



## 基于异构图卷积网络的小样本短文本分类方法

袁自勇<sup>1</sup>,高曙<sup>1</sup>,曹姣<sup>2</sup>,陈良臣<sup>1,3</sup>

(1. 武汉理工大学 计算机科学与技术学院, 武汉 430063; 2. 益阳医学高等专科学校 图书馆网络信息中心, 湖南 益阳 413046;  
3. 中国劳动关系学院 应用技术学院, 北京 100048)

**摘要:** 针对小样本短文本分类过程中出现的语义稀疏与过拟合问题,在异构图卷积网络中利用双重注意力机制学习不同相邻节点的重要性的不同节点类型对当前节点的重要性,构建小样本短文本分类模型HGCN-RN。利用BTM主题模型在短文本数据集中提取主题信息,构造一个集成实体和主题信息的短文本异构信息网络,用于解决短文本语义稀疏问题。在此基础上,构造基于随机去邻法和双重注意力机制的异构图卷积网络,提取短文本异构信息网络中的语义信息,同时利用随机去邻法进行数据增强,用于缓解过拟合问题。在3个短文本数据集上的实验结果表明,与LSTM、Text GCN、HGAT等基准模型相比,该模型在每个类别只有10个标记样本的情况下仍能达到最优性能。

**关键词:** 小样本短文本分类;异构图卷积网络;短文本异构信息网络;BTM主题模型;过拟合

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 袁自勇,高曙,曹姣,等.基于异构图卷积网络的小样本短文本分类方法[J].计算机工程,2021,47(12): 87-94.

**英文引用格式:** YUAN Z Y, GAO S, CAO J, et al. Method for few-shot short text classification based on heterogeneous graph convolutional network[J]. Computer Engineering, 2021, 47(12): 87-94.

## Method for Few-Shot Short Text Classification Based on Heterogeneous Graph Convolutional Network

YUAN Ziyong<sup>1</sup>, GAO Shu<sup>1</sup>, CAO Jiao<sup>2</sup>, CHEN Liangchen<sup>1,3</sup>

(1. College of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China;  
2. Library Network Information Center, Yiyang Medical College, Yiyang, Hunan 413046, China;  
3. Applied Technology College, China University of Labor Relations, Beijing 100048, China)

**[Abstract]** To solve the problem of semantic sparseness and overfitting in few-shot classification of short texts, this paper proposes a method for few-shot short text classification, which uses the dual-attention mechanism of a heterogeneous graph convolutional network to learn the importance of different neighbor nodes and the importance of different node types to the current node. The BTM is used to extract topic information from the short text datasets, and then a heterogeneous information network that can integrate entities and topic information is constructed for short texts to solve the problem of semantic sparseness. On this basis, a heterogeneous graph convolutional network using a dual-level attention mechanism and a method for random neighbor reduction is constructed to extract semantic information from the heterogeneous information network. At the same time, the method for random neighbor reduction is used for data enhancement to alleviate the problem of over-fitting. The experimental results on three short text datasets show that compared with the benchmark models such as LSTM, Text GCN and HGAT, the proposed model still achieves state-of-the-art performance when there are only ten labeled samples in per class.

**[Key words]** few-shot short text classification; heterogeneous graph convolution network; heterogeneous information network for short text; BTM topic model; over fitting

**DOI:** 10.19678/j.issn.1000-3428.0059920

**基金项目:** 国家自然科学基金(51679180);中国劳动关系学院中央高校基本科研业务费专项资金项目(21ZYJS017)。

**作者简介:** 袁自勇(1995—),男,硕士研究生,主研方向为自然语言处理;高曙,教授、博士;曹姣,讲师、硕士;陈良臣(通信作者),副教授、博士研究生。

**收稿日期:** 2020-11-05 **修回日期:** 2020-12-31 **E-mail:** 2931266461@qq.com

## 0 概述

随着社交网络规模不断扩大,每天以帖子或评论的形式产生数以百万计的短文本<sup>[1]</sup>,如何有效地分类和管理短文本数据是一个亟待解决的问题。尽管许多机器学习方法在自然语言处理中取得了较优的效果,但在直接应用于短文本分类时性能提升存在瓶颈,原因在于短文本存在稀疏性、不规则性等特点<sup>[2]</sup>,模型无法提取丰富的语义特征。此外,许多神经网络模型在训练时依赖大量标记样本,但在现实中,标记成本昂贵,有些标记需要专业知识,导致获得大量标记数据十分困难。因此,研究面向小样本的短文本分类方法具有重要的理论价值与现实意义。

目前,小样本短文本分类方法主要面临短文本缺乏足够的单词量和上下文信息、较少的标记数据只能提供有限的特征信息等挑战。如何充分利用少量标记样本和其他大量未标记样本是小样本短文本分类中的关键问题<sup>[3]</sup>。

本文构建一种基于异构图卷积网络(Heterogeneous Graph Convolutional Network, HGCN)的小样本短文本分类模型。针对短文本特点构造短文本异构信息网络 STHIN 实现多维语义增强,避免短文本语义稀疏,并构造异构图卷积网络 HGCN-RN 提取 STHIN 中的节点及邻居特征,利用少量标记样本和其他未标记样本中的特征信息实现小样本短文本分类。在此过程中,采用随机去邻法在 STHIN 中引入适量噪声以实现数据增强,解决过拟合问题,采用双重注意力机制分别从节点级别和类型级别捕获不同相邻节点的重要性,解决语义信息重要性不同的问题。

## 1 相关工作与问题定义

### 1.1 基于深度学习的短文本分类

近年来,针对短文本分类的研究获得了广泛的关注。在基于深度神经网络的研究中,常用模型是卷积神经网络<sup>[4]</sup>和递归神经网络<sup>[5]</sup>。学者们提出了多种适用于短文本分类的神经网络模型,如 KIM 等<sup>[6]</sup>提出一种基于词向量多卷积核的卷积神经网络,ZHANG 等<sup>[7]</sup>设计了 Char-CNN,在不使用预训练词向量和语法结构下,实现了文本分类。此外,增强短文本语义特征也是解决稀疏短文本问题的有效方法之一。WANG 等<sup>[8]</sup>通过知识库中的实体和概念来丰富短文本语义。但是这些方法依赖大量的训练数据,并且不包含语义关系。

### 1.2 基于小样本学习的文本分类

目前流行的小样本学习方法大致分为以下4类:

1) 基于迁移学习。此类方法在含有丰富标签的

大规模数据集上训练基础网络,并对特定领域的网络参数进行微调训练,如 HOWARD 等<sup>[9]</sup>在大型通用语料库上训练通用语言模型,并对其进行微调以实现分类。

2) 基于度量学习。此类方法是对样本间距离分布进行建模,使得同类样本靠近,异类样本远离,如 GENG 等<sup>[10]</sup>提出 Induction 网络,使用胶囊网络来动态的表征类向量,度量关系分数,进而完成分类。

3) 基于数据增强。运用增强技术扩充样本集,实现文本的同质异构变化,如 WEI 等<sup>[11]</sup>提出 EDA 技术,包括同义词替换、随机插入、随机交换和随机删除操作。

最近,图卷积网络在小样本学习领域中逐渐引起关注。YAO 等<sup>[12]</sup>基于单词共现和文档单词关系为语料库构建文本图,并训练 Text GCN 进行文本分类。HU 等<sup>[13]</sup>构建基于双重注意力机制的异构图注意力网络模型 HGAT,实现了半监督短文本分类。

综上,多数深度学习方法无法处理现实中仅有少量标记样本的学习场景,小样本学习方法较少关注短文本分类和因标记样本过少引发的过拟合问题。

### 1.3 问题定义

本文研究的问题定义如下:给定  $m$  篇短文本文档  $D = \{d_1, d_2, \dots, d_m\}$ ,  $n$  种类别标签  $C = \{c_1, c_2, \dots, c_n\}$ , 其中,  $s$  篇短文本文档带有对应的类别标记,  $0 < s \ll m$ 。求分类器  $f: D \rightarrow C, f(d_i) = c_j$  表示文档  $d_i$  属于类  $c_j$ 。

## 2 基于 HGCN 的小样本短文本分类模型

### 2.1 模型架构

本文构建基于异构图卷积网络的小样本短文本分类模型。如图1所示,该模型由输入层、特征提取层、语义附加层、图卷积层和输出层组成。

1) 输入层。一方面,将短文本数据集转换为向量形式,输入到语义附加层;另一方面,直接将短文本数据集输入特征提取层,用于提取短文本语义信息。

2) 特征提取层。通过 TagMe 获取实体概念和描述,作为实体特征;使用 BTM 获取文档-主题分布和主题-单词分布,作为主题特征。该层输入短文本数据集,输出为实体特征和主题特征。

3) 语义附加层。以文本、实体、主题为节点,融合来自特征提取层的实体特征、主题特征和来自输入层的短文本特征信息,从而构造短文本异构信息网络 STHIN。

4) 图卷积层。从 STHIN 中获得邻接矩阵、度矩阵和特征矩阵,其中特征矩阵  $X$  为包含  $n$  个节点及其特征的矩阵。将异构图卷积网络用于捕获 STHIN 中的节点信息及其关系,并通过双重注意力机制和随机去邻法捕获不同语义信息的重要性和缓解模型过

拟合问题。该层的输入为 STHIN, 输出为向量矩阵形式的卷积结果。

5) 输出层。通过 softmax 函数输出  $n$  个节点的类别标签,  $Y_i$  表示节点  $i(i=1, 2, \dots, n)$  的类别标签。该层输入为图卷积层的卷积结果, 输出为所有节点的类别标签。

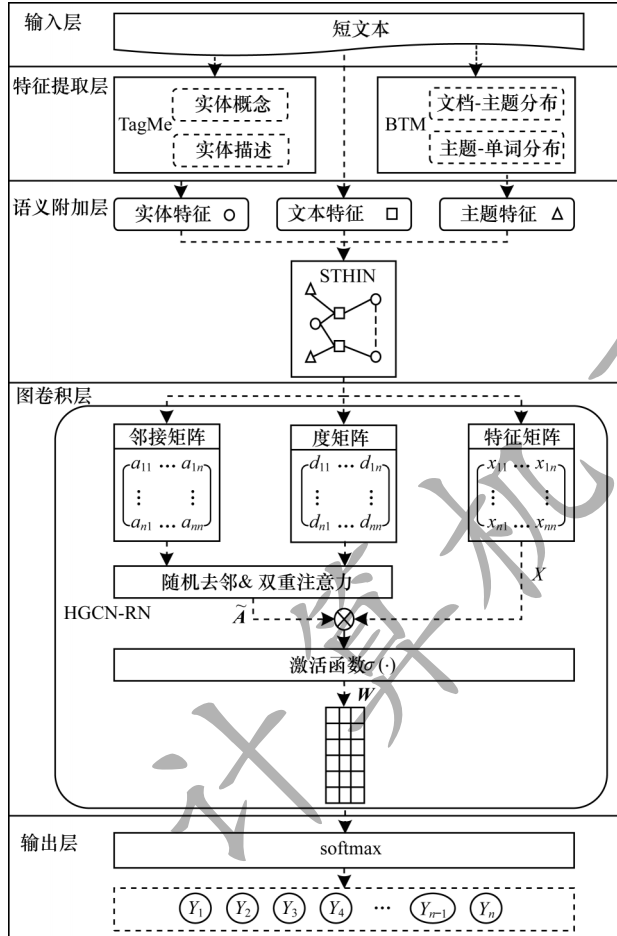


图1 基于 HGCN 的小样本短文本分类模型架构

Fig.1 Framework of few-shot short text classification model based on HGCN

## 2.2 短文本异构信息网络 STHIN

在特征提取层和语义附加层, 使用 BTM 主题模型<sup>[14]</sup>挖掘潜在主题信息并构造短文本异构信息网络 STHIN。

构造 STHIN 的具体过程如下:

1) 针对短文本特点, 使用 BTM 主题模型挖掘潜在主题  $T$ 。在获取文档-主题矩阵和主题-词矩阵后, 选择概率最高的前  $N$  个主题作为短文本  $D$  的潜在主题。因此, 每个短文本节点会有  $N$  条边与主题节点相连。

2) 使用实体链接工具 TagMe<sup>[15]</sup>将短文本  $D$  中的实体提及 (entity mention) 映射为 Wikipedia 中的实体  $E$ , 并使用基于 Wikipedia 语料库预训练的 word2vec 学习实体的词嵌入。如果短文本  $D$  包含  $K$  个实体,

则短文本节点会有  $K$  条边与实体节点相连。此外, 当实体节点之间的相似度高于预定阈值时, 则在该实体对之间建立边, 用于合并更多的语义信息。

STHIN 算法描述如下所示:

### 算法1 STHIN 构造算法

输入 训练集  $D_{\text{train}}$ , 阈值  $\delta$

输出 异构信息网络图 STHIN

从训练集  $D_{\text{train}}$  中的短文本  $D$  上提取实体信息  $E$  和潜在主题  $T$

建立图  $G$ , 以  $D, E, T$  为节点, 与  $D$  相关的  $E, T$  之间建立边  $D \leftrightarrow E, D \leftrightarrow T$

foreach ( $e_i, e_j \in E, i \neq j$ ) do

If  $\text{sim}(e_i, e_j) > \delta$  //  $\text{sim}$  为相似度计算函数

建立边:  $e_i \leftrightarrow e_j$

end

返回图  $G$ , 即生成的异构信息网络图 STHIN

STHIN 借助实体、主题等附加信息丰富了短文本语义特征, 其中潜在主题通过 BTM 主题模型直接在短文本语料库中训练获得, 与 LDA 相比, BTM 是在整个语料库中对词的共现模式进行建模, 解决了稀疏词的共现问题, 更适合于短文本。因此, 在短文本缺乏足够的单词量和上下文信息的情况时, 异构信息网络 STHIN 实现了短文本的多维语义增强, 有效地解决了短文本语义稀疏问题。

## 2.3 HGCN-RN 模型

### 2.3.1 异构图卷积网络 HGCN

语义附加层输出的 STHIN 中仅存在少量标记样本, 但未标记样本与标记样本之间以多种关系连接。本文改进传统图卷积网络 GCN 的传播规则, 将异构图卷积网络 (HGCN) 用于捕获 STHIN 中的各种类型节点信息及其关系, 从而充分利用标记样本和其他未标记样本的信息, 有效解决小样本问题。

将 2.2 节构建的 STHIN 表示为  $G=(V, E)$ , 其中  $V(|V|=n)$  和  $E$  是节点和边的集合, 获得邻接矩阵  $A$  和度矩阵  $D$ 。为了汇总节点本身的特征, 邻接矩阵  $A$  需要添加自环,  $A' = A + I$ 。设  $X \in \mathbb{R}^{n \times m}$  为包含所有节点及其特征的矩阵, 其中每行  $x_v \in \mathbb{R}^m$  是一个节点  $v$  的特征向量,  $m$  是特征向量的维数。则图卷积 GCN 的层传播规则为如式 (1) 和式 (2) 所示:

$$H^{(l+1)} = \sigma(\tilde{A} \cdot H^{(l)} \cdot W^{(l)}) \quad (1)$$

$$\tilde{A} = D^{-\frac{1}{2}} A' D^{-\frac{1}{2}} \quad (2)$$

其中:  $\tilde{A}$  代表归一化对称邻接矩阵;  $W^{(l)}$  是层特定的可训练权重矩阵;  $\sigma(\cdot)$  表示激活函数,  $\sigma(x) = \max(0, x)$ ;  $H^{(l)} \in \mathbb{R}^{n \times m}$  表示第  $l$  层中节点的隐藏表示,  $H^{(0)} = X$ 。

为了集成各种类型的节点信息并将它们各自的转换矩阵投影到一个公共空间中, 改进式 (1) 得到式 (3) 所示的异构图卷积传播规则:

$$H^{(l+1)} = \sigma\left(\sum_{\tau \in T} \tilde{A}_{\tau} \cdot H_{\tau}^{(l)} \cdot W_{\tau}^{(l)}\right) \quad (3)$$



其中:  $\tilde{A}_\tau \in \mathbb{R}^{n \times |V_\tau|}$  表示类型为  $\tau$  的归一化对称邻接矩阵;  $W_\tau^{(0)} \in \mathbb{R}^{m^{(0)} \times m^{(l+1)}}$  考虑了不同特征空间的差异,  $H_\tau^{(0)} \in X_\tau$ 。

本文利用异构图卷积网络捕获 STHIN 中的各种类型节点信息及关系, 充分利用有限的标记样本, 聚集节点邻域信息并进行前向传播, 从而解决小样本问题。

### 2.3.2 随机去邻法与双重注意力机制

图卷积层中将异构图卷积网络 HGCN 用于捕获 STHIN 中的节点信息, 但是由于不同类型节点的语义信息重要性不同, 需要结合双重注意力机制捕获不同类型节点的重要性。此外, 因标记样本过少引发的过拟合问题也会降低模型的性能, 因此本文采用随机去邻法, 在 STHIN 中引入适量噪声以实现数据增强。

1) 随机去邻法由于小样本学习中标记样本的数量有限, 导致模型学习时提取的数据特征较少, 削弱了模型的泛化能力, 易出现过拟合问题。文本数据增强技术可以有效解决小样本条件下样本不足的问题, 目前主要有回译法、EDA<sup>[11]</sup>、基于语言生成模型的文本数据增强等技术, 但是与图像的像素变换有着很大的差异性, 文本中单词的随机变换会导致语义的缺失或者歧义产生, 在对文本的同质异构变化过程中引入过多噪声。

为了实现数据增强并缓解过拟合现象, 本文在 STHIN 中随机删除边, 同时引入噪声信息, 降低模型收敛速度, 该方法<sup>[16]</sup>在基于图神经网络 GNN 的节点分类研究中已被证明有助于图卷积网络的加深和过拟合问题的缓解。为了防止引入过多噪声导致异构信息图的结构变形, 仅从 STHIN 中删除一定比例的边。

在图卷积层的每次训练时, 从短文本异构信息网络 STHIN 中随机丢弃一定比例的边, 即随机删除部分节点的邻居。邻接矩阵  $A'$  由  $A_{\text{drop}}$  表示, 如式(4)所示:

$$A_{\text{drop}} = A' - A_p \quad (4)$$

其中:  $A_p$  表示邻接矩阵中非零元素以概率  $p$  随机设置为 0 的位置。

从邻居节点信息聚合的角度分析, 异构图卷积网络 HGCN 聚合每个节点的所有邻居及其自身的信息进行消息传递, 以获取到节点的高阶特征, 而随机去邻法使得 HGCN 在训练时随机聚合邻居子集, 而非所有邻居信息, 即干扰图形连接, 引入适量的噪声信息, 进而导致图卷积层输入的 STHIN 信息具有随机性和多样性, 实现了数据增强。

此外, 以概率  $p$  随机去邻使得聚合邻居信息的期望与  $p$  相关, 而对权重进行归一化后则与  $p$  无关, 因此, 随机去邻法未改变邻居聚合的期望, 与旋转、裁剪等典型的图像数据增强方法相似, 随机去邻法

可以有效防止过拟合问题。

2) 双重注意力机制给定 STHIN 中一个特定节点, 不同类型的相邻节点对其具有不同的影响。例如, 当实体类型节点携带的信息多于主题类型节点时, 应设置实体类型的相邻节点高权重, 为异构图卷积网络的分类提供更有效的邻居信息。此外, 相同类型的不同相邻节点也具有不同的重要性, 例如同一种类型的不同主题节点中包含不同程度的潜在主题信息。

为了同时捕获节点级别和类型级别的不同重要性, HU 等<sup>[13]</sup>提出了双重注意力机制。其中: 类型级别的注意力机制用于学习不同类型的相邻节点的权重, 基于当前节点的词嵌入和类型嵌入计算类型级别的注意力分数; 节点级别的注意力机制用于捕获不同相邻节点的重要性并减少噪声节点的权重, 在类型级注意力权重的基础上, 获得节点的词嵌入, 并计算节点级别的注意力分数。双重注意力机制同时从类型级别和节点级别捕获不同相邻节点的重要性, 捕获关键语义信息, 降低噪声权重, 结合异构图卷积网络能聚合更有效的邻居信息, 提高分类性能。

### 2.4 模型训练

构造短文本异构信息网络 STHIN 后, 将其输入到基于双重注意力机制和随机去邻的异构图卷积网络 HGCN-RN 中, 传播规则如式(5)所示:

$$H^{(1)} = \text{softmax} \left( \sum_{\tau \in T} \tilde{A}_{\tau+\text{drop}} \cdot \text{RELU} \left( \sum_{\tau \in T} \tilde{A}_{\tau+\text{drop}} \cdot X_r^{(0)} \cdot W_r^{(0)} \right) \cdot W_r^{(1)} \right) \quad (5)$$

其中:  $\tilde{A}_{\tau+\text{drop}}$  代表随机去邻后类型  $\tau$  的标准化对称邻接矩阵。softmax 函数如式(6)所示:

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_i \exp(x_i)} \quad (6)$$

$\text{RELU}(x) = \max(0, x)$ ,  $W^{(0)}$  和  $W^{(1)}$  通过梯度下降法优化。所有带标签样本的损失函数定义为交叉熵损失, 并添加正则项  $L_2$  以减少模型的过拟合, 如式(7)所示:

$$L = - \sum_{s \in S_{\text{train}}} \sum_{c=1}^C Y_{sc} \ln Z_{sc} + \lambda \|\Theta\|_2 \quad (7)$$

其中:  $S_{\text{train}}$  是用于训练的短文本索引的集合;  $C$  表示类的数量;  $Y$  是相应的标签指示矩阵;  $\Theta$  是模型参数;  $\lambda$  表示正则化因子。

## 3 实验与结果分析

### 3.1 基准模型

实验采用以下 6 个基准模型:

1) CNN 模型。卷积神经网络<sup>[6]</sup>分为 2 种形式: 使用随机初始化词向量的 CNN-rand 和使用预训练词向量的 CNN-pre。

2) LSTM 模型。LSTM<sup>[17]</sup>使用最后一个隐藏状态作为整个文本的表示形式,构建随机初始化词向量的 LSTM-rand 和预训练词向量的 LSTM-pre2 种模型。

3) PTE 模型。预测性文本嵌入<sup>[18]</sup>,构造一个以单词、文档和标签作为节点的异构文本网络,并将平均词向量作为文档表示进行文本分类。

4) LEAM 模型。标签嵌入注意模型<sup>[19]</sup>,将单词和标签联合嵌入到同一空间,使用标签描述信息对词向量矩阵加权,并实现 MLP 分类器进行文本分类。

5) Text GCN 模型。文本图卷积网络<sup>[12]</sup>,将文本语料库建模为包含文档节点和单词节点的异构文本图,并应用 GCN 来实现文本分类。

6) HGAT 模型。异构图注意力网络<sup>[13]</sup>,该模型使用实体、主题和文档作为节点来构建异构信息图,其中主题通过 LDA 获取并使用概率分布表示,并将

GCN 与双重注意力机制相结合实现文本分类。

### 3.2 数据集

实验使用以下 3 个数据集:

1) AGNews 数据集<sup>[7]</sup>。本文实验从该数据集中随机选择 10 000 个样本,并将其分为 4 类。

2) Snippets 数据集<sup>[20]</sup>。使用多个领域的短语在搜索引擎中以短文本形式检索文本片段,该数据集共包含 12 340 个数据,涉及商业、计算机、教育等多个领域。

3) MR 数据集<sup>[21]</sup>。该数据集由电影评论及其情感标签组成,每条评论都由正负标记,用于二分情感分类。

表 1 显示了这 3 个数据集的统计信息。对于每个数据集,在每个类别中随机选择 10 个标记样本,其中一半用于训练,另一半用于验证。每个类别中剩余的样本作为无标签样本使用。

表 1 实验数据集信息

Table 1 Information of datasets for experiment

数据集	文档数	token 平均数	包含实体的文档比例/%	节点数	边数	类型数
AGNews	10 000	19.3	76	16 629	203 890	4
Snippets	12 340	14.5	94	21 404	346 210	8
MR	10 662	7.6	76	15 041	117 126	2

### 3.3 实验环境

模型训练的实验环境和配置如表 2 所示。Python 模块主要包括自然语言处理库 genism 3.8.3、机器学习库 scikit-learn 0.23.1 和数学运算库 numpy 1.18.5。对于深度学习网络的构建,本文实验使用 PyTorch 深度学习库,该库具有简洁高效的优势,支持动态计算图,并且在编写神经网络时具有高度的灵活性。

表 2 实验环境与配置

Table 2 Experimental environment and configuration

项目	配置
操作系统	Windows 10
CPU/内存	quad-core/8 GB
语言	Python 3.7.2
深度学习框架	PyTorch 1.4.0

### 3.4 评价指标

实验采用准确率和 F1 值作为模型分类效果的评估指标。

假设类别总数为  $C$ ,在类别  $c_i$  的分类结果中,正确分为该类别的样本数为  $a$ ,错误分类的样本数为  $b$ ,将属于  $c_i$  类的样本错误地分为其他类的样本数为  $c$ ,在其他类上正确分类的样本数为  $d$ 。

准确率代表所有正确预测的样本占总数的比例,计算公式如式(8)所示:

$$A_{\text{accuracy}} = \frac{a+d}{a+b+c+d} \quad (8)$$

精确率  $P = \frac{a}{a+b}$  衡量类别  $c_i$  的分类准确性,  $R = \frac{a}{a+c}$  衡量类别  $c_i$  的召回率。F1 值是召回率和准确率的加权平均,计算公式如式(9)所示:

$$F1 = \frac{2 \times PR}{P + R} \quad (9)$$

### 3.5 随机去邻实验及分析

为探索随机去邻法的有效性,在数据集上比较有无随机去邻对分类效果的影响。首先分别对各个数据集构造短文本异构信息网络 STHIN,其中主题数  $K$ 、最相关主题数  $P$  和实体间相似性阈值  $\delta$  选用通过调参实验得到的最优值,预训练词向量维度为 100;然后建立 2 个有无随机去邻的异构图卷积网络模型,隐藏层维度为 512,分别输入 STHIN,并统计分类准确率。限于篇幅,且考虑到 3 个数据集实验分析结论相似,本文仅选取 AGNews 数据集的随机去邻实验结论分析。实验结果如图 2 所示,可以看出,随机去邻的方法能够降低模型收敛速度,有效缓解过拟合问题。对 STHIN 中的边进行随机删除,增强了模型的鲁棒性,使分类平均准确率提高了 1%。

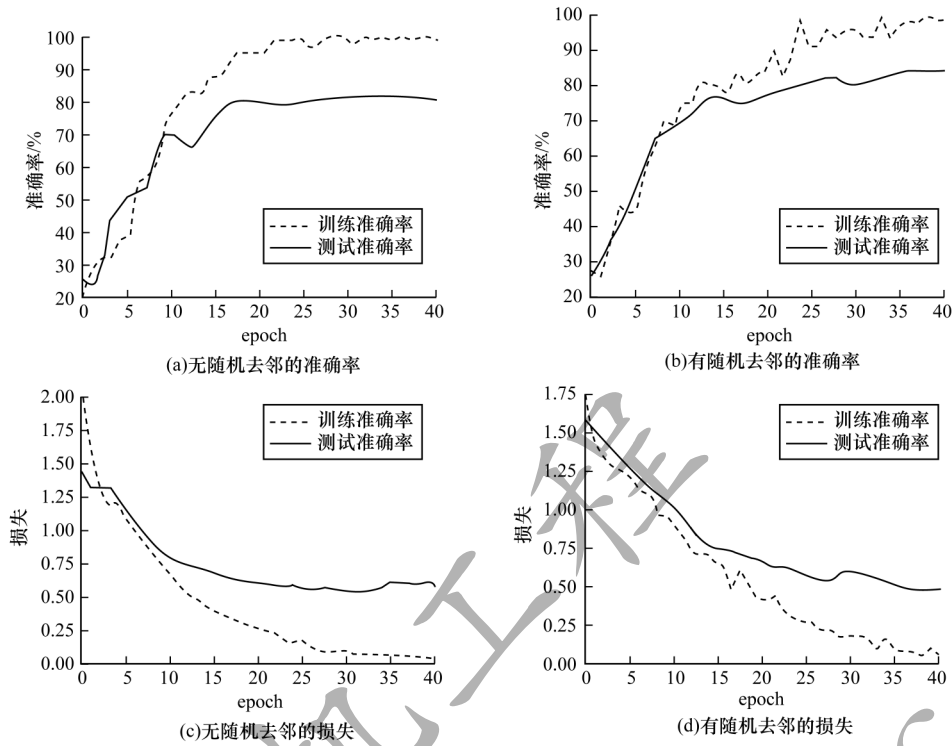


图2 AGNews上有/无随机去邻的准确率和损失

Fig.2 Accuracy and loss with/without reducing neighbors on AGNews

进一步地,探索随机删除边的比例对性能改进的影响。在AGNews数据集上不同比例的随机去邻对分类性能的影响如图3所示,可以看出,当5%左右的边被随机删除时可以获得最佳结果,而当删除更多的边时,由于破坏STHIN节点之间的传递性而引起的负面影响将会大于缓解过拟合问题的积极影响。

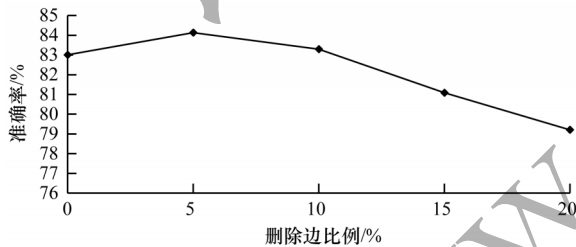


图3 AGNews上不同删除边比例条件下的准确率

Fig.3 Accuracy with different percentages of deleting edges on AGNews

### 3.6 标记样本数量实验及分析

选择Text GCN、HGAT和本文的HGCN-RN这三组性能较优的分类模型,在数据集上训练20次,取平均值作为最终结果,研究标记样本数量对分类效果的影响。限于篇幅且考虑3个数据集实验分析结论相似,本文仅选取AGNews数据集的标记样本数量实验结论分析。实验结果如图4所示,可以看出,随着标记样本数量的增加,所有模型在准确率方面都有所提升,其中本文的HGCN-RN模型性能最优。当每个类别中只有几十个标记样本时,其他模型的性能会显著下降,但是HGCN-

RN模型的准确率仍然优于所有其他模型。这表明基于图卷积网络的方法可以通过消息传播更好地利用有限的标记数据,而且STHIN、双重注意力机制和随机去邻法有效的缓解了短文本稀疏问题和过拟合问题,使HGCN-RN模型在小样本条件下依然具有较好的短文本分类效果。

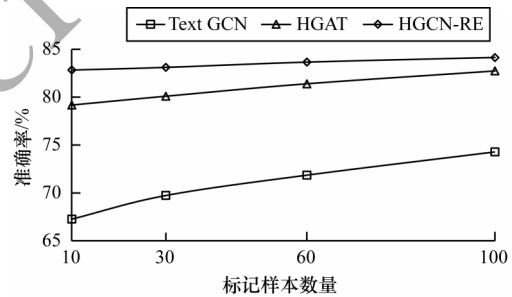


图4 AGNews上不同标记样本数量条件下的准确率

Fig.4 Accuracy with different number of labeled samples on AGNews

### 3.7 调参实验及分析

限于篇幅且考虑3个数据集实验分析结论相似,本文仅选取AGNews数据集的调参实验结论分析。

图5展示了AGNews数据集上主题数量 $K$ 、相关主题数 $P$ 和实体相似度阈值 $\delta$ 对分类性能的影响,分析如下:

1)图5(a)显示,当主题数达到15时,分类准确率最高,进一步增加 $K$ 值时,准确率会持续下



降。原因在于当主题数 $K$ 较小时,主题范围广,此时短文本可能仅属于一个主题。随着主题数的增加,将学习更多特定主题,为短文本选择最相关的 $P$ 个主题后,其他相关的主题信息将会丢失。

2)图5(b)显示分类准确率首先随着 $P$ 的增加而增加,而 $P>2$ 时性能下降。这表明模型对相关主题数 $P$ 较为敏感,当 $P$ 值较小时,与短文本语义有关的部分潜在主题信息丢失,导致模型只能提取有限信

息;当 $P$ 值较大时,模型会获取过多噪声主题信息,影响分类性能,因此 $P$ 值过大或过小都会对模型的分类效果产生消极影响。

3)图5(c)显示了实体之间相似性阈值 $\delta$ 对分类效果的影响。可以看出,当阈值设置得太高时,将导致STHIN中的边数大幅度减少,从而减少消息的传播。当阈值设置太低时,会连接一些语义相似度较低的实体,导致引入过多的噪音信息。

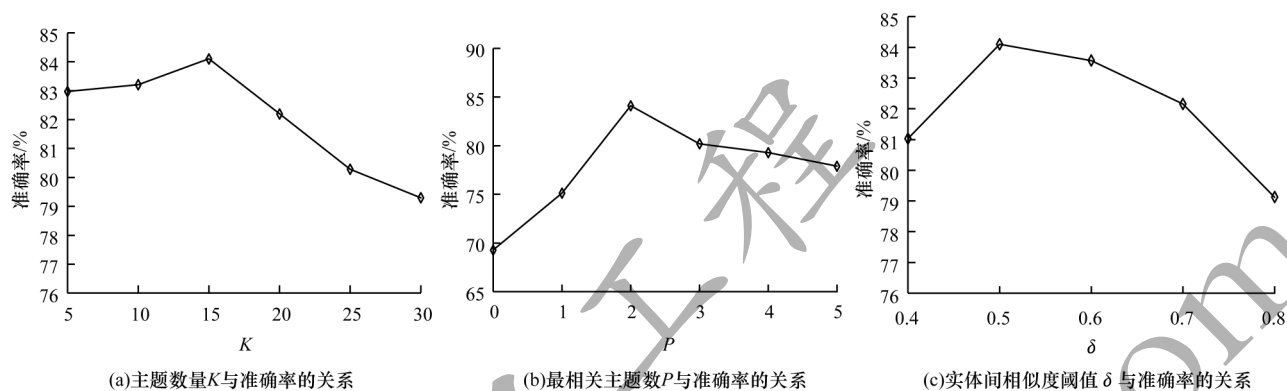


图5 AGNews上不同主题数量、相关主题数和实体间相似度阈值条件下的准确率

Fig.5 Accuracy with different topic numbers, related topics numbers and similarity threshold between entities on AGNews

### 3.8 模型评估及分析

根据以上实验结果及分析,选择合适的参数训练HGCN-RN模型。学习率设为0.005,dropout率为0.8,正则化因子 $\lambda=5e-6$ 。如果连续10个epoch关于验证集上的损失没有减少,则训练停止。

表3展示了3个基准数据集上不同模型的分类准确率和F1值,分析如下:

1)使用预训练向量的CNN-pre和LSTM-pre相比于使用随机初始化单词向量的CNN-rand和LSTM-rand的分类性能有了显著提升。原因在于它们根据不同的分类任务对词向量进行预训练,并且在模型训练过程中,词向量也被用作可优化的参数。

2)基于单词共现学习文本嵌入的PTE模型性能较差,原因是短文本中缺乏足够的单词量和上下文信息,而LEAM模型在3个数据集上的准确率比PTE提升了30.14%、10.10%和5.25%,证明了标签描述和嵌入的有效性。

3)基于图卷积网络的Text GCN和HGAT模型取得了竞争性的结果,这也证明了构造文本异构图并将其应用于GCN的方法可以有效提高文本分类性能。但是,由于它们未能充分考虑短文本特性或过拟合问题,导致分类性能低于HGCN-RN模型。

4)HGCN-RN模型的分类性能优于所有基准模型,相比于HGAT模型在AGNews、Snippets、MR数据集上的准确率分别提升了4.97%、7.81%、2.79%,在F1值上分别提升了5.38%、7.92%、2.82%,这表明HGCN-RN模型在小样本短文本分类中的有效性。

表3 3个标准数据集上的平均准确率和F1值

Table 3 Average accuracy and average F1 values on three standard datasets

模型	AGNews		Snippets		MR	
	准确率	F1值	准确率	F1值	准确率	F1值
CNN-rand	32.731	32.360	48.135	47.630	54.825	54.451
CNN-pre	67.074	67.574	77.030	76.897	58.352	57.970
LSTM-rand	31.023	30.623	26.336	26.218	52.692	52.421
LSTM-pre	66.278	65.800	75.896	75.490	59.899	59.486
PTE	36.837	36.818	63.251	62.832	54.027	53.725
LEAM	66.981	67.021	73.351	73.073	59.275	58.931
Text GCN	67.265	66.750	72.622	72.612	56.173	55.973
HGAT	79.167	78.719	75.275	74.892	59.531	59.197
HGCN-RN	84.136	84.095	83.087	82.813	62.317	62.020

HGCN-RN模型性能最优的原因有以下3点:

1)考虑短文本特征,通过BTM主题模型提取主题并构建短文本异构信息网络STHIN,可以同时捕获文本、主题和实体之间关系,丰富短文本语义信息。

2)结合小样本条件下标记样本少和未标记样本多的特点,应用异构图卷积聚集STHIN中节点邻域信息进行前向传播,充分利用了有限的特征信息。

3)应用双重注意力机制捕获STHIN中不同节点语义信息的重要性,并采用随机去邻法在图中引入适量噪声,降低了模型收敛速度,有效缓解了过拟合问题。

## 4 结束语

本文构建一种基于异构图卷积网络的小样本短文本分类模型。该模型充分考虑短文本特征,使用

BTM主题模型挖掘短文本潜在主题信息,建立短文本异构信息网络STHIN灵活地合并短文本语义信息及其关系。此外,结合小样本条件下标记样本少和未标记样本多的特点,应用异构图卷积和双重注意力机制捕捉STHIN中丰富的语义信息,并采用随机去邻法在异构图中引入适量噪声,缓解过拟合问题。后续将考虑构建堆叠的异构图卷积网络模型,在每个节点合并信息时,累积不同距离处的节点信息,进一步提高分类性能。

### 参考文献

- [1] ALSMADI I, GAN K H. Review of short-text classification[J]. International Journal of Web Information Systems, 2019, 15(2): 155-182.
- [2] 高云龙, 吴川, 朱明. 基于改进卷积神经网络的短文本分类模型[J]. 吉林大学学报(理学版), 2020, 58(4): 923-940. GAO Y L, WU C, ZHU M. Short text classification model based on improved convolutional neural network[J]. Journal of Jilin University( Science Edition ), 2020, 58(4): 923-940. (in Chinese)
- [3] JOULIN A, GRAVE E, BOJANOWSKI P, et al. Bag of tricks for efficient text classification[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. [S. l.]: ACL Press, 2017: 427-431.
- [4] SHIMURA K, LI J, FUKUMOTO F. HFT-CNN: learning hierarchical category structure for multi-label short text categorization[C]//Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. [S. l.]: ACL Press, 2018: 811-816.
- [5] SINHA K, DONG Y, CHEUNG J C K, et al. A hierarchical neural attention-based text classifier[C]// Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. [S. l.]: ACL Press, 2018: 817-823.
- [6] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of 2014 Conference on Empirical Methods in Natural Language Processing. [S. l.]: ACL Press, 2014: 1746-1751.
- [7] ZHANG X, ZHAO J B, LECUN Y. Character-level convolutional networks for text classification[C]// Proceedings of NIPS' 15. Cambridge, USA: MIT Press, 2015: 649-657.
- [8] WANG J, WANG Z Y, ZHANG D W, et al. Combining knowledge with deep convolutional neural networks for short text classification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York, USA: IJCAI Press, 2017: 2915-2921.
- [9] HOWARD J, RUDER S. Universal language model fine-tuning for text classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. [S. l.]: ACL Press, 2018: 328-339.
- [10] GENG R Y, LI B H, LI Y B, et al. Few-shot text classification with induction network[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. [S. l.]: ACL Press, 2019: 3904-3913.
- [11] JASON W, KAI Z. EDA: easy data augmentation techniques for boosting performance on text classification tasks[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. [S. l.]: ACL Press, 2019: 6381-6387.
- [12] YAO L, MAO C S, LUO Y. Graph convolutional networks for text classification[C]//Proceedings of 2019 AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2019: 7370-7377.
- [13] HU L M, YANG T C, SHI C, et al. Heterogeneous graph attention networks for semi-supervised short text classification[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. [S. l.]: ACL Press, 2019: 4821-4830.
- [14] YAN X H, GUO J F, LAN Y Y, et al. A biterm topic model for short texts[C]//Proceedings of the 22nd International Conference on World Wide Web. New York, USA: ACM Press, 2013: 1445-1456.
- [15] FERRAGINA P, SCAIELLA U. TAGME: on-the-fly annotation of short text fragments (by Wikipedia entities) [C]// Proceedings of the 19th ACM International Conference on Information and Knowledge Management. New York, USA: ACM Press, 2010: 1625-1628.
- [16] RONG Y, HUANG W B, XU T Y, et al. DropEdge: towards deep graph convolutional networks on node classification [EB/OL]. [2020-10-11]. <https://openreview.net/forum?id=Hkx1qkrKPr>.
- [17] LIU P F, QIU X P, HUANG X J. Recurrent neural network for text classification with multi-task learning [C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA: IJCAI/AAAI Press, 2016: 2873-2879.
- [18] TANG J, QU M, MEI Q Z. PTE: predictive text embedding through large-scale heterogeneous text networks [C]// Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2015: 1165-1174.
- [19] WANG G Y, LI C Y, WANG W L, et al. Joint embedding of words and labels for text classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. [S. l.]: ACL Press, 2018: 2321-2331.
- [20] PHAN X H, NGUYEN L M, HORIGUCHI S. Learning to classify short and sparse text & Web with hidden topics from large-scale data collections[C]//Proceedings of the 17th International Conference on World Wide Web. New York, USA: ACM Press, 2008: 91-100.
- [21] BO P, LEE L. Seeing Stars: exploiting class relationships for sentiment categorization with respect to rating scales[C]// Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics. [S. l.]: ACL Press, 2005: 115-124.