

融合多图卷积与层级池化的文本分类模型

魏子昂, 彭舰, 黄飞虎, 琚生根

引用本文

魏子昂, 彭舰, 黄飞虎, 琚生根. 融合多图卷积与层级池化的文本分类模型[J]. 计算机科学, 2024, 51(7): 303-309.

WEI Ziang, PENG Jian, HUANG Feihu, JU Shenggen. Text Classification Method Based on Multi Graph Convolution and Hierarchical Pooling [J]. Computer Science, 2024, 51(7): 303-309.

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)

[针对系统调用的基于语义特征的多方面信息融合的主机异常检测框架](#)

Host Anomaly Detection Framework Based on Multifaceted Information Fusion of Semantic Features for System Calls

计算机科学, 2024, 51(7): 380-388. <https://doi.org/10.11896/jsjcx.230400023>

[基于深度确定性策略梯度与注意力Critic的多智能体协同清障算法](#)

Multi-agent Cooperative Algorithm for Obstacle Clearance Based on Deep Deterministic Policy Gradient and Attention Critic

计算机科学, 2024, 51(7): 319-326. <https://doi.org/10.11896/jsjcx.230600129>

[基于联合学习的语言粒度融合的重叠事件抽取方法](#)

Overlap Event Extraction Method with Language Granularity Fusion Based on Joint Learning

计算机科学, 2024, 51(7): 287-295. <https://doi.org/10.11896/jsjcx.230700118>

[基于关键词异构图的生成式摘要研究](#)

KHGAS: Keywords Guided Heterogeneous Graph for Abstractive Summarization

计算机科学, 2024, 51(7): 278-286. <https://doi.org/10.11896/jsjcx.230500059>

[基于外部先验和自先验注意力的图像描述生成方法](#)

Image Captioning Generation Method Based on External Prior and Self-prior Attention

计算机科学, 2024, 51(7): 214-220. <https://doi.org/10.11896/jsjcx.230600167>

融合多图卷积与层级池化的文本分类模型

魏子昂 彭 舰 黄飞虎 琚生根

四川大学计算机学院 成都 610041

(weiziang@stu.scu.edu.cn)

摘要 文本分类是自然语言处理中的一个重要问题,其目的是将标签分配给输入的文档。在文本分类任务中,单词间的共现关系提供了文本特性及词汇分布的重要视角,而词嵌入信息能提供丰富的语义信息,并对全局词汇交互和潜在语义关系造成影响。然而,过去的研究未能有效整合这两方面,或过度关注其中一方面。在这样的背景下,文中提出了一种新的方法,用于自适应地融合这两类信息,在考虑结构关系和嵌入信息的同时,找到一个合理的平衡以提高模型效果。该模型首先从词汇共现模式和语义嵌入信息的角度将文本数据构建成文本共现图和文本嵌入图,利用图卷积来增强节点嵌入,图池化层融合节点嵌入并识别保留重要性更高的节点,遵循分层池化模式并按层学习文档级表示,并引入门控融合模块对两个图的嵌入进行自适应的融合。在5个公开的文本分类数据集上进行了大量实验,结果表明了HTGNN在文本分类任务上的优异性能。

关键词: 文本分类;图神经网络;图表示学习;图分类;注意力机制

中图分类号 TP183

Text Classification Method Based on Multi Graph Convolution and Hierarchical Pooling

WEI Ziang, PENG Jian, HUANG Feihu and JU Shenggen

College of Computer Science, Sichuan University, Chengdu 610041, China

Abstract Text classification, as a critical task in natural language processing, aims to assign labels to input documents. The Co-occurrence relationship between words offers key perspectives on text characteristics and vocabulary distribution, while word embeddings supply rich semantic information, influencing global vocabulary interaction and potential semantic relationships. Previous research has struggled to adequately incorporate both aspects or has disproportionately emphasized one over the other. To address this issue, a novel method is proposed in this paper that adaptively fuses these two types of information, aiming to strike a balance that can improve model performance while considering both structural relationships and embedded information. The method begins by constructing text data into text co-occurrence graphs and text embedding graphs, reflecting the context structure and semantic embedding information respectively. Graph convolution is then utilized to enhance node embeddings. In the graph pooling layer, node embeddings are fused and nodes of higher importance are identified by employing a hierarchical pooling model, learning document level representations layer by layer. Furthermore, we introduce a gated fusion module to adaptively fuse the embeddings of the two graphs. The proposed approach is validated with extensive experiments on five publicly available text classification datasets, and the experimental results show the superior performance of the HTGNN model in text classification tasks.

Keywords Text classification, Graph neural network, Graph representation learning, Graph classification, Attention mechanism

1 引言

文本分类是自然语言处理领域的关键任务,具有广泛的应用,如情感分析、新闻过滤和疾病诊断。早期的分类方法主要采用统计理论的特征提取方法,如TF-IDF和词袋模型,但这类方法存在抗噪声干扰性差、特征难以拟合等问题,且需要复杂的预处理步骤。相比之下,基于深度学习的模型能提供一种端到端的解决方案,可以用来解决以上问题。

神经网络的流行激发了一系列关于开发用于文本分类的

深度学习模型。深度学习模型架构已经证明它们比传统的基于统计特征的方法更有效^[1]。Jacovi等^[2]提出了TextCNN,使用共享参数的一维卷积核捕捉文本的局部特征,通过全局最大池化将局部特征转化为新的全局特征,从而获得不同抽象层次的语义信息。Kalchbrenner等^[3]提出了DCNNs,通过引入一种自适应的池化方法,在确定池化参数的同时动态合并句向量特征。Mikolov等^[4]提出了具有门控机制的LSTM,可以更好地捕获上下文之间的联系信息,利用遗忘门过滤信息,缓解了循环神经网络模型普遍存在的梯度消失

到稿日期:2023-04-25 返修日期:2023-09-15

基金项目:四川省重点研发计划(2022YFG0034,2023YFG0115);四川大学宜宾市合作项目(2020CDYB-30)

This work was supported by the Key R&D Program of Sichuan Province, China (2022YFG0034, 2023YFG0115) and Cooperative Program of Sichuan University and Yibin(2020CDYB-30).

通信作者:彭舰(jianpeng@scu.edu.cn)

问题, LSTM 最后一步的输出作为文本的表示向量。随着注意力机制的提出, 注意力方法也被应用于文本分类领域。Yang 等^[5]提出了 HAN, 该模型包括编码器和多层次的注意力层。注意力机制赋予模型区分不同句向量重要性的能力, 将词向量和句向量根据权重大小加权聚合成文本向量。Vaswani 等^[6]进一步提出了完全基于自注意力的 Transformer 模型, 对文本中不同位置的词向量之间的相关性进行建模, 进而表征整个文本序列, 并挖掘语义特征。

图神经网络(GNN)是一个可用于非欧几里得空间结构数据的连接模型, 它通过连接节点的边来传递消息, 以捕获图节点之间的依赖关系。GNN 可以缓解传统深度学习模型存在的距离依赖问题, 捕捉文本长距离的词向量之间的内在关系, 其重点在于如何有效构建文本图以及学习文本的图表示特征。Yao 等^[7]提出了 TextGCN 模型, TextGCN 将整个语料库构建为一张异构图, 文本和所有单词构成图节点, 边的权重是固定的, 单词节点间的边权重由单词的共现信息决定, 文档-单词节点间的边权重由 TF-IDF 决定, 对该图进行图卷积从而得到文本的嵌入表示, 然而固定权重限制了边的表达能力, 并且导致了巨大的内存消耗。Huang 等^[8]提出了 Text-LevelGNN, 为每个输入文本单独构建图, 单词作为节点, 每个词节点只与其左右的词相连。然而这种方法只考虑了文本的结构信息, 忽略了词向量的语义嵌入信息。Wang 等^[9]提出了 SHINE, 该模型是由基于词级别的子图构成的层次异构图, 从词性、互信息、实体关系 3 个角度构建子图, 分别在 3 个子图上利用 GCN 学习文本嵌入, 最后拼接子图特征通过全连接层预测分类。然而, 直接拼接特征无法捕捉到各子图特征之间的交互信息, 无法自适应提取更重要的特征。

在图池化方面, 现有的文本分类模型在生成文档级表示时通过平均池化、最大池化将节点表示聚合为文档级表示, 尽管在分类任务中的确有效, 但是图池化的方法还有很大的改进空间, 这些方法从信息聚合的角度来看在本质上是扁平的, 然而文本内部不同单词对于文档的贡献程度并不一样, 现有方法缺乏对重要性特征的筛选能力。

为了解决以上问题, 本文提出了新的 HTGNN 模型, 该模型首先从共现模式和语义嵌入信息的角度将文本数据构建成本文共现图和文本嵌入图, 利用图卷积来捕获高阶邻域信息, 所提出的图池化模块将图卷积得到的词级别表示聚合为文档级表示, 并引入门控融合模块对两个图的表示进行自适应的融合。本文的贡献如下:

- 1) 提出了一种多尺度的文本图构建方法, 用于有效捕获文本序列中的结构信息和语义信息。
- 2) 提出了一种新的图池化方法, 可以捕获到节点间的差异特征并有效找到重要性更高的节点, 解决了全局池化中信息冗余的问题。
- 3) 提出了一种自适应的特征融合机制, 通过门控机制增强表示能力。

2 相关工作

2.1 图神经网络

图神经网络已被广泛用于处理各种图结构数据, 包括

社交网络^[10]、生物网络^[11]、知识图^[12]等。在文本图分类领域, GNN 通过为文本构建一个词图, 并在图上进行卷积和池化操作, 得到图的全局特征并将其用于分类。对于图卷积(GCN), 目标是学习一个状态嵌入 $\mathbf{h}_i \in \mathbb{R}^{N \times d}$, 其中包含每个节点的邻域信息和自身信息。GCN 模型为每个节点聚合来自上一层邻域的信息。直观上, 如果堆叠多个 GCN 层, 感受野大小将随着深度增加。对于 l 层的 GCN, 传播规则具体定义为:

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

其中, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ 是附加了自连接的图 \mathcal{G} 的邻接矩阵, \mathbf{I}_N 是单位矩阵; $\tilde{\mathbf{D}}_i = \sum_j \tilde{\mathbf{A}}_{ij}$ 和 $\mathbf{W}^{(l)}$ 是可训练参数; $\sigma(\cdot)$ 表示激活函数; $\mathbf{H}^{(l)} \in \mathbb{R}^{N \times D}$ 是第 l 层的表示矩阵。

当处理图结构数据时, 不能总是存储和处理每个节点的所有邻域信息, 因此需要池化模块缩小图的规模。图池化方法分为直接池化和分层池化两类。在直接池化模型中, 将最大、均值、求和操作应用于节点特征以获得全局图形表示。MPNN^[13]使用 Set2set 方法作为读出函数来获取图形表示。SortPooling^[14]首先根据节点的结构对节点嵌入进行排序, 然后将排序后的嵌入输入 CNN 以获得表示。然而, 复杂和大规模的图通常具有丰富的层次结构, 这对于图级分类任务非常重要。早期的分层池化模型使用谱聚类算法, DiffPool^[15]通过可微分的软分配方法学习图的层次结构; EigenPooling^[16]使用图傅里叶变换提取局部子图特征; SAGPool^[17]用自注意力联合拓扑和嵌入学习图表示。

2.2 注意力机制

注意力机制^[18]已成为神经体系结构中常见的组成部分, 被广泛运用在文本分类任务中。Liu 等^[19]将双向 LSTM 对输入序列编码, 利用注意力机制计算不同特征对输出特征的权重。Shen 等^[20]提出了 Disan, 基于双向分块自注意力网络, 不使用基于 RNN 或 CNN 的序列编码模型, 而对输入序列中的每对元素应用注意力来生成包含上下文的表示。Yang 等提出了一种文本分类模型(HAM), 其采用双层注意力结构: 第一层基于词级别注意力, 用于将重要单词聚合成句子; 第二层基于句级别注意力, 将重要句子聚合为最终的文本表示。

注意力机制可与 GNN 结合, 用于计算特征聚合过程的权重系数。注意力为邻居分配不同的权重, 从而可以减轻噪音并取得更好的效果。Velickovic 等^[21]提出了 GAT, 将注意力机制纳入节点特征聚合, 通过分配邻居节点权重系数来确定邻居节点的重要性。Zhang 等^[22]提出了 GaAn, 模型应用多个独立的注意力头矩阵收集信息来代替 GAT 的平均操作。

2.3 文本分类任务描述

本文将文本数据构建为文本图网络, 从而转化为图分类问题。我们用 $\mathbf{G}_i = (\mathcal{V}_i, \mathcal{E}_i, \mathcal{A}_i)$ 表示一个文本图, 其中顶点集 $\mathbf{V}_i = \{v_1, v_2, \dots, v_n\}$ 表示文本实例中的单词节点, n 是实例中的单词数量, 而 \mathcal{E}_i 表示边的集合。邻接矩阵 $\mathcal{A}_i \in \mathbb{R}^{N \times N}$ 反映节点之间的关联信息。文本图分类的本质是学习一个映射函数 $F(\cdot)$, 将文本集 $\mathcal{G} = \{g_1, g_2, \dots, g_k\}$ 中的每一篇文本映射到标签集 \mathcal{Q} 中最可能的标签 y_i 。

3 融合多图卷积与层级池化的文本分类模型

3.1 总体结构

本文提出的 HTGNN 的结构如图 1 所示,该模型由 3 部分组成,依次为文本图构建层、文本表示学习层和分类层。首先将词嵌入编码器用于初始化文本单词表示,并用来构建

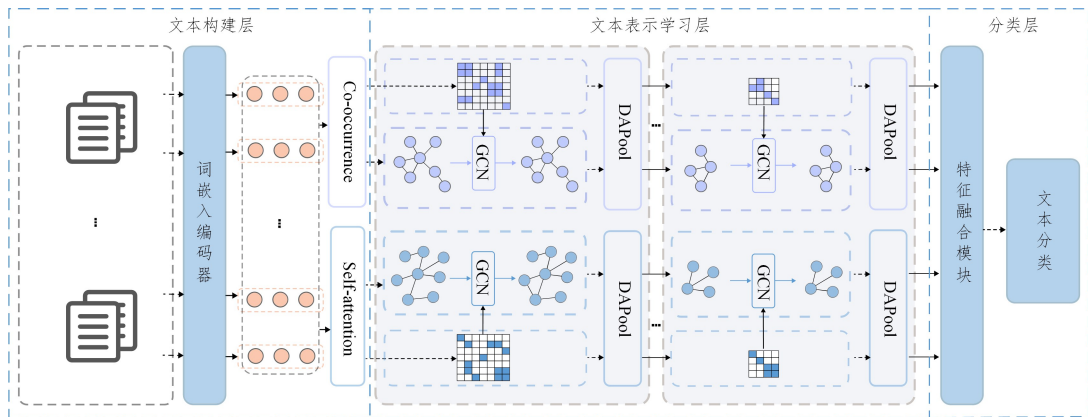


图 1 HTGNN 模型的结构图

Fig. 1 Diagram of HTGNN model structure

3.2 文本图构建

邻接矩阵反映的是词节点之间的关联性。本文分别基于词共现信息和词嵌入语义信息提出两种图构造的方法。文本共现图基于文本的结构信息,利用单词之间的共现信息构造文本图,可以捕捉到文本的固定结构信息;文本嵌入图基于语义嵌入信息,利用自注意力方法来捕获单词之间的隐藏动态关系。为了得到单词嵌入表示,本文以字为粒度对文本进行切分,利用 BERT 词嵌入来生成单词的向量化表示,每个词会被转换成一个 768 维的向量。

3.2.1 文本共现图

本节利用点间互信息计算文本共现图中词之间的关系权重。具体地,将给定的一个包含 l 个词的文本记为 $T = \{r_1, r_2, \dots, r_l\}$,其中 r_i 与文本中第 i 个单词对应。设定一个固定大小 m 的滑动窗口 W 来计算节点 i 和节点 j 间的权重。滑动窗口沿着序列 T 从左向右滑动,步长为 1,对于文本中任一单词 $word_i$,统计包含 $word_i$ 的窗口的数量 W_i ,计算 $word_i$ 在所有窗口中出现的频率。

$$p_i = \frac{W_i}{\sum W} \quad (2)$$

其中, $\sum W$ 是滑动窗口的总数,以及统计同时包含单词对 $(word_i, word_j)$ 的滑动窗口的次数 $W_{i,j}$,用 $p(i, j)$ 描述两个词同时出现在一个窗口内的频率,计算式如下:

$$p(i, j) = \frac{W_{i,j}}{\sum W} \quad (3)$$

其中, $p(i, j)$ 指单词间的共现关系,该值越大表明两个单词在文本中的关联性越高。进一步地,可以得到邻接矩阵元素 $A_{i,j}$,其表达式如下:

$$A_{i,j} = \text{ReLU} \left(\log \frac{p(i, j)}{p(i) p(j)} \right) \quad (4)$$

其中, $\text{ReLU}(\cdot)$ 是激活函数,邻接矩阵 A 中的所有元素的取值都大于 0。直观上,两个单词之间的权重越大,其关联度越高。

生成两个文本图,即基于单词共现信息的文本共现图,以及基于单词嵌入语义信息的文本嵌入图;其次,通过文本表示学习模块,经过独立的图卷积和层级池化,分别得到两个图的嵌入表示,通过特征融合模块得到文本最终的嵌入表示;最后在分类层利用 Softmax 分类器进行分类预测。更详细的结构和机制将在下文介绍。

3.2.2 文本嵌入图

本文设计了一个多头注意力网络,用于学习节点嵌入之间的关系特征。其将输入特征通过线性变换分成 3 个部分,分别为查询、键、值。对于计算节点 i, j 的注意力权重,具体表达式为:

$$\text{Att}_{i,j} = \text{softmax} \left(\frac{\mathbf{H}_i \mathbf{W}^q \cdot (\mathbf{H}_j \mathbf{W}^k)^T}{\sqrt{d_k}} \right) \quad (5)$$

根据计算好的注意力系数,将节点的邻居节点加权求和,其中单头的计算式如下:

$$\text{head}_i = \sigma \left(\sum_{j=1}^N \text{Att}_{i,j} \mathbf{H}_j \mathbf{W}^v \right) \quad (6)$$

式(5)、式(6)中, $\mathbf{H}_i, \mathbf{H}_j$ 分别为两个节点的嵌入特征, N 为节点数量; $\mathbf{W}^q \in \mathbb{R}^{d_e \times d_q}$, $\mathbf{W}^k \in \mathbb{R}^{d_e \times d_k}$, $\mathbf{W}^v \in \mathbb{R}^{d_e \times d_v}$ 是可学习的线性变换矩阵参数,在自注意力中, $d_q = d_k = d_v$ 。多头注意力中每个头的计算是独立的,最终将多个头的结果进行合并,邻接矩阵元素 $A_{i,j}$ 的计算式为:

$$A_{i,j} = [\text{head}_i \oplus \dots \oplus \text{head}_k] \cdot \mathbf{W}^o \quad (7)$$

其中, \mathbf{W}^o 为线性变换参数, \oplus 是拼接操作。

3.3 文本表示学习

3.3.1 多图卷积层

GCN 根据图的拓扑结构,通过聚合节点的邻居信息,得到增强节点的嵌入向量。一层 GCN 考虑一阶邻居信息,当更多的 GCN 层被堆叠在一起时,感受野随之增大。当 GCN 层数达到一定深度时,图中的每个节点都会与全图其他节点交互信息。

文本共现图和嵌入图从不同角度反映文本词节点的特征,然而单个图只能捕获局部细粒度的空间信息。因此,本文使用两个独立的图卷积模块处理多尺度的输入特征图的空间信息,以丰富特征空间,其表达式为:

$$\begin{aligned} \mathbf{X}_c' &= \sigma(\tilde{\mathbf{A}}_c \cdot \text{ReLU}(\tilde{\mathbf{A}}_c \mathbf{X} \mathbf{W}_c^0) \mathbf{W}_c^1) \\ \mathbf{X}_e' &= \sigma(\tilde{\mathbf{A}}_e \cdot \text{ReLU}(\tilde{\mathbf{A}}_e \mathbf{X} \mathbf{W}_e^0) \mathbf{W}_e^1) \end{aligned} \quad (8)$$

本文的图卷积模块由两层卷积层组成, \mathbf{A}_c 和 \mathbf{A}_e 是两个图的邻接矩阵, $\tilde{\mathbf{A}}$ 是包含自连接的拉普拉斯矩阵, σ 和 ReLU 为激活函数, \mathbf{X}_c' 和 \mathbf{X}_e' 是增强后的节点嵌入。

3.3.2 层级图池化模块

文本分类预测基于最终文档的图级表示, 因此需要图池化层将图中所有节点表示聚合成图级表示。大多数现有的 GNN 在生成文档级表示时应用均值池化或最大池化的全局池化策略, 这些方法本质上是扁平的, 忽略了图的层次结构, 在下采样的过程中缺乏区别单词优先级的能力。而层级池化通过筛选节点, 只保留对网络结构有用的节点, 能有效学习图的结构信息。层级图池化包含两个阶段: 第一个阶段考虑的是节点的特征聚合; 第二个阶段对更新后的节点进行评分, 保留重要性更高的节点。

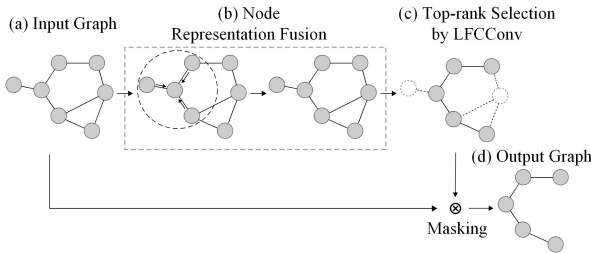
本节提出了一种新的层级图池化, 即 DAPool (Difference Aware Pooling)。DAPool 的架构如图 2 所示。DAPool 首先考虑节点间的信息融合。对于某个确定的节点, 与邻居节点的信息融合学习可以获得增强的嵌入。已有方法往往将各个邻居节点等地看待, 即平均聚合。然而, 在实际场景中, 不同的邻居节点可能对目标节点起着不同的作用, 因此提出了一种系数自适应匹配的方法来解决这个问题。注意力机制被应用于学习节点与邻居节点之间的系数; 注意力系数对应的是相邻节点的重要性权重, 用于节点表示聚合。研究重点在于如何有效地对节点信息进行聚合。考虑到注意力机制中存在重要性互斥问题, 即强调高相似度的节点间的特征聚合, 而没有考虑到节点差异性的重要性。如果一个节点的信息可以由其他节点的信息重构得出, 这意味该节点在图中是可以被其他节点替代的。相似节点可以互相替代, 而节点之间的差异性需要被保留在节点信息中。为了平衡相关性和差异性, 提出了一种自注意力方法 Variance2Token, 其表达式为:

$$\alpha_{i,j} = \text{softmax}(\bar{\mathbf{v}}^T \sigma(\mathbf{W}_1 \mathbf{h}_i \parallel \mathbf{W}_2 (\mathbf{h}_i - \mathbf{h}_j))) \quad (9)$$

其中, $\mathbf{h}_i, \mathbf{h}_j$ 为词节点表示; $\mathbf{h}_i - \mathbf{h}_j$ 展示了 \mathbf{h}_i 与 \mathbf{h}_j 的差异性, 可以通过考虑特征差异来调节影响注意力系数。得到注意力系数后, 节点表示融合可以被表示为:

$$\mathbf{h}_i' = \mathbf{h}_i + \sum_{j=1}^N \alpha_{i,j} \mathbf{h}_j \quad (10)$$

其中, \mathbf{h}_i' 是新的节点表示。



注: LFCCConv 用于对节点进行评分, 保留重要性高的节点。

图 2 DAPool 的结构

Fig. 2 Structure of DAPool

对于节点筛选, 需要计算单词的得分系数, 系数越大意味着该节点在网络中的重要性更高。本文设计了一种图卷积变体 LFCCConv (Local Features Coupling Convolution), 用于获得每个节点的得分, 分数是由节点嵌入和图拓扑共同确定的, 如式 (11) 所示:

$$\phi_i = \sigma(\mathbf{h}_i' \mathbf{w}_1 + \sum_{i=1}^N \mathbf{A}_{i,j}' (\mathbf{h}_i' - \mathbf{h}_j') \mathbf{w}_2) \quad (11)$$

其中, \mathbf{w}_1 和 \mathbf{w}_2 是学习参数, σ 是激活函数, \mathbf{A}' 是对应于当前图 G' 的邻接矩阵, \mathbf{h}_i' 是节点 i 的表示, \mathbf{h}_j' 是相邻节点的表示, ϕ_i 是节点 i 的得分。LFCCConv 通过线性变换获得节点本身的分数, 然后使用 $\mathbf{h}_i - \mathbf{h}_j$ 来捕捉节点之间的差异, 这是考虑到节点差异性对生成图级表示至关重要。拓扑不仅对图有巨大影响, 具有相同连通性的图中的不同节点表示也会使图发生变化。这种图卷积方法允许在对节点进行评分的同时考虑节点表示和图拓扑。节点得分向量 $\Phi = [\phi_1, \phi_2, \dots, \phi_N]^T$ 通过式 (11) 获得, 只保留输入图的一部分节点, 即使在输入不同大小和结构的图时也是如此。池化比率 $k \in (0, 1]$ 是一个超参数, 它决定了要保留的节点数。根据 Φ 的值选择排名靠前的 $\lceil kN \rceil$ 个节点。

$$\text{id}_{\text{idx}} = \text{top-}k(\Phi, \lceil kN \rceil), \Phi_{\text{mask}} = \Phi_{\text{id}_{\text{idx}}} \quad (12)$$

其中, $\text{top-}k$ 是一个函数, 对评分向量排序, 返回排名在前 $\lceil kN \rceil$ 的节点的索引; Φ_{mask} 是节点表示掩码。新的池化图被构建为:

$$\dot{\mathbf{X}} = \mathbf{X}_{\text{id}_{\text{idx}}}, \mathbf{X}_p = \dot{\mathbf{X}} \odot \Phi_{\text{mask}}, \mathbf{A}_p = \mathbf{A}_{\text{id}_{\text{idx}}} \quad (13)$$

式 (13) 表示提取节点表示的行或列和邻接矩阵, 用于重建节点之间的连接; 其中, \odot 是广播哈达玛积, \mathbf{X}_p 是所选保留节点的节点表示矩阵, \mathbf{A}_p 是相应的修剪邻接矩阵。

3.4 分类预测

经过图池化, 分别得到文本共现图的图级表示 \mathbf{H}_c 和文本嵌入图的图级表示 \mathbf{H}_e , 这两个特征从不同的角度表示文本特征信息。但是, 对于不同文本, \mathbf{H}_c 和 \mathbf{H}_e 对于文本并非同等重要, 输入一个捕捉无用表示的输出会误导模型的最终预测, 因此需要自适应的机制来提取文本级表示。为了增强表示能力, 我们使用门控机制来融合两个图上提取的特征, 门控机制可以控制神经网络中信息的传递。具体来说, 我们定义了一个门来控制两个特征的重要性。

$$\mathbf{G} = \text{sigmoid}(\mathbf{W}^g [\mathbf{H}_c, \mathbf{H}_e] + \mathbf{b}^g) \quad (14)$$

其中, $[\cdot, \cdot]$ 表示拼接操作; $\text{sigmoid}(\cdot)$ 是激活函数; $\mathbf{W}^g \in \mathbb{R}^{2d \times d}$ 和 $\mathbf{b}^g \in \mathbb{R}^d$ 是可学习参数; $G \in (0, 1]$, 为每个特征分配不同的重要性。G 的值越大, 表明该特征的重要性越高。基于该门, 融合两个特征:

$$\mathbf{Z}_g = \mathbf{G} \odot \mathbf{H}_c + (1 - \mathbf{G}) \odot \mathbf{H}_e \quad (15)$$

因此, $\mathbf{Z}^g \in \mathbb{R}^{N \times d}$ 表示 N 个节点的更新表示, 其聚集了来自两种图的特征信息。最终的输出集成了不同尺度的图级表示, 并可以自适应地调整不同图对结果的影响。最后, 将 \mathbf{Z}^g 输入全连接层来预测文本的标签。

$$y_i = \text{softmax}(\mathbf{W} \mathbf{Z}^g + \mathbf{b}) \quad (16)$$

其中, \mathbf{W} 和 \mathbf{b} 是学习参数。损失函数为:

$$\text{loss} = -\mathbf{g}; \log y_i \quad (17)$$

其中, \mathbf{g}_i 是真实标签的 one-hot 向量表示。

4 实验

4.1 数据集

本文实验使用了 20NG, R8, MR, R52, Ohsumed 5 个文本分类领域内广泛使用的公开数据集, 统计信息如表 1 所列。

表1 数据集统计信息

Table 1 Description of datasets

Dataset	Training	Test	Classes	Average Length
R8	5485	2189	8	65.72
R52	6532	2568	52	69.82
Ohsumed	3357	4043	23	135.82
20NG	11314	7532	20	221.26
MR	7108	3554	2	20.39

R8 和 R52 是来自于路透社新闻数据集的文本分类集合。R8 包含了 8 个不同类别的新闻文本,而 R52 则包含了 52 个不同类别的新闻文本,且每个数据只与一个类别相关。

Ohsumed 数据集源自医学文献数据库 MEDLINE,其中每篇文本都是一篇医学研究摘要,并且都可以被归类到 23 种疾病类别之一。

MR 是一个电影评论数据集,用于二元情感分类,其中每个评论只包含一句话。

20NG 是一个新闻组文章数据集,这些文档均匀地分布在 20 个不同的类别中。

4.2 实验设置

实验使用准确率作为模型性能的评估指标,准确率越高,模型的性能越好,具体计算式如下:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (18)$$

其中, TP 表示真阳性, TN 表示真阴性, FN 表示假阴性, FP 表示假阳性。

在模型训练的实验过程中,本文模型的具体结构如图 1 所示。本文选择 BERT 作为预训练模型进行词向量的初始化。对于构建好的文本图,默认堆叠两层 GCN 更新节点的特征信息。在节点的提取过程中,本文使用 DAPool 模块提取当前文本图中的关键节点并构建新子图。对于文本分类器,则搭建全连接层,并使用 ReLU 作为激活函数,最终通过 Softmax 分类器预测文本标签。

对于涉及的 5 个数据集,训练集和测试集的划分遵循了与基准模型相同的策略。在模型参数的优化上,本文使用 Adam^[23] 优化器,并设置初始学习率为 5×10^{-4} , L2 权重衰减率为 10^{-4} 。在模型的超参数初始化过程中,使用 Xavier^[24] 方法进行模型初始化。为防止模型过拟合,采用 Dropout^[25] 技术并设置 Dropout 比率为 0.5。为捕捉单词之间的共现模式,将滑动窗口的长度设置为 15,并将其作为默认设置。DAPool 模块中的节点池化率 p 为 0.5,即文本图每经过一次 DAPool 模块,就提取注意力分数排名前 50% 的关键节点,并建立新的子图作为下一个模块的输入。在上述超参数设置环境下,训练批次大小和迭代次数分别设置为 128 和 100,并设置了早停机制控制迭代次数。

4.3 对比实验

为验证所提模型在文本分类任务中的效能,本文选取了一系列具有代表性的基线模型进行对比。这些基线模型不仅包括经典的文本分类算法,也涵盖了当前领域先进的基于图神经网络的模型。参与对比的基准模型如下:

1) CNN^[2]: 对单词嵌入执行卷积和最大池运算,以获得文本的表示。

2) LSTM^[4]: 一种循环神经网络模型,使用最后一个隐藏状态表示文本。

3) fastText^[26]: 将单词嵌入的平均值作为文档嵌入,通过全连接层分类。

4) SWEM^[27]: 一种词嵌入模型,采用将预训练词向量简单池化的策略得到文本表示。

5) TextGCN^[7]: 一种用于文本分类的图卷积网络模型,构建了基于文本和词的异构图,然后使用 GCN 学习文本的表示。

6) TextLevelGNN^[8]: 为单个文档构建文本图,提出了一种消息传递机制(MPM)的非频谱方法,被用于卷积。

7) TensorGCN^[28]: 对语料使用 LSTM 进行编码,且提出了一种多角度的建图方式,使用改进图卷积融合异构图特征。

本文模型与上述基准方法的性能比较如表 2 所列,其中最佳结果以粗体突出显示。

表2 实验结果比较

Table 2 Comparison of experimental results

Model	Ohsumed	MR	R8	R52	20NG
CNN	58.44	77.75	85.71	87.59	82.15
LSTM	51.10	77.33	96.09	90.48	75.43
fastText	57.70	75.14	96.13	92.81	79.38
SWEM	63.12	76.65	95.32	92.94	85.16
TextGCN	68.36	76.74	97.07	93.56	86.34
TextLevelGNN	69.40	—	97.80	84.60	—
TensorGCN	70.11	77.91	98.04	95.05	87.74
HTGNN	71.55	82.84	98.12	96.37	87.91

由表 2 可知,基于图神经网络的模型优于传统深度学习模型 CNN 和 LSTM,表明了 GNN 结构的有用性;其次,TextGCN 优于 fastText,这可能是得益于图卷积能够通过融合邻域信息获得网络中节点的增强表示。词节点迭代地聚合来自邻居的表示,然后将聚合的信息与邻域的信息集成,从而将信息传播到整个图,这表明图神经网络能有效地学习到全局语义特征。

与基准模型相比,本文模型在 5 个数据集上取得了最佳性能。HTGNN 的这种优势可能归功于其先进的机制:首先是多尺度的图构建方式,GCN 从两个角度通过图边传播学习到了隐藏的高阶语义和结构信息;其次是提出的图池化方法,能有效地捕获邻域信息的相关性和差异性,用于融合学习。所提出的 LFCCConv 基于节点重要性对节点进行评分,保留重要性高的节点,有助于减少非重点节点对准确率的影响。

4.4 消融实验

本节通过消融实验来验证模型各模块的有效性,基于原始模型建立以下几个变体进行对比实验。

4.4.1 Variance2Token 的有效性

为了更好地验证图池化层中的注意力模块 Variance2-Token(V2T) 在模型中的作用,将 V2T 分别替换为 3 种之前提出的注意力方法,包括 Token2Token(T2T)^[20], Source2Token(S2T)^[20], Master2Token(M2T)^[29]。如表 3 所列,注意力采用 V2T 明显优于其他注意力,这证明了 V2T 用于节点表示聚合的有效性。

表3 Variance2Token 的有效性
Table 3 Validity of Variance2Token

Model	Ohsumed	MR	R8	R52	20NG
HTGNN-T2T	70.90	82.28	97.22	95.49	87.31
HTGNN-S2T	71.07	81.02	97.47	95.79	87.23
HTGNN-M2T	71.22	81.48	97.71	96.13	87.49
HTGNN-V2T	71.55	82.84	98.12	96.37	87.91

当注意力用于节点表示增强时,之前的方法强调相似度高的节点之间的聚合特征,却忽略了重要性互斥的问题,导致不同节点信息趋于单调,存在过多特征相似的节点会加剧图的过平滑现象,造成语义信息的损失。而 V2T 关注于高差异的节点信息,节点的差异特征对于生成有效的图级表示来说是不可缺的。对于两个有同样的拓扑结构的图,节点表示不同会导致图标签相异。从实验结果可以证明,将差异特征引入信息融合,能很好地解决重要性互斥的问题。

4.4.2 LFCCConv 的有效性

本小节进行了一组消融研究,以评估 LFCCConv 作为重要性评分函数在节点选择中的贡献。将 LFCCConv 分别替换为 3 个之前的图卷积运算,包括 GCN^[30], GAT^[31] 和 GraphSAGE^[32]。如表 4 所列,与基准模型的最好结果相比,模型准确率在各数据集上相比基准模型最佳性能分别增加了 0.33, 1.40, 0.41, 0.24 和 0.51。

表4 基于图卷积的不同评分函数的性能比较

Table 4 Comparison of different scoring functions based on graph convolution

Model	Ohsumed	MR	R8	R52	20NG
HTGNN-GCN	70.94	81.24	97.68	95.79	87.06
HTGNN-GAT	71.22	81.47	97.47	96.13	87.40
HTGNN-GraphSAGE	71.08	81.63	97.71	95.74	87.2
HTGNN-LFCCConv	71.55	82.84	98.12	96.37	87.91

GCN 灵活性差,不支持动态学习邻居权重,这导致在选择节点时存在偏差; GAT 是 T2T 注意力的实例,考虑了各个邻居节点的权重,对节点表示给予重要性,但缺乏捕捉节点差异性的方法;而 GraphSAGE 对邻居特征平均聚合,没有考虑邻居节点重要性不同的问题。LFCCConv 解决了上述问题,通过将节点特征差异纳入考虑,实验结果验证了 LFCCConv 的有效性。

4.4.3 文本图构建的有效性

为验证联合训练文本共线图 A_c 和文本嵌入图 A_e 对实验结果的提升作用,消融实验主要验证稳定空间的有效性建模,保持动态空间建模机制不变,具体设计如下:

- 1) HTGNN- A_c : 去掉 A_e , 仅保留 A_c 。
- 2) HTGNN- A_e : 去掉 A_c , 仅保留 A_e 。
- 3) HTGNN- A_p : 去掉 A_c 和 A_e , 使用皮尔逊相关系数构图。

基于不同图构造方法的性能如表 5 所列,模型的性能优于其他几种图构建方法。皮尔逊相关性通过度量节点嵌入的相似度来定义节点接近度,然而不能包含有关文本语义的完整信息,并且不与预测任务直接相关,这可能会导致很大的偏差。无论是 HTGNN- A_c 还是 HTGNN- A_e , 在缺少结构或是语义信息时都会造成精度下降。而本文方法确实带来了更好

的预测能力,可以更准确地捕捉词节点间的相互依赖关系。

表5 基于不同图构造方法的性能比较

Table 5 Comparison of different graph construction methods

Model	Ohsumed	MR	R8	R52	20NG
HTGNN- A_c	70.47	80.98	97.35	94.77	86.42
HTGNN- A_e	70.11	81.53	97.80	95.18	87.63
HTGNN- A_p	70.52	80.83	97.80	94.74	86.85
HTGNN	71.55	82.84	98.12	96.37	87.91

4.5 超参分析实验

不同池化率在数据集上的表现如表 6 所列。池化率代表了保留节点的比例,当池化率 $p=1$ 表明所有节点都将保留。由表 6 可知,随池化率下降,模型性能在各数据集上都取得了不同程度的提升,这表明池化模块有效地减少了信息冗余;在 Ohsumed 和 20NG 上池化率 $p=0.5$ 取得了最高准确率,可能是由于这两个数据集的文本长度较长,较低的池化率可以过滤掉更多的低重要性的节点;但是,当池化率进一步降低时,性能会变得稍差,这种下降可能是特征信息损失导致的。

表6 不同池化率的性能比较

Table 6 Comparison of different pooling ratios

Pool Ratio	Ohsumed	MR	R8	R52	20NG
0.25	70.96	82.36	97.45	95.23	86.79
0.50	71.55	82.84	98.12	96.37	87.91
0.75	71.20	82.65	98.20	96.46	87.63
1.00	70.59	81.90	97.54	95.73	87.27

结束语 文本分类问题可以被看作一个文档与标签之间的匹配问题,因此提升此类模型性能的一个方向就是学习文本的嵌入表达和进行高效的特征组合。本文提出了一个融合多图卷积与层级池化的文本分类模型(HTGNN),该模型首先从单词共现模式和语义嵌入信息的角度将文本数据构建成文本共现图和文本嵌入图,利用图卷积和图池化层遵循分层模式并按层学习文档级表示,并引入门控融合模块对两个图的嵌入进行自适应的融合。在文本表示学习模块中,本文提出了一种新的图池化方法 DAPool。DAPool 首先考虑节点的信息融合,提出了 V2T 自注意用来捕获节点与一阶邻居的关联,然后用 LFCCConv 对更新后的节点进行评分,保留小部分重要性更高的节点作为新的池化图,将文档的图级表示输入多层感知机,以分类预测。在 5 个公开数据集上的实验结果表明,相比基准算法,本文模型取得了最佳性能表现。未来的工作会考虑引入外部知识以增强语义表达,进一步提升模型的性能。

参考文献

- [1] KOLANU N, BROWN A S, BEECH A, et al. Natural language processing of radiology reports for the identification of patients with fracture[J]. Archives of Osteoporosis, 2021, 16: 1-8.
- [2] JACOVI A, SHALOM O S, GOLDBERG Y. Understanding convolutional neural networks for text classification[J]. arXiv: 1809.08037, 2018.
- [3] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences[J]. arXiv: 1404.2188, 2014.
- [4] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent

- neural network based language model [C]//Interspeech. 2010; 1045-1048.
- [5] 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.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems, 2017;6000-6010.
- [7] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019;7370-7377.
- [8] HUANG L, MA D, LI S, et al. Text level graph neural network for text classification[J]. arXiv:1910.02356, 2019.
- [9] WANG Y, WANG S, YAO Q, et al. Hierarchical heterogeneous graph representation learning for short text classification[J]. arXiv:2111.00180, 2021.
- [10] CHEN Y Z, LIU X S, SUN L T, et al. Social Network Influence Prediction Algorithm Based on Graph Neural Network[J]. Journal of Nanjing University(Natural Science), 2022, 58(3): 386-397.
- [11] MA Y Q, CAI M L, CHEN M, et al. Drug Interaction Prediction Method Based on Graph Neural Network[J]. Computer Knowledge and Technology, 2022, 18(18):61-63.
- [12] WANG S, LI Y, ZHANG J, et al. Pm2.5-gnn: A domain knowledge enhanced graph neural network for pm2.5 forecasting [C]//Proceedings of the 28th International Conference on Advances in Geographic Information Systems. 2020;163-166.
- [13] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry[C]//International Conference on Machine Learning. PMLR, 2017;1263-1272.
- [14] ZHANG M, CUI Z, NEUMANN M, et al. An end-to-end deep learning architecture for graph classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018.
- [15] YING Z, YOU J, MORRIS C, et al. Hierarchical graph representation learning with differentiable pooling [C]//Advances in Neural Information Processing Systems, 2018;4805-4815.
- [16] MA Y, WANG S, AGGARWAL C C, et al. Graph convolutional networks with eigenpooling [C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery Data Mining. 2019;723-731.
- [17] LEE J, LEE I, KANG J. Self-attention graph pooling [C]//International Conference on Machine Learning. PMLR, 2019; 3734-3743.
- [18] ZHU Z L, RAO Y, WU Y, et al. Research Progress of Attention Mechanism in Deep Learning[J]. Journal of Chinese Information Processing, 2019, 33(6): 1-11.
- [19] LIU G, GUO J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification[J]. Neurocomputing, 2019, 337: 325338.
- [20] SHEN T, ZHOU T, LONG G, et al. Disan: Directional self-attention network for rnn/cnn-free language understanding [C]//Proceedings of the AAAI conference on Artificial Intelligence. 2018.
- [21] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv:1710.10903, 2017.
- [22] ZHANG J, SHI X, XIE J, et al. Gaan: Gated attention networks for learning on large and spatiotemporal graphs[J]. arXiv:1803.07294, 2018.
- [23] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv:1412.6980, 2014.
- [24] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings, 2010;249-256.
- [25] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv:1207.0580, 2012.
- [26] JOULIN A, GRAVE E, BOJANOWSKI P, et al. Bag of tricks for efficient text classification[J]. arXiv:1607.01759, 2016.
- [27] SHEN D, WANG G, WANG W, et al. Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms[J]. arXiv:1805.09843, 2018.
- [28] LIU X, YOU X, ZHANG X, et al. Tensor graph convolutional networks for text classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020;8409-8416.
- [29] RANJAN E, SANYAL S, TALUKDAR P. Asap: Adaptive structure aware pooling for learning hierarchical graph representations [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020;5470-5477.
- [30] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [31] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C]//Advances in Neural Information Processing Systems. 2017;1025-1035.
- [32] QIAO L, ZHANG L, CHEN S, et al. Data-driven graph construction and graph learning: A review[J]. Neurocomputing, 2018, 312: 336-351.



WEI Ziang, born in 1999, postgraduate, is a student member of CCF (No. J8855G). His main research interests include graph neural network and natural language processing.



PENG Jian, born in 1970, Ph.D, professor, Ph.D supervisor, is an outstanding member of CCF (No. 22761D). His main research interests include big data and wireless sensor network.