据集开展了对比实验,从表 3 中可以看出 GCNRRN 模型在 ogbn-arxiv、Citseer 和 PubMed 数据集中 *accuracy*、*valid*、*macro\_avg* 和 *weight\_avg* 指标均优于其他模型,有着明显的提升.说明本文提出的模型是有效的,并且能够很好地缓解过平滑问题,从而最终得到更好的节点表征.

表 3 模型实验结果

Dataset	Model	<i>accuracy</i>	<i>valid</i>	<i>macro_avg</i>	<i>weighted_avg</i>
ogbn-arxiv	MLP	0.5550±0.0023	0.5763±0.0012	0.4977±0.0037	0.5663±0.0041
	GCN	0.7174±0.0029	0.7280±0.0016	0.5527±0.0036	0.7132±0.0031
	DeeperGCN	0.7112±0.0016	0.7262±0.0014	0.5548±0.0021	0.7111±0.0030
	JKNet	0.7219±0.0021	0.7335±0.0010	0.5732±0.0023	0.7251±0.0017
	GraphSAGE	0.7149±0.0016	0.7176±0.0010	0.5552±0.0022	0.7103±0.0023
	GCNRRN	<b>0.7301±0.0024</b>	<b>0.7374±0.0022</b>	<b>0.6050±0.0021</b>	<b>0.7296±0.0014</b>
Cora	MLP	0.5972±0.0101	0.6110±0.0091	0.5842±0.0077	0.5798±0.0064
	GCN	0.8226±0.0054	0.8165±0.0027	0.8210±0.0013	0.8102±0.0017
	DeeperGCN	0.8561±0.0028	0.8755±0.0031	0.8588±0.0037	0.8614±0.0041
	JKNet	0.8672±0.0017	0.8835±0.0022	0.8731±0.0010	0.8689±0.0016
	GraphSAGE	<b>0.8819±0.0029</b>	<b>0.8875±0.0013</b>	0.8770±0.0023	<b>0.8772±0.0022</b>
	GCNRRN	0.8796±0.0023	0.8810±0.0025	<b>0.8808±0.0021</b>	0.8733±0.0017
Citeseer	MLP	0.5775±0.0105	0.5960±0.0092	0.5864±0.0061	0.5897±0.0088
	GCN	0.7113±0.0112	0.7110±0.0122	0.7083±0.0030	0.6988±0.0119
	DeeperGCN	0.7342±0.0037	0.7561±0.0033	0.7477±0.0026	0.7470±0.0039
	JKNet	0.7477±0.0016	0.7553±0.0037	0.7482±0.0023	0.7571±0.0025
	GraphSAGE	0.7598±0.0015	0.7683±0.0026	0.7538±0.0017	0.7624±0.0013
	GCNRRN	<b>0.7613±0.0035</b>	<b>0.7768±0.0028</b>	<b>0.7642±0.0018</b>	<b>0.7627±0.0017</b>
PubMed	MLP	0.7315±0.0068	0.7339±0.0062	0.7330±0.0055	0.7357±0.0031
	GCN	0.7889±0.0071	0.7891±0.0063	0.7727±0.0038	0.7887±0.0083
	DeeperGCN	0.7979±0.0047	0.8008±0.0055	0.7960±0.0049	0.7998±0.0041
	JKNet	0.8732±0.0030	0.8782±0.0033	0.8677±0.0023	0.8738±0.0029
	GraphSAGE	0.8801±0.0029	0.8830±0.0041	0.8811±0.0015	0.8825±0.0019
	GCNRRN	<b>0.8878±0.0027</b>	<b>0.8913±0.0022</b>	<b>0.8891±0.0013</b>	<b>0.8857±0.0022</b>

但是存在对于 Cora 数据集出现 GraphSAGE 模型的部分指标优于 GCNRRN 的问题.主要是因为 Cora 数

据集的样本量较少,只有 2708 个节点,而 GraphSAGE 在进行邻居聚合的时候是选取少量数量的节点.对于

GCNRN 相比 GraphSAGE 聚合所有邻居节点, 对于小数据集会出现过拟合现象, 但部分指标只是略低于 GraphSAGE, 且仍然优于其他模型. 对于出现小数据集过拟合的现象, 本文在后面章节使用模糊积分融合的方法将 GraphSAGE 模型优点与 GCNRN 模型融合, 并且取得最终分类准确率的提升.

如图 3 所示, 本文提出的 GCNRN 模型在所有数据集上都是在 100 轮次之内就开始收敛, 速度较快, 尤其是在 ogbn-arxiv 密集型数据集上收敛速度最快, 而对于 Cora、Citeseer 和 PubMed 这 3 个相对稀疏的数据集, GCNRN 模型在 PubMed 数据集上的收敛速度依然高于其他 2 个数据集, 这也证明了本文提出的模型在密集的数据集收敛比稀疏数据集收敛更快.

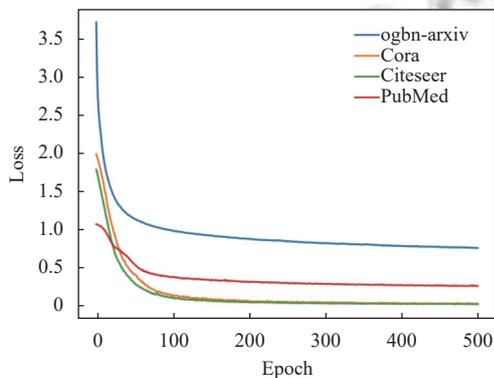


图 3 GCNRN 模型收敛曲线图

本文使用 GCNRN 模型对 ogbn-arxiv 数据集做节点分类实验, 并且将节点预测的类别清晰的区分开并且进行可视化, 如图 4 所示, 可以看出 GCNRN 模型的分可视化效果相比于其他模型图更平滑, 并且形成一种聚集的状态.

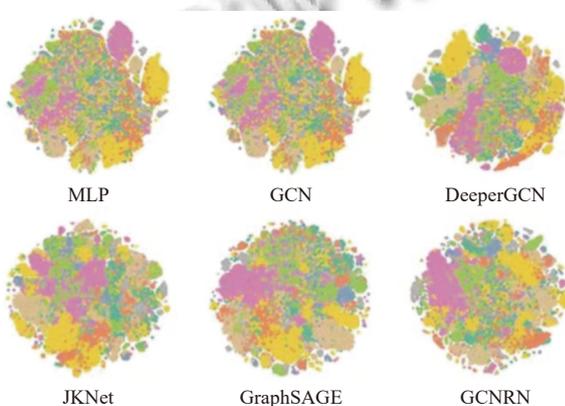


图 4 分类可视化

除此之外, 在提出的分类框架中, 有效的模型融合可以取长补短各模型的优点, 本文采用了模糊积分融合机制, GCNRN 的基础上融合了多个分类模型, 以提高分类框架整体分类性能. 本文采用融合模块的节点分类模型分别为 GraphSAGE 和 JKNet. 并且使用模糊积分对比常用的平均法、最大值法和投票法这 3 种融合方法, 并且使用分类的准确率作为衡量指标.

在上述中出现 GCNRN 模型在 Cora 小数据集中部分指标略差与 GraphSAGE 模型. 为了解决这个问题, 在模型融合中, 本文需要将 GraphSAGE 模型对于小数据集缓解过平滑的优点迁移到 GCNRN 模型, 因此本文使用模型融合对分别使用 GCNRN、GraphSAGE 和 JKNet 模型进行两两融合与 3 个模型整体融合.

在表 4 模型融合结果中, 对于 ogbn-arxiv 数据集, 模型融合结果中可以看出, 在 GCNRN、GraphSAGE 和 JKNet 中无论是两两融合还是 3 个模型一起融合, 使用模糊积分进行融合的方法, 其分类准确率都是优于其他 3 种方法的. 从两两融合模型中来看, GCNRN 模型与 GraphSAGE 模型融合是比 GCNRN 模型与 JKNet 模型融合的分类准确率更高的, 因为 GCNRN 模型融合了 GraphSAGE 随机采样的优点, 从而使得最终的融合结果更优.

另外, 在 Cora 数据集中使用融合后性能提升幅度很大, 主要原因是这个数据集较小, 那么模型融合后就会取得更优的模型融合效果. 与 Cora 数据集大小相近的数据集 Citeseer, 同样发现使用模型融合后分类准确率提升幅度很大.

所以, 从上述的分析与表的模型融合结果可以得出以下结论, 对于数量级相对大的 ogbn-arxiv、PubMed 数据集, 经过模糊积分融合后都是有提升的, 但是 Cora、Citeseer 小数据量的数据集来看, 它们经过模型融合后提升的效果是更高的, 因为对于小数据集来说, 模型的划分的测试数据集是比较小的, 那么每个模型预测对这个结果的影响是更大的.

其次, 本文使用模糊积分融合相比于其他融合方法取得更好的效果, 使用模糊积分的优点是能使分类器的准确性、描述的质量和可区分性具有可解释性, 通过式 (8) 确定模糊测度后, 将分类精度作为确定模糊测度的基础, 从分类器的多个方面对其进行了检验. 这使得基于模糊积分融合技术进行多分类器融合相比于传统融合方法具有更优的效果.

表4 模型融合结果

数据集	融合方式	平均法	最大值法	投票法	模糊积分
ogbn-arxiv	<GCNRRN, JKNet, GraphSAGE>	0.7174±0.0017	0.7156±0.0026	0.7184±0.0022	<b>0.7335±0.0025</b>
	<GCNRRN, JKNet>	0.7126±0.0019	0.7058±0.0016	0.7014±0.0021	<b>0.7312±0.0016</b>
	<JKNet, GraphSAGE>	0.7219±0.0023	0.7186±0.0018	0.7050±0.0033	<b>0.7260±0.0019</b>
	<GCNRRN, GraphSAGE>	0.7219±0.0014	0.7186±0.0019	0.7050±0.0017	<b>0.7283±0.0013</b>
Cora	<GCNRRN, JKNet, GraphSAGE>	0.8763±0.0022	0.8800±0.0022	0.8726±0.0022	<b>0.8982±0.0027</b>
	<GCNRRN, JKNet>	0.8745±0.0023	0.8763±0.0027	0.8735±0.0017	<b>0.8835±0.0022</b>
	<JKNet, GraphSAGE>	0.8671±0.0027	0.8653±0.0026	0.8726±0.0024	<b>0.8892±0.0029</b>
	<GCNRRN, GraphSAGE>	0.8819±0.0029	0.8819±0.0023	0.8837±0.0022	<b>0.8929±0.0037</b>
Citeseer	<GCNRRN, JKNet, GraphSAGE>	0.7657±0.0011	0.7657±0.0014	0.7672±0.0026	<b>0.7867±0.0025</b>
	<GCNRRN, JKNet>	0.7687±0.0023	0.7612±0.0031	0.7522±0.0032	<b>0.7792±0.0026</b>
	<JKNet, GraphSAGE>	0.7687±0.0021	0.7687±0.0023	0.7537±0.0027	<b>0.7822±0.0024</b>
	<GCNRRN, GraphSAGE>	0.7657±0.0031	0.7627±0.0029	0.7552±0.0033	<b>0.7837±0.0032</b>
PubMed	<GCNRRN, JKNet, GraphSAGE>	0.8871±0.0023	0.8884±0.0020	0.8846±0.0016	<b>0.8945±0.0017</b>
	<GCNRRN, JKNet>	0.8823±0.0025	0.8841±0.0023	0.8785±0.0025	<b>0.8904±0.0015</b>
	<JKNet, GraphSAGE>	0.8820±0.0021	0.8831±0.0023	0.8742±0.0026	<b>0.8846±0.0025</b>
	<GCNRRN, GraphSAGE>	0.8891±0.0019	0.8899±0.0019	0.8803±0.0027	<b>0.8939±0.0024</b>

如图5所示,以数据集 ogbn-arxiv 为例,将不同融合方法作用的结果进行了可视化分析,可以清晰看出使用模糊积分的模型融合机制的分类框架准确率明显优于其他方法,这体现了模糊积分做多分类融合的优势,相比于其他常见的方法能有更好的融合能力。蓝色标识的3种模型融合的效果优于融合2种模型。

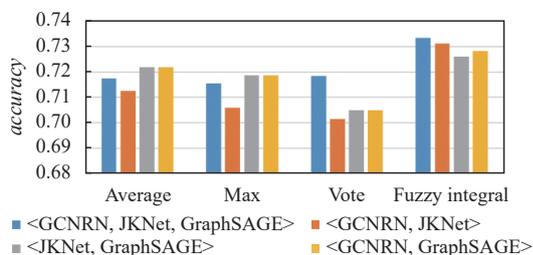


图5 作用于 ogbn-arxiv 数据集模型融合统计图

## 5 总结

目前大部分基于图卷积神经网络的分类算法使用各种数据增强手段提升最终分类准确度,本文研究并总结了现阶段已经存在的基于 GCN 的常见模型与常见数据增强方法并用于分类任务。受 JKNet 和 DeeperGCN 等模型的启发,我们提出了基于残差连接结构同时结合层自注意力方案的 GCNRRN 模型,使用自注意力优化了图卷积部分嵌入向量的最终表达,GCNRRN 模型比其他以 GCN 为核的模型取得了更好的效果。

为了进一步提取深层次的信息,本文采用模糊集

成的方法进行多分类模型融合,并在实验结果表明,模糊积分融合比常用的模型融合能更有效地融合分类器的互补信息。在4个基准数据集上进行实验验证,结果表明,本文提出的方法对提升分类效果有一定的作用。未来将在节点表征上采用更先进的框架,采用不同的特征融合机制,探究不同的特征融合机制对分类效果的影响。

## 参考文献

- 1 Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2016.
- 2 Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. arXiv:1710.10903, 2017.
- 3 Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.
- 4 Xu KL, Li CT, Tian YL, et al. Representation learning on graphs with jumping knowledge networks. Proceedings of the 35th International Conference on Machine Learning. Stockholmsmässan: PMLR, 2018. 5453–5462.
- 5 Chen M, Wei Z, Huang Z, et al. Simple and deep graph convolutional networks. Proceedings of the 2020 International Conference on Machine Learning. PMLR, 2020. 1725–1735.
- 6 Li G, Xiong C, Thabet A, et al. DeeperGCN: All you need to train deeper GCNs. arXiv:2006.07739, 2020.

- 7 胡毅, 瞿博阳, 梁静, 等. 进化集成学习算法综述. 智能科学与技术学报, 2021, 3(1): 18–33. [doi: [10.11959/j.issn.2096-6652.02103](https://doi.org/10.11959/j.issn.2096-6652.02103)]
- 8 Wang WY, Tran D, Feiszli M. What makes training multi-modal classification networks hard. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 12692–12702.
- 9 Izmailov P, Podoprikin D, Garipov T, *et al.* Averaging weights leads to wider optima and better generalization. Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence. Monterey: AUAI Press, 2018. 876–885.
- 10 Micheli A. Neural network for graphs: A contextual constructive approach. IEEE Transactions on Neural Networks, 2009, 20(3): 498–511. [doi: [10.1109/TNN.2008.2010350](https://doi.org/10.1109/TNN.2008.2010350)]
- 11 Bruna J, Zaremba W, Szlam A, *et al.* Spectral networks and locally connected networks on graphs. arXiv:1312.6203, 2013.
- 12 Lou SH, Feng YX, Li ZW, *et al.* Two-additive fuzzy measure-based information integration approach to product design alternative evaluation. Journal of Industrial Information Integration, 2022, 25: 100247. [doi: [10.1016/j.jii.2021.100247](https://doi.org/10.1016/j.jii.2021.100247)]
- 13 Wu YX, He KM. Group normalization. arXiv:1803.08494. 2018. 3–19.
- 14 Shen ZW, Yang HZ, Zhang SJ. Optimal approximation rate of ReLU networks in terms of width and depth. Journal de Mathématiques Pures et Appliquées, 2022, 157: 101–135.
- 15 Kong XT, Liu XN, Gu JJ, *et al.* Re-flash dropout in image super-resolution. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 5992–6002.
- 16 Grabisch M, Labreuche C. Bi-capacities—II: The Choquet integral. Fuzzy Sets and Systems, 2005, 151(2): 237–259. [doi: [10.1016/j.fss.2004.08.013](https://doi.org/10.1016/j.fss.2004.08.013)]
- 17 王晓君. 基于模糊积分的多分类器融合研究 [硕士学位论文]. 保定: 河北大学, 2005.
- 18 王业琳, 宁新宝, 尹义龙. 指纹图像细化算法的研究. 南京大学学报(自然科学), 2003, 39(4): 468–475.
- 19 Hu W, Fey M, Zitnik M, *et al.* Open graph benchmark: Datasets for machine learning on graphs. Advances in Neural Information Processing Systems, 2020, 33: 22118–22133.
- 20 Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations. Cambridge: MIT Press, 1987.

(校对责编: 牛欣悦)