



研究与开发

基于多通道的边学习图卷积网络

杨帅, 王瑞琴, 马辉

(湖州师范学院信息工程学院, 浙江 湖州 313000)

摘要: 通常图的边包含了图的重要信息, 然而目前大多数用于图学习的深度学习模型(如图卷积网络(graph convolutional network, GCN)和图注意力网络(graph attention network, GAT))没有充分利用多维边特征的特性; 另一个问题是图中可能存在噪声, 影响图学习的性能。使用多层感知机对图数据进行去噪优化处理, 在 GCN 的基础上引入了多通道学习边特征的方法, 对图的多维边属性进行编码, 按原始图所包含的属性分别建模为多通道, 每个通道对应一种边特征属性对图节点进行约束训练, 可以让算法更合理地学习图中多维边特征, 在 Cora、Tox21、Freesolv 等数据集上的实验证明了去噪方法与多通道方法的有效性。

关键词: 图卷积网络; 边特征; 图去噪; 多通道; 边学习

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2022250

Multi-channel based edge-learning graph convolutional network

YANG Shuai, WANG Ruiqin, MA Hui

School of Information Engineering, Huzhou University, Huzhou 313000, China

Abstract: Usually the edges of the graph contain important information of the graph. However, most of deep learning models for graph learning, such as graph convolutional network (GCN) and graph attention network (GAT), do not fully utilize the characteristics of multi-dimensional edge features. Another problem is that there may be noise in the graph that affects the performance of graph learning. Multilayer perceptron (MLP) was used to denoise and optimize the graph data, and a multi-channel learning edge feature method was introduced on the basis of GCN. The multi-dimensional edge attributes of the graph were encoded, and the attributes contained in the original graph were modeled as multi-channel. Each channel corresponds to an edge feature attribute to constrain the training of graph nodes, which allows the algorithm to learn multi-dimensional edge features in the graph more reasonably. Experiments based on Cora, Tox21, Freesolv and other datasets had proved the effectiveness of denoising methods and multi-channel methods.

Key words: graph convolutional network, edge feature, graph denoising, multi-channel, edge-learning

收稿日期: 2022-04-05; 修回日期: 2022-08-22

通信作者: 王瑞琴, wrq@zjhu.edu.cn

基金项目: 国家社会科学基金资助项目 (No.20BTQ093); 浙江省自然科学基金资助项目 (No.LY20F020006)

Foundation Items: The National Social Science Foundation of China (No.20BTQ093), The Natural Science Foundation of Zhejiang Province (No.LY20F020006)



0 引言

近年来,神经网络方法在各种顶级会议与期刊上成为最热门的机器学习技术之一。例如,卷积神经网络(convolutional neural network, CNN)^[1]在图像识别领域、循环神经网络(recurrent neural network, RNN)在自然语言处理领域等方面,解决实际需求的问题时以优异的性能受到研究人员的青睐。在现实世界中许多问题可以直接用图代替传统的表格、图像或时间序列建模,并且比传统模型更加直观高效。一般来说,图由节点和边组成,在现实生活中真实存在的部分可以作为节点,其间存在的联系或互动作为边。例如,路上的公交车站可以作为节点,公交车的运行路线可以作为边。对于节点,一般可以收集其信息建模为特征向量;对于路线,可以简化为节点间的直线边进行建模。由于图一般具有复杂的结构信息,图学习的一个挑战就是找到一个行之有效的方法在众多信息中学习一个高效的模型。这几年来,研究者们已经开发了数种针对图学习的神经网络模型,相比传统的神经网络模型,其性能更优异,应用也更广泛。受图傅里叶变换的启发,Defferrard 等^[2]将卷积神经网络中的卷积操作作为原型提出了一种图卷积运算,Kipf 等^[3]使用重新归一化的一阶邻接矩阵逼近多项式,以获得图节点分类任务的可比较结果,这些图卷积网络(graph convolutional network, GCN)^[2-3]是结合图节点特征和图拓扑结构信息进行预测的。Veličković 等^[4]将注意力机制引入图学习,提出了图注意力网络(graph attention network, GAT)。

目前,许多图神经网络模型存在的一个问题是未能充分利用图中的边特征。在图卷积神经网络中,图被简化为无边属性的连通图,在图神经网络中同样只需要使用节点间是否连通的特性。然而,实际上图的边通常拥有丰富的特征,如强度、类型、各种离散数值等多维变量。合理地利用边特征可能会帮助许多图学习方法获得更高的

精度。图神经网络的另一个问题就是每个神经层是根据作为输入给出的原始邻接矩阵聚合节点特征,但是原始的图网络有可能存在噪声问题,这将限制信息聚合与预测任务的有效性。本文在标准 GCN 的基础上以更合理的方式利用边特征并优化噪声问题,提出边学习图卷积网络(edge-learning graph convolution network, EGCN)。本文在几个引文网络数据集和分子数据集上进行了实验。通过利用多通道学习边特征,本文的方法与其他的最新方法相比,获得了不错的结果。实验结果表明,边特征对图学习具有重要帮助,并证明 EGCN 可以有效地学习边特征。总之,本文的特点包括两点:提出了针对数据中存在的噪声的优化方法,该方法通过多层感知机(multilayer perceptron, MLP)计算节点间的类相似度,忽略部分类相似度值较低的噪声边的影响,从而提升模型性能;提出多通道边特征学习的方法,对图中边的每一种属性进行编码,使模型在训练中更合理地学习边的不同属性的隐藏特征。

1 相关工作

图学习的一个关键问题是图数据的复杂非欧结构。为了解决这一问题,传统的机器学习方法用图表统计、核函数或其他手动制作的特征提取邻域局部结构特征。这些方法缺乏灵活性,且手动设计合理的功能非常耗时,需要大量的实验才能将其推广到不同的任务或数据集中。图表示学习尝试使用数据导向的方法将图或图节点嵌入低维向量空间,较为流行的一类为基于矩阵分解的嵌入方法,如拉普拉斯特征映射、图分解和高阶邻近保持嵌入(high-order proximity preserved embedding, HOPE);另一类方法侧重于使用基于节点相似性度量的随机游走,如 DeepWalk、node2vec 和大规模信息网络嵌入(large-scale information network embedding, LINE)。基于矩阵分解和基于随机游走的图学习方法均存在一定的

局限性。例如,映射到低维向量空间的嵌入函数是过于简单的线性函数,可能无法捕捉复杂的信息,而且它们通常不包含节点特性;它们需要从邻域聚合信息,所以需要对整个图的结构进行训练。

最近,随着深度学习的发展,图学习中的部分问题有所解决。基于神经网络的深度学习可以表示复杂的映射函数,并通过梯度下降方法进行有效优化。随着 CNN 在图像识别方面的成功,人们自然而然地想将卷积应用于图形学习中。在文献[5]中,卷积运算是在图拉普拉斯变换的谱空间中定义的。Defferrard 等^[2]提出利用图拉普拉斯算子的切比雪夫展开式近似滤波器,从而产生空间域局部化的滤波器,同时避免了计算特征向量的拉普拉斯算子。Veličković 等^[4]在图学习中引入了注意机制,提出了 GAT。但 GAT 与 GCN 仍重点关注节点特征,未能充分考虑节点间边的多维信息特征,仅使用了边的连通性,同时他们的模型没有考虑任务中部分节点的邻居节点有误导影响,没有考虑噪声的局限性。

为了有效利用图上的边特征,边增强图神经网络(edge enhanced graph neural network, EGNN)^[6]对边特征使用双重随机归一化以及跨网络自适应方法以学习边特征,并且在 GCN 与 GAT 的基础上提出了 EGNN(C)与 EGNN(A)的改进模型。Chen 等^[7]则通过提取边与节点的互信息(mutual information),提出边信息最大化图神经网络(edge information maximized graph neural network,

EIGNN)。Gilmer 等^[8]研究近年来的众多图学习模型的共性后,提出了消息传递神经网络(message passing neural network, MPNN),图神经网络的本质是特征聚合,即节点根据周围邻居节点更新自身特征信息,这个方法即消息传递。上述图神经网络将部分边信息引入图学习,但仍存在未充分利用边的多维信息和原始节点间可能存在噪声影响的局限性。

2 多通道图卷积神经网络的结构

传统的 GNN 有许多变体,如 GCN^[3]、GAT^[4]、Graphs AGE^[9]等。这些神经网络模型侧重于学习节点状态,可以为邻居分配权重,但不能处理各种边特征。其中,通用的邻居节点聚合方案为:

$$h_v^{(l+1)} = \sigma \left(\sum_{w \in N_v} a_{vw} W_l^{(l)} h_w^{(l)} + W_0^{(l)} h_v^{(l)} \right) \quad (1)$$

其中, l 表示第 l 层神经网络层, σ 为非线性激活函数; a_{vw} 是一个标准化常数或学习注意系数; h 是一个将节点特征从输入空间映射到输出空间的变换, W^l 是一个 $F^l \times F^{l+1}$ 的可学习权重矩阵。这些聚合方法尚不能满足处理边特征的需求。例如,对于分子分析的任务来说,原子键上的属性对于分子属性具有决定性影响,但目前的图神经网络大多是针对单维边属性设计的,且不适合分子分析这种单个图比较小、总体上图的数量多的任务。

在传统图卷积神经网络的基础上,本文的多通道边学习图卷积网络模型总体结构如图1所示。

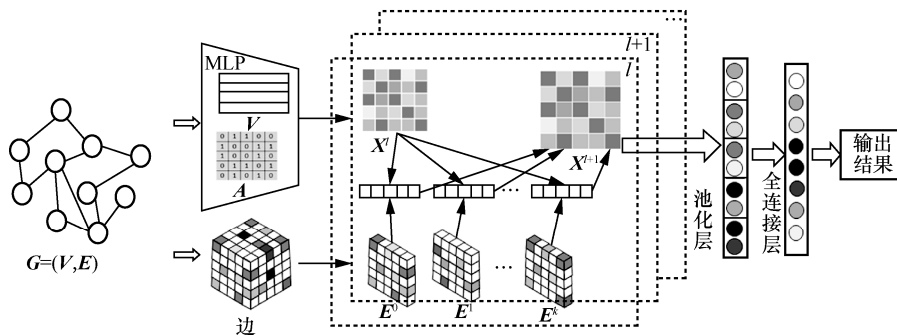


图1 多通道边学习图卷积网络模型总体结构



设一个图 $G=(V, E)$, G 表示有 N 个节点的图, 其中 V 是一组 N 个数据的节点表示, E 是一组 N 个数据的边表示。图 1 中的节点编码为特征矩阵 $X \in \mathbb{R}^{N \times f}$, f 表示一个节点的维度, X^0 是第一层神经网络的输入, E^k 表示通道 k 中的边矩阵。在图 1 中, 上标 l 表示第 l 层的输出。首先, 用 MLP 层对数据进行预处理, 经过预处理后获得更适合本文模型需要的节点集 X^0 与边集 E^0 并将其输入下一步的神经网络层。在信息聚合层, X^0 和 E^0 表示神经网络的输入。在第一层聚合层将边特征 E^0 应用于 X^0 , 经过聚合邻域产生一个 $N \times F^1$ 的新节点特征矩阵 X^1 。经过调整的 E^1 作为边特征送到下一层。这个步骤在以后的每一层都重复。节点特征 X^l 可被认为是图节点在 F^l 维空间中的嵌入。对于节点分类问题, 将 softmax 函数应用于最后一维的每个节点嵌入向量 X^L 。对于全图预测 (分类或回归) 问题, 将池化层应用于 X^L 的第一维, 从而将特征矩阵简化为整个图的单个向量嵌入。然后将全连接层应用于图向量, 其输出可用于预测回归, 也可用于分类。

2.1 去噪处理

现有图网络的初始邻接矩阵可能有噪声且不是最优的, 这将影响邻域聚合操作的有效性。例如, 引文网络中机器学习论文有时引用数学论文或其他理论论文, 然而, 数学论文可能很少引用机器学习论文。在此之前已有类似的工作, 例如, 文献[10]使用标签传播和单个训练过的图神经网络模型扩大图神经网络的训练集, 文献[11]采用 DeepCluster^[12]等无监督学习技术帮助图神经网络的自训练, 文献[10]和文献[11]可被看作是图神经网络自训练或协同训练的特定算法, 文献[13]使用正则化项对每个节点提供新的监督信号进行训练。

节点间的类相关度可以通过直接使用标签加节点数据进行计算获得, 但是在稀疏分割条件下, 只有小部分节点和标签可用, 很难直接

从节点和标签计算其类相关度, 所以本文使用多层感知机从节点特征中学习获得类相关度。具体来说, 首先用多层感知机从原始节点属性中提取类感知信息, 多层感知机的第 l 层的定义为:

$$Z_m^{(l)} = \sigma(Z_m^{(l-1)} W_m^{(l)}) \quad (2)$$

其中, $Z_m^{(0)} = X^0$, $W_m^{(l)}$ 为 MLP 的可学习权重矩阵, σ 为激活函数, MLP 的最后输出为 Z_m , 对 Z_m 使用 softmax 函数可获得概率矩阵 P , 如式 (3) 所示, 矩阵 P 中每个元素 P_{ic} 表示属于类 C 的概率。矩阵 P 经处理可得类相似度矩阵 B , 将类相似度矩阵 B 中的一些比例的低相似度的边置为 0, 获得分配矩阵, 在后续训练中起到删除该边的作用。

$$P = \text{softmax}(Z_m) \quad (3)$$

总的来说, 本步骤基本思想为基于阈值的节点间类相似度的边筛选方法, 基于 MLP 的分类标签生成候选边集。在训练过程中计算节点间的类相似度, 将类相似度值低于阈值的节点对的边划入低类相关度集合, 选择集合中的一些比例的最低值的边置为 0, 即经过算法处理可得分配矩阵, 该矩阵在网络中的作用为按比例删除假定为噪声的边。为了不过度破坏数据集原有的结构以致影响任务的精度, 实验中对于节点低相关性的边进行保守删除, 过高的删除比例虽然会获得更高的准度, 但会出现训练严重过拟合、精度不高等问题。

2.2 基于多通道边学习的图卷积网络

通常实际数据中图的边上包含有关图的重要信息。例如, 在引文网络中, 引用关系应该是有方向的, 对于节点分类任务来说, 论文的引用方向可能有十分重要的作用, 因此在建模引文网络数据集时, 将论文的引用方向作为一组边特征编码为一个通道; 在分子性质分析任务中, 对分子建模时, 由于其性质十分复杂, 不能简单地将其视为无属性连通图。本文根据图数据天然存在的属性将边编码为多维矩阵, 边的每一个属性将对

应一个通道,让模型在各通道中根据不同的神经网络参数学习节点特征与边特征。在引文网络数据集中将论文的引用方向和引用次数分解为双通道进行建模,将分子数据利用不同的边属性分解为 5 个通道建模。图的边特征作为输入进入卷积层将被转化为边约束,卷积层 l 中的图在基于边约束的条件下转换为约束图信号,这个图信号由边约束矩阵 $\{A_1, A_2, \dots, A_k\}$ 组成。图卷积聚合来自每个通道的所有一阶邻居的节点信息。在第 l 层中,约束图信号的图卷积式为:

$$X_k^{l+1} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A}_k^l \tilde{D}^{-\frac{1}{2}} X^l W_k^l \right) \quad (4)$$

图卷积的一个重要工具是图拉普拉斯矩阵^[3]。经过证明,归一化的图拉普拉斯矩阵定义为 $L = I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$, 是图上聚合邻域信息的合理方法。在这个基础上引入从边学习的权值 w , 即 $L = I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \cdot w$ 。多通道的边学习图卷积算法如算法 1 所示,该权值可被认为是该通道中边特征在卷积中所起作用的权重,这个参数的加入可以让每个通道卷积时获得边信息。每个通道中在边权值 w 的约束下聚合每个节点及其所有邻居的特征,在每层卷积层完成后将每个通道的输出 X_k^{l+1} 进行拼接获得 X^{l+1} , 如式 (5) 所示,当权值 w 为 1 时,本文方法退化为传统的图卷积网络。

$$X^{l+1} = [X_0^{l+1}, X_1^{l+1}, \dots, X_k^{l+1}] = \sigma \left[\left\| \left(\tilde{D}^{-\frac{1}{2}} \tilde{A}_k^l \tilde{D}^{-\frac{1}{2}} X^l W_k^l \right) \right\| \right] \quad (5)$$

其中, $\tilde{A}_k^l = A_k^l w_k + I$, D 为归一化度矩阵, w_k 为通道 k 中可学习的边权值, W_k^l 为卷积层中可学习权重, σ 为激活函数, $\|$ 表示拼接操作。所有节点嵌入拼接形成完整的特征矩阵 X_k^{l+1} 。 X^{l+1} 进入下一个卷积层或输出至池化层。

算法 1 多通道的边学习图卷积算法

输入 通道 k ; 模型 $f(\cdot)$; 节点张量 X ; 边张量 E ; 邻接矩阵 A

可学习参数: 权重 W , 权重 w

多通道卷积方法

for $i=1$ to k **do**

$w_k = \text{sigmoid}(\text{Conv2d}(E_k))$

$\tilde{A}_k^l = A_k^l \cdot w_k + I$

$X_k^{l+1} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A}_k^l \tilde{D}^{-\frac{1}{2}} X^l W_k^l \right)$

end for

输出 $X^{l+1} = \sigma \left[\left\| \left(\tilde{D}^{-\frac{1}{2}} \tilde{A}_k^l \tilde{D}^{-\frac{1}{2}} X^l W_k^l \right) \right\| \right]$

总之,本文的工作是基于—阶近似的切比雪夫 (Chebyshev) 卷积,在此基础上使用多通道学习方法对图卷积进行多次使用。多通道机制在模型学习过程中会产生一组边权值 w_k , 这些参数提供了比传统图卷积网络更具表现力的网络模型,图卷积网络可以学习节点之间多种边类型,根据不同通道的可学习参数,让模型学习不同边属性的特征。

2.3 可微分池化

本文模型使用了可微分池化 (Diffpool)^[14] 以层级化学习图表示。可微分池化设计了一个图池化神经网络来生成节点的赋值矩阵 S :

$$S = \text{softmax}(\text{GCN}(V, E)) \in \mathbf{R}^{n_l \times n_{l+1}} \quad (6)$$

softmax 函数的输入是特征矩阵 X^l 和邻接矩阵 A , GCN_l^l 是第 l 层的传统图卷积网络^[3]。在可微分池化层的第一层 $n_l = n$, 由于加入了多通道学习方法,第一个可微分池化层的输入是 $A^l = \sum_{i=1}^k \tilde{A}_i^{l-1}$, 然后使用 GCN_2^l 得到新的聚类邻接矩阵 X^{l+1} 和聚类节点的特征矩阵 A^{l+1} , 如式 (7)、式 (8) 所示。

$$X^{l+1} = (S^l)^T \text{GCN}_2^l(A^l, X^l) \quad (7)$$

$$A^{l+1} = (S^l)^T A^l S^l \quad (8)$$

最后的输出是图表示 $X \in \mathbf{R}^{1 \times G}$, f_G 表示图的维数。Diffpool 定义了新的层级化池化层以抽取图的信息,通过学习赋值矩阵将不同的节点分配到



不同的聚类中, 结合正则化的边预测来共同优化学习卷积参数, 从而学习到更丰富的层级化网络信息。

2.4 复杂度分析

与传统的卷积神经网络不同, 由于不同的节点与边的复杂性不同, 图卷积的成本是不稳定的, 无法准确给出每一步计算的时间复杂度, 在计算机中数据的加减运算的成本较低, 故训练算法的时间复杂度可以通过乘法运算的次数进行推算。GCN、GAT 与 EGCN 的时间复杂度与内存复杂度对比见表 1。

在表 1 中, n 是所有节点的数量, m 是所有边的数量, K 是网络层数, c 是通道数量, 节点隐含特征维数为 d 。传统的 GCN^[3]的时间复杂度为 $O(Kmd+Knd^2)$, GAT 较 GCN 的计算量更少, 但是需要处理的数据规模是相同的, 而本文同样使用传统 GCN 的基于一阶近似的切比雪夫卷积, 所以增加的时间复杂度与计算复杂度为增加的通道数量。

3 实验结果

本文所有的算法与实验都基于 Python 3.7 平台上的 torch-1.6.0 实现。在所有的实验中, 模型都用 Intel core i5-10500 处理器进行处理。实验的任务包括验证本文方法对几种基础方法的改进。由于引文网络数据集需要使用边特征来建模, 因此使用了 Cora 和 Citeseer 的原始版本。分子数据集使用了 Tox21、Freesolv、Lipophilicity 和 eSOL 数据集。节点特征和边特征是使用 RDKit 提取的, RDKit 是一个开源的化学信息学软件包。RDKit 将 SMILES 字符串转换为“mol”格式, 其中包含用于构建分子图的分子结构信息。

表 1 GCN、GAT 与 EGCN 的时间复杂度与内存复杂度对比

模型	GCN	GAT	EGCN
时间复杂度	$O(Kmd+Knd^2)$	$O(Kmd+Knd^2)$	$O(cKmd+cKnd^2)$
内存复杂度	$O(Knd+Kd^2)$	$O(Knd+Kd^2)$	$O(cKnd+cKd^2)$

3.1 数据集

为了测试本文提出 EGCN 的有效性, 将其应用于多种常见的数据集。本节测试了两个节点分类任务引文网络数据集——Cora、Citeseer; 一个全图分类任务的分子分析数据集, Tox21; 3 个预测分子性质的分子分析数据集, Lipophilicity、Freesolv 和 eSOL。引文网络中的 Cora 与 Citeseer 均用论文中关键字是否出现以二进制指标表示节点特征, 根据论文的引用方向关系编码为多个离散值的边特征矩阵。Tox21 的原始数据来自 21 世纪的毒理学研究计划, 它包含 7 831 个环境化合物和药物, 以及 12 个标签的生物学结果, 是全图分类任务。回归任务有 3 个数据集: Freesolv 是一个含有 642 个水中小中性分子的水合自由能的数据库; Lipophilicity (Lipo) 由 ChEMBL 数据库整理, 包含 4 200 个化合物; eSOL 提供了 1 128 种化合物的水溶性数据。各数据集的基础统计数据见表 2。分子数据集均从 MoleculeNet^[15]下载。实验中忽略了结构图没有边的 SMILES 样本。

表 2 各数据集的基础统计数据

数据集	节点数 (分子数) /个	边/个	标签数/个	任务类型
Cora	2 708	5 429	7	分类
CiteSeer	3 327	4 732	6	分类
Tox21	7 433	—	13	分类
Freesolv	642	—	2	回归
eSOL	1 125	—	2	回归
Lipo	4 190	—	9	回归

3.2 评价指标

对于分类任务, Cora 与 Citeseer 数据集的任务为节点分类任务, 评价指标为准确率, 由预测标签和实际标签计算得来。Tox21 数据集的任务为分子分类, 即全图分类, 采用 ROC 曲线下面积

(area under curve, AUC) 作为评价指标。其中, ROC 曲线全称为“受试者工作特征曲线(receiver operating characteristic curve)”,它是根据分类得到的二值分类结果,以真阳性率(true positive rate, TPR)为纵坐标、假阳性率(false positive rate, FPR)为横坐标绘制, AUC 值可通过对 ROC 曲线下各部分的面积求和而得。由定义可知 AUC 取值在 0.5 和 1.0 之间,越接近 1.0,模型可靠性越高。

Lipo、Freesolv 与 eSOL 数据集为回归任务,比较模型在不同数据集上的预测误差,采用均方根误差(root mean square error, RMSE)作为评价指标,由式(9)可知 RMSE 值由算法的预测值 y_i 和实际值 $\hat{y}_i$ 计算得来, RMSE 值越小,模型性能越好。

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (9)$$

3.3 对比算法

本文将提出的基于多通道边学习的图卷积网络与传统的图卷积网络、图注意力网络等分别做了相应的对比,几种对比算法的简介如下。

(1) 图卷积网络^[3]使用一阶近似的切比雪夫卷积实现谱域卷积。

(2) 图注意力网络^[4]将注意力机制引入基于空间域,通过一阶邻居节点的不同权重的表征更新节点特征。

(3) 随机森林(random forest, RF)算法^[15]通过训练多个决策树,生成模型,然后综合利用多个决策树进行分类。

(4) Weave^[15]模型原理类似于图卷积,Weave 特征化编码了局部化学环境和分子中原子的连通性,专门用于分子分析。

(5) 边增强图神经网络^[6]是通过多种边处理方法以学习边信息的图神经网络。

(6) 边信息最大化图神经网络^[7]通过最大化边特征与消息传递通道间的互信息来学习边的多维属性。

(7) 消息传递神经网络^[8]是通过关注节点之

间信息的传递,定义聚合函数而提出的一种通用的 GNN 框架。

3.4 结果分析

3.4.1 引文网络实验

对于引文网络数据集 Cora 和 Citeseer,任务为对论文进行分类,是节点分类任务。本文将整个数据集划分为训练集、验证集和测试集,参照文献[6]将数据集分为稀疏分割(Sparse)和密集分割(Dense)两种模式,稀疏分割划分为 5%训练、15%验证和 80%测试的 3 个子集;密集分割划分为 60%训练、20%验证和 20%测试的 3 个子集。对于类相似度估计,使用一个隐藏层有 512 个单元和另一个隐藏层有 32 个单元的双层 MLP。对于神经网络层,按文献[6]的参数设置,使用两层的卷积层,隐藏层设置输出维度为 64, dropout 率设为 0.5,对权值 W 进行 L2 正则化,权值衰减为 0.000 5,并采用指数线性单元(exponential linear unit, ELU)作为隐藏层的非线性激活。在本次实验中,将标准的 GCN^[3]与 GAT^[4]作为基础对照方法,并将文献[6]中最佳性能的 EGNN 加入对比试验。同时,为了研究每个功能的有效性,在实验中对模型进行了消融实验,EGCN (N)表示保留图去噪方法,EGCN (M)表示保留多通道边学习方法,EGCN (MN)表示使用完整的模型。本节总共测试了 6 个模型在 Cora 和 Citeseer 两个数据集的两种分割方法下的表现,每个模型分别运行 10 次,取其精度的均值和标准差。各模型基于引文网络数据集的结果和 EGCN 模型在引文网络数据集上消融实验的结果分别见表 3 和表 4,其中粗体数字表示不同模型得到的最好性能。

表 3 各模型基于引文网络数据集的结果

模型	Cora		Citeseer	
	Sparse	Dense	Sparse	Dense
GCN	0.729±0.008	0.720±1.002	0.692±0.007	0.753±0.004
GAT	0.755±1.001	0.790±1.000	0.695±0.005	0.749±0.005
EGNN	0.834±0.003	0.888±0.003	0.706±0.003	0.771±0.004
EGCN	0.849±0.005	0.883±0.005	0.747±0.009	0.791±0.004



表 4 EGCN 模型在引文网络数据集上消融实验的结果

模型	Cora		Citeseer	
	Sparse	Dense	Sparse	Dense
EGCN(N)	0.827±0.004	0.861±0.005	0.725±0.002	0.74.9±0.005
EGCN(M)	0.781±0.004	0.802±0.004	0.698±0.004	0.78.7±0.003
EGCN(MN)	0.849±0.005	0.883±0.005	0.747±0.009	0.791±0.004

分析表 3 和表 4 可知, EGCN 总体上比基础的 GCN 与 GAT 模型有更好的性能, 保留去噪方法对分类结果的准确率提升效果明显, 而保留多通道边学习方法提升较小, 可能是引用网络数据集的可用边特征较少, 对算法的影响有限, 密集分割条件下, EGCN 在 Cora 数据集上表现为第二优。这些结果表明 EGCN 模型在数据集较为简单的情况下仍能保持相当的性能。考虑到实际应用中可能存在实体标签缺失的问题, 如何提高模型在稀疏分割或半监督任务下的性能是未来研究的一个方向。

3.4.2 分子分析实验

本文在 4 个分子数据集: Tox21、Freesolv、eSOL 及 Lipo 上进行测试, 数据集均使用 RDKit 将分子转换为图数据, 将原子键信息编码为边特征矩阵。对于模型, 本文采用两层图卷积层, 每个通道的各图处理层的输出维度分别为 60、100, 一层微分池化层, 输出维度为 128, 一层全连接层, 输出维度为 64。对于分类任务, 全连接层使用 sigmoid 交叉熵损失函数输出 logits。回归任务的全连接层使用均方误差损失函数输出 logits。对权值 W 进行 L2

正则化, 权值衰减为 0.000 1, dropout 率设为 0.6, 采用指数线性单元作为隐含层的非线性激活。在分子分析任务中, 将实验中最佳性能的 EGNN^[6]模型加上 GCN^[3]、GAT^[4]、EIGNN^[7]和 MPNN^[8]作为参照, 另外也加上 RF 和 Weave 进行了测试。参照文献[6]的设置, 这些数据集以 80%、10%、10%的比例分为训练、验证和测试子集。上述模型每个运行 5 次, 取评价得分的平均值和标准偏差。对于分类任务 (Tox21), 采用 ROC 曲线下方面积作为评价指标; 对于回归任务 (Lipo 等), 采用均方根误差作为评价指标; 分子数据集下各神经网络的结果见表 5, 其中粗体数字表示不同模型得到的最好性能。

EGCN 在 Freesolv 与 eSOL 分子数据集上的消融实验结果见表 6, 验证了多通道学习在模型中的必要性。基于分子分析数据集的 EGCN 中共有 5 个通道, 分别代表分子图数据集中的 Atom Pair Type、Bond Order、Aromaticity、Conjugation 和 Ring Status 的属性。消融实验基于 Freesolv 与 eSOL 数据集, k 表示 EGCN 中保留的通道数量, 当 $k=1$ 时 EGCN 退化为传统 GCN 模型, 故将 GCN 引入对比测试。每个 EGCN _{k} 均为记录多次实验平均值加标准差, 例如, EGCN₄ 为保留 4 个通道, 具体来说就是每次删除一个通道, 运行 5 次, 然后删除其他数个通道重复该操作, 共记录 25 条数据, 取其均值。

表 5 分子数据集下各神经网络的结果

模型	Tox21 (AUC-ROC)		Lipo (RMSE)		Freesolv (RMSE)		eSOL (RMSE)	
	val	test	val	test	val	test	val	test
RF	0.78±0.01	0.75±0.02	0.87±0.02	0.86±0.03	1.98±0.82	1.62±0.14	1.27±0.07	1.18±0.11
Weave	0.79±0.02	0.80±0.01	0.88±0.11	0.89±0.09	1.35±0.25	1.37±0.19	0.90±0.08	0.85±0.03
GCN	0.82±0.01	0.83±0.02	0.70±0.03	0.68±0.03	1.24±0.26	1.14±0.10	0.76±0.08	0.80±0.17
GAT	0.83±0.02	0.83±0.01	0.85±0.02	0.84±0.01	1.73±0.25	1.45±0.25	0.86±0.02	1.01±0.03
MPNN	—	—	0.70±0.04	0.68±0.04	1.38±0.42	1.40±0.08	0.78±0.07	0.80±0.06
EIGNN	—	—	0.63±0.02	0.65±0.02	1.24±0.13	1.27±0.14	0.75±0.08	0.77±0.07
EGNN	0.82±0.01	0.82±0.01	0.79±0.02	0.75±0.01	1.09±0.08	1.01±0.12	0.70±0.03	0.72±0.02
EGCN	0.85±0.01	0.86±0.01	0.66±0.08	0.65±0.04	0.99±0.12	0.97±0.13	0.63±0.06	0.69±0.10

表 6 EGCN 在 Freesolv 与 eSOL 分子数据集上的消融实验结果

模型	Freesolv		eSOL	
	val	test	val	test
EGCN ₅	0.99±0.13	0.97±0.12	0.65±0.06	0.68±0.07
EGCN ₄	1.08±0.16	1.04±0.14	0.65±0.06	0.69±0.08
EGCN ₃	1.11±0.08	1.10±0.13	0.67±0.07	0.71±0.04
EGCN ₂	1.16±0.18	1.22±0.09	0.68±0.08	0.73±0.07
GCN	1.24±0.26	1.14±0.10	0.76±0.08	0.80±0.17

从表 6 中可以清楚地看到删除数个通道后的结果与原模型的实验结果, 比较分析后可知通道数量增加使模型性能有明显提升。

3.5 参数敏感性分析

在去噪处理这部分, 计算节点间的类相关度使用 MLP 从节点特征中学习获得类相关度矩阵, 再将类相关度值最低的边删除, 删除一定数量的噪声边后根据剩余边比例获得数据结果, EGCN 在 Cora 数据集上基于剩余边的比例的准确率如图 2 所示。

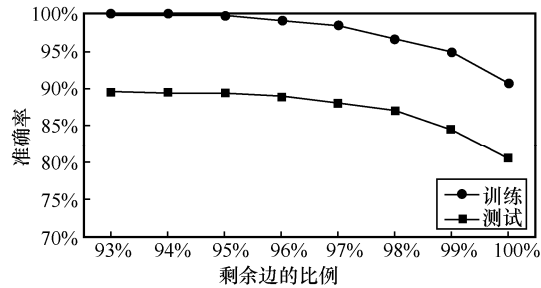


图 2 EGCN 在 Cora 数据集上基于剩余边的比例的准确率

从图 2 可以看出, 经过去噪处理后分类准确率提升明显, 说明图中的部分噪声边影响明显, 剔除噪声后 EGCN 能有效提升分类性能。被删除的边达到一定数量后, 继续删除图中的噪声边虽然可以获得略高的准确率, 但是提升幅度不明显, 并且训练准确率已达 100%, 表明过拟合效应严重。这里可以认为进一步删除边会造成数据失真以致训练过拟合, 所以选择保留 96% 的边作为后续实验的阈值。

在传统的 CNN^[1]中扩大感受野可以明显提高模型性能。图卷积神经网络在引入多通道学习后, 为了探究更大的感受野对模型训练是否有增益, 本节在一阶近似的切比雪夫卷积的基础上对模型进

行 k 阶切比雪夫多项式改造。高阶邻域对于图卷积神经网络意味着更大的感受野, 对于分子分析来说, 单个分子作为一个小型图, 对其使用高阶卷积可能并不适合。本节实验基于 Cora 与 Citeseer 数据集进行对比。这里将 k 的阶数设为 1~4, 基于 Cora 与 Citeseer 的 k 阶邻域对比实验结果如图 3 所示。

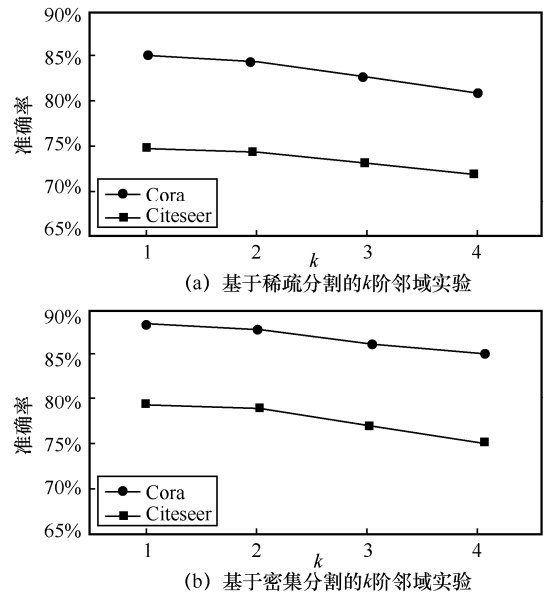
图 3 基于 Cora 与 Citeseer 的 k 阶邻域对比实验结果

图 3 中, 在 Cora 与 Citeseer 数据集上, 与 $k=1$ 相比, 当 $k=2$ 时准确率有略微下降, 当 $k=3$ 或 $k=4$ 时准确率均出现明显降低。基于这些结果分析可知, 对于图卷积网络, 每次消息传递来自其 k 阶邻域, 通常需要一个合适的感受野捕捉邻域信息, 直接邻域足以提供卷积所需的特征信息, 来自高阶邻域的大量信息会放大噪声的干扰, 并对节点预测任务造成负面影响。在实验中发现, 使用高阶切比雪夫卷积时会增大模型的计算量, 且明显增加训练的时间, 所以 EGCN 采用一阶近似的切比雪夫卷积。

4 结束语

本文首先总结了现有的图学习神经网络, 分析了各种图神经网络的优点、缺点。接着针对传统的图卷积网络存在的噪声问题与边信息利用不充分问题, 提出了使用 MLP 的去噪方法和基于多通道



边学习的图卷积神经网络。通过大量的实验证明多通道边学习用于图卷积的可行性和有效性。

基于多通道的边学习图卷积网络相较于传统的图神经网络方法有所改进，但是仍存在问题亟须解决，如加深网络层数产生的过度平滑、特征聚合过程中计算量较大等。在未来的工作中，将尝试更多的先进思想解决本文中 EGCN 存在的问题并加以改进。

参考文献：

- [1] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [2] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]//Proceedings of 30th Conference on Neural Information Processing Systems. New York: ACM Press, 2016.
- [3] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of 5th International Conference on Learning Representations(ICLR 2017). [S.l.:s.n.], 2017.
- [4] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]//Proceedings of 6th International Conference on Learning Representations(ICLR 2018). [S.l.:s.n.], 2018.
- [5] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and deep locally connected networks on graphs[C]//Proceedings of 2nd International Conference on Learning Representations. Netherlands: Elsevier Press, 2014.
- [6] GONG L, CHENG Q. Exploiting edge features for graph neural networks[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019.
- [7] CHEN P, LIU W, HSIEH C Y, et al. Utilizing edge features in graph neural networks via variational information maximization[J]. arXiv preprint arXiv: 1906. 05488, 2019.
- [8] 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.
- [9] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of 31st Conference on Neural Information Processing Systems. New York: ACM Press, 2017.
- [10] LI Q, HAN Z, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]//Proceedings of 32nd AAAI Conference on Artificial Intelligence. New York: ACM Press, 2018.
- [11] SUN K, LIN Z, ZHU Z. Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 5892-5899.
- [12] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features[C]//Proceedings of the European Conference on Computer Vision. New York: Springer Press, 2018: 132-149.
- [13] YANG H, MA K, CHENG J. Rethinking graph regularization for graph neural networks[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2021: 4573-4581.
- [14] YING Z, YOU J, MORRIS C, et al. Hierarchical graph representation learning with differentiable pooling[C]//Proceedings of 32nd Conference on Neural Information Processing Systems. New York: ACM Press, 2018.
- [15] WU Z, RAMSUNDAR B, FEINBERG E N, et al. MoleculeNet: a benchmark for molecular machine learning[J]. Chemical science, 2018, 9(2): 513-530.

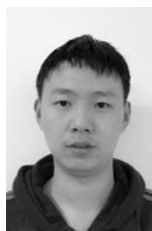
[作者简介]



杨帅（1996—），男，湖州师范学院信息工程学院硕士生，主要研究方向为机器学习、个性化推荐等。



王瑞琴（1979—），女，博士，湖州师范学院信息工程学院教授，主要研究方向为机器学习、数据挖掘、社交网络分析以及个性化推荐等。



马辉（1997—），男，湖州师范学院信息工程学院硕士生，主要研究方向为聚类、神经网络等。