# 基于双超图神经网络特征融合的文本分类

郑诚[1,2]，李鹏飞[1,2]

（1.安徽大学计算机科学与技术学院，安徽 合肥 230601；2.计算智能与信号处理教育部重点实验室，安徽 合肥 230601）

**摘 要：**近年来，图神经网络（GNN）在文本分类任务中受到广泛应用。当前基于 GNN 的文本分类模型首先将文本建模为图，然后使用 GNN 对文本图进行特征传播与聚合，但是此类方法有两点不足：一是现有模型由于图结构的限制无法捕获单词之间的高阶语义关系；二是现有模型无法捕获文本中的关键语义信息。为了解决上述问题，提出一种基于双超图卷积网络特征融合的文本分类模型。一方面，使用原始文本建立文本超图；另一方面，为短文本引入外部知识，使用基于 SenticNet 词库的外部知识对文本进行语义增强，构建语义超图。经过超图卷积后通过注意力机制对双超图特征进行融合，实现短文本分类。在 4 个文本分类数据集上的实验结果表明，该模型优于基线模型，具有优越的文本分类性能。

**关键词：**文本分类；超图；特征融合；SenticNet 词库；自然语言处理

**中图分类号：**TP391 **文献标志码：**A **DOI：**10.19678/j.issn.1000-3428.0068324

# Text Classification Based on Feature Fusion of Dual Hypergraph Neural Networks

ZHENG Cheng[1,2], LI Pengfei[1,2]

(1. School of Computer Science and Technology, Anhui University, Hefei 230601, Anhui, China;
2. Key Laboratory of Computational Intelligence and Signal Processing, Ministry of Education, Hefei 230601, Anhui, China)

**【Abstract】** In recent years, Graph Neural Networks (GNNs) have been widely used for text classification tasks. Current models based on GNNs first model the text as a graph and then use GNNs to propagate and aggregate the features of the text graph. However, these methods have two notable limitations. First, existing models cannot capture high-order semantic relationships between words because of the limitations of graph structures. Second, existing models cannot capture key semantic information from the text. To address these issues, this paper proposes a text classification model based on the feature fusion of dual hypergraph convolutional networks. On one hand, the original text is used to construct a text hypergraph; on the other hand, external knowledge is introduced for short texts. The text is semantically enhanced using external knowledge based on the SenticNet lexicon, and a semantic hypergraph is constructed. After hypergraph convolution, an attention mechanism is used to fuse the features of the dual hypergraphs for short-text classification. Experimental results on four text classification datasets show that the proposed model outperforms the baseline methods and demonstrates superior text classification performance.

**【Key words】** text classification; hypergraph; feature fusion; SenticNet lexicon; natural language processing

## 0 引言

文本分类是将一篇文档自动归属到一个类别中，在情感分析、主题标记、自然语言推理等方面得到广泛应用。

文本分类方法根据其发展历程可以分为两类，一类是基于浅层学习的文本分类方法，另一类是基于深度学习的文本分类方法。浅层学习是指基于统计的模型，例如支持向量机（SVM）[1]、$k$-近邻（$k$-NN）[2]等。这类方法首先通过文本特征提取算法将文本转化为向量，然后使用文本分类算法对这些向量进行分类。常用的文本特征提取算法有词袋模型[3]、向量空间模型[4]、Word2Vec[5]等。与之前基于规则的方法相比，这些文本分类方法具有更高的准确度并且更加稳定，但此类文本分类方法也有着诸多限制，比如特征工程通常需要人工完成，并且十分依赖领域知识。自 2010 年以来，文本分类领域的研究重点已从浅层学习模型转换到深度学习模型。基于深度学习的文本分类方法可以自动地获得文本特征。例如卷积神经网络（CNN）[6]和循环神经网络（RNN）[7]，此类深度学习模型考虑到了文本中隐含的序列信息，能够从单词序列中捕获单词之间的语义信

**收稿日期：**2023-09-05 **修回日期：**2024-01-24

**基金项目：**安徽省重点研究与开发计划项目（202004d07020009）。

**通信作者 E-mail：**csahu@126.com

息,但是此类模型忽略了文本之间的全局联系和远距离语义关系,无法捕获单词之间的全局语义信息。

图神经网络(GNN)在文本分类领域的应用较好地解决了上述问题。YAO 等[8]提出的用于文本分类的图卷积网络(TextGCN)模型将 GCN 应用到文本分类任务中,该模型首先通过图结构对文本进行建模,使用图中节点表示单词和文档,然后使用 GCN 进行特征传播与聚合。TextGCN 模型将整个语料库构建为一个大型无向图,因为图中的节点是构建时确定的,该模型无法为新样本进行分类。HUANG 等[9]提出的文档级图神经网络(Text-Level GNN)解决了该问题,该模型为每个文本构建一个图,以单词为节点,使用滑动窗口建立节点之间的边,并通过消息传播机制更新节点信息。上述这种 GNN 文本分类方法受制于图结构的约束,只能在单词之间建立一对一的联系,而不能对单词之间的高阶关系进行建模。DING 等[10]提出的超图注意力网络(HyperGAT)模型则通过把文本转化为超图[11],从而对单词之间的高阶关系进行建模,然而其中使用主题模型构建的语义超边无法充分捕获文本中的语义信息。

本文针对上述问题,提出了一种利用双超图融合的方法来进行文本分类,在保留文本超图的基础上引入外部知识库 SenticNet[12]丰富语义信息并构建语义超图。首先通过超图神经网络进行特征传播与聚合,分别得到文本超图特征和语义超图特征;然后通过注意力机制进行特征融合[13]。

本文主要贡献如下:

1)提出一种基于双超图神经网络融合的文本分类模型,使用超图对文本进行建模能够捕获自然语言中的高阶关系。

2)通过注意力机制对双超图特征进行融合,提高文本表示中关键信息的重要性,有效增强模型的文本表示能力。

3)在 4 个短文本分类数据集上的实验结果表明,该模型优于其他的基线模型。

## 1 相关工作

随着深度学习的发展,深度学习模型已经广泛应用于文本分类任务。KIM 等[14]设计了一种用于文本分类的卷积神经网络(TextCNN)。该模型首先将文本转化为特征矩阵,并利用多个不同大小的卷积核对特征矩阵进行卷积,从而能够更好地捕捉文本的局部语义关系,然后通过特定的池化层捕获文本的显著特征,最后经过全连接层进行分类。TextCNN 模型进行文本分类时,其输入维度和卷积核大小是固定的,这使得模型不能捕获文本中的远距离依赖信息和全局信息。在 TextCNN 之后,ZHANG 等[15]提出了用于文本分类的字符级卷积神经网络(CharCNN)模型,与 TextCNN 不同的是,CharCNN 模型是更小粒度的文本分类模型,输入为字符级别的文本数据,并采用卷积加池化操作来提取有意义的特征并分类。LIU 等[16]提出了用于文本分类的循环神经网络(TextRNN)模型,将 RNN 用于文本分类。与 TextCNN 模型相比,TextRNN 可以处理不定长的输入,能够建模更长的序列信息,捕获文本中的远距离依赖信息和全局语义关系。另外,用于文本分类的双向长短期记忆(LSTM)网络[17]模型取得了较单向 LSTM 网络模型更出色的效果。尽管上述方法取得了不错的结果,但它们具有局限性:主要关注局部空间下的连续单词序列,无法捕获单词之间的远距离关系。

近年来,图卷积神经网络(GCNN)[18]被广泛应用在文本分类领域中。基于 GCNN 的文本分类方法通过把文本转化成图的方式可以建模单词之间的复杂语义关系,这不但能捕获文本中的局部上下文信息,而且还能够保留文本中的全局语义信息。近年来,大量研究验证了 GCNN 在文本分类任务中的有效性。

根据学习方法的不同,现有的模型可以分为直推式文本分类模型和归纳式文本分类模型。TextGCN 模型[8]将图卷积神经网络应用到文本分类任务中,为整个语料库构建一个大型异构图。该图的节点是单词节点和文档节点,边是使用词共现信息、单词与文档包含关系构造的,节点使用独热向量初始化,使用 GNN 进行特征学习得到文档表示。LIU 等[19]提出了用于文本分类的张量图卷积网络(TensorGCN)。该模型基于语义信息、语法信息和序列信息构造了 3 个图,3 个图分别建模单词之间的依赖关系、序列信息以及语义信息。为了编码来自多个图的异构信息,该模型分别执行两种传播方式,首先是图内传播,用于对单个图内的节点进行特征传播与聚合,然后是图间传播,用于协调 3 个图分别表达的依赖信息、序列信息以及语义信息。然而,上述两种方法本质上均是直推式学习,图的结构和参数在训练之前确定,训练后无法修改,无法对新样本进行测试,同时上述两种方法均关注文本的全局关系,忽略了文本中的局部语义信息和上下文信息。

为了解决这个问题,相关研究者提出了一系列归纳式文本分类模型。HUANG 等[9]提出了一种

用于文本分类的 Text-Level GNN 模型，该模型把文本转化为有向图，并通过消息传播机制更新节点。该模型构建的有向图以单词为节点，使用较小的滑动窗口为单词节点创建边。该模型属于归纳式学习，具有较小的内存消耗并且对新样本具有良好的泛化能力。ZHANG 等[20]将提出的 TextING 模型用于文本分类，该模型首先把文本转化为文本级图，把图中的单词作为节点，使用滑动窗口捕获的单词间的共现关系作为边，并使用门控图神经网络(GGNN)[21]进行特征传播与聚合，每个节点可以从它的邻居节点处获得信息并与自身的表示合并来更新，然后在读出层对图中的节点表示进行池化来获得最终文档表示并分类。与直推式学习模型相比，该模型更注重局部的单词交互，因此能够捕获到文档内的单词上下文信息。如图 1 所示，对于“Don′t be a wet blanket.”这句话，其中“wet blanket”词组表达的意思是“扫兴的人”，如果将这句话构建为普通图，则是在“wet”与“blanket”之间建立二元关系，这会导致模型把“blanket”理解为“毯子”，得出错误的分类结果。

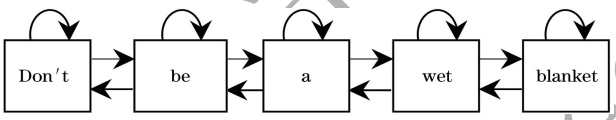


**图 1 常规文本图**

**Fig. 1 Conventional text graph**

DING 等[10]提出的 HyperGAT 模型将超图神经网络[22]用于文本分类任务。该模型使用文档级超图对文本进行建模，能够捕获单词之间的高阶语义关系。HyperGAT 模型构建了顺序超边、语义超边来描述并捕获单词间的高阶关系，其中，顺序超边连接句子中的所有单词，能够捕获文档的结构信息，语义超边使用潜在狄利克雷分布(LDA)[23]主题模型从文档中挖掘潜在的主题。然而，这些方法在处理短文本时仍然面临数据稀疏性。

为了解决短文本数据稀疏的问题，部分研究者通过外部知识或者模型融合的方法获取更丰富的文本表示。对于短文本，杨世刚等[24]首先从语料库中提取节点特征和边特征，然后将其作为全局特征构建图注意力网络模型。YUAN 等[25]提出了一种异构图注意力网络(HGAT)模型，该模型引入了主题和实体以获得更丰富的语义，将主题、实体等信息作为异构图的节点，再挖掘各节点之间的关系并将其作为边构建异构图，从而提出一种在异构图上进行特征传播的方法，然而该模型仍然属于直推式文本分类，无法测试新样本。DAI 等[26]提出 GFN 模型用于文本分类，该模型将每个文档根据词共现、余弦相似性、欧氏距离等信息构造 4 个图，分别进行图学习后融合。ZHANG 等[27]提出 HINT 模型，该模型首先使用依赖关系解析把文档转化为树形结构，然后通过学习编码树的表示来对整个文档进行分类，这种学习方式可以捕获文本中的层次信息。HUANG 等[28]提出 ConTextING 模型，该模型将基于 Transformer 的双向编码器表示(BERT)[29]用于文本分类任务，融合文档上下文特征和图特征并结合 BERT 分类器和图分类器进行分类。

为了更好地建模句子中的语义关系，本文首先为短文本分类引入知识，结合 GNN 和外部知识的优势，对于给定的文本，分别构建文本超图和语义超图，采用两种不同的构图方法，学习不同的文本特征和语义特征，然后通过注意力机制[30]对双超图特征进行融合。

## 2 文本分类模型

针对现有文本分类方法中普通图无法建模自然语言中的单词的高阶关系的问题，本文提出的模型整体结构如图2所示。模型处理流程为：首先进行

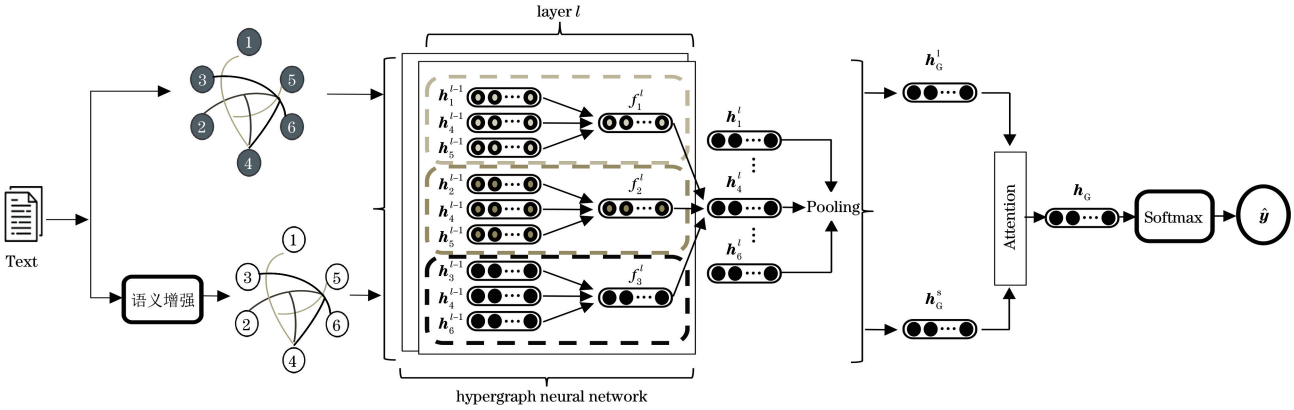


**图 2 模型结构**

**Fig. 2 Model structure**

文本预处理，包括分词、去除停用词等；然后将文本构建为超图，具体为：一方面使用原始文本建立原始超图，另一方面为短文本引入外部知识，使用基于SenticNet[12]词库的外部知识进行语义增强后构建语义超图；接着经过超图卷积网络对双超图进行特征传播与聚合，得到双超图特征；最后对双超图特征进行融合，得到最终的文本表示，并对文档标签进行预测。

### 2.1 超图

与常规图不同的是，超图中的一条超边可以同时连接多个节点。如图 3 所示，节点 $N_3$、$N_4$、$N_6$ 共同连接至一条超边 $S_3$，体现了节点之间的高阶关系。在文本分类任务中，单词间的联系并不都是一对一的，更多的是更高阶的关系，超图可以有效地表示单词之间的高阶关系。

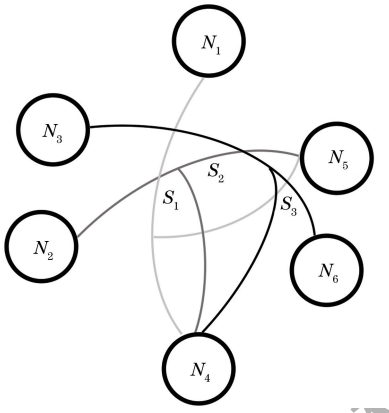


**图 3 超图结构示意图**

**Fig. 3 Schematic diagram of hypergraph structure**

**定义 1**（超图） 超图 $\boldsymbol{G}=(\boldsymbol{N},\boldsymbol{S})$，其中，$\boldsymbol{N}=(\boldsymbol{n}_1,\boldsymbol{n}_2,\cdots,\boldsymbol{n}_n)$ 是节点集，$\boldsymbol{S}=(\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_m)$ 是超边集。每条超边可以连接至少一个节点。超图 $\boldsymbol{G}$ 的关联矩阵 $\boldsymbol{H}\in\mathbb{R}^{n\times m}$ 构造方法如式(1)所示：

$$\boldsymbol{H}_{i,j}=\begin{cases}1, & \boldsymbol{n}_i\in\boldsymbol{s}_j\\ 0, & \boldsymbol{n}_i\notin\boldsymbol{s}_j\end{cases} \tag{1}$$

### 2.2 双超图构建

对于给定的文本，分别构建文本超图和语义超图。

#### 2.2.1 文本超图

对于语料库中给定的文本 $D=(w_1,w_2,\cdots,w_n)$，本文要构建的超图可以表示为 $\boldsymbol{G}=(\boldsymbol{N},\boldsymbol{S})$，其中，$\boldsymbol{N}=(\boldsymbol{n}_1,\boldsymbol{n}_2,\cdots,\boldsymbol{n}_n)$ 表示超图中节点的集合，由 $D$ 中的 $n$ 个不同的单词组成，每个单词对应一个节点，$\boldsymbol{S}=(\boldsymbol{s}_1,\boldsymbol{s}_2,\cdots,\boldsymbol{s}_m)$ 表示超图中超边的集合，由 $D$ 中的 $m$ 个句子组成。将每个句子视为一个超边，它连接了这个句子中的所有单词，使用句子作为超边使模型不仅能捕获单词之间的局部共现信息，而且还能建模文本的结构信息。超图中的节点信息使用 GloVe[31] 词向量进行初始化。

#### 2.2.2 语义超图

构建语义超图首先要对文本进行语义增强，即使用预先确定的单词列表中的替代单词替换给定文本中的关键词。替代单词从外部数据库 SenticNet[12] 中提取，SenticNet 提供了将语义、情感、极性关联的 1 000 个自然语言概念。本文基于 SenticNet 知识库构建了一个用于替换单词的样本集。该样本集由 6 个部分组成：原始单词，语义词 1，语义词 2，语义词 3，语义词 4，语义词 5。例如：（“baseball”，“team”，“baseball team”，“type of sport”，“sport”，“team up”）。本文使用语义词 1～语义词 5 作为关键词的替代单词，若单词多次出现则使用不同的语义词进行替换。

在选择给定文本中的关键词进行单词替换时，遵循一个约束条件：不会选择文本中已经被替换的单词或停用词。对每个文本的替换关键词数量进行计算，如式(2)所示：

$$n=\lfloor\alpha\times l\rfloor \tag{2}$$

式中：$l$ 是文本长度；$\lfloor\cdot\rfloor$ 表示向下取整；$\alpha$ 是超参数，控制文本中被替换单词的比例。

在确定文本中要替换的单词数量之后，使用 TextRank[32] 算法进行关键词提取。选择提取关键词而不是随机选择单词进行替换的原因是文本中的关键词对分类任务起着重要的作用，一个文本的类别往往是由几个关键词决定的。增强关键词的语义信息更有利于分类正确。语义增强算法的伪代码如算法 1 所示。

**算法 1** semantic enhancement

**输入** text $D$, hyperparameter $\alpha$, sample set $K$

**输出** updated text $C$

```
C←D
n←α×len(D) //计算要提取的关键词数量
s←keyextraction(n)//提取关键词
for keyword in s://进行关键词替换
    if keyword has already appeared:
        replace keywords of C with new word
    else:replace keywords of C
return C
```

在进行语义增强后，把更新后的文本转化为语义超图，在语义超图和文本超图中具有不同的节点，如图 4 所示，使用原始文本和进行单词替换后的文本分别建立文本超图和语义超图，其中，文本超图和

语义超图中节点的颜色不同表示双超图中的节点代表不同的单词。

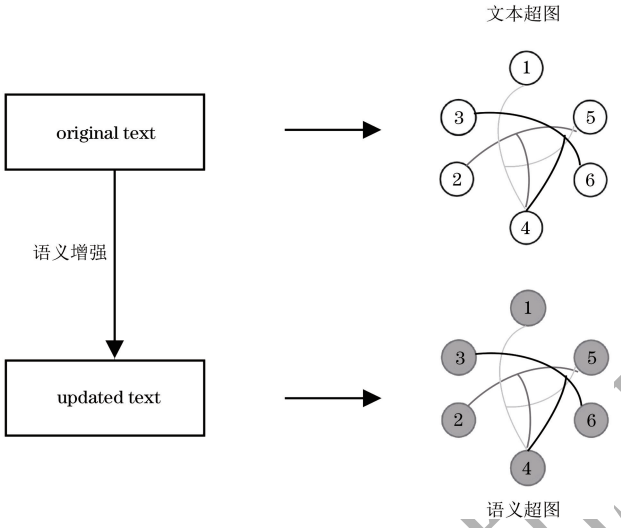


图 4 双超图构建

Fig. 4 Dual hypergraph construction

### 2.3 超图神经网络

将文本转化为超图，文本中的每个单词作为超图中的节点，该节点初始化为一个 $d$ 维向量，文本超图的节点集 $\boldsymbol{X}=[\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n]^{\mathrm{T}}\in\mathbb{R}^{n\times d}$。由此，一个文本超图可以表示为 $\boldsymbol{G}=(\boldsymbol{H},\boldsymbol{X})$，其中，$\boldsymbol{X}$ 是节点集，$\boldsymbol{H}$ 是超图对应的关联矩阵。在超图构造完成之后，使用式(3)对超图进行特征传播学习，得到最终的文本表示。

$$\boldsymbol{X}^{l+1}=\mathrm{LeakyReLU}(\boldsymbol{D}_{\mathrm{v}}^{-\frac{1}{2}}\boldsymbol{H}\boldsymbol{W}\boldsymbol{D}_{\mathrm{e}}^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{D}_{\mathrm{v}}^{-\frac{1}{2}}\boldsymbol{X}^{l}\boldsymbol{\theta}^{l}) \tag{3}$$

式中：$\boldsymbol{X}^{l}\in\mathbb{R}^{n\times d}$ 是超图第 $l$ 层的节点特征矩阵；$\boldsymbol{D}_{\mathrm{e}}$ 表示超边度的对角矩阵；$\boldsymbol{D}_{\mathrm{v}}$ 表示顶点度的对角矩阵；$\boldsymbol{W}$ 和 $\boldsymbol{\theta}$ 是可训练的参数。

超图卷积过程如图 5 所示，对于给定的节点特征矩阵 $\boldsymbol{X}^{l}$，经过 $L$ 层的特征传播与聚合后得到文本特征 $\boldsymbol{h}_v^L$。

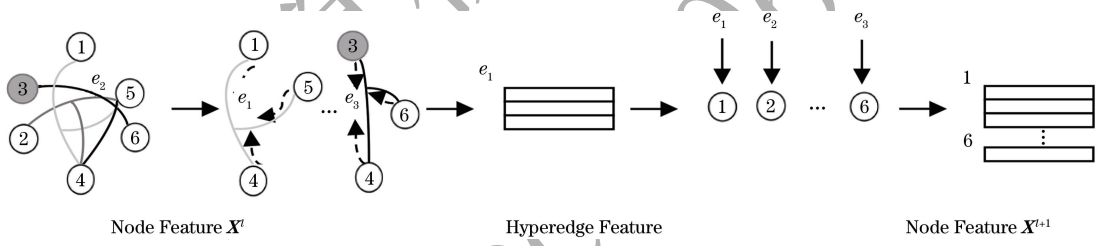


图 5 超图卷积

Fig. 5 Hypergraph convolution

对于每个文档，在经过 $L$ 层超图卷积之后能够计算得到所构建的文本超图上的所有节点特征。文档表示的计算方法如式(4)所示。本文基于学习的节点表示 $\boldsymbol{h}_v^L$，应用 Mean Pooling 来获得文档表示 $\boldsymbol{h}_{\mathrm{G}}$。

$$\boldsymbol{h}_{\mathrm{G}}=\frac{1}{|V|}\sum_{v\in V}\boldsymbol{h}_v^L \tag{4}$$

### 2.4 特征融合

假设 $\boldsymbol{h}_{\mathrm{G}}^{\mathrm{l}}$ 是文本超图进行特征传播与聚合后得到的文本表示，$\boldsymbol{h}_{\mathrm{G}}^{\mathrm{s}}$ 是语义超图进行特征传播与聚合后得到的文本表示。如图 6 所示，为了充分利用双超图特征，使用基于注意力的特征融合方法对双超图特征进行融合，计算过程如下：

$$\boldsymbol{Y}=\mathrm{LeakyReLU}(\boldsymbol{W}_{\mathrm{A}}\boldsymbol{h}_{\mathrm{G}}^{\mathrm{s}}) \tag{5}$$

$$\boldsymbol{\alpha}_j=\frac{\exp(\boldsymbol{Y}\boldsymbol{v}_{\mathrm{A}})}{\sum_{j\in n}\exp(\boldsymbol{Y}\boldsymbol{v}_{\mathrm{A}})} \tag{6}$$

$$\boldsymbol{h}_{\mathrm{G}}=\sum_{j\in n}\boldsymbol{\alpha}_j\boldsymbol{h}_{\mathrm{G}}^{\mathrm{l}} \tag{7}$$

式中：$\boldsymbol{W}_{\mathrm{A}}$ 和 $\boldsymbol{v}_{\mathrm{A}}$ 为训练过程中要学习的参数；$\boldsymbol{\alpha}_j$ 为双超图特征的注意力。

经过双超图特征结合之后得到最终文本表示 $\boldsymbol{h}_{\mathrm{G}}$。

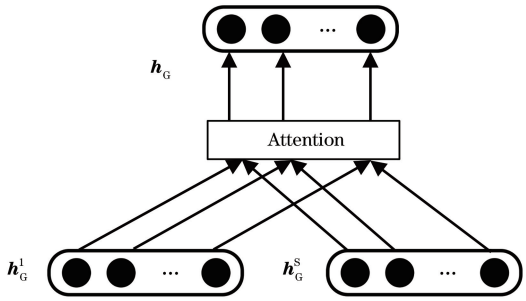


图 6 特征融合

Fig. 6 Feature fusion

### 2.5 分类

在特征融合之后，对融合后的特征进行分类，计算过程如式(8)所示：

$$\hat{\boldsymbol{y}}=\mathrm{Softmax}(\boldsymbol{W}\boldsymbol{h}_{\mathrm{G}}+\boldsymbol{b}) \tag{8}$$

式中：$\boldsymbol{W}$ 是参数矩阵；$\boldsymbol{b}$ 是偏置。

通过交叉熵损失函数最小化损失，计算过程如式(9)所示：

$$\mathcal{L}=-\sum_k\boldsymbol{y}_k\log_a(\hat{\boldsymbol{y}}_k) \tag{9}$$

式中：$\boldsymbol{y}_k$ 表示第 $k$ 个文档的类别；$\hat{\boldsymbol{y}}_k$ 表示第 $k$ 个文档的分类结果。

## 3 实验

在 4 个常用的短文本分类数据集上进行实验，并将本文模型与其他的基线模型进行对比分析。

### 3.1 数据集介绍

表 1 给出了所用的 4 个数据集的详细信息，其中，#Doc 是指该数据集的总样本数量，#Train 是指该数据集中训练样本的数量，#Test 是指该数据集中测试样本的数量，#Classes 是指该数据集共有多少类别，Avg Length 表示该数据集的平均文本长度。

1）R8 和 R52 是英文新闻数据集，来自文献[8]，其中，R8 分为 8 个类别，训练集样本数量为 5 485，测试集样本数量为 2 189，R52 分为 52 个类别，训练集样本数量为 6 532，测试集样本数量为 2 568。

2）Ohsumed[33]，其内容是心血管疾病的摘要，该数据集共分为 23 个类别，训练集样本数量为 3 357，测试集样本数量为 4 043。

3）MR[34]，其内容是对电影的点评，包括消极或积极两个类别，其中，训练文档 7 108 个，测试文档 3 554 个。

**表 1 数据集信息**

**Table 1 Information of datasets**

| Dataset | #Doc | #Train | #Test | #Classes | Avg Length |
|---|---|---|---|---|---|
| R8 | 7 674 | 5 485 | 2 189 | 8 | 65.72 |
| R52 | 9 100 | 6 532 | 2 568 | 52 | 69.82 |
| Ohsumed | 7 400 | 3 357 | 4 043 | 23 | 135.82 |
| MR | 10 662 | 7 108 | 3 554 | 2 | 20.39 |

### 3.2 实验设置

采用数据集中的 90%的样本作为训练集，10%的样本作为验证集。使用预训练的 GloVe[31] 词向量对单词节点进行初始化。学习率设置为 0.001，超参数 $\alpha$ 设置为 0.1，Dropout 设置为 0.5，将准确率($A$)作为模型性能的评价指标，其定义如式(10)所示：

$$A=\frac{N_{\mathrm{PT}}}{N_{\mathrm{Total}}} \tag{10}$$

式中：$N_{\mathrm{PT}}$ 是测试集中分类正确的文档个数；$N_{\mathrm{Total}}$ 是测试集中的所有文档个数。

### 3.3 基线模型

将本文模型与下列文本分类经典模型进行性能对比：

1）CNN[12]：用于文本分类任务的卷积神经网络模型。

2）LSTM[16]：用于文本分类任务的长短期记忆网络模型。

3）Bi-LSTM[17]：用于文本分类的双向长短期记忆网络模型。

4）TextGCN[8]：用于文本分类的语料库级图神经网络模型。图的节点是单词节点和文档节点，边是基于词共现信息、单词与文档包含关系构造的，节点使用独热向量初始化。

5）Text-Level GNN[9]：一种文档级文本分类模型。该模型把每个文档构建成一个文本图，图中把单词作为节点，词共现关系作为边。

6）$S^2GC$[35]：使用马尔可夫扩散核推导出的 GCN 变体模型。

7）HyperGAT[10]：把文本转化为超图，使用超图神经网络进行特征学习和分类。

### 3.4 实验结果与分析

本文在 4 个短文本数据集上进行了实验，结果如表 2 所示，其中，最优指标值用加粗字体标示，下同。

**表 2 分类准确率结果对比**

**Table 2 Comparision of classification accuracy results**

| Model | MR | R8 | R52 | Ohsumed |
|---|---|---|---|---|
| CNN | 0.777 5 | 0.957 1 | 0.875 9 | 0.583 3 |
| LSTM | 0.750 6 | 0.936 8 | 0.855 4 | 0.411 3 |
| Bi-LSTM | 0.776 8 | 0.963 1 | 0.905 4 | 0.492 7 |
| TextGCN | 0.766 4 | 0.970 7 | 0.935 6 | 0.683 6 |
| Text-Level GNN | 0.754 7 | 0.978 9 | 0.946 0 | 0.694 3 |
| $S^2GC$ | 0.767 0 | 0.974 0 | 0.945 0 | 0.685 0 |
| HyperGAT | 0.783 2 | 0.979 7 | 0.949 8 | 0.699 0 |
| Ours | **0.783 8** | **0.981 1** | **0.950 7** | **0.701 3** |

在平均文本长度较短的 R8 数据集和 MR 数据集上，本文模型的结果相比于 CNN、LSTM 和 Bi-LSTM 模型具有约 2 百分点的准确率提升，在平均文本长度较长的 Ohsumed 数据集和分类类别数较多的 R52 数据集上，本文模型有着更大的提升。这说明 CNN、LSTM 和 Bi-LSTM 模型在面对较长的文本时会丢失部分远距离上下文信息，而超图可以建模单词间的远距离依赖信息。

与图神经网络模型比较，本文模型具有约 1 百分点的准确率提升，这表示本文构建的文档级超图能够更详细地建模文本内的局部信息和单词间的上下文信息，通过构建语义超图的方法获得了文本中的丰富语义信息，基于注意力的双超图特征融合方法提高了语义特征在最终文本特征中的重要性。相比于普通图，超图模型能够建模文本中的复杂结构

信息和单词之间的高阶关系，捕获文本之间的长距离依赖信息。

与 HyperGAT 相比，在 4 个数据集上本文模型有一定的性能提升。这说明本文模型把文本构建为文本超图，在获得文本信息的同时通过构建语义超图的方法增强了文本中的重要语义信息，并通过基于注意力的特征融合方法提升了模型对关键信息的提取能力。实验结果表明，本文模型的性能优于上述文本分类基线模型。

### 3.5 消融实验

为了进一步分析各模块对模型性能的影响，本文进行了消融实验。实验对比了单独使用文本超图和同时使用文本超图与语义超图的实验效果，结果如表 3 所示，其中，w/o SH 表示删除语义超图，仅构建文本超图进行分类，w/o TH 表示删除文本超图，仅构建语义超图进行分类，w/o ATT 表示不使用基于注意力的双图特征融合方法，仅对双超图特征取平均值进行融合。从表 3 中可以看出，虽然各模块对模型的贡献程度不同，但删除各模块后总会引起性能的下降，这说明了各模块的有效性。

**表 3 消融实验**

**Table 3 Ablation experiment**

| Model | MR | R8 | R52 | Ohsumed |
|---|---|---|---|---|
| w/o SH | 0.776 8 | 0.975 4 | 0.943 9 | 0.683 3 |
| w/o TH | 0.772 9 | 0.971 3 | 0.942 1 | 0.677 9 |
| w/o ATT | 0.781 5 | 0.980 3 | 0.948 7 | 0.699 5 |
| Ours | **0.783 8** | **0.981 1** | **0.950 7** | **0.701 3** |

从表 3 中还可以看出：

在删除语义超图后，仅使用文本超图进行分类会造成模型性能小幅度下降，这说明本文模型通过构建语义超图增强了文本的重要语义信息，提升了模型分类性能。

在删除文本超图后，仅使用语义超图进行分类同样造成了模型性能下降，仅使用语义超图的效果比仅使用文本超图进行分类的效果略差，这是因为语义超图是为了辅助文本超图而构造的，构造语义超图时为了提升关键词的语义信息而丢弃了一些不重要的语义信息，在仅使用语义超图进行分类时造成了模型性能下降。

在删除基于注意力的双图特征融合方法后模型性能小幅度下降，这是因为利用注意力机制可以捕获文本中的重要特征。经过实验发现，模型使用基于注意力的特征融合方法在 4 个数据集上的表现总是优于简单的平均值特征融合方法。这验证了注意力机制对特征融合的有效性。

### 3.6 参数分析

#### 3.6.1 单词嵌入维度对性能的影响

为了探究单词嵌入维度对文本分类效果的影响，在 R8 和 R52 数据集上使用不同的单词嵌入维度进行实验。图 7 和图 8 分别展示了在 R8 和 R52 数据集上使用不同单词嵌入维度对模型分类效果的影响。

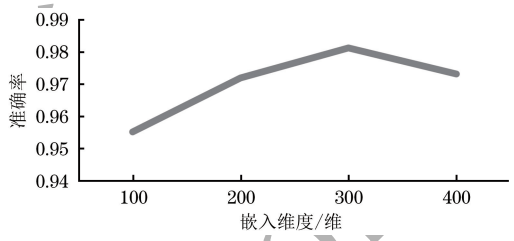


**图 7 嵌入维度对 R8 分类效果的影响**

**Fig. 7 Classification effect of embedding dimension on the R8**

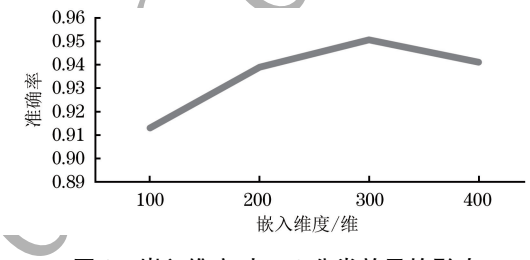


**图 8 嵌入维度对 R52 分类效果的影响**

**Fig. 8 Classification effect of embedding dimension on the R52**

#### 3.6.2 训练集比率对性能的影响

本文在 R8 和 R52 数据集上选取了不同比率的训练集做了对比实验，用来评估不同比率的训练集对模型分类效果的影响。实验结果如图 9 和图 10 所示。

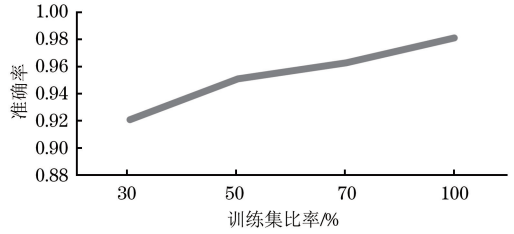


**图 9 训练集比率对 R8 分类效果的影响**

**Fig. 9 Classification effect of training set ratio on the R8**

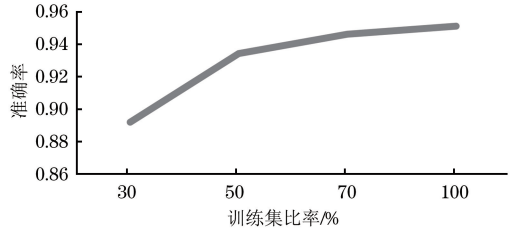


**图 10 训练集比率对 R52 分类效果的影响**

**Fig. 10 Classification effect of training set ratio on the R52**

图 9 和图 10 展示了 R8 和 R52 数据集使用不同比率训练集时模型的准确率变化。在使用 30%

训练集的情况下，对 R8 的测试准确率为 0.921 0，对 R52 的测试准确率为 0.891 3，随着训练集比率的增加，模型的测试准确率也随之增加。当训练集比率小于 50%时，随着训练集比率的增加，模型的测试准确率增加速度较快；当训练集比率大于 50%时，随着训练集比率的增加，模型的测试准确率增加速度较慢。

## 4 结束语

本文提出了一种基于双超图特征融合的文本分类模型，一方面使用原始文本建立文本超图，另一方面为短文本引入外部知识，使用基于 SenticNet 词库的外部知识对文本进行语义增强后构建语义超图，经过超图卷积后对双超图特征进行融合得到最终文本表示并进行分类。双超图在保留了文本超图的基础上，通过语义超图对文本中的重要语义信息进行挖掘，同时利用基于注意力机制的特征融合方法突出重要特征，增强模型的文本表示能力。在 4 个常用的短文本数据集上的实验结果表明，本文模型性能优于其他基线模型。

## 参考文献


[1] NOBLE W S. What is a support vector machine? [J]. Nature Biotechnology, 2006, 24: 1565-1567.

[2] ALFEILAT H A A, HASSANAT A B A, LASASSMEH O, et al. Effects of distance measure choice on *k*-nearest neighbor classifier performance: a review[J]. Big Data, 2019, 7(4): 221-248.

[3] HARRIS Z S. Distributional structure[J]. Word, 1954, 10(2/3): 146-162.

[4] SALTON G, WONG A, YANG C S. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.

[5] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. [2023-08-04]. https://arxiv.org/abs/1301.3781v3.

[6] ALBAWI S, MOHAMMED T A, AL-ZAWI S. Understanding of a convolutional neural network[C]// Proceedings of the International Conference on Engineering and Technology (ICET). Washington D.C., USA: IEEE Press, 2017: 1-6.

[7] MEDSKER L, JAIN L C. Recurrent neural networks: design and applications[M]. Boca Raton, USA: CRC Press, 1999.

[8] YAO L, MAO C S, LUO Y, et al. Graph convolutional networks for text classification[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence and 31st Innovative Applications of Artificial Intelligence Conference and 9th AAAI Symposium on Educational Advances in Artificial Intelligence. New York, USA: ACM Press, 2019: 7370-7377.

[9] HUANG L Z, MA D H, LI S J, et al. Text level graph neural network for 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 (EMNLP-IJCNLP). Stroudsburg, USA: ACL Press, 2019: 3444-3450.

[10] DING K Z, WANG J L, LI J D, 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 (EMNLP). Stroudsburg, USA: ACL Press, 2020: 4927-4936.

[11] 陈杰. 基于多方面特征表示与图卷积网络的短文本分类研究[D]. 合肥:安徽大学,2022.
CHEN J. Research on short text classification based on multifaceted feature representation and graph convolution network[D]. Hefei: Anhui University, 2022. (in Chinese)

[12] CAMBRIA E, SPEER R, HAVASI C, et al. SenticNet: a publicly available semantic resource for opinion mining[EB/OL]. [2023-08-04]. https://www.semanticscholar.org/paper/SenticNet%3A-A-Publicly-Available-Semantic-Resource-Cambria-Speer/8ae2d6c78c067acaa17713614c0d7d6b0a53baa8.

[13] 闫佳丹, 贾彩燕. 基于双图神经网络信息融合的文本分类方法[J]. 计算机科学, 2022, 49(8): 230-236.
YAN J D, JIA C Y. Text classification method based on information fusion of dual-graph neural network[J]. Computer Science, 2022, 49(8): 230-236. (in Chinese)

[14] KIM Y. Convolutional neural networks for sentence classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, USA: ACL Press, 2014: 1746-1751.

[15] ZHANG X, ZHAO J B, LECUN Y, et al. Character-level convolutional networks for text classification[C]// Proceedings of the 29th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2015: 649-657.

[16] 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. Palo Alto, USA: AAAI Press, 2016: 2873-2879.

[17] JANG B, KIM M, HARERIMANA G, et al. Bi-LSTM model to increase accuracy in text classification: combining Word2Vec CNN and attention mechanism[J]. Applied Sciences, 2020, 10(17): 5841.

[18] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. [2023-08-04]. https://arxiv.org/abs/1609.02907v4.

[19] LIU X E, YOU X X, ZHANG X, et al. Tensor graph convolutional networks for text classification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 8409-8416.

[20] ZHANG Y F, YU X L, CUI Z Y, et al. Every document owns its structure: inductive text classification via graph neural networks[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL Press, 2020: 334-339.

[21] RUIZ L, GAMA F, RIBEIRO A, et al. Gated graph sequence neural networks[J]. IEEE Transactions on Signal Processing, 2020, 68: 6303-6318.

[22] FENG Y F, YOU H X, ZHANG Z Z, et al. Hypergraph neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2019: 3558-3565.

[23] JELODAR H, WANG Y L, YUAN C, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey[J]. Multimedia Tools and Applications, 2019, 78(11): 15169-15211.

[24] 杨世刚, 刘勇国. 融合语料库特征与图注意力网络的短文本分类方法[J]. 计算机应用, 2022, 42(5): 1324-1329.
YANG S G, LIU Y G. Short text classification method by

fusing corpus features and graph attention network[J]. Journal of Computer Applications, 2022, 42(5): 1324-1329. (in Chinese)
[25] YANG T C, HU L M, SHI C, et al. HGAT: heterogeneous graph attention networks for semi-supervised short text classification [J]. ACM Transactions on Information Systems, 2021, 39(3): 1-29.
[26] DAI Y, SHOU L J, GONG M, et al. Graph fusion network for text classification[J]. Knowledge-Based Systems, 2022, 236: 107659.
[27] ZHANG C, ZHU H, PENG X, et al. Hierarchical information matters: text classification via tree based graph neural network[C]// Proceedings of the 29th International Conference on Computational Linguistics. Stroudsburg, USA: ACL Press, 2022: 950-959.
[28] HUANG Y H, CHEN Y H, CHEN Y S. ConTextING: granting document-wise contextual embeddings to graph neural networks for inductive text classification C]// Proceedings of the 29th International Conference on Computational Linguistics. Stroudsburg, USA: ACL Press, 2022: 1163-1168.
[29] KENTON J D M W C, TOUTANOVA L K. BERT: pre-training of deep bidirectional Transformers for language understanding [C] // Proceedings of NAACL-HLT '19. Stroudsburg, USA: ACL Press, 2019: 4171-4186.
[30] VASWANI A, SHAZEER N, PARMERN, et al. Attention is all you need[C]// Proceedings of the 31th International Conference on Neural Information Processing Systems. Cambridge, USA: MIT Press, 2017: 6000-6010.
[31] PENNINGTON J, SOCHER R, MANNING C. GloVe: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, USA: ACL Press, 2014: 1532-1543.
[32] ZHANG M X, LI X M, YUE S B, et al. An empirical study of TextRank for keyword extraction[J]. IEEE Access, 2020, 8: 178849-178858.
[33] WANG K Z, HAN S C, POON J. InducT-GCN: inductive graph convolutional networks for text classification[C]// Proceedings of the 26th International Conference on Pattern Recognition (ICPR). Washington D.C., USA: IEEE Press, 2022: 1243-1249.
[34] LIN C H, HE Y L, LIN C H, et al. Joint sentiment/topic model for sentiment analysis[C]//Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York, USA: ACM Press, 2009: 375-384.
[35] ZHU H, KONIUSZ P. Simple spectral graph convolution[C]// Proceedings of the International Conference on Learning Representation. New York, USA: ACM Press, 2021:151-163.

编辑 陆燕菲