



# 计算机科学

COMPUTER SCIENCE

## 基于神经网络和依存句法分析的文本分类

杨旭华, 金鑫, 陶进, 毛剑飞

引用本文

杨旭华, 金鑫, 陶进, 毛剑飞. 基于神经网络和依存句法分析的文本分类[J]. 计算机科学, 2022, 49(12): 293-300.

YANG Xu-hua, JIN Xin, TAO Jin, MAO Jian-fei. [Text Classification Based on Graph Neural Networks and Dependency Parsing](#) [J]. Computer Science, 2022, 49(12): 293-300.

---

## 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)

### [基于语义导向的软件在线升级功能逆向定位](#)

Reverse Location of Software Online Upgrade Function Based on Semantic Guidance  
计算机科学, 2022, 49(12): 353-361. <https://doi.org/10.11896/jsjx.211000059>

### [一种改进的特征选择算法在邮件过滤中的应用](#)

Application of Improved Feature Selection Algorithm in Spam Filtering  
计算机科学, 2022, 49(11A): 211000028-5. <https://doi.org/10.11896/jsjx.211000028>

### [融合多层次视觉信息的人物交互动作识别](#)

Human-Object Interaction Recognition Integrating Multi-level Visual Features  
计算机科学, 2022, 49(11A): 220700012-8. <https://doi.org/10.11896/jsjx.220700012>

### [基于属性图注意力网络的电影推荐模型](#)

Movie Recommendation Model Based on Attribute Graph Attention Network  
计算机科学, 2022, 49(11A): 211100106-8. <https://doi.org/10.11896/jsjx.211100106>

### [文档级实体关系抽取方法研究综述](#)

Survey of Document-level Entity Relation Extraction Methods  
计算机科学, 2022, 49(10): 224-242. <https://doi.org/10.11896/jsjx.211000057>

# 基于图神经网络和依存句法分析的文本分类

杨旭华 金鑫 陶进 毛剑飞

浙江工业大学计算机科学与技术学院 杭州 310023

(xhyang@zjut.edu.cn)

**摘要** 文本分类被广泛应用于新闻分类、话题标记和情感分析等语言处理场景中,是自然语言处理中的一个基本而重要的任务。目前的文本分类模型一般没有同时考虑文本单词的共现关系和文本自身的句法特性,从而限制了文本分类的效果。因此,提出了一个基于图卷积神经网络的文本分类模型(Mix-GCN)。首先基于文本单词之间的共现关系和句法依存关系,将文本数据构建成文本共现图和句法依存图;接着,利用GCN模型对文本图和句法依赖图进行表示学习,得到单词的嵌入向量;然后通过图池化方法以及自适应融合的方法得到文本的嵌入向量;最后通过图分类方法完成文本分类。Mix-GCN模型同时考虑了文本中相邻单词之间的关系和文本单词之间存在的句法依存关系,提升了文本分类性能。在6个基准数据集上与8种知名文本分类方法进行了比较,实验结果表明Mix-GCN具有良好的文本分类效果。

**关键词:** 文本分类;图神经网络;依存句法分析;图分类

中图法分类号 TP391

## Text Classification Based on Graph Neural Networks and Dependency Parsing

YANG Xu-hua, JIN Xin, TAO Jin and MAO Jian-fei

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

**Abstract** Text classification is a basic and important task in natural language processing. It is widely used in language processing scenarios such as news classification, topic tagging and sentiment analysis. The current text classification models generally do not consider the co-occurrence relationship of text words and the syntactic characteristics of the text itself, thus limiting the effect of text classification. Therefore, a text classification model based on graph convolutional neural network (Mix-GCN) is proposed. Firstly, based on the co-occurrence relationship and syntactic dependency between text words, the text data is constructed into a text co-occurrence graph and a syntactic dependency graph. Then the GCN model is used to perform representation learning on the text graph and syntactic dependency graph, and the embedding vector of the word is obtained. Then the embedding vector of the text is obtained by graph pooling method and adaptive fusion method, and the text classification is completed by the graph classification method. Mix-GCN model simultaneously considers the relationship between adjacent words in the text and the syntactic dependencies existing between text words, which improves the performance of text classification. On 6 benchmark datasets, compared to 8 well-known text classification methods, experimental results show that Mix-GCN has a good text classification effect.

**Keywords** Text classification, Graph neural network, Dependency parsing, Graph classification

## 1 引言

文本是传递信息的重要途径和载体,也是人们学习、生活、研究的重要组成部分。如何快速、高效地处理这些文本信息资源是当前重要的研究课题。文本分类作为自然语言处理的基本任务和核心技术,能够快速、准确地判别文本所属类别,从而更好地服务自然语言处理类型的任务,被广泛应用于垃圾邮件判别<sup>[1]</sup>、新闻筛选<sup>[2]</sup>和舆情分析<sup>[3]</sup>等任务中。

方法包括特征工程和分类器构建两个部分。特征工程需要人工挑选信息量大的特征项,效率低下,文本分类提取特征的方法主要是使用词频(TF)、词频-逆文档频率指数(TF-IDF)<sup>[4]</sup>或者词袋模型等作为特征。分类器通常采用LR<sup>[5]</sup>、朴素贝叶斯(Naive Bayes)<sup>[6]</sup>、KNN<sup>[7]</sup>、支持向量机(Support Vector Machines, SVM)<sup>[8]</sup>等机器学习方法作为分类函数,但是这类基于机器学习的模型存在特征稀疏、模型表达能力不足等问题。

2013年Mikolov等<sup>[9]</sup>提出了word2vec模型,该模型用

早期的文本分类模型大多基于传统机器学习方法,该

到稿日期:2022-03-21 返修日期:2022-05-17

基金项目:国家自然科学基金(62176236)

This work was supported by the National Natural Science Foundation of China(62176236).

通信作者:毛剑飞(mjf@zjut.edu.cn)

一个定长的稠密向量来表示单词,这为深度学习应用于自然语言处理奠定了坚实的基础。接着研究者们将 CNN(Convolutional Neural Network)<sup>[10-11]</sup>、RNN(Recurrent Neural Network)<sup>[12-13]</sup>、胶囊网络<sup>[14]</sup>等深度学习<sup>[15]</sup>模型应用于文本分类等自然语言处理领域,并取得了较好结果。Kim<sup>[10]</sup>使用 CNN 模型来捕获文本抽象特征,通过权值共享、局部感知野的卷积核取代了全连接层,提高了模型的精度,并且降低了参数量。Facebook 人工智能研究院提出了 FastText 模型<sup>[16]</sup>,用于文本的快速分类,该模型主要由输入层、隐藏层和输出层组成。将输入文本向量化之后由隐藏层计算句子的词向量的平均值,最后在输出层实现分类。Zhang 等<sup>[17]</sup>提出了一种更小粒度的文本分类模型 Char-CNN,该模型可以不考虑单词的内在含义、语义和语法的信息,可以跨语言使用,并且能有效解决拼写错误和表情符号等问题。Tang 等<sup>[18]</sup>在 GRU 的基础上提出了一种基于文本组成原则的方法,它通过句子或文本的各个组成部分的意义来理解整个句子或文本。该方法首先基于词的表达使用 CNN 或 LSTM(Long Short-Term Memory)来构建句子的表达,然后基于句子及其内在关系使用 GNN 来构建文本的表达。TextRNN<sup>[19]</sup>模型是在 RNN 基础之上提出的,具有短期记忆功能,适合处理带有时序关系的任务,该方法提升了文本分类的效果。Wang 等<sup>[20]</sup>将卷积神经网络和循环神经网络相结合,提出了一种新的模型,它既具有卷积神经网络从文本中提取局部特征的优点,又具有长短期记忆网络(LSTM)连接提取特征的优点。

近几年,由于图神经网络(Graph Neural Network, GNN)<sup>[21]</sup>捕捉数据中的拓扑信息的能力出色,非常适合用来处理不规则图结构数据,在计算机视觉、自然语言处理领域都取得了优秀的效果,因此受到了研究者的广泛关注。Gao 等<sup>[22]</sup>提出用于图池化的 gpool 层和结合 GCN(Graph Convolutional Networks)和常规卷积运算的混合卷积(hConv)层。gpool 层采用可训练的投影向量来衡量图中节点的重要性,hConv 层可以快速增加感受野并自动计算特征。Yao 等提出了 Text-GCN 模型<sup>[23]</sup>,该模型将整个语料库根据文本和单词的共现关系构建一张异构图,训练文本、测试文本以及文本内的单词都作为图网络的节点,然后对文本图网络进行嵌入学习,获得文档节点的嵌入表示向量进行分类。Peng 等<sup>[24]</sup>提出了一种基于图网络的深度学习模型 Graph-CNN,利用文本单词之间的共现关系建立文本共现图,然后利用 CNN 对图网络进行学习,捕获文本的拓扑结构信息和远距离依赖关系。Wu 等<sup>[25]</sup>提出了 DAGNN 模型用于跨域文本分类,将文档建模成图,捕获非连续和长距离的语义,使用来自不同领域的图来联合训练分层图神经网络,从每个图中提取特征以进行有效的文本分类。上述模型中,有的 GNN 模型需要为整个语料库构建一张图,空间复杂度非常高;有的模型仅仅从邻域单词共现的角度构建文本网络,忽略了文本语言本身的规律关系,即没有从语言本身的句法依存关系<sup>[26]</sup>层面考虑文本信息。为了解决以上问题,我们提出了新的 Mix-GCN 模型,将文本数据构建成文本图和句法依存图,对两个图的特征进行融合,提升了文本分类效果。

本文的主要贡献如下:

(1)将文本数据构建成文本共现图和句法依存图,分别从两个图中提取特征,并进行自适应融合。

(2)将文本分类转化成图分类问题,提出了一种用于文本分类的新型图卷积神经网络,用于学习文本网络的拓扑结构关系。

(3)Mix-GCN 利用两个独立的 GCN 学习文本共现图和句法依存图,分别获得两种文本图的嵌入向量,然后对两种嵌入向量进行自适应融合,提升了文本分类效果。

本文第 2 节介绍相关工作;第 3 节描述本文提出的新图卷积神经网络模型;第 4 节进行实验仿真和结果分析,最后总结全文。

## 2 相关工作

目前已有大量关于文本分类的算法,而我们提出的新型图神经网络模型和已有工作存在明显的不同。本文提出的模型将文本构建图数据结构,将文本分类转化成图分类的问题,并引入文本的句法依存关系信息,提升了文本分类效果。

### 2.1 文本表示

词是自然语言处理的基本单位,但是计算机无法对其进行直接处理,因此也无法直接使用各种机器学习、深度学习模型,为此,学者们提出了各种文本表示模型。常见的文本表示模型有:词袋模型<sup>[27]</sup>(Bag of Words, BOW)、主题模型<sup>[28]</sup>(Topic Model)、词嵌入模型<sup>[29]</sup>(Word Embedding)等。目前流行的文本模型主要是词嵌入模型。

Bengio 等<sup>[30]</sup>于 2003 年提出了神经网络语言模型(Neural Network Language Model),该模型在学习语言模型的同时训练词向量。Mikolov 等<sup>[9]</sup>于 2013 年提出了 CBOW(Continuous Bag-of-Words Model)模型和 Skip-Gram(Continuous Skip-gram Model)模型。Skip-gram 的主要思想是用当前词预测其上下文,而 CBOW 模型与 Skip-gram 模型的训练方式刚好相反,它通过词的上下文预测当前词。2014 年,Mikolov 等<sup>[31]</sup>基于 Word2vec 的模型提出了 Doc2vec 模型,通过计算句子/篇章向量之间的距离来判断句子/篇章之间的相似度并得到句子和篇章向量的表示。2014 年 Pennington 等<sup>[32]</sup>提出了 Glove 词向量方法,该方法基于全局词汇信息来学习词向量的嵌入表示,相比 word2vec,它在很大程度上提升了词向量的嵌入表示效果。Facebook 于 2016 年开源了 FastText<sup>[16]</sup>,它是一个词向量与文本分类工具,利用简单而高效的文本分类和表征学习的方法进行文本分类,并且在保持高精度的同时提升了训练速度,使用  $n$ -grams 来缩小线性模型和深度模型的准确度差值。Tan 等<sup>[33]</sup>提出使用余弦相似度(Cosine Similarity)代替点积(Dot Product)来训练文本嵌入表示,提升了分类效果。

### 2.2 图神经网络

将文本分类任务转换为图分类任务中,非常重要的一步就是对文本网络进行嵌入学习。而网络嵌入的目的是用稠密的低维向量表示高维、稀疏的网络空间,同时最大程度地保留网络的拓扑结构信息。对于如何对网络进行嵌入表示,学者们已经提出了许多模型,比如 Perozzi 等提出的无监督模型 DeepWalk<sup>[34]</sup>,Goyal 等提出的监督学习模型 SemiGraph<sup>[35]</sup>。

一般情况下,网络中相似或相近的节点在映射空间中有相似的嵌入表示。图神经网络可以处理网络类型的非欧几里得数据,通过聚合邻域节点信息来学习网络的结构特征,从而得到网络的嵌入表示。

图卷积网络(GCN)是 GNN 中表现出众和应用最广的一类图神经网络,一般被用来处理非规则的图结构数据。自 Kipf 等提出的图卷积神经网络在许多基准图数据集达到了最优分类效果后,GCN 便受到了广泛的关注,并在计算机视觉、知识图谱、自然语言处理等领域成功应用,比如在图像分割<sup>[36]</sup>、语义角色标签<sup>[37]</sup>、关系分类<sup>[38]</sup>和机器翻译<sup>[39]</sup>等任务上。当前图卷积网络大致上可以分成两类:基于谱域的图卷积方法(Spectral-based GCN Methods)和基于空域的图卷积(Spatial-based GCN Methods)。基于谱域的图卷积方法是从图信号中去除噪声,基于空域的图卷积则通过信息聚合完成消息传递。GCN 采用基于谱域的卷积滤波器,节点可以通过其聚集邻域的特征进行表示学习。在图卷积神经网络的基础上,研究者们提出了多个具有更高效信息聚合器的 GNN 模型,例如, Velićković 等提出了带有节点重要性权重的图注意力网络(Graph Attention Network, GAT)<sup>[40]</sup>,该模型为各个节点分配不同的重要性权重,以便节点可以在聚合特征时突出显示重要邻域;Hamilton 等提出了 GraphSAGE<sup>[41]</sup>,每个节点学习一组聚合函数,灵活聚合来自不同 hop 中邻域的信息。Kim 等<sup>[42]</sup>提出了自监督图注意力网络(SuperGAT),该模型利用自监督机制来预测边的存在和缺失,并且通过对边的编码,SuperGAT 在区分错误连边的邻居时会获得更具表现力的注意力,从而提升了图神经网络在图噪声中的学习能力。Alon 等<sup>[43]</sup>针对远距离节点之间传播信息的 over-squarshing 问题提出了解决方法,该方法无须任何调整或额外的权重即可突破瓶颈得到改善后的结果。Yao 等<sup>[44]</sup>提出了 GAUG 图数据扩充框架,该框架利用图数据增强的理论和策略来提升基于 GNN 的节点分类性能。在多个基准上进行的实验表明,通过 GAUG 框架进行的增强可提升 GNN 架构和数据集的性能。

在文本分类任务中,图神经网络可以有效提取文本特征,在保留全局结构信息的同时,捕获文本的结构信息和单词间的长距离依赖关系,从而获得良好的分类效果。比如,基于图神经网络的方法 TextGCN 将整个语料库中的词和文档作为节点,建立了整个语料库的一张异构图,将文本分类问题转换成节点分类问题,在长文本分类任务中取得了很好的效果。

### 2.3 注意力机制

注意力机制起源于神经机器翻译<sup>[45]</sup>(Neural Machine Translation, NMT),并被广泛应用于自然语言处理(Natural Language Processing, NLP)相关任务,如实体抽取以及角色标注等。从更高层次上来说,通过允许解码器在多个向量上提取特征,注意力将编码器从输入嵌入到单个固定长度向量的负担中解脱出来,从而允许保留更多的信息。NLP 领域中的注意力机制通常结合编码器-解码器(Encoder-Decoder)模型<sup>[46]</sup>来实现。编码器-解码器模型利用序列标注方法将一组输入转化为一组输出,核心是通过编码器对输入序列编码,将编码后的序列输入到解码器中进行解码,得到特定的输出序列。

因为注意力机制的效果良好<sup>[47]</sup>,所以学者们提出了许多注意力机制的变体,以便处理更复杂的任务。Sukhbaatar 等<sup>[48]</sup>提出了一种利用循环注意力模型发现复杂 NLP 任务中潜在依赖关系的端到端的记忆网络。Wang 等<sup>[49]</sup>提出了多层注意力机制来捕捉术语之间的间接关系,获得更深层次的信息。Yang 等<sup>[50]</sup>提出了一种用于文档分类的分级注意力机制,该机制在单词级和句子级对文档进行建模,从而可以提取局部和全局的重要信息。谷歌团队提出了自注意力机制<sup>[51]</sup>来捕获句子中的深层上下文信息,可以捕获长距离依赖,同时降低了时间复杂度。Sun 等<sup>[52]</sup>提出了一种操纵多头注意力机制的不同头部来生成不同翻译的方法,该方法与反向翻译技术结合使用,显著提升了翻译任务的性能。Misra 等<sup>[53]</sup>提出了一种轻量并且有效的注意力机制(Triplet Attention),该注意力机制通过使用 Triplet Branch 结构捕获跨维度交互来计算注意力权重,并且可以通过旋转操作和残差变换建立维度间的依存关系。此外, Triplet Attention 还能以可忽略的计算开销来对通道和空间信息进行编码。

## 3 基于图神经网络和句法依存关系的文本分类

针对现有的文本分类方法没有考虑文本数据自身的句法依赖信息和拓扑结构信息的问题,我们将文本数据转化成图结构数据,提出新型图卷积神经网络模型。该模型同时学习文本的拓扑结构信息以及文本的句法信息,丰富文本的特征表示,用图分类方法完成文本分类任务,提高了分类效果。

### 3.1 问题描述

给定文本语料库的标签集合  $L = \{l_1, l_2, \dots, l_{|L|}\}$  词表集合  $V = \{v_1, v_2, \dots, v_{|V|}\}$ , 文本集合  $X = \{D_1, D_2, \dots, D_N\}$ , 单标签文本分类的任务就是给文本集合中的每篇文本  $D_i = \{\omega_{i1}, \omega_{i2}, \dots\}$  分配一个最有可能的预测标签  $y_i$ , 其中文本的单词  $\omega_{ij} \in V$ , 预测标签  $y_i \in L$ 。

我们的目标是将文本内容构建成 2 个文本网络,将文本分类问题转变成图分类问题,提出新型图神经网络模型 Mix-GCN,用 Mix-GCN 对文本网络进行图分类,提高文本分类效果。

### 3.2 Mix-GCN 模型框架

本文提出的 Mix-GCN 文本分类模型主要由 3 个部分构成,分别为文本共现图和句法依赖图的构建、文本网络嵌入学习模块、文本分类模块。模型结构如图 1 所示。

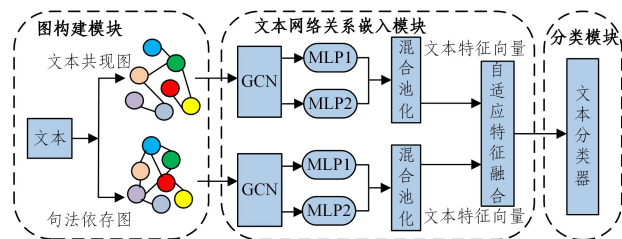


图 1 Mix-GCN 模型结构图

Fig. 1 Diagram of Mix-GCN model structure

文本数据需要经过预处理,然后将处理后的文本用来构建 2 个文本网络:一个是依据文本邻域单词的共现关系构建的文本共现图,另一个是依据文本自身句法依存关系构建的

句法依存图。使用两个独立的图卷积网络分别学习文本共现图和句法依存图的邻近单词关系和句法依存关系,分别得到两个图的单词嵌入表示,然后使用软注意力机制对文本的单词嵌入进行线性变换,并分别进行混合池化,得到对应的文本特征向量,最后依据自适应特征融合策略将两个文本向量变成一个定长的文本特征向量,使用 softmax 网络层进行文本分类。

### 3.3 文本共现图和句法依存图的构建

图的构建过程是把每个单词当作节点,根据单词之间的某种规则或者联系来确定两个单词节点之间是否存在连边。本文利用单词共现的方式和句法依存关系的方式来确定单词节点之间的连边,构建了两张文本图网络,丰富了文本特征信息,提升了嵌入效果。

#### 3.3.1 文本共现图

假定一段长度为  $n$  的文本序列,我们用  $W = \{\omega_1, \omega_2, \dots, \omega_n\}$  表示,设定滑动窗口大小为  $m$ ,共现文本图的构建步骤如下:

(1)从给定文本序列中提取单词集合,这个单词集合就是共现图的节点集合,也就是说每一个单词代表一个节点。

(2)滑动窗口沿着文本序列从左向右滑动,窗口中心单词是  $\omega_i$ ,如果  $\omega_j$  和  $\omega_i$  在一个窗口内,就构建  $\omega_i$  和  $\omega_j$  单词节点之间的连边,用  $c_{ij}$  表示两个节点在滑动窗口内共现出现的次数。

(3)计算两个节点连边的权重  $w_{ij} = \frac{c_{ij}}{\sum c_{ij}}$ 。

以“Beckham is a football star who has won many awards”这个句子为例。图 2 给出了把一个文本转换成文本共现图的例子。

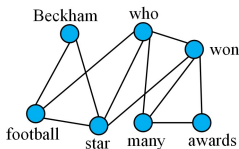


图 2 文本共现图的构建

Fig. 2 Construction of text co-occurrence graph

在进行文本分类前,往往需要对文本进行预处理,把一些停用词去掉,因此在图 2 的例子中,只选取了“Beckham”“football”“star”“who”“won”“many”“awards”作为图中的节点,同时设定滑动窗口大小为 3。因此,“Beckham”和“football”在同一个滑动窗口中,“Beckham”和“football”两个单词节点间存在一条无向边。

#### 3.3.2 句法依存图

句法分析是自然语言理解的基础,用来分析句子中的成分以及各个成分之间的关系。句法依存分析是按照依存关系对句子中的词语进行分析,句子中的任意两个词之间都存在着某种支配和被支配关系。本文采用斯坦福的句法分析器 stanfordCoreNLP 对文本中的句子进行句法依存分析,将文本转变成为一个图网络数据结构。比如,给定一个句子“Jack runs very fast”,利用 stanfordCoreNLP 分析该句子,结果如图 3 所示。

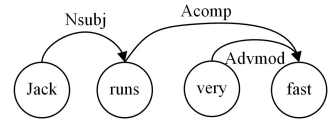


图 3 句法依存图的构建

Fig. 3 Construction of syntax dependency graph

从图 3 可以看出,句法分析的结果中不仅标出了词语之间的依赖关系,还包括依赖关系类型,如“runs”和“fast”之间是“acomp”关系,表示形容词的补充。我们按照句法分析结果构建文本的句法图,通过句法分析器分析词语之间的关系能较好地表达词语之间的依赖关系和整个文本的结构。通过句法分析得到的文本实际上是一个有向图,我们统一将有向图转换成无向图来考虑。

句法依存图和文本共现图各有优缺点,它们对文本不同方面的特征进行建模。文本共现图的构建过程简单,它是从邻域单词的共现角度考虑;而句法依存图相对比较复杂,它是从文本自身的语法特性进行分析。因此,本文同时使用两种图数据,同时考虑单词之间的相邻关系和语法关系,以捕获更多的文本特征信息。

### 3.4 文本图嵌入

通过图的构建,得到两张文本图:文本共现图  $G_1 = (V, E_1)$  和句法依存图  $G_2 = (V, E_2)$ ,用  $X$  表示文本单词节点的初始化嵌入表示。其中,  $V$  是图中的节点的集合,也就是文档中预处理后文本的单词集合;  $E_1$  是文本共现图中边的集合,代表两个单词之间存在关系;  $E_2$  是句法依存图的边的集合。

使用两个图卷积神经网络来分别学习两个文本网络的拓扑结构,依据两个文本网络的单词连接关系进行邻域聚合操作,从而得到文本单词的嵌入表示,计算式如下:

$$\mathbf{H}_M = \sigma(\mathbf{A}_M \mathbf{H}_M \mathbf{W}_M) \quad (1)$$

$$\mathbf{H}_L = \sigma(\mathbf{A}_L, \mathbf{H}_L, \mathbf{W}_L) \quad (2)$$

其中,  $\mathbf{H}_M$  表示文本共现图的嵌入矩阵,  $\mathbf{H}_L$  表示句法依存图的嵌入矩阵。  $\mathbf{H}_L^i = X$ ,  $X \in R^{N \times L}$  是图卷积网络的第一层输入,  $N$  代表单词节点的个数,  $L$  表示单词节点的嵌入维度。  $\mathbf{A} \in R^{N \times N}$  是邻接矩阵,元素都是 0 到 1 之间的小数,表示两个单词节点之间的连边权重。  $\mathbf{W} \in R^{D \times K}$  表示权重参数矩阵,  $\sigma(\cdot)$  表示非线性激活函数。若要捕获单词节点的高阶领域信息,则需要叠加多层图卷积网络。

### 3.5 软注意力机制和混合池化

经过图卷积层之后,我们得到了文本中每个单词节点的嵌入表示。为了从图数据结构中得到文本的表示向量,需要将这些节点的表示向量聚合成文档的图级表示,基于这个图级表示向量,进行最后的分类。图池化部分主要包含软注意力机制和混合池化两部分。软注意力机制包含两个多层感知机,其中一个用于对文本嵌入单词向量的非线性变换,另一个用于计算文本每一个单词向量的权重系数,这个权重系数代表了文档各个单词的重要性,计算过程如式(3)所示:

$$\mathbf{h}_L = \sigma(mlp_1(\mathbf{h}_L^i)) \odot \tanh(mlp_2(\mathbf{h}_L^i)) \quad (3)$$

$$\mathbf{h}_M = \sigma(mlp_3(\mathbf{h}_M^i)) \odot \tanh(mlp_4(\mathbf{h}_M^i)) \quad (4)$$

其中,  $\sigma$  表示激活函数;  $mlp_1, mlp_2, mlp_3, mlp_4$  表示多层感知机,  $mlp_1$  和  $mlp_3$  计算权重系数,  $mlp_2$  和  $mlp_4$  对节点表示

进行非线性变换; $\mathbf{h}_i$ 表示文本共现图中每一个节点的嵌入向量, $\mathbf{h}_M$ 表示句法依存图中每一个节点的嵌入向量。

图池化操作对于文本特征向量的提取非常重要,因为其可以降低特征输出维度,捕获更大范围的信息,从而获得图级别的特征向量。均值池化和最大池化是最常用的两种池化方式,它们各有优缺点。因此,我们将两种方式结合,在对各个节点进行聚合时,同时使用均值池化和最大池化,兼顾了它们的优点。计算式如下:

$$\mathbf{h}_{G1} = \left( \frac{1}{|\mathcal{V}|} \sum_{v \in G1} \mathbf{h}_v + \text{MaxPooling}(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_v) \right) / 2 \quad (5)$$

$$\mathbf{h}_{G2} = \left( \frac{1}{|\mathcal{V}|} \sum_{v \in G2} \mathbf{h}_v + \text{MaxPooling}(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_v) \right) / 2 \quad (6)$$

其中, $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_v$ 表示文本单词节点的嵌入表示,公式前半部分表示均值池化,后半部分表示最大池化。

### 3.6 自适应特征融合

实际上,两个独立的图神经网络提取的特征向量是不同的。从文本共现图中提取的是单词共现关系,从句法依存图中提取的是文本的句法依存关系。我们需要解决的问题是两个不同文本的特征向量如何融合,特征越重要,占的比重越大。为此,本文提出了自适应的特征融合策略,使模型自身根据数据集的不同来学习不同特征的重要性。计算式如下:

$$\mathbf{h} = a * \mathbf{h}_{G1} + (1-a)\mathbf{h}_{G2} \quad (7)$$

其中, $a$ 是一个可训练参数,在模型训练过程中,它会自动寻找优化的特征融合参数。

最后,我们将得到的文本表示向量 $\mathbf{h}$ 输入softmax层进行文本分类,如式(8)所示:

$$y = \text{softmax}(\mathbf{W}\mathbf{h} + \mathbf{b}) \quad (8)$$

通过最小化交叉熵损失函数来达到优化目标:

$$\mathcal{L} = - \sum_i y_i \log(\hat{y}_i) \quad (9)$$

在式(8)和式(9)中, $\mathbf{W}$ 是权重参数矩阵; $\mathbf{b}$ 是偏置向量; $y_i$ 是 $y$ 的第 $i$ 个元素,表示第 $i$ 个样本的预测结果。 $\hat{y}_i$ 是第 $i$ 个样本的真实标签。模型中的所有参数采用梯度下降算法Adam进行更新优化。

### 3.7 模型实施步骤

本算法的伪代码如算法1所示。算法输入为文本数据 $D$ ,输出是文本分类的结果,即文本的标签 $y$ 。

#### 算法1 Mix-GCN

Input:  $D, y$

Input Parameters:  $k, \text{epoch}$

1.  $D \leftarrow \text{preprocess}(D)$  {文本预处理}
2.  $A_1 \leftarrow \text{construct1}(D, k)$  {构建文本共现图}
3.  $A_2 \leftarrow \text{construct2}(D)$  {构建句法依存图}
4.  $X \leftarrow \text{get}(D)$  {得到预测训练词向量}
5. for in range(epoch)
6.  $H_M \leftarrow \text{GCN1}(X, A_1)$  {文本共现网络嵌入学习}
7.  $H_L \leftarrow \text{GCN2}(X, A_2)$  {句法依赖网络嵌入学习}
8.  $h_{\theta 1} \leftarrow \text{Mixpooling}(H_M)$
9.  $h_{\theta 2} \leftarrow \text{Mixpooling}(H_L)$
10.  $\mathbf{h} \leftarrow \text{AdaptiveFusion}(h_{\theta 1}, h_{\theta 2})$
11.  $y_{\text{pred}} \leftarrow \text{softmax}(\mathbf{m})$
12. minimize(loss( $y_{\text{pred}}, y$ ))

13. end

14. Output:  $y \leftarrow \text{Mix-GCN}()$

### 3.8 时间复杂度分析

文本共现图和句法依存图的节点都是单词,假设两张图的节点数为 $N$ ,两张图的连边数分别为 $E_1$ 和 $E_2$ 。Mix-GCN模型中包含两个GCN,一个GCN计算文本共现图的时间复杂度为 $O(E_1)^{[54]}$ ,另一个GCN计算句法依存图的时间复杂度为 $O(E_2)$ 。图池化部分的时间复杂度为 $O(2N)$ ,因此,Mix-GCN模型的时间复杂度为 $O(E_1 + E_2)$ 。

## 4 实验

为了验证模型的有效性,我们在6个公开数据集上做了文本分类实验,包括Ohsumed, R52, R8和Movie Review(MR)等。使用正确率作为评价模型性能的指标。

$$\text{acc} = M/N \quad (10)$$

其中, $M$ 表示预测正确的样本数量, $N$ 表示样本总数。

### 4.1 数据集

R8<sup>[55]</sup>, MR<sup>[56]</sup>, R52<sup>[55]</sup>, Ohsumed<sup>[57]</sup>, AG News<sup>[58]</sup>, SST1<sup>[59]</sup>数据集的信息如表1所列。

表1 实验数据集信息

Dataset	Training	Test	Classes	Average Length
R8	5485	2189	8	65.72
MR	7180	3554	2	20.39
R52	6532	2568	52	69.00
Ohsumed	3357	4043	23	135.00
AG News	120000	7600	4	45.00
SST1	8544	3311	5	20.00

MR:MR是一个电影评论数据集,每个样本只有一句话,每个样本的情感标签只能是消极或积极。

Ohsumed:Ohsumed数据集来源于MEDLINE数据库系统,该数据库系统主要是重要生物医学文章的书目数据库系统。Ohsumed数据集集中的每个文档样本都有23个疾病中的一个或多个标记,去除拥有多个标签的文档后,ohsumed数据集分为3357个训练文档和4043个测试文档。

R8和R52:R8和R52是路透社21578个数据集中的两个子集。R8有8个类别,包括5485个训练文档和2189个测试文档;R52有52个类别,并被拆分为6532个训练文档和2568个测试文档。

AG news:AG News数据集共包含496835条数据(新闻文章),它们都来自AG新闻语料库,每个文档样本包含标题和描述字段,数据集被分成4个类别。

SST1:Stanford Sentiment Treebank是一个电影评论数据集,是MR数据集的扩展。评论根据句子的情绪水平分为5类,包括非常正面、正面、中立、负面和非常负面。

### 4.2 对比算法

LR<sup>[5]</sup>:线性回归模型是机器学习中的一种分类模型,具有简单、高效的优点。

SVM<sup>[8]</sup>:SVM表示支持向量机模型,属于监督学习机器学习算法,旨在找到一个超平面使得分割最大化。

FastText<sup>[16]</sup>: Facebook 提出的一种快速文本分类方法, 模型只有 3 个网络层, 模型简单, 训练速度快。

CNN<sup>[10]</sup>: 该模型使用预训练的 GloVe 词向量和 CNN 网络对输入文本进行嵌入和卷积操作, 最后通过全连接层进行分类。

LSTM<sup>[12]</sup>: 该方法使用 LSTM 网络提取文本语义特征, 然后使用 softmax 进行文本分类。

Text-GCN<sup>[23]</sup>: 该方法使用紧耦合的方式将整个语料库构建成为一张异构图, 然后使用图卷积神经网络学习文本的嵌入表示。

SWEM<sup>[60]</sup>: 一种简单快速的文本分类方法, 把预训练词向量简单池化后直接进行文本分类。

#### 4.3 实验参数设置

在生成文本图的过程中, 滑动窗口大小设置为 5, 滑动窗口的步长设为 1; 图中最大节点数设置为 200, 如果超过 200, 则只截取前 200 个; 单词嵌入维度设为 300, 使用 Adam 训练器, 进行  $epochs=200$  的训练, 学习率初始化为 0.01, dropout 率设为 0.55, 批处理大小设为 4096; 在模型中使用 RELU 作为非线性激活函数。其中图卷积模块使用两层卷积, 单词嵌入维度为 300。

#### 4.4 实验结果

表 2 列出了本文提出的模型以及对比算法的性能。可以看出, Mix-GCN 在 6 个数据集集中的 4 个数据集上取得了最优的结果。在没有取得最优结果的 R8 数据集上, Mix-GCN 模型的分类正确率排名第 2。在 AG News 数据集上, Mix-GCN 模型的正确率排名第 3。综上所述, 在 9 个文本分类模型中, Mix-GCN 模型的总体性能最佳。这表明将图神经网络应用于文本分类并引入文本的句法依存信息是有效的。其原因可能是 GCN 模型能够对单词的全局共现关系进行建模, 并且在嵌入表示时综合考虑了拓扑信息和单词共现信息。此外, 我们发现 Mix-GCN 模型在 R52 和 Ohsumed 数据集上的提升效果更明显, 这是因为 R52 和 Ohsumed 数据集的文档较长, 长文本的语法关系更为丰富, 包含的信息也更多。

表 2 不同算法的文本分类正确率

Table 2 Text classification accuracy of different algorithms

方法	MR	R8	R52	Ohsumed	AG News	SST1
CNN	0.7775	0.9571	0.8759	0.5844	0.9081	0.4364
RNN	0.7768	0.9631	0.9054	0.4927	0.9014	0.4225
FastText	0.7514	0.9613	0.9281	0.5770	<b>0.9150</b>	0.4261
SWEM	0.7665	0.9532	0.9294	0.6312	0.9106	—
LR+TFIDF	0.7446	0.9531	0.8728	0.6218	0.8527	0.4035
SVM+TFIDF	0.7569	0.9535	0.9231	0.6345	0.8829	0.3829
Bi-LSTM	0.7689	0.9631	0.9051	0.4935	0.8982	0.4765
TextGCN	0.7674	<b>0.9707</b>	0.9356	0.6836	0.9142	—
Mix-GCN	<b>0.7850</b>	0.9683	<b>0.9417</b>	<b>0.6983</b>	0.9126	<b>0.4780</b>

#### 4.5 Mix-GCN 的参数敏感度实验

图 4—图 6 分别给出了 Mix-GCN 模型在 3 个数据集上使用不同滑动窗口大小对分类的影响。从 3 个图中可以看出, 在 MR, R8, Ohsumed 这 3 个数据集上, 随着滑动窗口大小的增加, Mix-GCN 模型分类正确率总体呈现先提高后下降的趋势。因此, 将滑动窗口设置得过大或过小都不是最优的策略。如果滑动窗口设置过大, 一个直接的影响就是窗口中

的单词数量会增加, 导致文本图中节点连边数量增多, 在为节点增加更多邻域信息的同时会引入与分类无关的信息; 而滑动窗口设置过小意味着窗口中的单词数量会减少, 构建得到的文本图中的节点连边数减少, 可能会丢失节点之间一些重要的连边, 从而损失重要的信息, 降低文本分类性能。

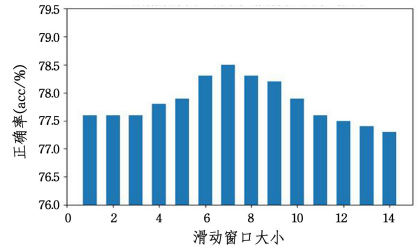


图 4 MR 数据集分类正确率和滑动窗口大小的关系

Fig. 4 Relationship between classification accuracy and sliding window size in MR dataset

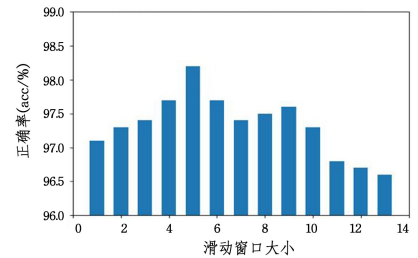


图 5 R8 数据集分类正确率和滑动窗口大小的关系

Fig. 5 Relationship between classification accuracy and sliding window size in R8 dataset

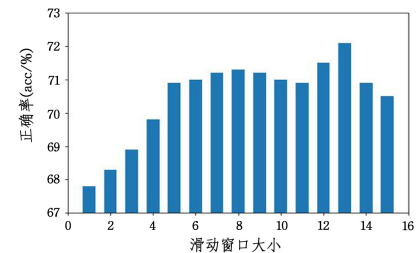


图 6 Ohsumed 数据集分类正确率和滑动窗口大小的关系

Fig. 6 Relationship between classification accuracy and sliding window size in Ohsumed dataset

通过以上分析可以得知, 滑动窗口大小决定了构建文本图的语义信息丰富程度, 对构建文本图具有重要的作用, 滑动窗口应当设置为适当数值, 太大或太小都不合适。

**结束语** 本文提出了一种新颖的基于图神经网络的文本分类方法, 将传统文本分类问题转变成一个图分类问题。首先根据文本单词的邻域共现关系和文本单词的句法依存关系分别将文本转换成文本共现图和句法依存图; 然后使用图卷积神经来学习文本单词节点的嵌入表示, 使用软注意力机制和混合池化得到两个文本网络的嵌入表示; 最后根据自适应融合策略, 得到一个定长的特征向量进行文本分类。实验结果表明了在模型中考虑文本语言自身的语法特性和引入文本句法依存信息对提升文本分类效果的有效性。未来, 我们将考虑在文本嵌入学习中增加更丰富的特征信息, 如单词之间的位置信息以及外部的常识信息; 并且, 在将文本分类问题转化

为图分类问题和节点分类问题的基础上,我们将进一步研究如何对文本网络进行更有效的嵌入学习表示,从而更加有效地提升文本分类的准确度。

### 参考文献

- [1] ROY P K, SINGH J P, BANERJEE S. Deep learning to filter SMS Spam [J]. *Future Generation Computer Systems*, 2020, 102:524-533.
- [2] MINAE S, KALCHBRENNER N, CAMBRIA E, et al. Deep Learning-based Text Classification: A Comprehensive Review [J]. *ACM Computing Surveys*, 2021, 54(3):1-40.
- [3] XIONG S F, JI D H. A short text sentiment topic model for product review analysis [J]. *Acta Automatica Sinica*, 2016, 42(8):1227-1237.
- [4] JIN L, HE T T. Chinese text classification model based on improved TF-IDF and ABLCNN [J]. *Computer Science*, 2021, 48(11A):170-175.
- [5] HARRELL F E. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis [M]. New York: Springer, 2015.
- [6] RISH I. An empirical study of the naive Bayes classifier [C] // IJ-CAI 2001 Workshop on Empirical Methods in Artificial Intelligence. 2001:41-46.
- [7] KELLER J M, GRAY M R, GIVENS J A. A fuzzy k-nearest neighbor algorithm [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1985(4):580-585.
- [8] SANCHEZ AV D. Advanced support vector machines and kernel methods [J]. *Neurocomputing*, 2003, 55(1/2):5-20.
- [9] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [J]. *arXiv*:1301.3781, 2013.
- [10] KIM Y. Convolutional neural networks for sentence classification [J]. *arXiv*:1408.5882, 2014.
- [11] JOHNSON R, TONG Z. Deep Pyramid Convolutional Neural Networks for Text Categorization [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. 2017:562-570.
- [12] MIKOLOV T, KARAFIAT M, BURGET L, et al. Recurrent neural network based language model [C] // Proceedings of the Annual Conference of the International Speech Communication Association. 2010:1045-1048.
- [13] ZHANG Y, LIU Q, SONG L. Sentence-State LSTM for Text Representation [C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018:317-327.
- [14] YANG M, ZHAO W, YE J, et al. Investigating capsule networks with dynamic routing for text classification [C] // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:3110-3119.
- [15] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. *Nature*, 2015, 521(7553):436-444.
- [16] 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: Volume 2, Short Papers. 2017:427-431.
- [17] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification [J]. *Advances in Neural Information Processing Systems*, 2015, 28:649-657.
- [18] TANG D, QIN B, LIU T. Document modeling with gated recurrent neural network for sentiment classification [C] // Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2015:1422-1432.
- [19] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification [C] // Twenty-ninth AAAI Conference on Artificial Intelligence. 2015:2267-2273.
- [20] WANG R, LI Z, CAO J, et al. Convolutional Recurrent Neural Networks for Text Classification [C] // 2019 International Joint Conference on Neural Networks. 2019:1-6.
- [21] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks [J]. *arXiv*:1609.02907, 2016.
- [22] GAO H, CHEN Y, JI S. Learning Graph Pooling and Hybrid Convolutional Operations for Text Representations [C] // World Wide Web Conference. 2019:2743-2749.
- [23] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019:7370-7377.
- [24] PENG H, LI J, HE Y, et al. Large-scale hierarchical text classification with recursively regularized deep graph-cnn [C] // Proceedings of the 2018 World Wide Web Conference. 2018:1063-1072.
- [25] WU M, PAN S, ZHU X, et al. Domain-adversarial graph neural networks for text classification [C] // 2019 IEEE International Conference on Data Mining. IEEE, 2019:648-657.
- [26] IVANOVA A, OEPEN S, ØVRELID L. Survey on parsing three dependency representations for English [C] // 51st Annual Meeting of the Association for Computational Linguistics Proceedings of the Student Research Workshop. 2013:31-37.
- [27] CHEN M, WEINBERGER K Q, SHA F. An alternative text representation to TF-IDF and Bag-of-Words [J]. *arXiv*:1301.6770, 2013.
- [28] ZHU Y, LI L, LUO L. Learning to classify short text with topic model and external knowledge [C] // International Conference on Knowledge Science, Engineering and Management. 2013:493-503.
- [29] MA X, JIN R, PAIK J Y, et al. LargeScale Text Classification with Efficient Word Embedding [C] // International Conference on Mobile and Wireless Technology. 2017:465-469.
- [30] BENGIO Y, DUCHARME R, VINCENT P, et al. A neural probabilistic language model [J]. *The Journal of Machine Learning Research*, 2003, 3:1137-1155.
- [31] LE Q, MIKOLOV T. Distributed representations of sentences and documents [C] // International Conference on Machine Learning. PMLR, 2014:1188-1196.
- [32] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation [C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1532-1543.
- [33] TAN T, PHIENTHRAKUL T. Sentiment Classification Using

- Document Embeddings Trained with Cosine Similarity [C] // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Student Research Workshop. 2019: 407-414.
- [34] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk; Online learning of social representations [C] // Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014:701-710.
- [35] GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: A survey [J]. Knowledge-Based Systems, 2018, 151:78-94.
- [36] ZHANG R, LI J T. A Survey of Research on Scene Segmentation Algorithms Based on Deep Learning [J]. Computer Research and Development, 2020, 57(4): 859-875.
- [37] FEI H, LI F, LI B, et al. Encoder-decoder based unified semantic role labeling with label-aware syntax [C] // Proceedings of AAAI National Conference on Artificial Intelligence. 2021:1479-1488.
- [38] LI Y, JIN R, LUO Y. Classifying relations in clinical narratives using segment graph convolutional and recurrent neural networks (Seg-GCRNs) [J]. Journal of the American Medical Informatics Association, 2019, 26(3): 262-268.
- [39] BASTINGS J, TITOV I, AZIZ W, et al. Graph convolutional encoders for syntax-aware neural machine translation [J]. arXiv: 1704. 04675, 2017.
- [40] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [J]. arXiv: 1710. 10903, 2017.
- [41] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs [J]. arXiv: 1706. 02216, 2017.
- [42] KIM D, OH A. How to find your friendly neighborhood: Graph attention design with self-supervision [J]. arXiv: 2204. 04879, 2022.
- [43] ALON U, YAHAV E. On the bottleneck of graph neural networks and its practical implications [J]. arXiv: 2006. 05205, 2020.
- [44] ZHAO T, LIU Y, NEVES L, et al. Data Augmentation for Graph Neural Networks [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2021:11015-11023.
- [45] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation [J]. arXiv: 1508. 04025, 2015.
- [46] CHO K, COURVILLE A, BENGIO Y. Describing multimedia content using attention-based encoder-decoder networks [J]. IEEE Transactions on Multimedia, 2015, 17(11): 1875-1886.
- [47] ZENG Y F, LAN T, WU Z F, et al. Aspect-level sentiment classification model based on dual-memory attention [J]. Chinese Journal of Computers, 2019, 8: 1845-1857.
- [48] SUKHBAAATAR S, WESTON J, FERGUS R. End-to-end memory networks [C] // Proceedings of the Advances in Neural Information Processing Systems. 2015:2440-2448.
- [49] WANG W, PAN S J, DAHLMEIER D, et al. Coupled multi-layer attentions for co-extraction of aspect and opinion terms [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2017.
- [50] 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.
- [51] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] // Proceedings of the Advances in Neural Information Processing Systems. 2017:5998-6008.
- [52] SUN Z, HUANG S, WEI H R, et al. Generating diverse translation by manipulating multi-head attention [C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 8976-8983.
- [53] MISRA D, NALAMADA T, ARASANIPALAI A U, et al. Rotate to attend: Convolutional triplet attention module [C] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021:3139-3148.
- [54] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24.
- [55] DAI Y, SHOU L, GONG M, et al. Graph Fusion Network for Text Classification [J]. Knowledge-Based Systems, 2022, 236: 107659.
- [56] GUO B, ZHANG C, LIU J, et al. Improving text classification with weighted word embeddings via a multi-channel TextCNN model [J]. Neurocomputing, 2019, 363: 366-374.
- [57] LIU H, CHEN G, LI P, et al. Multi-label text classification via joint learning from label embedding and label correlation [J]. Neurocomputing, 2021, 460: 385-398.
- [58] XU J, CAI Y, WU X, et al. Incorporating context-relevant concepts into convolutional neural networks for short text classification [J]. Neurocomputing, 2020, 386: 42-53.
- [59] WU Y, LI J, WU J, et al. Siamese capsule networks with global and local features for text classification [J]. Neurocomputing, 2020, 390: 88-98.
- [60] 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.



**YANG Xu-hua**, born in 1971, Ph.D, professor, is a senior member of China Computer Federation. His main research interests include machine learning and network science.



**MAO Jian-fei**, born in 1976, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include machine learning and natural language processing.