



# 计算机科学

COMPUTER SCIENCE

## 基于图卷积神经网络的文本分类方法研究综述

檀莹莹, 王俊丽, 张超波

引用本文

檀莹莹, 王俊丽, 张超波. [基于图卷积神经网络的文本分类方法研究综述](#)[J]. 计算机科学, 2022, 49(8): 205-216.

TAN Ying-ying, WANG Jun-li, ZHANG Chao-bo. [Review of Text Classification Methods Based on Graph Convolutional Network](#)[J]. Computer Science, 2022, 49(8): 205-216.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

### [监督和半监督学习下的多标签分类综述](#)

Survey of Multi-label Classification Based on Supervised and Semi-supervised Learning

计算机科学, 2022, 49(8): 12-25. <https://doi.org/10.11896/jsjcx.210700111>

### [基于双图神经网络信息融合的文本分类方法](#)

Text Classification Method Based on Information Fusion of Dual-graph Neural Network

计算机科学, 2022, 49(8): 230-236. <https://doi.org/10.11896/jsjcx.210600042>

### [基于多智能体强化学习的端到端合作的自适应奖励方法](#)

Adaptive Reward Method for End-to-End Cooperation Based on Multi-agent Reinforcement Learning

计算机科学, 2022, 49(8): 247-256. <https://doi.org/10.11896/jsjcx.210700100>

### [基于可变形图卷积的点云表征学习](#)

Deformable Graph Convolutional Networks Based Point Cloud Representation Learning

计算机科学, 2022, 49(8): 273-278. <https://doi.org/10.11896/jsjcx.210900023>

### [面向文本分类的类别区分式通用对抗攻击方法](#)

Class Discriminative Universal Adversarial Attack for Text Classification

计算机科学, 2022, 49(8): 323-329. <https://doi.org/10.11896/jsjcx.220200077>

# 基于图卷积神经网络的文本分类方法研究综述

檀莹莹 王绚丽 张超波

嵌入式系统与计算教育部重点实验室(同济大学) 上海 201804

(tansyka@tongji.edu.cn)

**摘要** 文本分类是自然语言处理领域中常见的任务,机器学习和深度学习在该任务中已有较多研究并取得了很大进展,然而,这些传统方法只能处理欧氏空间的数据,不能完全有效地表达出文本的语义信息。为了打破传统的学习模式,诸多研究开始尝试用图表示文本中各实体间的丰富关系,并利用图卷积神经网络学习文本表示。文中对基于图卷积神经网络的文本分类方法进行了综述,首先概述了图卷积神经网络的背景与原理;其次,利用不同类型的图网络详细阐述了基于图卷积神经网络的文本分类方法,同时分析了图卷积神经网络在网络深度上的局限性,并介绍了深层网络在文本分类任务上的最新进展;最后,通过实验比较了各模型的性能,并探讨了该领域的难点与未来的发展方向。

**关键词** 文本分类;图卷积神经网络;图注意力网络;过平滑;非欧空间

中图分类号 TP391.1

## Review of Text Classification Methods Based on Graph Convolutional Network

TAN Ying-ying, WANG Jun-li and ZHANG Chao-bo

Key Laboratory of Embedded System and Service Computing(Tongji University), Ministry of Education, Shanghai 201804, China

**Abstract** Text classification is a common task in natural language processing, in which there are a lot of research and progress based on machine learning and deep learning. However, these traditional methods can only process Euclidean spatial data, and cannot express the semantic information of the document effectively. To break the traditional learning mode, many recent studies start to use graphs to represent complicated relationships among entities in the document, and explore graph convolutional neural network for text representation. This paper reviews the text classification methods based on graph convolutional network. Firstly, the background and principle of graph convolutional network are summarized. Then, text classification methods based on graph convolutional network are described in detail according to different types of graph-based networks. Meanwhile, it analyzes the limitation of graph convolutional network in the depth of networks, and introduces the latest developments of deep networks in text classification. Finally, the classification performance of models involved in this paper is compared through some experiments, and the challenges and future research direction in this field are discussed.

**Keywords** Text classification, Graph convolutional network, Graph attention network, Over-smoothing, Non-Euclidean space

### 1 引言

文本分类是自然语言处理领域中的一个重要且具有挑战性的应用分支,该任务是利用文本处理技术对一段给定的文本内容进行分析,进而对其进行分类,并为其分配一个或多个类别标签。常见的应用场景有文档主题分类<sup>[1]</sup>、垃圾邮件识别<sup>[2]</sup>、情感分析<sup>[3]</sup>等。文本分类的处理流程大致分为文本预处理、文本特征提取以及模型的构建等,其难点在于文本特征的提取,即如何在保留文本语义信息的前提下对文本信息进行向量化表示。

传统的文本分类方法主要分为两种,一种是基于机器

学习的文本分类方法<sup>[4]</sup>,另一种是基于深度学习的文本分类方法<sup>[5]</sup>。机器学习方法在发展初期常被用于解决文本分类问题,常见的分类器有支持向量机(Support Vector Machine, SVM)、朴素贝叶斯(Naive Bayes, NB)、K近邻算法(K-Nearest Neighbor, KNN)<sup>[6]</sup>等。其基本思路是利用特征工程进行文本表示,再通过分类器训练预测文本的类别。其中,特征工程主要包括特征提取和特征选择两种操作,例如,Joachims<sup>[7]</sup>提出将SVM模型应用于文本分类任务,其原理是采用词袋模型将文本映射成固定长度的向量,再通过信息增益准则选择特征以实现降维,最后利用SVM分类器对处理后的特征向量进行迭代训练,达到了不错的分类效果。除词袋模型外,

到稿日期:2021-08-09 返修日期:2021-10-19

基金项目:国家重点研发计划(2017YFA0700602);国家自然科学基金(61672381)

This work was supported by the National Key Research and Development Project of China(2017YFA0700602) and National Natural Science Foundation of China(61672381).

通信作者:王绚丽(junliwang@tongji.edu.cn)

$n$ -grams 模型<sup>[8]</sup>也常被用于特征提取。然而传统的特征提取方法存在很多弊端:其一,在大型语料库中容易产生较多的参数;其二,最终映射的向量过于稀疏。

相较于机器学习,深度学习的优势在于其利用自身的网络结构可以自动学习数据的特征表示,而无须人工进行特征工程。常用的模型有卷积神经网络(Convolutional Neural Networks, CNNs)<sup>[9]</sup>、循环神经网络(Recurrent Neural Networks, RNNs)<sup>[10]</sup>等。例如,TextCNN 模型<sup>[11]</sup>解决了句子级别的文本分类任务,该方法将句子视为由多个词向量构成的特征矩阵,并利用不同大小的卷积核对其进行一维卷积,随后通过最大池化层提取每个特征图中最重要的特征。而 Zhang 等<sup>[12]</sup>使用了字符级别的 CNN 结构,也取得了可观的分类效果。Liu 等<sup>[13]</sup>实现了 RNN 模型在文本分类上的应用。TextRNN 与 CNN 不同,TextRNN 模型将文本看成一段时间序列,并结合上下文信息学习文本的特征表示。HAN(Hierarchical Attention Networks)模型<sup>[14]</sup>在 TextRNN 模型的基础上引入注意力机制<sup>[15]</sup>,提出“单词-句子-文本”的文本层次化结构,并分别给予其不同的注意力权重,从而具有不同程度的表达能力。

然而,机器学习和传统的神经网络只能处理欧几里得数据。例如,图像和视频就属于欧氏空间的数据,前者是由像素

点构成的规则的二维数据,而后者则是时间序列上的图像数据。传统神经网络如 CNN 通常将这类数据作为输入,利用其平移不变性有效地获取数据的局部特征信息。对于文本这种非欧几里得数据而言,传统方法也是先将文本数据转化为由词、短语或句子等语言单位组成的二维向量矩阵,再利用网络模型加以训练,但是非网格化结构的数据常常限制了神经网络的表达能力,尤其是包含复杂语法结构的文本。在这样的背景下,图卷积神经网络(Graph Convolutional Networks, GCNs)<sup>[16]</sup>的出现引起了文本分类领域的广泛关注。图卷积神经网络以图的形式表示文本,图可以很直观地表达出文本中各元素间的丰富关系,并且图网络能够利用节点间的连接关系保留全局的图信息。近年来,大量研究利用该特点努力探索将图卷积神经网络应用于文本分类任务,以表达文本中的语义关系。

虽然目前已有一些综述文章分别围绕图卷积神经网络<sup>[17-19]</sup>和文本分类研究领域<sup>[20-21]</sup>展开,但目前缺少对这两者结合的综述。为此,本文总结并分析了近 5 年内图卷积神经网络在文本分类任务上的发展和趋势,并通过实验分析比较了它们的分类性能。所提及的主流方法均出自水平较高的人工智能期刊或会议,其相关模型信息如表 1 所列,包括发表年份、模型名称、发表的出处、具体的应用场景以及代码链接。

表 1 本文提及的基于图卷积模型的文本分类方法

Table 1 Graph convolutional network models for text classification in this paper

年份	模型	出处	应用	代码
2017	GCN	ICLR	引文分类	[22]
2017	GraphSage	NeurIPS	引文分类	[23]
2018	GAT	ICLR	引文分类	[24]
2018	FastGCN	ICLR	引文分类	[25]
2018	JKNet	ICML	引文分类	[26]
2019	SGC	ICML	引文分类、主题分类、情感分析	[27]
2019	Textgen	AAAI	主题分类、情感分析	[28]
2019	hConv-gPool-Net	WWW	主题分类、情感分析	[29]
2019	APPNP	ICLR	引文分类	[30]
2019	Text-level GCN	EMNLP	主题分类	[31]
2019	HGAT	EMNLP	主题分类、情感分析	[32]
2019	TD-GAT	EMNLP	情感分析	—
2020	TensorGCN	AAAI	主题分类、情感分析	[33]
2020	ReGNN	AAAI	主题分类、情感分析	—
2020	DropEdge	ICLR	引文分类	[34]
2020	GresNet	arXiv	引文分类	[35]
2020	PairNorm	ICLR	引文分类	[36]
2020	TG-Transformer	EMNLP	主题分类、情感分析	—
2020	D-GCN	COLING	情感分析	[37]
2020	HyperGAT	EMNLP	主题分类、情感分析	[38]
2020	TextING	ACL	主题分类、情感分析	[39]
2020	G-ATT	ACL	情感分析	[40]
2020	GRAND	NeurIPS	引文分类	[41]
2020	DGN	NeurIPS	引文分类	[42]
2020	GCNII	ICML	引文分类	[43]

注:“—”表示代码未开源

本文第 2 节详细介绍了图卷积神经网络的演变过程;第 3 节主要从 4 种不同类型的图模型分别阐述近年来图卷积神经网络在文本分类领域中的发展与改进;第 4 节主要概述图卷积神经网络的深度局限性以及改进方法;第 5 节主要通过实验对一些相关模型进行分析对比;第 6 节探

讨了未来的发展方向。

## 2 图卷积神经网络

图卷积神经网络,顾名思义,其灵感源于深度学习中的卷积神经网络 CNN。CNN 具有局部连接和权值共享两大

特性,其利用参数共享的卷积核对感受野进行卷积,从而达到提取特征的目的,现已成为图像处理领域中的一大利器。受到卷积网络在计算机视觉应用中的启发,研究者开始研究如何在图上构建卷积算子进行图嵌入。

图卷积神经网络(GCN)是由频谱卷积神经网络(Spectral CNN)和切比雪夫网络(ChebNet)演变而来的模型。频谱卷积神经网络<sup>[44]</sup>作为第一代图卷积神经网络,最早将卷积神经网络应用在图数据上。该模型在谱图理论<sup>[45]</sup>以及图信号处理<sup>[46]</sup>的基础上,根据卷积定理<sup>[47]</sup>实现节点信息与卷积核在谱域中的图卷积操作,以获取节点嵌入。假定图 $G$ 存在 $N$ 个节点,其图卷积的定义为节点特征 $x \in R^N$ 与卷积核 $g_\theta$ (卷积核参数 $\theta \in R^N$ )在傅立叶域中的乘积,如式(1)所示:

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

其中, $*$ 为卷积操作, $U$ 为归一化拉普拉斯图矩阵 $L$ 的特征向量矩阵, $\Lambda$ 是其特征值构成的对角矩阵,而卷积核 $g_\theta$ 可以被理解为关于矩阵 $\Lambda$ 的函数,即 $g_\theta(\Lambda)$ 。然而,该模型的计算成本较高,并且不具备局部连接的特性。因此,Defferrard等<sup>[48]</sup>提出了切比雪夫网络。其原理是在第一代图卷积神经网络的基础上,利用切比雪夫多项式的 $K$ 阶截断展开式来拟合卷积核 $g_\theta(\Lambda)$ <sup>[49]</sup>。其图卷积如式(2)所示:

$$g_\theta * x \approx \sum_{k=0}^K \theta_k' T_k(\tilde{L}) x \quad (2)$$

其中, $\tilde{L} = \frac{2}{\lambda_{\max}} L - I \in [-1, 1]$ , $\lambda_{\max}$ 表示 $L$ 最大的特征值; $\theta' \in R^{K+1}$ 为切比雪夫系数向量,而切比雪夫多项式被定义为 $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ ,其中 $T_0(x) = 1$ , $T_1(x) = x$ 。相较于Spectral CNN,该模型避免了拉普拉斯图算子 $L$ 的特征分解,从而降低了计算的成本与难度;同时由于该运算是关于拉普拉斯图矩阵的 $K$ 阶多项式,因此图卷积的结果仅取决于距离目标节点步长不超过 $K$ 阶的所有节点,这足以说明ChebNet具有很好的 $K$ -局部连接性。

在Spectral CNN和ChebNet的基础上,Kipf等<sup>[50]</sup>提出了GCN模型,以更加巧妙的方法简化了节点信息的传播规则。该模型利用切比雪夫多项式的一阶截断展开式来拟合卷积核,即 $K=1$ ,并通过减少参数来缓解过拟合以及简化卷积层的运算。假设 $\mathcal{L}^{(j)}$ 为第 $j$ 层GCN模型输出的节点表示,那么第 $j+1$ 层的节点表示如下:

$$\mathcal{L}^{(j+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \mathcal{L}^{(j)} W_j) \quad (3)$$

其中, $\mathcal{L}^{(0)} = X$ 为节点的初始特征; $\tilde{A} = A + I$ 为带自环的邻接矩阵, $I$ 为单位矩阵;度矩阵 $\tilde{D}$ 为对角矩阵,其中对角线上的元素 $\tilde{D}_i = \sum_j \tilde{A}_{ij}$ ; $\sigma(\cdot)$ 表示激活函数; $W_j$ 为可训练的权重矩阵。

一般而言,一张图由若干个节点和边构成,节点代表现实世界中的实体,边表示实体之间的关系。图卷积神经网络通过节点间边的连接,聚合邻居节点的信息,从而更新当前节点的特征表示。尽管该网络可以利用卷积操作提取节点和图的信息,但仍然存在一些局限性。为了进一步提升网络的泛化能力,相继出现了很多变体模型,如图注意力网络(Graph Attention Network, GAT)、图门控神经网络(Graph Gated Neural Network, GGNN)<sup>[51]</sup>、图变换网络(Graph Transformer Network, GTN)<sup>[52]</sup>等。

图卷积网络和图注意力网络的主要区别在于对边权的设计,前者以相等或预设的边权聚合邻居节点信息,但对于中心节点而言,不同的邻居应该具有不同的影响力。因此,图注意力网络引入注意力机制,以训练的方式学习相邻节点间的边权(即实体关系);同时,该网络增加了模型的可解释性。与图卷积网络相比,图门控神经网络不仅可以捕捉图中的递归或序列信息,还可以利用门控机制选择性地记忆来自邻居节点和本身的信息。然而,上述网络只能在多跳范围内进行节点间的交互,当图网络具有 $L$ 层时,中心节点只能与 $L$ 跳以内的所有节点实现消息传递,该范围之外的信息却无法获取。这意味着上述网络大多只关注局部信息,无法捕获长距离依赖关系或全局信息。受Transformer的启发,图变换网络将其应用于图数据上。该网络将输入图抽象成一张全连通图,图中每个节点都能与所有节点进行信息交互,有效地缓解了这一问题。表2总结并比较了这4种图网络的主要差异和优势。

表2 4种图网络的主要差异和优势

Table 2 Main differences and advantages of four graph-based networks

模型	核心组件	可解释性	优势
GCN	Graph-based CNN		可处理非欧空间的数据
GAT	Graph-based Attention	✓	能自适应地给予每个邻居节点不同的影响力
GGNN	Graph-based RNN		捕捉递归或序列信息,并选择性地记忆不同来源的信息
GTN	Graph-based Transformer	✓	能关注全局信息或长距离依赖关系

### 3 基于图卷积的文本分类方法

不同于图像、文本等,非网格化结构的数据常常使传统神经网络的表达能力受到限制,而GCN模型及其变体的出现极大地推动了自然语言处理领域的发展,尤其是文本分类任务。本节主要从不同的图模型类型分别阐述图卷积神经网络在文本分类任务中的发展与改进。

#### 3.1 图卷积网络

Textgcn模型<sup>[53]</sup>首次将图卷积网络应用于文本分类任务,该模型将单词和文本均视为节点,为整个语料库构建一个无向加权异质图,同时学习词嵌入和文本嵌入。在该模型中,词节点间的关系取决于词共现信息,其边权由点互信息算法(Pointwise Mutual Information, PMI)<sup>[54]</sup>计算得到。而文本与词节点间的边权定义为TF-IDF值<sup>[55]</sup>,用于评估单词在文本中的重要程度。该模型的邻接矩阵 $\tilde{A}$ (含自环)和PMI的计算式如下:

$$\tilde{A}_{ij} = \tilde{A}_{ij} = \begin{cases} PMI(i, j), & i, j \text{ are words, } PMI(i, j) > 0 \\ TFIDF_{ij}, & i \text{ is a word, } j \text{ is a doc} \\ 1, & i = j \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

$$PMI(i, j) = \log \frac{P(i, j)}{P(i)P(j)} \quad (5)$$

其中,  $\tilde{\mathbf{A}}$  为对称矩阵,  $P(i)$  表示单词  $i$  在语料库中出现的概率,  $P(i, j)$  表示在语料库中单词  $i$  和单词  $j$  同时出现的概率。如图 1 所示, 输入图中包含的单词 (Word) 和文本 (Doc) 两类节点, Textgcn 模型通过两层 GCN 模型<sup>[50]</sup> 的学习, 获取到节点的最终表示  $R(\cdot)$ , 以实现多分类任务。与传统方法 (如 TextRNN<sup>[33]</sup>) 相比, 该模型可以捕捉到全局语义信息。除词共现关系外, TensorGCN 模型<sup>[56]</sup> 还引入了长短期记忆网络

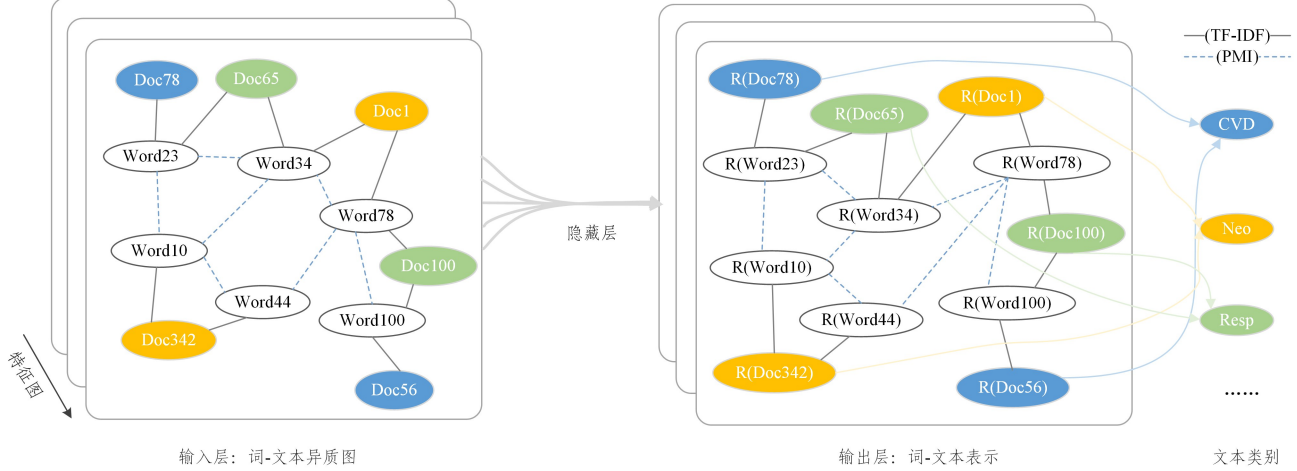


图 1 Textgcn 模型的流程示意图<sup>[53]</sup>

Fig. 1 Schematic diagram of Textgcn<sup>[53]</sup>

然而, 现有模型均为整个语料库构建输入图, 却忽略了文本内的单词交互。因此, Huang 等<sup>[58]</sup> 提出了 Text-level GCN 模型, 该模型为每一篇文本构建一张有向图, 并设计出全局共享的节点特征矩阵和边权矩阵, 如图 2 所示。该模型通过消息传递机制<sup>[59]</sup> 不断更新节点表示和边权, 以实现信息的有向传递。Text-level GCN 模型可以有效地解决高内存消耗的问题, 且对新样本具有较好的泛化能力。此外, 为了获取上下文关系, Gao 等<sup>[60]</sup> 结合 GCN 图卷积与 CNN 一维卷积, 提出了一种混合卷积操作 (hConv), 该操作不仅能快速增大感受野, 还能提取文本的语序信息。

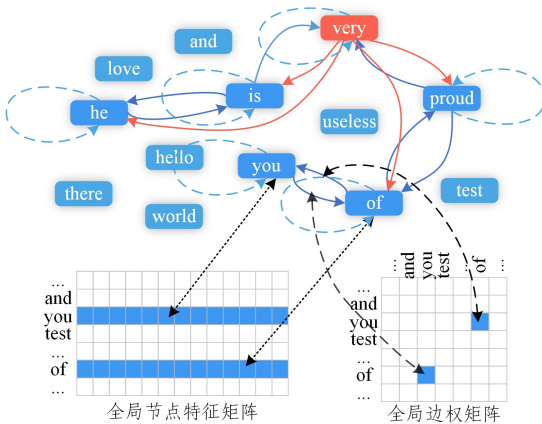


图 2 Text-level GCN 模型的输入图<sup>[58]</sup>

Fig. 2 Input graph of Text-level GCN<sup>[58]</sup>

根据方法适用的网络类型, 将本节提到的方法进行总结, 如表 3 所列, 包括节点类型、边的关系属性、边权状态以及转换的任务类型。

(Long Short-Term Memory, LSTM) 和单词间的句法依赖, 用于表达单词间的语义与句法关系。该模型根据这 3 种关系分别构建 3 种异质图, 并且采用图内和图间两种信息传播方式, 分别用于单图中节点的信息聚合以及协调各图之间的异构信息。SGC 模型<sup>[57]</sup> 提出消除隐藏层之间的激活操作, 将中间过程转换为简单的线性变换, 以减少模型的复杂度和冗余计算。

表 3 图卷积网络方法总结

Table 3 Summary of methods based on graph convolutional network

模型	节点类型	边关系属性	边权状态	任务
Textgcn <sup>[53]</sup>	文档+词	文档-词包含关系、词共现关系	固定不变	节点分类
TensorGCN <sup>[56]</sup>	文档+词	混合多元关系	固定不变	节点分类
SGC <sup>[57]</sup>	文档+词	文档-词包含关系、词共现关系	固定不变	节点分类
Text-level GCN <sup>[58]</sup>	词	词共现关系	动态更新	图分类
hConv-gPool-Net <sup>[60]</sup>	词	词共现关系	固定不变	图分类

### 3.2 图注意力网络

在 GCN 模型<sup>[50]</sup> 中, 每个邻居节点都具有同等的重要性, 即拥有相等的权重。与 GCN 模型不同, GAT 模型<sup>[61]</sup> 认为不同的节点具有不同的影响力, 并实现了引文分类任务。该模型的输入图是引文网络, 图中的节点对应学术论文, 其特征由词袋模型进行初始化, 而节点间的边对应着论文间的引用关系。其采用一层前馈神经网络来计算节点间注意力分数, 而该分数作为节点间的邻接权值, 表达了它们之间的重要程度。如图 3 所示, 在 GAT 模型中, 左图表示以节点  $i$  和其邻居节点  $j \in N_i$  的特征作为输入, 计算其注意力分数并归一化处理; 而右图表示 GAT 模型以当前节点  $h_i$  为中心, 利用注意力分数加权求和, 实现邻居节点的信息聚合, 从而更新自身的中间状态  $h_i'$ 。其具体计算式如下:

$$h_i' = \sigma \left( \sum_{j \in N_i} \alpha_{ij} \mathbf{W} h_j \right) \quad (6)$$

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [\mathbf{W} h_i \parallel \mathbf{W} h_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(a^T [\mathbf{W} h_i \parallel \mathbf{W} h_k]))} \quad (7)$$

其中,  $\sigma(\cdot)$  和  $\text{LeakyReLU}(\cdot)$  均为激活函数,  $\parallel$  为拼接操作,  $N_i$  表示节点  $i$  的邻居集合,  $h_i$  和  $h_i'$  分别表示节点  $i$  当前和更新后的特征表示, 而  $\mathbf{W}$  为权重矩阵,  $\alpha_{ij}$  是节点  $i, j$  间的注意力分数,  $a$  为上下文向量。

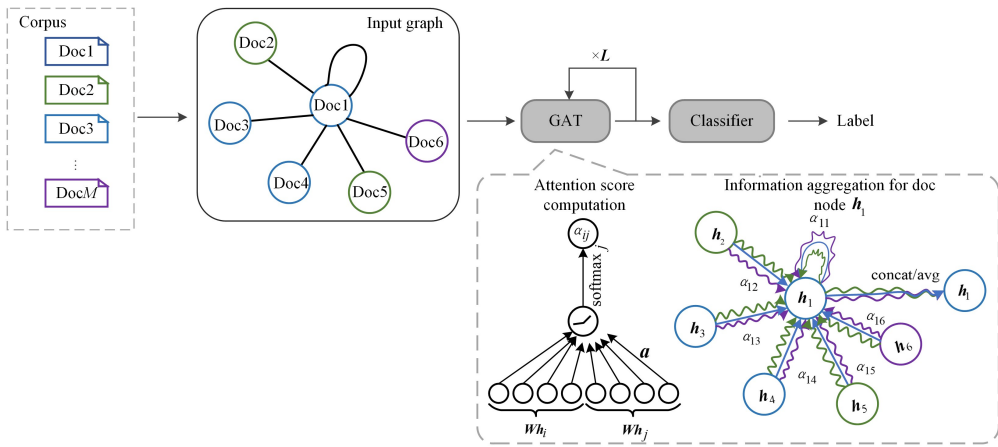


图3 GAT模型的信息聚合机制

Fig. 3 Information aggregation mechanism of GAT

HGAT (Heterogeneous Graph Attention Networks) 模型<sup>[62]</sup>利用图注意力网络实现了短文本分类,解决了短文本中语义稀疏、存在歧义的问题。该模型为短文本语料库构建了一张异构图,图中包含短文档、概念实体以及相关主题3种

类型的节点。对于中心节点而言,邻居节点的各种类型可能会对其产生不同的作用,并且相同类型的不同节点也具有不同的重要性。如图4所示,文档节点*i*存在主题和概念实体两种不同类型的邻居节点 $N_i$ 。

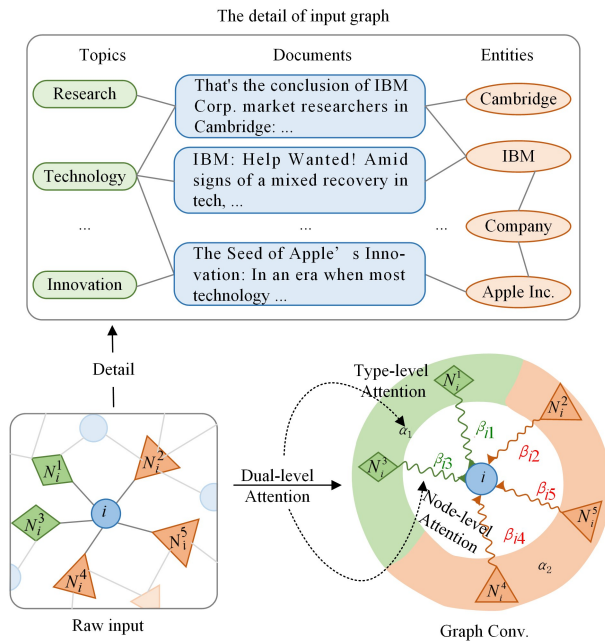


图4 具有双层注意力的HGAT模型<sup>[62]</sup>

Fig. 4 HGAT model with dual-level attention

针对不同的节点类型和节点,该模型提出具有类型级别和节点级别的双层注意力网络,同时捕捉节点类型和其本身所带来的不同影响,而Yuan等<sup>[63]</sup>采用注意力网络处理方面级的情感分析任务,即G-ATT模型,该模型为每个句子构建一棵无向依赖树,用于表达单词的角色和语法依赖,其利用图注意力衡量节点的重要性,同时采用一维卷积获取局部的上下文信息,最终将这两种信息通过记忆网络加以融合。

传统的GCN模型仅应用于无向图,与之相比,GAT模型可以应用于任何图结构。例如,Chen等<sup>[64]</sup>为每篇文本构建一个有向的依存句法图,图上的节点为文本中的单词,单词间的连接为有向边,代表着左边、右边以及自环的相对位置关系,不同位置方向的单词可能会产生不同的影响。因此,该模型引入位置建模和图注意力机制,以获取文本中的上下文

信息。Ding等<sup>[65]</sup>采用超图的思想处理文本中“单词-句子-文本”的层次化结构,即HyperGAT模型。相较于简单图中节点间成对的连接关系,超图往往更能直观地表达出更高阶的节点联系。该模型将每篇文档表示成一张超图,而将句子视作超边,超边上的节点对应着出现在句子中的单词;同时,HyperGAT模型还引入文本主题模型<sup>[66]</sup>,将文本主题也看作超边,其上节点均为与该主题相关的单词。如图5所示,该模型在超图上使用双层图注意力网络:节点注意力和超边注意力。节点注意力通过计算节点对超边的注意力分数,突出对超边更重要的节点信息,从而获取超边嵌入;而超边注意力则强调更有意义、更具有信息量的超边特征,通过聚合这些信息来更新节点特征,这样不仅可以捕捉到高阶的单词间的互动,还可以突出节点在不同粒度下的关键信息。

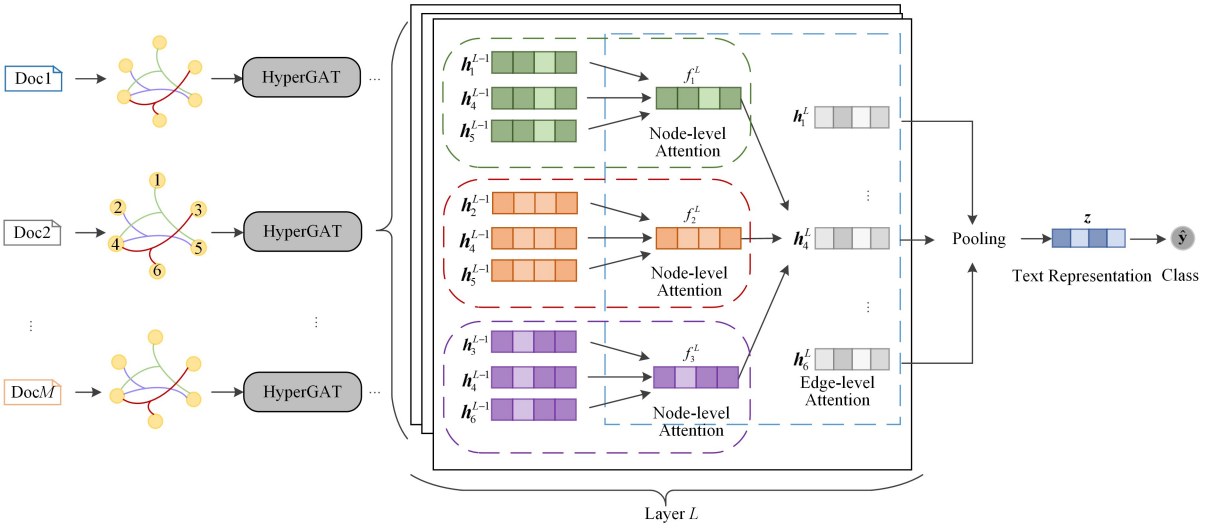


图 5 HyperGAT 模型的结构<sup>[65]</sup>

Fig. 5 Architecture of HyperGAT<sup>[65]</sup>

根据方法适用的网络类型,将本节提到的方法进行总结,如表 4 所列,包括节点类型、边的关系属性、输入图类型以及转换的任务类型。

表 4 图注意力网络方法总结

Table 4 Summary of methods based on graph attention network

模型	节点类型	边关系属性	输入图	任务
GAT <sup>[61]</sup>	文档	引用关系	无向图	节点分类
HGAT <sup>[62]</sup>	文档+词+主题	混合多元关系	无向图	节点分类
G-ATT <sup>[63]</sup>	词	句法依赖关系	无向图	图分类
D-GCN <sup>[64]</sup>	词	句法依赖关系	有向图	图分类
HyperGAT <sup>[65]</sup>	词	句子+主题	超图	图分类

### 3.3 图门控网络

现有研究也将图门控神经网络应用于文本分类任务中。例如,TextING 模型<sup>[67]</sup>采用门控图卷积网络解决文本分类任务,该模型提出每篇文档都应该构建自己特有的图结构,因此该模型为每篇文档构建一张输入图,将文本分类视为图分类任务。如图 6 所示,该模型通过图卷积网络聚合邻居信息,并利用门控循环单元(Gated Recurrent Unit,GRU)更新自身的当前状态。在读出层中,其利用简单的注意力机制和池化操作获取最终的文档表示。与现有模型相比,TextING 模型更关注单词在每篇文档中的交互,因此可以捕捉到细粒度的上下文关系。

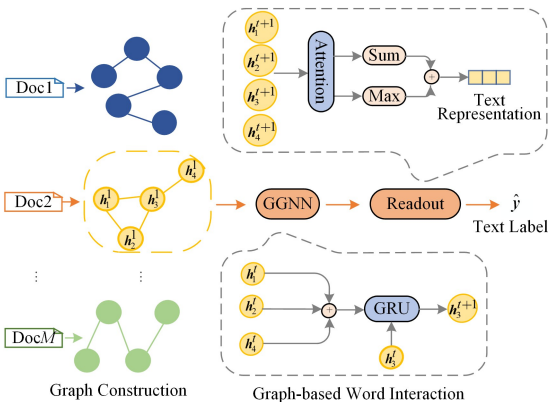


图 6 用于文本分类的 TextING 模型<sup>[67]</sup>

Fig. 6 TextING model for text classification

然而,语料库级和文档级的输入图各有优势,前者可以捕捉全局信息,而后者更关注单词间的局部交互,如何结合两者的优势成为了研究者思考的问题。受到 S-LSTM 模型<sup>[68]</sup>的启发,ReGNN 模型<sup>[69]</sup>在每张输入图上引入一个全局节点(即图节点),并使用门控图注意力网络实现文本分类。如图 7 所示,图节点  $g$  与所有的词节点  $h$  均进行交互。该模型以节点上一状态  $h^{l-1}$ 、初始状态  $x$ 、图节点状态  $g^{l-1}$  以及邻居信息  $N^{l-1}$  作为输入,利用 LSTM 模型控制信息的流动,从而更新词嵌入。与 S-LSTM 模型的不同之处在于,该模型通过图注意力机制聚合当前节点的所有邻居信息,而 S-LSTM 仅通过拼接的方式聚合左右两个邻居节点的信息。此外,考虑到各文档中的语法结构,TD-GAT 模型<sup>[70]</sup>利用单词之间的语法相关性,为每个句子构建句法依赖树,其通过多头图注意力网络更新节点的中间状态,并使用 LSTM 模型明确目标信息,减少噪声信息。

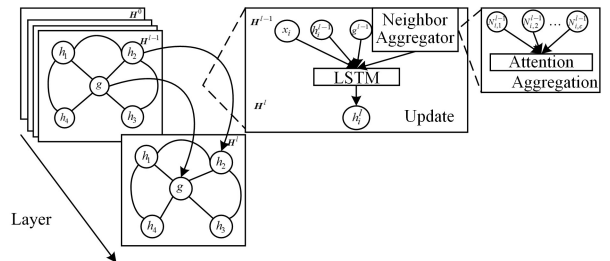


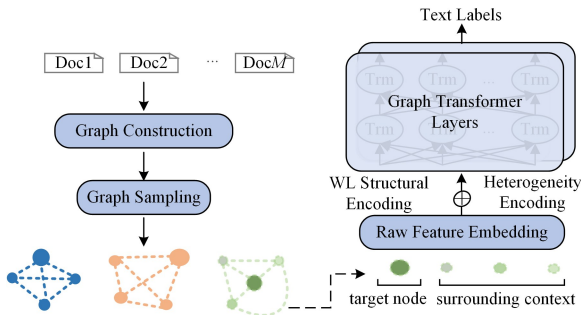
图 7 ReGNN 的示意图<sup>[69]</sup>

Fig. 7 Schematic diagram of ReGNN<sup>[69]</sup>

### 3.4 图变换网络

Transformer 系列模型在自然语言处理领域中取得了突破性的进展,因此很多研究者尝试将该模型应用于图数据,以解决文本分类任务。其中以 TG-Transformer 模型<sup>[71]</sup>为代表,其利用图变换网络解决了文本分类任务中的两个问题:1) 现有的研究常将文本图构建成为异质图,但卷积过程中却没有区分节点间的异构性;2) 每次都需要训练整张图,计算和存储成本高。因此,如图 8 所示,TG-Transformer 模型提出了子图采样策略,在文档-词的异质图上为每个目标文档节点

采样一定数量的邻居词节点,构成以文档节点为中心的子图,这种做法将原来的节点分类转为子图分类,并且子图的小批量处理可以大大节省内存消耗和计算成本。同时,该模型在各子图上利用图变换网络聚合信息,学习目标节点表示,以实现文本分类。

图8 TG-Transformer的示意图<sup>[71]</sup>Fig. 8 Schematic diagram of TG-Transformer<sup>[71]</sup>

#### 4 深度局限性及改进

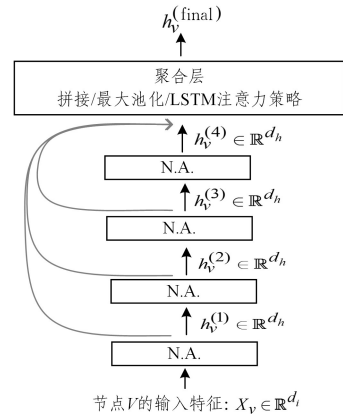
图卷积模型在发展初期多为浅层网络,采用2~4层的网络就可实现很好的分类效果,但随着网络结构进一步加深,其性能会出现明显下降的情况。据观察分析,这种性能衰退的现象主要和两个因素有关,分别是过拟合和过平滑<sup>[72]</sup>。一方面,加深网络会引入大量额外的参数,这可能会削弱模型的泛化能力,从而造成过拟合问题;另一方面,图卷积神经网络的本质是一种聚合器,通过聚合邻居节点的信息来更新当前节点的中间状态,当卷积层不断叠加,其聚合的邻域也不断扩大,使得每个节点获得的信息出现重复,最终导致所有节点的特征表示收敛至不可区分的向量表示,从而引发过平滑问题。为了构建深层图卷积网络模型,如何防止过拟合和过平滑已经成为目前急需解决的问题。

为了缓解上述问题,Klicpera等<sup>[73]</sup>根据GCN模型和PageRank模型<sup>[74]</sup>间的联系,提出了基于个性化PageRank的传播方案,确保PageRank分数能对每个根节点的局部邻域进行编码。DropEdge模型<sup>[75]</sup>在每次训练的过程中以一定比例随机丢弃输入图中的边,而GRAND模型<sup>[76]</sup>选择随机丢弃节点的特征属性,为节点特征矩阵增加一定的扰动性。从本质上看,两种方法均为数据扩充和增强技术,通过增强输入数据的随机性和多样性防止过拟合和过平滑。基于ResNet模型<sup>[77]</sup>中残差连接的思想,JKNet模型<sup>[78]</sup>以层级聚合的方式构建深层图卷积模型。如图9所示,将多个中间层状态传递到最后一层,并提出拼接、最大池化以及基于LSTM的注意力策略3种聚合方式。而GResNet模型<sup>[79]</sup>考虑到图的拓扑结构,引入了4种不同的图残差项,如表5所列。Chen等<sup>[80]</sup>也深受启发并采用初始残差和恒等映射两种简单技术来缓解过平滑问题,即GCNII模型。计算式如式(8)所示:

$$\mathcal{L}^{(j+1)} = \sigma((1 - \alpha_j) \hat{\mathbf{A}} \mathbf{L}^{(j)} + \alpha_j \mathcal{L}^{(0)}) ((1 - \beta_j) \mathbf{I} + \beta_j \mathbf{W}_j) \quad (8)$$

其中, $\sigma(\cdot)$ 为激活函数; $\alpha_j$ 与 $\beta_j$ 均为超参数; $\mathbf{W}_j$ 为可学习的权重矩阵; $\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}}$ 表示归一化后的邻接矩阵, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$

为带自环的邻接矩阵, $\mathbf{I}$ 为单位矩阵;度矩阵 $\tilde{\mathbf{D}}$ 为对角矩阵,其对角线上的元素 $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ ;  $\mathcal{L}^{(j)}$ 表示第 $j$ 层的节点表示,而 $\mathcal{L}^{(0)} = \mathbf{X}$ 为节点的初始表示。

图9 JKNet示意图<sup>[78]</sup>Fig. 9 Schematic diagram of JKNet<sup>[78]</sup>表5 GResNet模型的图残差项<sup>[79]</sup>Table 5 Summary of graph residual terms in GResNet<sup>[79]</sup>

名称	残差项	描述
naive residual	$\mathcal{L}^{(j-1)}$	假设各节点残差项之间相互独立,且仅由当前状态决定
graph-naive residual	$\hat{\mathbf{A}} \mathcal{L}^{(j-1)}$	各节点残差项之间根据网络结构进行关联,且由当前状态决定
raw residual	$\mathbf{X}$	假设各节点残差项之间相互独立,且仅由初始的输入特征决定
graph-raw residual	$\hat{\mathbf{A}} \mathbf{X}$	各节点残差项之间根据网络结构进行关联,且由初始的输入特征决定

相比传统的GCN,GCNII模型有两点改进:1)利用初始残差连接,将平滑表示 $\hat{\mathbf{A}} \mathcal{L}^{(j)}$ 与初始表示 $\mathcal{L}^{(0)}$ 相结合,确保每个节点的最终表示至少有一定比例的 $\alpha_j$ 来自输入层的特征;2)在每层的权重矩阵 $\mathbf{W}_j$ 中增加恒等映射 $\mathbf{I}$ ,确保深层GCNII模型至少可实现其浅层网络的性能效果,并能减少信息损失。

此外,还有很多方法是基于节点对的距离来衡量过平滑问题<sup>[81-82]</sup>。理想情况下,相连的节点对应该平滑化,即在嵌入空间中节点对相互靠近;而不相连的节点对应该避免过平滑,即节点对应该相互远离。Zhao等<sup>[83]</sup>提出的PairNorm方法无须引入额外参数。如图10所示,该网络结构的每层操作均由图卷积和PairNorm共同组成。PairNorm通过中心化和缩放两种简单操作,以确保图中所有节点对的总平方特征距离在每层之间保持恒定不变。由于图卷积操作可以使相连的节点对具有更相似的特征表示,导致所有相连节点对的总距离缩短,因此基于所有节点对的总距离恒定不变,不相连节点对的总距离将会增大,即节点相互远离,以达到防止过平滑的目的。然而这些方法往往忽略了全局图的结构信息,并且未能充分利用类别的特性。因此,Zhou等<sup>[84]</sup>提出可微分的群组归一化方法,即DGN方法。其不再以节点的互连关系作为研究对象,而是通过学习特定的群组分配策略,将具有相似嵌入表示的节点以较高的概率映射到同一个群组中;同时,该方法还需对每个群组进行图归一化处理,通过调整平均值和方差

使同组内节点的嵌入分布相似,并分散不同组的节点分布。

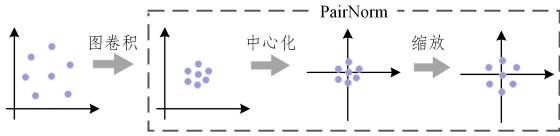


图 10 PairNorm 中心化和缩放操作的示意图<sup>[83]</sup>

Fig. 10 Illustration of Pairnorm(centering and rescaling steps)<sup>[83]</sup>

## 5 基于图卷积的文本分类实验

为了评估图卷积神经网络在文本分类任务中的性能,本节采用主题分类数据集和引文网络两种不同的数据集进行对比,以提供更客观的评价。数据集的相关信息如表 6 所列。

**主题分类数据集:**R8, R52 和 Ohsumed 拥有丰富的语义和句法关系,可用于文本分类任务。根据不同的模型,文本分类任务可看作节点分类或图分类任务。R8 和 R52 均为路透社新闻数据集 Reuters 的子集,其每篇文本对应一个新闻主题;而 Ohsumed 来源于医学文献数据库 Medline,其包含 7400 篇关于心血管疾病的医学摘要,且每篇摘要的标签来自于 23 种疾病类别。

**引文网络:**Cora, Citeseer 和 Pubmed<sup>[85]</sup>作为引文网络数据集,可用于半监督的节点分类任务。在该网络中,节点和边分别表示论文和引用关系,节点的特征为论文的词袋表示,其标签为该论文所属的学术主题。

表 6 数据集信息

Table 6 Summary statistics of datasets

Category	Dataset	# Docs	# Words	Max. Length	Avg. Length	# Classes
主题分类数据集	R8	7 674	7 688	520	65.72	8
	R52	9 100	8 892	612	69.82	52
	Ohsumed	7 400	14 157	476	135.82	23
引文网络	Cora	1	2 708	5 429	1 433	7
	Citeseer	1	3 327	4 732	3 703	6
	Pubmed	1	19 717	44 338	500	3

### 5.1 文本分类性能分析

图卷积神经网络通常为浅层网络,仅采用 2~4 层图卷积层就可解决分类问题。因此,本节的文本分类实验均采用浅层图卷积网络,并将其分别应用于主题分类数据集和引文网络数据集上。

#### (1) 主题分类

传统的文本分类方法分为两种,一种为基于机器学习的文本分类方法,另一种是基于深度学习的文本分类方法。为了更客观地评价图卷积模型的性能,选用 3 种传统的文本分类方法作为基线模型与其进行对比分析。对于图卷积神经网络,该实验选用 Textgcn, SGC, Text-level GCN 和 TensorGCN 这 4 种模型实现主题分类任务。Textgcn 和 SGC 均为整个语料库构建一张无向异质图,文本和词均被视为节点,即文本分类转为节点分类任务;Text-level GCN 为每篇文本构建一张有向图,而节点仅表示文本中的词,属于图分类任务;而 TensorGCN 为整个语料库构建 3 种不同的图,分别基于语义、句法以及上下文信息,其中每张图都是包含文本和词

两类节点的无向异质图。实验以每个模型运行 10 次后的平均准确率为性能的衡量指标,如表 7 所列。

表 7 文本分类准确率的比较

Table 7 Comparison of text classification accuracy (单位:%)

模型	数据集		
	R8	R52	Ohsumed
SVM	95.2	88.2	48.3
CNN <sup>[53]</sup>	95.7	87.6	58.4
LSTM <sup>[53]</sup>	93.7	85.5	41.1
Textgcn	97.0	92.8	68.5
SGC	96.1	93.2	65.9
Text-level GCN <sup>[58]</sup>	97.8	94.6	69.4
TensorGCN <sup>[56]</sup>	98.0	95.1	70.1

机器学习方法主要依赖于人工提取特征,往往会产生具有高维、高稀疏性的特征矩阵,造成较大的冗余甚至过拟合。传统的神经网络方法虽然能在局部连续的词序列中获取到语义信息,但却忽视了语料库中的全局信息;而图卷积神经网络模型不仅具有自动提取特征的能力,还可利用自身结构的优势保持全局图信息。与传统的文本分类方法相比,图卷积神经网络在 3 种数据集上的分类准确率更高。同时,Text-level GCN 和 TensorGCN 的分类效果优于其他图卷积模型,原因可能是:1)两种模型均考虑了文本语序的重要性;2)两种模型均采用预训练后的词嵌入作为节点的特征表示,有利于下游的分类任务。

#### (2) 引文分类

这部分选取 GCN, GraphSage<sup>[86]</sup>, FastGCN<sup>[87]</sup> 以及 SGC 这 4 种模型在 3 种引文网络上实现文本分类任务,每个模型将引文网络转成无向图,并对其进行训练集/验证集/测试集的划分<sup>[88]</sup>。该实验以准确率和训练时间两个维度作为评价指标,所有实验结果均为每个模型运行 10 次后的平均值。如表 8 所列, FastGCN 在 3 种引文数据集中分类准确率最高,但在训练时间上并不占优势。GraphSage 通过固定每个节点的邻居大小,实现了以节点为中心的批量训练,然而其节点的邻域随层数的增加呈指数增长,导致该模型的计算成本和时间复杂度也随之增大。因此,相较于其他模型而言, GraphSage 的训练时间最长,而 SGC 模型的训练时间最短,进一步证实了消除了隐藏层间的非线性变换可以减少冗余计算。

表 8 文本分类准确率和训练时间的比较

Table 8 Comparison of text classification accuracy and training time

数据集	模型	准确率/%	训练时间/ (10 <sup>-2</sup> s/epoch)
Cora	GCN	79.8	0.78
	GraphSage	80.3	215.28
	FastGCN	85.6	12.55
	SGC	81.1	0.18
Citeseer	GCN	71.3	0.81
	GraphSage	70.4	277.42
	FastGCN	76.6	21.70
	SGC	69.8	0.20
Pubmed	GCN	78.8	0.84
	GraphSage	77.9	115.62
	FastGCN	87.5	235.85
	SGC	78.5	0.20

## 5.2 网络深度性能分析

图卷积模型具有网络深度的局限性,因此很多研究正尝试将其拓展到深层网络。本节实验采用引文网络数据集分析了各深层图卷积模型在不同网络深度中的文本分类性能,结果如表9所列。实验结果表明,随着网络层数的加深,GCN的分类准确率急速下降,而其他5种模型却能延缓这种性能衰退的现象,尤其是GCNII模型。通过加深模型的网络深度,大多模型可以超越两层GCN的分类效果,甚至,Cora的分类准确率最高可提升5.8%,Citeseer最高可提升8.4%,Pubmed最高可提升5.1%。

表9 各模型在不同网络深度中的文本分类准确率

Table 9 Text classification accuracy of each model in different network depths (单位:%)

数据集	模型	层数				
		2	4	8	16	32
Cora	GCN	79.5	14.8	10.8	15.9	14.7
	DropEdge	81.9	80.0	77.8	74.3	35.3
	GCNII	81.6	<b>83.2</b>	<b>84.3</b>	<b>85.3</b>	<b>84.8</b>
	GResNet	80.2	80.1	79.6	79.2	75.1
	PairNorm	80.1	76.2	74.9	73.4	41.5
	DGN	<b>82.3</b>	81.1	75.2	73.2	71.0
Citeseer	GCN	65.8	42.4	14.9	22.7	22.9
	DropEdge	<b>70.6</b>	<b>69.3</b>	57.8	42.6	36.6
	GCNII	69.3	69.0	<b>71.7</b>	<b>73.7</b>	<b>74.2</b>
	GResNet	64.4	60.5	56.7	44.6	57.8
	PairNorm	65.2	66.0	60.8	61.8	44.9
	DGN	69.1	59.4	55.1	54.2	47.9
Pubmed	GCN	75.2	60.6	30.9	34.0	24.5
	DropEdge	<b>79.2</b>	75.8	74.5	67.2	40.8
	GCNII	77.9	<b>76.7</b>	<b>79.6</b>	<b>80.3</b>	<b>79.9</b>
	GResNet	76.1	75.2	76.8	77.0	72.9
	PairNorm	76.5	76.5	78.1	78.2	69.5
	DGN	<b>79.2</b>	76.5	75.6	75.0	75.5

**结束语** 面对互联网上海量的文本信息,文本分类已成为自然语言处理领域中的研究热点。尽管机器学习和传统神经网络在该任务中已取得较大进展,但仍存在弊端。机器学习提取特征的计算成本较高,而传统神经网络也只能提取局部的词序列信息,相比之下,基于图卷积神经网络的文本分类方法具有一定的优势,其不仅可以用图表达文本中复杂的语义关系,而且还可以捕获全局的图信息。

本文在分析传统文本分类方法不足的基础上,对基于图卷积神经网络的文本分类方法进行了较为全面的综述。文中以早期谱图卷积 Spectral CNN 和切比雪夫网络 ChebNet 作为出发点,引出 GCN 的相关原理;根据不同类型的图模型,对基于图卷积神经网络的文本分类方法进行归纳和分类,包括基于图卷积网络、基于图注意力网络、基于图门控网络以及基于图变换网络,并对比其优缺点;分析图卷积神经网络在网络深度上的局限性,并梳理了深层网络在文本分类任务上的最新进展;同时,在6种不同的文本数据集上进行分类实验,分析各模型的性能。

综上所述,近年来,图卷积神经网络在文本分类任务中已取得较大进展,但该领域依然面临着一些挑战,具有极大的发展空间。

### (1) 基于大规模图的研究

在处理大型语料库的过程中,其复杂的语义关系常常会导致输入图的规模过大,而图卷积神经网络大多采用全批量的训练方式,需要将所有节点的中间状态全部放入内存中,造成了很高的内存消耗。尽管已有模型提出了小批量处理的方法,但其计算成本依然较高,因此,如何处理大型文本数据是图卷积神经网络需要解决的难题。

### (2) 基于深层图卷积网络的研究

基于谱域的图卷积神经网络本质上是拉普拉斯平滑的一种特殊形式<sup>[72]</sup>,因此随着网络加深,其将面临过平滑的问题。尽管现有的深层网络可以缓解该问题,但大多也只能暂时延缓过平滑的速度,并不能从根本意义上提升模型的性能。同时,目前基于深层图卷积网络的文本分类方法仅应用于无向图上,而没有泛化到其他类型的输入图上。因此,如何在文本分类任务中构建有效的深层图卷积模型也是值得思考的方向。

### (3) 基于动态图卷积网络的研究

目前基于图卷积神经网络的文本分类方法大多定位在静态图上,其节点和边是固定不变的。然而,一些实际情景无法离开动态图。例如在引文网络中,若某一领域突然引起广泛关注,则相关论文也会被频繁发表,这意味着一篇文章的引用量或者其他信息会随着时间发生改变。在情感分析任务中,用户对同种产品的好感度也会随着时间的沉淀和市场的更新换代而发生变化,因此,基于动态图的图卷积神经网络在文本分类任务中也是值得深究的领域。

## 参考文献

- [1] WANG J, WANG Z, ZHANG D, et al. Combining knowledge with deep convolutional neural networks for short text classification[C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. 2017:2915-2921.
- [2] HU X J, LIU L, QIU N J. A Novel Spam Categorization Algorithm Based on Active Learning Method and Negative Selection Algorithm[J]. Dian Zi Xue Bao/Acta Electronica Sinica, 2018, 46(1):203-209.
- [3] MAAS A, DALY R E, PHAM P T, et al. Learning word vectors for sentiment analysis[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011:142-150.
- [4] SU J S, ZHANG B F, XU X. Advances in Machine Learning Based Text Categorization[J]. Ruan Jian Xue Bao/Journal of Software, 2006, 17(9):1848-1859.
- [5] WU Y J, LI J, SONG C F, et al. High Utility Neural Networks for Text Classification[J]. Dian Zi Xue Bao: Acta Electronica Sinica, 2020, 48(2):279-284.
- [6] COVER T, HART P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1):21-27.
- [7] JOACHIMS T. Text categorization with support vector machines: Learning with many relevant features[C]// European

- Conference on Machine Learning. Berlin: Springer, 1998: 137-142.
- [8] WANG S I, MANNING C D. Baselines and bigrams: Simple, good sentiment and topic classification[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. 2012:90-94.
- [9] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. 2014:655-665.
- [10] ZHU X, SOBHANI P, GUO H. Long short-term memory over recursive structures[C]//Proceedings of the 32nd International Conference on Machine Learning. 2015:1604-1612.
- [11] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1746-1751.
- [12] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[C]//Advances in Neural Information Processing Systems. 2015:649-657.
- [13] LIU P, QIU X, HUANG X. Recurrent Neural Network for Text Classification with Multi-Task Learning[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence. 2016:2873-2879.
- [14] YANG Z, YANG D, DYER C, et al. Hierarchical Attention Networks for Document Classification[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2016:1480-1489.
- [15] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. arXiv:1409.0473, 2014.
- [16] XU B B, CEN K T, HUANG J J, et al. A Survey on Graph Convolutional Neural Network [J]. Ji Suan Ji Xue Bao/Chinese Journal of Computers, 2020, 43(5): 755-780.
- [17] WU Z, PAN S, CHEN F, et al. A Comprehensive Survey on Graph Neural Networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
- [18] ZHANG Z, CUI P, ZHU W. Deep Learning on Graphs: A Survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 249-270.
- [19] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57-81.
- [20] MINAE S, KALCHBRENNER N, CAMBRIA E, et al. Deep Learning Based Text Classification: A Comprehensive Review [J]. arXiv:2004.03705v1, 2020.
- [21] LI Q, PENG H, LI J, et al. A Survey on Text Classification: From Shallow to Deep Learning[J]. arXiv:2008.00364, 2020.
- [22] KIPF T N, WELING M. A PyTorch implementation of GCN [EB/OL]. <https://github.com/tkipf/pygn>.
- [23] HAMILTON W, YING Z, LESKOVEC J. A PyTorch implementation of GraphSage [EB/OL]. <https://github.com/williamleif/graphsage-simple/>.
- [24] VELIKOVI P, CUCURULL G, CASANOVA A, et al. A Tensorflow implementation of GAT [EB/OL]. <https://github.com/PetarV-/GAT>.
- [25] CHEN J, MA T, XIAO C. A Tensorflow implementation of FastGCN[EB/OL]. <https://github.com/matenure/FastGCN>.
- [26] WANG M, ZHENG D, YE Z, et al. A PyTorch implementation of JKNet [EB/OL]. <https://github.com/dmlc/dgl/tree/master/examples/pytorch/jknet>.
- [27] WU F, ZHANG T, SOUZA A, et al. A PyTorch implementation of SGC[EB/OL]. <https://github.com/Tiiiger/SGC>.
- [28] YAO L, MAO C, LUO Y. A Tensorflow implementation of Text-gcn[EB/OL]. [https://github.com/yao8839836/text\\_gcn](https://github.com/yao8839836/text_gcn).
- [29] GAO H, CHEN Y, JI S. A Tensorflow implementation of hConv-gPool-Net [EB/OL]. <https://github.com/HongyangGao/hConv-gPool-Net>.
- [30] KLICPERA J, BOJCHEVSKI A, GUNNEMANN S. TensorFlow and PyTorch implementations of PPNP and APPNP[EB/OL]. <https://github.com/klicperajo/ppnp>.
- [31] HUANG L, MA D, LI S, et al. A PyTorch implementation of Text-level GCN [EB/OL]. <https://github.com/mojave-pku/TextLevelGCN>.
- [32] HU L, YANG T, SHI C, et al. A PyTorch implementation of HAGT[EB/OL]. <https://github.com/ytc272098215/HGAT>.
- [33] LIU X, YOU X, ZHANG X, et al. A Tensorflow implementation of TensorGCN[EB/OL]. <https://github.com/THUMLP/TensorGCN>.
- [34] RONG Y, HUANG W, XU T, et al. A PyTorch implementation of DropEdge [EB/OL]. <https://github.com/DropEdge/DropEdge>.
- [35] ZHANG J, MENG L. A PyTorch implementation of GResNet [EB/OL]. <https://github.com/jwzhanggy/GResNet>.
- [36] ZHAO L, AKOGLU L. A PyTorch implementation of PairNorm [EB/OL]. <https://github.com/LingxiaoShawn/PairNorm>.
- [37] CHEN G, TIAN Y, SONG Y. A PyTorch implementation of D-GCN[EB/OL]. <https://github.com/cuhkksz-nlp/DGSA>.
- [38] DING K, WANG J, LI J, et al. A PyTorch implementation of HyperGAT [EB/OL]. [https://github.com/kaize0409/HyperGAT\\_TextClassification](https://github.com/kaize0409/HyperGAT_TextClassification).
- [39] ZHANG Y, YU X, CUI Z, et al. A Tensorflow implementation of TextING[EB/OL]. <https://github.com/CRIPAC-DIG/Text-ING>.
- [40] YUAN L, WANG J, YU L, et al. A PyTorch implementation of G-ATT[EB/OL]. <https://github.com/YuanLi95/GATT-For-Aspect>.
- [41] FENG W, ZHANG J, DONG Y, et al. A PyTorch implementation of GRAND [EB/OL]. <https://github.com/THUDM/GRAND>.
- [42] ZHOU K, HUANG X, LI Y, et al. A PyTorch implementation of DGN[EB/OL]. <https://github.com/Kaixiong-Zhou/DGN>.
- [43] CHEN M, WEI Z, HUANG Z, et al. A PyTorch implementation of GCNII[EB/OL]. <https://github.com/chennnM/GCNII>.
- [44] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks

- and locally connected networks on graphs[J]. arXiv:1312.6203, 2013.
- [45] CHUNG F R K. Spectral Graph Theory [EB/OL]. <http://www.ams.org/books/cbms/092/>.
- [46] SANDRYHAILA A, MOURA J M F. Discrete Signal Processing on Graphs[J]. *IEEE Trans. on Signal Processing*, 2014, 62(12):3042-3054.
- [47] SHUMAN DI, NARANG S K, FROSSARD P, et al. The Emerging Field of Signal Processing on Graphs; Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains [J]. *IEEE Signal Processing Magazine*, 2013, 30(3): 83-98.
- [48] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering[C]// *Advances in Neural Information Processing Systems*. 2016:3844-3852.
- [49] HAMMOND D K, VANDERGHEYNST P, GRIBONVAL R. Wavelets on graphs via spectral graph theory[J]. *Applied and Computational Harmonic Analysis*, 2011, 30(2): 129-150.
- [50] KIPF T N, WELING M. Semi-Supervised Classification with Graph Convolutional Networks[C]// *Proceedings of the 5th International Conference on Learning Representations*. 2017:1-13.
- [51] LI Y, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks[J]. arXiv:1511.05493, 2016.
- [52] YUN S, JEONG M, KIM R, et al. Graph Transformer Networks [C]// *Advances in Neural Information Processing Systems*. 2019:11960-11970.
- [53] YAO L, MAO C, LUO Y. Graph Convolutional Networks for Text Classification[C]// *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. 2019:7370-7377.
- [54] CHURCH K W, HANKS P. Word association norms, mutual information, and lexicography [J]. *Computational linguistics*, 1990, 16(1):22-29.
- [55] ZAAANEN M, KANTERS P. Automatic Mood Classification using TF \* IDF based on Lyrics[C]// *Proceedings of the 11th International Society for Music Information Retrieval Conference*. 2010:75-80.
- [56] LIU X, YOU X, ZHANG X, et al. Tensor graph convolutional networks for text classification[C]// *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. 2020:8409-8416.
- [57] WU F, ZHANG T, SOUZA A, et al. Simplifying graph convolutional networks[C]// *Proceedings of 36th International Conference on Machine Learning*. 2019:6861-6871.
- [58] HUANG L, MA D, LI S, et al. Text Level Graph Neural Network for Text Classification[C]// *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 2019:3442-3448.
- [59] 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*. 2017:1263-1272.
- [60] GAO H, CHEN Y, JI S. Learning graph pooling and hybrid convolutional operations for text representations[C]// *Proceedings of the World Wide Web Conference*. 2019:2743-2749.
- [61] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[J]. arXiv:1710.10903, 2018.
- [62] HU L, YANG T, SHI C, et al. Heterogeneous Graph Attention Networks for Semi-supervised Short Text Classification[C]// *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. 2019:4821-4830.
- [63] YUAN L, WANG J, YU L, et al. Graph Attention Network with Memory Fusion for Aspect-level Sentiment Analysis[C]// *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*. 2020:27-36.
- [64] CHEN G, TIAN Y, SONG Y. Joint Aspect Extraction and Sentiment Analysis with Directional Graph Convolutional Networks [C]// *Proceedings of the 28th International Conference on Computational Linguistics*. 2020:272-279.
- [65] DING K, WANG J, LI J, et al. Be More with Less: Hypergraph Attention Networks for Inductive Text Classification[C]// *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 2020:4927-4936.
- [66] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3:993-1022.
- [67] ZHANG Y, YU X, CUI Z, et al. Every Document Owns Its Structure: Inductive Text Classification via Graph Neural Networks[C]// *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2020:334-339.
- [68] YUE Z, QI L, SONG L. Sentence-State LSTM for Text Representation[C]// *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. 2018:317-327.
- [69] LI W, LI S, MA S, et al. Recursive Graphical Neural Networks for Text Classification[J]. arXiv:1909.08166, 2020.
- [70] HUANG B, CARLEY K M. Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks[C]// *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing*. 2019:5469-5477.
- [71] ZHANG H, ZHANG J. Text Graph Transformer for Document Classification[C]// *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. 2020: 8322-8327.
- [72] LI Q, HAN Z, WU X. Deeper insights into graph convolutional networks for semi-supervised learning[C]// *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 2018: 3538-3545.
- [73] KLICPERA J, BOJCHEVSKI A, GUNNEMANN S. Predict Then Propagate: Graph Neural Networks Meet Personalized Pagerank[J]. arXiv:1810.05997, 2019.
- [74] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: Bringing order to the web [EB/OL]. <http://ilpubs>.

stanford.edu;8090/422/1/1999-66.pdf.

- [75] RONG Y, HUANG W, XU T, et al. DropEdge: Towards Deep Graph Convolutional Networks on Node Classification[C] // Proceedings of the International Conference on Learning Representations. 2020;1-17.
- [76] FENG W, ZHANG J, DONG Y, et al. Graph Random Neural Network for Semi-Supervised Learning on Graphs[C] // Advances in Neural Information Processing Systems. 2020;22092-22103.
- [77] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C] // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016;770-778.
- [78] XU K, LI C, TIAN Y, et al. Representation Learning on Graphs with Jumping Knowledge Networks[C] // Proceedings of the 35th International Conference on Machine Learning. 2018;5449-5458.
- [79] ZHANG J, MENG L. GResNet: Graph Residual Network for Revisiting Deep GNNs from Suspended Animation[J]. arXiv:1909.05729, 2020.
- [80] CHEN M, WEI Z, HUANG Z, et al. Simple and Deep Graph Convolutional Networks[C] // Proceedings of 37th International Conference on Machine Learning. 2020;1725-1735.
- [81] HEN D, LIN Y, LI W, et al. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view[C] // Proceedings of the 34th AAAI Conference on Artificial Intelligence. 2020;3438-3445.
- [82] DWIVEDI V, JOSHI C, LAURENT T, et al. Benchmarking graph neural networks[J]. arXiv:2003.00982, 2020.
- [83] ZHAO L, AKOGLU L. Pairnorm: Tackling oversmoothing in GNNS[J]. arXiv:1909.12223v1, 2020.
- [84] ZHOU K, HUANG X, LI Y, et al. Towards Deeper Graph Neural Networks with Differentiable Group Normalization[C] // Advances in Neural Information Processing Systems Conference. 2020;1-12.
- [85] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI magazine, 2008, 29(3):93-106.
- [86] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C] // Advances in Neural Information Processing Systems. 2017;1024-1034.
- [87] CHEN J, MA T, XIAO C. FastGCN: fast learning with graph convolutional networks via importance sampling [J]. arXiv:1801.10247, 2018.
- [88] YANG Z, COHEN W, SALAKHUTDINOV R. Revisiting semi-supervised learning with graph embeddings[C] // Proceedings of the International Conference on Machine Learning. 2016;40-48.



**TAN Ying-ying**, born in 1998, postgraduate. Her main research interests include natural language processing and deep learning.



**WANG Jun-li**, born in 1978, Ph.D, associate researcher. Her main research interests include text data analysis, deep learning and artificial intelligence.

(责任编辑:何杨)