



# 计算机科学

COMPUTER SCIENCE

## 基于注意力门控图神经网络的文本分类

邓朝阳, 仲国强, 王栋

### 引用本文

邓朝阳, 仲国强, 王栋. 基于注意力门控图神经网络的文本分类[J]. 计算机科学, 2022, 49(6): 326-334.

DENG Zhao-yang, ZHONG Guo-qiang, WANG Dong. [Text Classification Based on Attention Gated Graph Neural Network](#)[J]. Computer Science, 2022, 49(6): 326-334.

---

### 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)

#### [基于注意力机制和门控网络相结合的混合推荐系统](#)

Hybrid Recommender System Based on Attention Mechanisms and Gating Network

计算机科学, 2022, 49(6): 158-164. <https://doi.org/10.11896/jsjcx.210500013>

#### [融合用户偏好的图神经网络推荐模型](#)

Graph Neural Network Recommendation Model Integrating User Preferences

计算机科学, 2022, 49(6): 165-171. <https://doi.org/10.11896/jsjcx.210400276>

#### [基于解耦-检索-生成的图像风格化描述生成模型](#)

Stylized Image Captioning Model Based on Disentangle-Retrieve-Generate

计算机科学, 2022, 49(6): 180-186. <https://doi.org/10.11896/jsjcx.211100129>

#### [基于特征注意力融合网络的遥感变化检测研究](#)

Remote Sensing Change Detection Based on Feature Fusion and Attention Network

计算机科学, 2022, 49(6): 193-198. <https://doi.org/10.11896/jsjcx.210500058>

#### [基于样本分布损失的图像多标签分类研究](#)

Study on Multi-label Image Classification Based on Sample Distribution Loss

计算机科学, 2022, 49(6): 210-216. <https://doi.org/10.11896/jsjcx.210300267>

# 基于注意力门控图神经网络的文本分类

邓朝阳<sup>1</sup> 仲国强<sup>1</sup> 王 栋<sup>2</sup>

1 中国海洋大学信息科学与工程学院 山东 青岛 266100

2 中国海洋大学图书馆 山东 青岛 266100

(zydeng@stu.ouc.edu.cn)

**摘 要** 针对现有的文本分类工作在生成文本表示时通常忽略单词之间语义交互的问题,提出了一种新的基于注意力门控图神经网络的文本分类模型,有效地利用单词的语义特征并在充分语义交互的基础上提高了文本分类的准确率。首先,将每个输入文本转换为独立图结构数据并提取单词节点的语义特征;其次,利用注意力门控图神经网络对单词节点的语义特征进行交互和更新;然后,使用基于注意力机制的文本池化模块提取语义特征具有判别性的单词节点,以构建文本图表示;最后,基于文本图表示实现有效的文本分类。实验结果表明,所提方法在文本数据集 Ohsumed, R8, R52 和 MR 上的准确率分别达到了 70.83%, 98.18%, 94.72% 与 80.03%, 优于现有的方法。

**关键词**: 深度学习; 文本分类; 图神经网络; 注意力机制

中图法分类号 TP183

## Text Classification Based on Attention Gated Graph Neural Network

DENG Zhao-yang<sup>1</sup>, ZHONG Guo-qiang<sup>1</sup> and WANG Dong<sup>2</sup>

1 School of Information Science and Engineering, Ocean University of China, Qingdao, Shandong 266100, China

2 Library of Ocean University of China, Qingdao, Shandong 266100, China

**Abstract** To address the problem that the existing text classification work usually ignores the semantic interaction between words when generating text representation, this paper proposes a novel text classification model based on attention gated graph neural network. It makes effective use of the semantic features of words and improves the accuracy of text classification based on the adequate semantic interaction. Firstly, each input text is converted to a single graph-structured data and the semantic features of word nodes are extracted. Secondly, attention gated graph neural network is used to interact and update the semantic features of word nodes. In addition, the attention-based text pooling module is used to extract the word nodes with discriminative semantic features to construct text graph representation. Finally, effective text classification is implemented based on the text graph representation. Experimental results show that the proposed method achieves an accuracy of 70.83%, 98.18%, 94.72% and 80.03% on Ohsumed, R8, R52 and MR datasets, respectively, and outperforms existing methods.

**Keywords** Deep learning, Text classification, Graph neural network, Attention mechanism

## 1 引言

随着互联网技术的兴盛,大量的数据正在源源不断地产生,其中文本数据占绝大部分。面对体量庞大的文本数据,人们往往会遭遇信息灾难,一个典型的例子是从铺天盖地的邮件中筛选垃圾邮件。为了能够高效地获取和利用所需要的信息,就需要先对文本进行分类,根据文本所具有的特征信息将其归类到相应的一个或多个类别<sup>[1]</sup>。文本分类(Text Classi-

fication)是一项非常重要且富有挑战性的工作,也是垃圾邮件过滤<sup>[2]</sup>、话题标注<sup>[3]</sup>、情感分类<sup>[4]</sup>和摘要模型<sup>[5]</sup>等许多其他任务的基础。因此,准确地进行文本分类在一定程度上能够缩短用户筛选文本信息需要的时间,减小搜索负担。

在自然语言处理领域中,许多研究人员尝试从文档所包含的字符、单词及句子入手,利用它们的特征来整合学习文档表示并最终实现文本分类。一般来讲,文本分类可以总结为两种不同的实现方式:人工分类和自动分类。人工分类方式

到稿日期:2021-04-20 返修日期:2021-06-12

基金项目:科技创新 2030——“新一代人工智能”重大项目(2018AAA0100400);装备预研教育部联合基金项目(6141A020337);山东省自然科学基金项目(ZR2020MF131);青岛市科技计划项目(21-1-4-ny-19-nsh)

This work was supported by the Major Project for New Generation of AI (2018AAA0100400), Joint Fund of the Equipments Pre-Research and Ministry of Education of China (6141A020337), Natural Science Foundation of Shandong Province (ZR2020MF131) and Science and Technology Program of Qingdao (21-1-4-ny-19-nsh).

通信作者:王栋(wangdong@ouc.edu.cn)

通常需要注释人员使用一套预先编写好的语言规则对文本内容进行解释翻译,然后对其分类。其中,常见的方法主要有粗糙集(Rough Set)<sup>[6]</sup>和关联规则(Association Rules)<sup>[7]</sup>等,这类方法具有很明显的缺陷,通常会造成大量的时间消耗及昂贵的费用。自动分类方法分为基于机器学习的方法和基于深度学习的方法。基于机器学习的分类方法主要包括朴素贝叶斯(Naive Bayesian Model)<sup>[8]</sup>、支持向量机(Support Vector Machine)<sup>[9]</sup>、 $k$ 近邻( $k$ -nearest Neighbor)<sup>[10]</sup>以及TF-IWF算法<sup>[11]</sup>等。虽然这类方法能够带来显著的性能提升,但仍然受限于手动设计的特征且效率低下。随着深度学习技术的不断发展,神经网络为解决各类复杂问题提供了新的途径,具有代表性的是卷积神经网络(Convolutional Neural Network, CNN)<sup>[12]</sup>和循环神经网络(Recurrent Neural Network, RNN)<sup>[13]</sup>。2014年, Kim<sup>[14]</sup>提出了基于CNN的文本分类模型,该模型使用预训练的词嵌入向量作为输入,并利用一维卷积捕捉句子中的词序信息。随后, Liu等<sup>[15]</sup>提出了基于RNN的文本分类方法,将目标文本看作单词序列,进而使用RNN提取序列中的上下文信息。基于神经网络构建文本分类模型是一次成功的尝试,这为其他研究人员提供了文本分类的新思路。Wang等<sup>[16]</sup>将正则化层次Softmax和CNN相结合,构建了有效的文本分类模型。Ma等<sup>[17]</sup>将神经网络成功应用于旅游问句的文本分类任务中。尽管这些方法减少了大量与特征工程相关的工作,但它们主要关注于单词的局部特征,缺少长距离单词之间以及非连续单词之间的语义交互,导致模型在训练阶段无法有效地利用单词之间潜在的语义关系。事实上,单词之间存在的语义关系能够为文本的分类提供更多信息,如果合理利用,将会提高模型分类准确率。近年来,图神经网络(Graph Neural Network, GNN)因能够构建有效的关系结构并能捕捉全局结构信息而被广泛应用于解决该类问题。Yao等<sup>[18]</sup>基于图卷积网络(Graph Convolutional Network, GCN)<sup>[19]</sup>提出了TextGCN,该网络通过为文本和单词构建单一的异构图将文本分类任务转换为节点分类任务。然而,这种图构建方式存在两个主要的缺点,一方面是需要考虑整个语料库中的所有单词,从而导致大量的内存资源被用于存储边信息,另一方面是图结构中边的表达能力受到严重限制,从而影响了最终的分类效果。与TextGCN不同, Huang等<sup>[20]</sup>为每个输入文本构建独立的图结构,并采用全局参数共享的方式更新单词节点的特征信息。虽然文献[19]中的方法大大减少了模型所需要存储的边的数量,但其单词节点之间的语义信息没有得到充分的交互。Velickovic等<sup>[21]</sup>将注意力机制<sup>[22]</sup>和图神经网络相结合,提出了图注意力网络(Graph Attention Network, GAT),该模型通过计算注意力分数为每个节点的邻居节点分配不同的权重,从而提取重要节点的语义特征。受到文献[20]的启发,本文将每个输入的目标文本转换为独立的图结构数据,将文本中的单词作为节点,在节省内存资源消耗的同时保证了节点的特征表达能力。同时,为了有效地利用单词的语义特征信息,并在单词节点之间进行充分的语义信息交互,本文提出了注意力门控图神经网络AGGNN,并基于AGGNN搭建了有效的文本分类模型。该模型将文本图数据输入AGGNN中,以捕捉单词

之间的依赖关系,尤其是非连续单词之间的长距离依赖关系以及上下文语义信息。除此之外,该模型使用新提出的基于注意力机制的文本池化模块TextPool来计算每个单词节点的注意力分数,在此基础上提取当前层中具有判别性语义信息的节点,构建有效的文本图嵌入表示,以完成最终的文本分类。

本文的主要贡献与创新点如下:1)提出了新的注意力门控图神经网络AGGNN,AGGNN在门控循环单元<sup>[23]</sup>的基础上设计了注意力门,加强了邻域内不同单词节点之间的特征信息交互;2)提出了基于注意力机制的文本池化模块TextPool,TextPool通过设计的全局注意力机制和局部注意力机制来提取当前层中具有判别性语义特征的单词节点。不同的注意力机制能够针对不同尺度大小的数据集作适应性调整;3)将本文提出的模型分别应用于Ohsumed, R8, R52和MR这4个文本数据集,其得到的分类结果优于现有的其他模型。

## 2 基于注意力门控图神经网络的文本分类模型

本节主要介绍提出的基于注意力门控图神经网络的文本分类模型及其各模块的具体结构。该模型主要包括文本图构建、注意力门控图神经网络AGGNN、基于注意力的文本池化模块TextPool以及图读出模块Readout。首先,将每个输入的文本数据转换为独立的图结构数据;其次,将构建的图结构数据(文本图)送入AGGNN以更新单词节点的特征信息;然后,TextPool根据当前层中的所有单词节点,提取出其中具有判别性特征的关键节点;最后,Readout基于提取出的关键节点构建文本嵌入表示,将结果送入文本分类器以完成文本分类。整个模型的处理流程如图1所示。

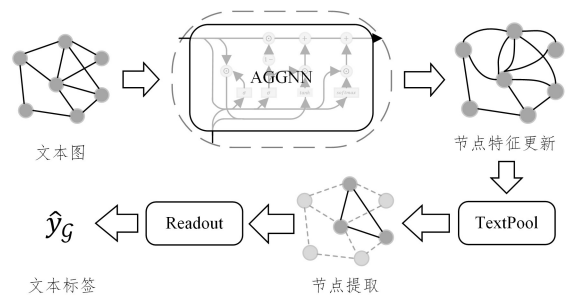


图1 基于注意力门控图神经网络的文本分类模型

Fig. 1 Text classification model based on attention gated graph neural network

### 2.1 文本图构建

首先对文本图中涉及的符号进行详细描述。给定文本图 $G=(V, E)$ ,其中 $V, E$ 分别为图中顶点、边的集合,  $|V|=n$ 表示节点的数目。 $X \in \mathbb{R}^{n \times d}$ 表示 $n$ 个单词节点的特征矩阵,每个单词使用 $d$ 维的词向量表示。对于特征矩阵 $X$ ,本文以隐藏状态 $h^t \in \mathbb{R}^{n \times d}$ 表示其在训练过程中每个时间步 $t$ 的更新状态,并使用 $X$ 对 $h^0$ 进行初始化。此外,本文使用预训练的GloVe词向量<sup>[24]</sup>构建单词特征矩阵 $X$ 。 $A$ 为 $G$ 的邻接矩阵,当存在边 $(v_i, v_j) \in E$ 时, $A_{ij}=1$ ;否则 $A_{ij}=0$ 。 $D$ 为图 $G$ 的对角化度矩阵,其中 $D_{ii}=\sum_j A_{ij}$ ,且 $\hat{A}=D^{-1/2}AD^{-1/2}$ 为对称归一化邻接矩阵。在文本图的构建过程中,本文针对每个输入的目标文本构建独立的图结构,文本中的每个独立的单词被

表示成节点,单词节点之间的边通过使用固定尺寸的滑动窗口捕捉不同单词之间的共现关系来建立。滑动窗口的尺寸默认设置为3,图中的边统一为无向边。在文本图G构建完成后,将根据图1所示的流程对其进行处理。

## 2.2 注意力门控图神经网络

现有的文本分类模型通常会忽略单词之间的长距离依赖关系,尤其是离散的非连续单词之间的语义关系。Kyunghyun等<sup>[25]</sup>提出了门控循环单元(Gated Recurrent Unit, GRU),以门控的形式捕捉句子中单词之间的依赖关系。受其思想的启发,为使文本图中的每个单词节点能够在充分考虑与其他单词之间依赖关系的基础上更新自身特征,本文提出注意力门控图神经网络AGGNN来捕捉文本图中单词节点之间的语义关系,以达到控制每个节点特征信息传递的目的。AGGNN通过门控机制对不同节点之间的语义信息进行聚合和更新,以获得有效的文本嵌入表示,其主要由3个门控组成:注意力门、更新门和重置门。其网络结构如图2所示,图2中两曲线交汇的地方表示矩阵的拼接操作,两曲线分叉的地方表示矩阵的复制操作。

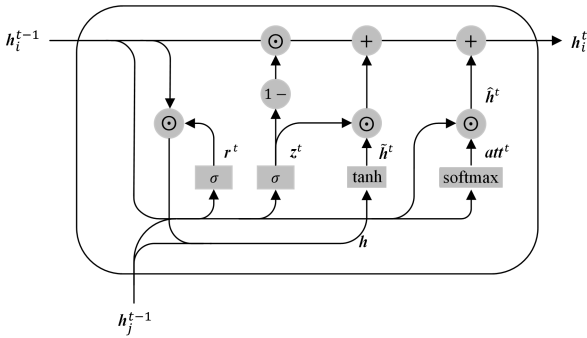


图2 AGGNN网络结构图

Fig.2 Network structure of AGGNN

注意力门由注意力机制自动学习单词节点之间语义信息的流动,在保留节点间重要语义特征信息并剔除冗余信息的同时,提高了节点特征更新过程的合理性及完整性。具体地,注意力门主要用于捕捉每个节点与其一阶邻居节点之间的语义关联程度,其中注意力分数决定着每个邻居节点中可以保留下来并整合到当前节点中的特征信息量。对于图中的每个节点,其注意力分数的计算式如下:

$$att_i^t = \text{softmax}(h_i^{t-1}) = \frac{\exp(h_i^{t-1}W)}{\sum_{j \in N} \exp(h_j^{t-1}W)} \quad (1)$$

其中,  $W \in \mathbb{R}^{d \times 1}$  表示线性转换矩阵,  $N$  表示当前节点及其一阶邻居节点的集合, softmax 函数用于对注意力向量进行归一化处理。

针对每个节点,根据式(1)计算出其邻居节点的注意力分数,并将其特征信息加权聚合。对于每个节点  $v_i$ ,其特征向量通过注意力加权之后为:

$$\hat{h}_i^t = \sum_{j \in N} att_j^t h_j^{t-1} \quad (2)$$

其中,  $N$  表示当前节点  $z^t$  及其一阶邻居节点的集合。

在引入注意力门后,AGGNN的完整更新公式如下:

$$a^t = \hat{A}^{t-1} h^{t-1} W_a \quad (3)$$

$$z^t = \sigma(W_z a^t + U_z h^{t-1} + b_z) \quad (4)$$

$$r^t = \sigma(W_r a^t + U_r h^{t-1} + b_r) \quad (5)$$

$$\tilde{h}^t = \tanh(W_h a^t + U_h (r^t \odot h^{t-1}) + b_h) \quad (6)$$

$$h^t = \tilde{h}^t \odot z^t + h^{t-1} \odot (1 - z^t) + \hat{h}_i^t \quad (7)$$

其中,  $z^t$  代表更新门,  $r^t$  代表重置门,  $W$  和  $U$  是可以训练的权重矩阵,  $b$  为偏置向量,  $\sigma$  表示 sigmoid 激活函数,  $\odot$  表示点积运算。

文本图的隐藏状态  $h^{t-1}$  在经过 AGGNN 后更新为  $h^t$ 。在该过程中,每个节点除了使用更新门和重置门聚合周围邻居节点的特征信息外,还通过注意力门加强与当前节点关联的语义信息,提高了后续文本分类的准确率。

## 2.3 基于注意力的文本池化模块

AGGNN 使得每个单词节点能够在信息传递和聚合的过程中不断更新自身的特征表示。当每个节点的特征表示更新完成后,本文希望利用这些节点的特征提取出当前层文本级别的特征表示,这类似于 CNN 中的特征图提取操作。CNN 中因为存在规则的空间局部性概念,所以可以使用平均池化或最大池化来扩大感受野,从而完成特征图的提取<sup>[26-27]</sup>。然而,文本图因不满足空间局部性条件,导致直接照搬 CNN 中的池化操作难以达到理想的效果。

为了将 CNN 中的池化操作泛化到文本图中,本文提出了基于注意力的文本池化模块 TextPool。TextPool 使用注意力机制重点提取对当前文本具有判别性语义特征的关键节点。事实上,每个文本的标签由其中的关键单词决定,通过提取具有判别性语义特征的节点,便能为目标文档分配期望的标签。TextPool 完整的处理过程包含 3 个步骤:1)通过注意力函数  $F(\cdot)$  计算每个节点的注意力分数;2)将每个节点的注意力分数按从高到低的顺序排序,然后选择出前  $k$  个分数最高的节点,即语义信息最为重要的关键节点;3)将邻居节点的特征信息聚合到被选择出的  $k$  个节点上。由此,TextPool 能够提取出当前图中的关键节点,从而达到特征图提取的目的,且其不受文本图尺寸及结构的限制。

注意力函数  $F(\cdot)$  会为每个单词节点分配一个正值的注意力分数,该函数包括两个输入部分:当前层的文本隐藏状态  $h^t$  和邻接矩阵  $\hat{A}^t$ 。为确定基于当前层节点总数  $n^t$  需要提取出的关键节点数目  $k$ ,引入池化比例  $\alpha$ ,其中  $\alpha \in (0, 1], k = \lceil \alpha n^t \rceil$ 。该过程的计算式如下:

$$S^t = F(h^t) \quad (8)$$

$$idx^t = \text{top-rank}(S^t, k) \quad (9)$$

其中,  $S^t \in \mathbb{R}^{n^t \times 1}$  是计算得出的注意力向量,  $\text{top-rank}(\cdot)$  为排序函数,可返回前  $k$  个节点的索引。

最终,当前文本图的特征将根据计算得出的注意力分数来进行进一步的更新。

$$h^{t'} = S^t \odot h^t, h^{t'} \in \mathbb{R}^{n^t \times d} \quad (10)$$

基于提取出的关键节点,新子图将被重新构建,其邻接矩阵通过关键节点的索引进行提取,即  $A_i^{t'} = \hat{A}_i^t, i \in idx^t$ 。新子图的隐藏状态的计算式为:

$$h^{t+1} = A^{t'} h^{t'}, h^{t+1} \in \mathbb{R}^{k \times d} \quad (11)$$

回顾整个节点的提取过程不难发现,新子图中节点之间的距离在原始文本图中表示关键节点所处的聚簇与聚簇之间

的距离。基于此,新子图的邻接矩阵更新公式为:

$$\hat{\mathbf{A}}^{t+1} = \mathbf{A}^{t'} \hat{\mathbf{A}}^t (\mathbf{A}^t)^\top, \hat{\mathbf{A}}^{t+1} \in \mathbb{R}^{k \times k} \quad (12)$$

最终通过上述的文本池化操作便可得到新的子图,新子

图可通过隐藏状态  $\mathbf{h}^{t+1}$  及邻接矩阵  $\hat{\mathbf{A}}^{t+1}$  表示。

为了将 TextPool 泛化到多种不同尺度的文本数据集中以提高其鲁棒性,本文提出了两种不同的注意力实现机制,分别是全局注意力机制和局部注意力机制。

### 2.3.1 全局注意力机制

全局注意力机制主要考虑当前文本图中每一个单词节点对最终文本标签的贡献程度,即在当前文本图中直接计算每个节点的注意力分数。全局注意力分数的计算式如下:

$$F_i(\mathbf{h}^i) = \text{softmax}_i(\mathbf{h}^i \mathbf{W}) = \frac{\exp(\mathbf{h}^i \mathbf{W})}{\sum_{j \in n} \exp(\mathbf{h}^j \mathbf{W})} \quad (13)$$

其中,  $\mathbf{W}$  为线性转换矩阵。

### 2.3.2 局部注意力机制

对于尺度较小的文本图数据,全局注意力机制能够快速有效地提取出当前文本图中具有判别性语义信息的关键节点,从而加速模型的训练过程。然而,对于大尺度的文本图数据,全局注意力机制通常得不到最好的性能表现。一方面,复杂的图结构往往比简单图包含更多的节点数目以及更复杂的边结构信息,使用全局注意力机制会消耗大量的计算资源,另一方面,节点的信息经过传递和聚合得到了充分的更新,导致每个节点与其一阶邻居节点的特征表示逐渐变得相似,这一现象被称为拉普拉斯平滑(Laplacian smoothing)<sup>[28]</sup>。在这种情况下,每个节点的注意力分数会与其一阶邻居节点非常接近,导致注意力被限定在某一个特定的狭窄区域内,造成信息冗余。为缓解该问题,本文选择在大尺度文本图中针对每个节点及其所处的一阶邻域内计算中心节点的注意力分数,即局部注意力分数。对于每个节点  $v_i$ ,其局部注意力分数的计算式如下:

$$F_i(\mathbf{h}^i) = \frac{\hat{\mathbf{A}}_{ii} \exp(\mathbf{h}^i \mathbf{W} \cdot \theta)}{\sum_{j \in \mathcal{A}} (\hat{\mathbf{A}}_{ij} \exp(\mathbf{h}^j \mathbf{W} \cdot \theta))^{j \in \mathcal{A}}} \quad (14)$$

其中,  $\mathbf{W}$  为线性转换矩阵,  $N$  表示中心节点  $v_i$  及其一阶邻居节点  $v_j$  的集合,  $\theta$  为可学习的起调节作用的参数。

事实上,式(14)是一个加权的 softmax 函数,每个节点的注意力分数由节点  $v_i$  和节点  $v_j$  之间的距离产生。同时,在注意力分数的计算过程中,本文使用度矩阵参与乘积运算以减少由邻居节点的数目不同而产生的负面影响。

## 2.4 图读出模块

当提取出关键节点并建立新的子图后,图读出模块 Readout 需要根据每个节点的特征表示构建文本图的嵌入表示  $\mathbf{h}_G$ 。在该过程中,TextPool 模块提取出的每个节点将贡献出自身的部分特征,同时最具代表性的节点应为文档表示做出更多贡献。因此,Readout 使用均值函数及最大化函数来整合每个关键节点的特征表示,并设计出图读出计算式,具体如下:

$$\mathbf{h}_G = \frac{1}{k} \sum_k \mathbf{h}_v^{t+1} \parallel \max_k \mathbf{h}_v^{t+1} \quad (15)$$

其中,  $\parallel$  表示拼接操作。

在获得文本图嵌入表示  $\mathbf{h}_G$  后,本文使用如下文本分类器来实现目标文本的分类。

$$\hat{\mathbf{y}} = \text{softmax}(\text{ReLU}(\mathbf{W}\mathbf{h}_G + \mathbf{b})) \quad (16)$$

其中,  $\mathbf{W}$  为线性转换矩阵,它将文本图嵌入表示映射到输出空间中,  $\mathbf{b}$  为偏置向量。

经过文本分类器后便生成了目标文本的预测标签  $\hat{\mathbf{y}}$ ,在此基础上利用交叉熵损失函数(cross-entropy loss)<sup>[29]</sup>优化真实标签  $\mathbf{y}$  和预测标签  $\hat{\mathbf{y}}$  之间的距离。

$$\mathbf{L} = - \sum_i \mathbf{y}_i \log(\hat{\mathbf{y}}_i) \quad (17)$$

其中,  $\mathbf{y}_i$  为真值标签独热编码中的第  $i$  个元素。

## 3 实验

### 3.1 数据集

为验证本文模型的有效性,本文在 4 个基准文本数据集上进行实验验证,分别为 Ohsumed, R8, R52 和 MR 数据集。数据集的相关统计信息如表 1 所列。

表 1 数据集统计  
Table 1 Dataset statistics

Datasets	# Docs	# Training	# Test	# words	# Nodes	# Classes	Average Length
Ohsumed	7400	3357	4043	14157	21557	23	135.82
R8	7674	5485	2189	7688	15362	8	65.72
R52	9100	6532	2568	8892	17992	52	69.82
MR	10662	7108	3554	18764	29426	2	20.39

Ohsumed 数据集<sup>[30]</sup>包含来自医药数据库 MEDLINE10 的 270 个医药类杂志的摘要文本和标题。本文选择其中的 23 种心血管疾病医学摘要的文档集合作为文本分类数据集,并筛选出 7400 个具有单一文本标签的文档,其中训练集包含 3357 个文档,测试集包含 4043 个文档。

R8 和 R52 数据集都是 Reuters-21578 数据集<sup>[31-32]</sup>的子集。Reuters-21578 数据集是一个多类、多标签的数据集,总共包含 90 个类别、12902 个文档,文本选择其子集 R8 和 R52 分别进行分类性能的验证。R8 数据集包含 8 种类别标签,由 5485 个文档构成训练集,由 2189 个文档构成测试集;R52 数据集包含 52 种类别标签,由 6532 个文档构成训练集,由 2568 个文档构成测试集。

MR(Movie Review)数据集<sup>[33]</sup>是一个整合了电影评论的文档集合,每个文档与一个积极或消极的情感标签关联。整个数据集包含 5331 个积极的评论文档和 5331 个消极的评论文档。本文按照文献<sup>[34]</sup>中的方法将 MR 数据集分为训练集和测试集。

在模型训练之前需要先对以上数据集进行预处理,本文使用文献<sup>[14]</sup>中的预处理方法处理文本数据,这些方法包括分词(tokenization)<sup>[35]</sup>和去除停用词(stopword removal)<sup>[36]</sup>等。因 MR 数据集的文本长度较短,所以本文不对其进行去除停用词处理。

### 3.2 对比模型

实验中使用的对比模型主要有 9 个,分别是 TF-IDF + LR, TextCNN, LSTM, Bi-LSTM, fastText, SWEM, GraphCNN, TextGCN 及 Textlevel-GNN。TF-IDF + LR 表示传统

的词袋模型,它使用 TF-IDF 方法计算每个单词的权重,并使用对数几率回归对文本分类。TextCNN<sup>[14]</sup>为基于 CNN 的文本分类模型,通过一维卷积捕捉句子中的词序信息。LSTM<sup>[15]</sup>为基于预训练和信息共享机制的 LSTM 文本分类模型。Bi-LSTM 为基于预训练模型的双向 LSTM 模型。fastText<sup>[37]</sup>为快速文本分类器,通过 N-gram 向量对文本的类别进行预测。SWEM<sup>[38]</sup>全称为简单词嵌入模型,其直接在词向量上使用简单的池化操作完成文本分类。Graph-CNN<sup>[39]</sup>为 Defferrard 等于 2016 年提出的 GCN 模型,其将卷积操作泛化到图谱域上,并利用切比雪夫多项式对卷积公式进行简化。TextGCN<sup>[18]</sup>为 Yao 等于 2019 年提出的图网络模型,它为整个文本语料库构建异构图,将数据集中的文档和单词作为节点,使用点间互信息及 TF-IDF 方法分别计算单词节点之间以及单词节点和文档节点之间的边权重。Textlevel-GNN<sup>[20]</sup>为 Huang 等于 2019 年提出的图网络模型,其直接为每个输入文本单独建立文本图,图中仅包含单词节点。

### 3.3 实验设置

在实验过程中,本文模型的具体结构如图 3 所示。对于构建好的文本图,默认堆叠两层 AGGNN 更新节点的特征信息。当面对不同尺度的数据集时,AGGNN 的堆叠层数应根据数据集的内在属性进行相应调整。在节点的提取过程中,本文使用一层 TextPool 模块提取当前文本图中的关键节点并构建新子图。至于文本分类器,则搭建两层全连接层(Fully Connected Layer, FC),并使用线性整流函数(Rectified Linear Unit, ReLU)作为激活函数,最终通过 softmax 分类器预测文本标签。

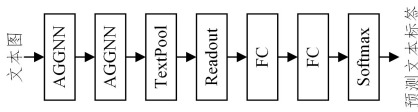


图 3 模型结构图

Fig. 3 Model architecture

对于涉及的 4 个数据集,训练集和测试集的划分比例依据 3.1 节中的方式进行划分,实验中所有模型统一将训练集中 90% 的文本数据随机筛选出来用于模型训练,剩下 10% 的数据用于模型验证。在模型参数的优化上,本文使用 Adam<sup>[40]</sup>优化器,并设置初始学习率为 0.01,  $L_2$  权重衰减率为  $10^{-4}$ 。在模型的超参数初始化过程中,使用 Xavier<sup>[41]</sup>方法进行模型初始化。训练过程采用 Dropout<sup>[42]</sup>技术防止模型过拟合,并设置 Dropout 比率为 0.5。为捕捉单词与单词之间的语义共现关系,本文通过对参数进行实验验证,最终将滑动窗口的长度设置为 3,并将其作为默认设置。TextPool 模块中的节点池化比例  $\alpha$  为 0.5,即文本图数据每经过一次 TextPool 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,将本文提出的模型训练 400 个批次,使模型达到最优性能。对于实验部分的其他对比模型,将依据其原有文献的设置进行实施。此外,对于所有使用预训练词向量的文本分类模型,统一使用维度为 300 的 GloVe 词嵌入向量构建单词的特征表示。

### 3.4 实验结果

为了验证本文模型在文本分类任务中的有效性,本文从多个角度对其进行了综合实验分析,包括对比实验、消融实验及模型参数验证 3 个部分。

#### 3.4.1 对比实验

为减小各模型的对比误差,本文将每个模型分别在相应的数据集上运行 10 次,然后采用均值加减标准差的形式对实验结果进行展现。各模型在 Ohsumed, R8, R52 和 MR 数据集上的分类准确率如表 2 所列(对比模型的实验结果与其原有文献一致),其中粗体部分表示各数据集上取得的最优分类结果。表 2 中,Graph-CNN, TextGCN, Textlevel-GNN 和本文方法均为图网络模型,其余为其他类型的分类模型。

表 2 不同方法的分类准确率对比

Table 2 Classification accuracy comparison of different methods

(单位: %)

Models	Ohsumed	R8	R52	MR
TF-IDF+LR	54.66±0.00	93.74±0.00	86.95±0.00	74.59±0.00
TextCNN <sup>[12]</sup>	58.44±1.06	95.71±0.52	87.59±0.48	77.75±0.72
LSTM <sup>[13]</sup>	51.10±1.50	96.09±0.19	90.48±0.86	77.73±0.89
Bi-LSTM	49.27±1.07	96.31±0.33	90.54±0.91	77.68±0.86
fastText <sup>[33]</sup>	57.70±0.49	96.13±0.21	92.81±0.09	75.14±0.20
SWEM <sup>[34]</sup>	63.12±0.55	95.32±0.26	92.94±0.24	76.65±0.63
Graph-CNN <sup>[35]</sup>	63.86±0.53	96.99±0.12	92.75±0.22	77.22±0.27
TextGCN <sup>[14]</sup>	68.36±0.56	97.07±0.10	93.56±0.18	76.74±0.20
Textlevel-GNN <sup>[16]</sup>	69.40±0.60	97.80±0.20	94.60±0.30	78.86±0.34
Ours	<b>70.83±0.38</b>	<b>98.18±0.10</b>	<b>94.72±0.29</b>	<b>80.03±0.22</b>

由表 2 可以看出,基于图网络的模型明显比其他类型的模型在性能表现上更加出色,这是因为文档中所涉及的数据都是语义间具有一定关联关系的单词,基于神经网络的方法,如 CNN, LSTM 等,虽然能够捕捉到短距离或序列中的连续单词之间的关系,但是对于相距较远的单词,特别是非连续单词之间,往往不能进行有效的信息交互。而在图结构中,无论在原文档中相聚多远的两个单词,都能够通过捕捉单词之间的共现信息来为其建立语义交互结构,并以此实现不同节点之间语义信息的更新,这正是其他类型的文本分类模型不具有的优势。

与其他基于图网络模型的实验结果进行对比后发现,本文方法在分类效果上具有显著的性能提升。Graph-CNN 在实验中使用大尺度的窗口搭建词袋模型以表示文本特征,这使得它们的方法难以有效地区分不同单词节点之间的特征信息。TextGCN 为整个语料库构建异构图,使图结构在展现全局关系时既囊括了节点与节点之间的关系,又包含了节点与文档之间的关系,多种关系杂糅削弱了边的信息表征能力,同时在边上标签信息的传递过程中又造成了信息冗余,使得模型无法发挥出应有的表现效果。Textlevel-GNN 在 TextGCN 的基础上充分吸取教训,在构建文本图时不再为整个语料库搭建图结构,而是对每一个文档都单独建立文本图,通过捕捉单词之间的共现信息建立节点之间的边关系。Textlevel-GNN 与本文方法中构建文本图的思路一致,而实验结果也表明,这种思路能够剔除文档与语料库之间的冗余信息,保证边只在单词节点之间传递有效的语义信息比在整个语料库上构造图更为有效。然而,与本文方法进行对比发现,Textlevel-

GNN 使用基于语料库的全局共享的边权重优化模型,导致模型需要额外存储一个权重矩阵,而该矩阵的尺寸是由图中节点的数目决定的,为  $|V| \times |V|$ 。因此,综合来看,与其他现有的方法相比,本文方法具有明显优势且效果更好。

### 3.4.2 消融实验

为了验证模型结构的合理性,本文对提出的模型在 Ohsumed, R8, R52 和 MR 数据集上进行了消融实验,包括 AGGNN 模块中的注意力门、TextPool 模块中针对不同数据集的注意力机制实现方式以及 Readout 函数。表 3 列出了本文模型在各文本数据集上的消融实验结果,其中“TextPool-G”和“TextPool-L”分别表示 TextPool 模块的实现方式为全局注意力机制及局部注意力机制,而“w/o A”和“w/ A”分别表示 AGGNN 模块不使用注意力门以及使用注意力门。表 3 中粗体部分表示的实验数据代表最优分类准确率。

表 3 消融实验结果

Table 3 Results of ablation experiment

(单位:%)

Models	Ohsumed	R8	R52	MR
TextPool-G w/o A	67.89	96.72	92.09	77.84
TextPool-G w/ A	69.05	97.82	93.16	<b>80.03</b>
TextPool-L w/o A	68.60	96.37	92.69	76.51
TextPool-L w/ A	<b>70.83</b>	<b>98.18</b>	<b>94.70</b>	78.65

#### (1) AGGNN 模块中的注意力门

通过观察表 3 中的数据可以发现,不论 TextPool 模块的实现方式是全局注意力机制还是局部注意力机制,在 AGGNN 模块中使用注意力门的模型比没有使用注意力门的模型的表现效果更好。其原因是,在 AGGNN 模块中,注意力门能够弥补更新门和重置门对语义信息学习能力不足的缺陷,当在邻居节点与中心节点之间进行语义信息传递时,注意力门能够对单词节点的特征信息进行充分的更新,为后续建立图的嵌入表示做出良好的铺垫。综合表中的结果可以得知,AGGNN 模块中的注意力门是必不可少的。

#### (2) TextPool 模块中的注意力机制

如表 3 所列,本文模型针对不同尺度的数据集在 TextPool 模块中使用了不同的注意力实现机制。通过横向和纵向的实验数据对比可以发现,在较小尺度的数据集上,如 MR 数据集,通常使用全局注意力机制便可以得到理想的实验结果。这是因为小尺度的数据集本身包含的单词数量较少,所以在对应的文本图中节点的数目也相对较少,图结构相对简单且没有太多复杂的边关系,使用全局注意力机制既简便又能够使模型快速收敛,得出的实验结果较为出色。而对于尺度比较大的数据集,如 Ohsumed, R8 和 R52,其文本图结构较为复杂,全局注意力机制容易将注意力限定在某一狭窄的局部区域中,导致模型只关注于一种语义的单词节点,造成了信息的冗余而且容易忽略其他单词节点,因此在大尺度文本数据集上的分类结果差强人意。为了使模型能够适用于不同尺度的数据集,本文针对注意力机制设计了不同的实现方式,使得模型在多种情况下都能取得不错的分类结果。

#### (3) Readout 函数

为验证该模型中 Readout 模块的有效性,本文对不同

形式的 Readout 函数进行了对比分析,包括基于均值函数的 Readout(Mean-R)、基于最大值函数的 Readout(Max-R)以及基于均值与最大值相结合的 Readout(Ours)。其实验结果如表 4 所列。

表 4 采取不同 Readout 函数的实验结果

Table 4 Experimental results with different Readout functions

Readout	Ohsumed	R8	R52	MR
Mean-R	68.62	97.75	93.54	78.31
Max-R	54.51	95.37	91.86	75.42
Ours	70.83	98.18	94.72	<b>80.03</b>

通过实验结果可以发现,仅仅提取节点特征的最大值并不能使模型获得理想的分类效果。基于均值函数的 Readout 能够平均图中各关键节点的特征,获得了相对有效的分类结果,但并不是最优结果。而基于均值函数和最大值函数相结合的 Readout 则在 Mean-R 的基础上强化了对文本标签贡献最为突出的节点特征,进一步提升了模型分类准确率。因此,本文选择基于均值和最大值函数相结合的方式,整合得到最终的文本图嵌入表示,以获得最优分类效果。

#### 3.4.3 参数敏感性

为进一步检验模型的鲁棒性,本文对模型中涉及的相关参数进行了讨论,主要包括 3 个方面:AGGNN 堆叠的层数、模型中使用的滑动窗口尺寸及 TextPool 模块的池化比例  $\alpha$ 。

##### (1) AGGNN 模块的堆叠层数

依照实践经验,传统的神经网络通过不断加深网络的层数可以使模型学习到丰富的特征信息,使其在分类或者预测任务上表现出优异的性能。而这一搭建深层网络的思路在图网络模型中似乎不可行,如果将图网络模型的层数堆叠得过深,则会使节点之间的特征表示逐渐变得平滑,即拉普拉斯平滑,最终导致模型过拟合。因此,为了在不同的数据集上获得最优的分类性能并防止模型过拟合,本文通过堆叠不同层数的 AGGNN 对模型的性能进行讨论。关于 MR 和 Ohsumed 数据集的实验结果如图 4 所示。

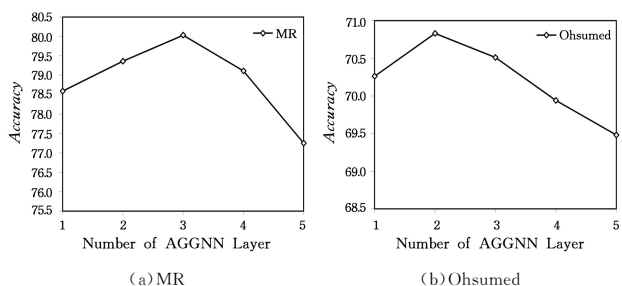


图 4 在 MR 和 Ohsumed 数据集上堆叠不同层数 AGGNN 的实验结果

Fig. 4 Experimental results of different layers of AGGNN on datasets MR and Ohsumed

通过堆叠不同层数的 AGGNN,可以使节点在特征信息传递和聚合的过程中不只局限于一阶邻域,而是能扩散到多阶邻域。例如堆叠两层 AGGNN 模块,每个节点的信息传播范围便能扩散到二阶邻域。由图 4 可以看出,当模型在 MR 数据集上不断堆叠 AGGNN 模块时,模型分类性能不断提升。但是,当 AGGNN 模块的堆叠层数超过 3 层以后,模型

性能便开始不断下降。在 Ohsumed 数据集上的折线情况变化类似,不同的是模型性能在 AGGNN 模块堆叠了两层以后开始下降。

综合分析以上情况发现,AGGNN 模块与文献[43]阐述的理论一致,当 AGGNN 堆叠一层时,能够在中心节点和一阶邻居节点之间进行信息的传递和更新;当堆叠两层时,便能够在二阶邻居节点之间间接地进行信息的传递和更新;而随着图网络层数的不断叠加,节点的特征信息判别性变差,模型性能也逐渐变差。因此,在实验中应根据不同文本数据集的固有特征对模型的结构进行相应调整,从而使模型具有更好的适用性及鲁棒性。

### (2) 滑动窗口的尺寸

为验证滑动窗口的尺寸在构建文本图结构时是否会对最终的模型分类性能造成影响,本文对其进行了相应的调整和讨论。本文分别在 MR 和 Ohsumed 数据集上对滑动窗口的尺寸进行了参数调整,其实验结果如图 5 所示。

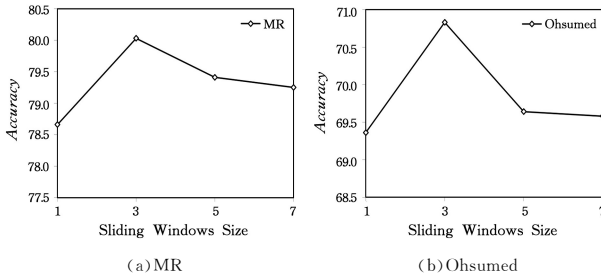


图 5 在 MR 和 Ohsumed 数据集上调整不同滑动窗口尺寸的实验结果

Fig. 5 Experimental results of adjusting the sizes of sliding windows on datasets MR and Ohsumed

可以看出,针对 MR 和 Ohsumed 两个数据集,当滑动窗口的尺寸为 3 时,模型可以拥有最好的分类性能。当滑动窗口的尺寸小于 3 时,构建的文本图无法有效捕捉单词间的语义共现关系,模型表现不佳;而当尺寸大于 3 时,文本图中会包含太多的边,不仅造成了信息冗余,还会削弱边的表达能力,降低模型分类准确率。通过综合对比分析,本文在实验中统一使用尺度为 3 的滑动窗口捕捉单词之间的语义共现关系。

### (3) TextPool 模块的池化比例

TextPool 模块相当于传统 CNN 模型中的池化层,CNN 使用固定的尺寸以及步长来控制生成特征图的大小,TextPool 则通过调整池化比例  $\alpha$  来达到相应的目的。为探究模型的最优结构,本文在 MR 和 Ohsumed 数据集上通过调整参数对其进行讨论,结果如图 6 和图 7 所示。

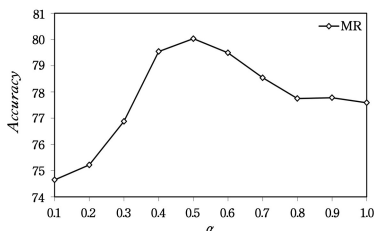


图 6 在 MR 数据集上调整不同比例  $\alpha$  的实验结果

Fig. 6 Experimental results of different  $\alpha$  on dataset MR

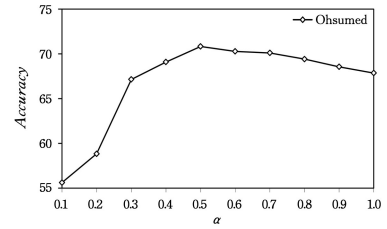


图 7 在 Ohsumed 数据集上调整不同比例  $\alpha$  的实验结果

Fig. 7 Experimental results of different  $\alpha$  on dataset Ohsumed

从图中 MR 和 Ohsumed 数据集的实验结果来看,当  $\alpha$  设置为较小的比例时(如 0.1, 0.2),模型因在经过 TextPool 模块后筛选掉了过多的关键节点,导致无法有效地构建图的嵌入表示,影响了模型的性能。随着  $\alpha$  数值的增大,TextPool 模块能够筛选出最关键的节点,并剔除掉多余的节点,因此模型能够获得有效的分类准确率。当  $\alpha$  数值逐渐接近于 1 时,TextPool 模块起到的作用逐渐变弱,导致与文本标签关联性不高的节点也被用于构建图的嵌入表示,因此在最终对文本进行分类时,模型的性能有所下降。综合图 6、图 7 中的数据,本文在实验时统一设置  $\alpha$  的数值为 0.5,以使模型能够表现出最佳的分类性能。

**结束语** 现有的文本分类模型通常无法有效地捕捉单词节点之间的语义信息,本文针对该问题提出了一种新的基于注意力门控图神经网络的文本分类模型。该模型在训练阶段为每个输入文本构建独立的文本图,其中 AGGNN 模块能够对每个节点和其所在的一阶邻域进行节点特征的加权聚合,使节点的语义特征得到充分更新;TextPool 模块利用全局注意力机制或局部注意力机制提取当前文本图中具有判别性语义特征的关键节点;Readout 模块融合关键节点的特征得到最终的图嵌入表示,以完成文本分类。相比现有的模型,本文模型在多个文本数据集上的准确率更高,且其网络结构更为合理。

本文模型还有改进的空间,未来将继续对模型中的注意力机制进行探索,包括在保持模型分类准确率的基础上简化 AGGNN 模块以减少整个模型的参数量等工作。

## 参考文献

- [1] CHARU C A, ZHAI C X. A survey of text classification algorithms [M] // Mining Text Data. Berlin: Springer, 2012: 163-222.
- [2] YAN M. Research on spam filtering algorithm based on fast-Text [D]. Guangzhou: South China University of Technology, 2020.
- [3] GRIFFITHS T L, STEYVERS M. Finding scientific topics [J]. Proceedings of the National Academy of Sciences, 2004, 101(suppl. 1): 5228-5235.
- [4] XIA R, ZONG C Q, LI S S. Ensemble of feature sets and classification algorithms for sentiment classification [J]. Information Sciences, 2011, 181(6): 1138-1152.
- [5] ZHOU W X, LAN W F. Summarization Model Using Multi-Task Learning Fused with Text Classification [J]. Computer Engineering, 2021, 47(4): 48-55.

- [6] SWINIARSKI R W, SKOWRON A. Rough set methods in feature selection and recognition [J]. *Pattern Recognition Letters*, 2003, 24(6): 833-849.
- [7] WONG P C, WHITNEY P, THOMAS J. Visualizing association rules for text mining [C]// *Proceedings of the IEEE Symposium on Information Visualization*. Piscataway, NJ: IEEE Press, 1999: 120-123.
- [8] PENG F C, SCHUURMANS D. Combining naive Bayes and n-gram language models for text classification [C]// *Proceedings of the European Conference on Information Retrieval*. Berlin: Springer, 2003: 335-350.
- [9] JOACHIMS T. Text categorization with support vector machines: learning with many relevant features [C]// *Proceedings of the European Conference on Machine Learning*. Berlin: Springer, 1998: 137-142.
- [10] KANG H S, NAM K, KIM S. The decomposed k-nearest neighbor algorithm for imbalanced text classification [C]// *Proceedings of the International Conference on Future Generation Information Technology*. Berlin: Springer, 2012: 87-94.
- [11] LI C B, DUAN Q J, JI C H, et al. Method of Short Text Classification Based on CHI and TF-IWF Feature Selection [J]. *Journal of Chongqing University of Technology (Nature Science)*, 2021, 35(5): 135-140.
- [12] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
- [13] MIKOLOV T, KARAFIAT M, BURGET L, et al. Recurrent neural network based language model [C]// *Proceedings of the 11th Annual Conference of the International Speech Communication Association*. 2010: 1045-1048.
- [14] KIM Y. Convolutional neural networks for sentence classification [C]// *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2014: 1746-1751.
- [15] 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, CA: AAAI Press, 2016: 2873-2879.
- [16] WANG Y, HE Y M, CHEN H X, et al. RHS-CNN: A CNN Text Classification Model Based on Regularized Hierarchical Softmax [J]. *Journal of Chongqing University of Technology (Nature Science)*, 2020, 34(5): 187-195.
- [17] MA Z K, DILIYAER P, ZAO KERE K, et al. A Classification Algorithm for Tourist Question Texts Integrated with Deep Learning Models [J]. *Computer Engineering*, 2020, 46(11): 70-76.
- [18] YAO L, MAO C S, LUO Y. Graph convolutional networks for text classification [C]// *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2019: 7370-7377.
- [19] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [C]// *Proceedings of the 5th International Conference on Learning Representations*. 2017.
- [20] HUANG L Z, MA D, LI S J, et al. Text level graph neural network for text classification [C]// *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA: Association for Computational Linguistics, 2019: 3444-3450.
- [21] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks [C]// *Proceedings of the 6th International Conference on Learning Representations*. 2018.
- [22] 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, MA: MIT Press, 2017: 6000-6010.
- [23] TANG D Y, 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.
- [24] PENNINGTON J, SOCHER R, MANNING C D. Glove: global vectors for word representation [C]// *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2014: 1532-1543.
- [25] CHOK, VAN MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation [C]// *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 2014: 1724-1734.
- [26] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv: 1409. 1556*, 2014.
- [27] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE Press, 2015: 1-9.
- [28] LI Q M, HAN Z C, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press, 2018: 3538-3545.
- [29] DE BOER P T, KROESE D P, MANNOR S, et al. A tutorial on the cross-entropy method [J]. *Annals of Operations Research*, 2005, 134(1): 19-67.
- [30] HERSH W, BUCKLEY C, LEONE T J, et al. Ohsumed: an interactive retrieval evaluation and new large test collection for research [C]// *Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR '94)*. London: Springer, 1994: 192-201.
- [31] APTE C, DAMERAU F, WEISS S M. Automated learning of decision rules for text categorization [J]. *ACM Transactions on Information Systems*, 1994, 12(3): 233-251.
- [32] APTE C, DAMERAU F, WEISS S M. Towards language independent automated learning of text categorization models [C]// *Proceedings of the International Conference on Research and Development in Information Retrieval (SIGIR '94)*. London: Springer, 1994: 23-30.

- [33] PANG B, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques [C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2002:79-86.
- [34] TANG J, QU M, MEI Q Z. Pte: predictive text embedding through large-scale heterogeneous text networks [C]// Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2015:1165-1174.
- [35] ROUSSEAU F, KIAGIASE, VAZIRGIANNIS M. Text categorization as a graph classification problem [C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics. 2015:1702-1712.
- [36] BLANCO R, LIOMA C. Graph-based term weighting for information retrieval [J]. Information Retrieval, 2012, 15(1): 54-92.
- [37] JOULIN A, GRAVE E, BOJANOWSKI P, et al. Bag of tricks for efficient text classification [J]. arXiv:1607.01759, 2016.
- [38] SHEN D H, WANG G Y, WANG W L, et al. Baseline needs more love: on simple word-embedding-based models and associated pooling mechanisms [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 440-450.
- [39] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering [C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2016:3844-3852.
- [40] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv:1412.6980, 2014.
- [41] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks [C]// Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. 2010:249-256.
- [42] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. arXiv:1207.0580, 2012.
- [43] LEE J, LEE I, KANG J. Self-attention graph pooling [C]// Proceedings of the International Conference on Machine Learning. 2019:3734-3743.



**DENG Zhao-yang**, born in 1995, post-graduate. His main research interests include deep learning and graph neural network.



**WANG Dong**, born in 1979, Ph.D, senior engineer. His main research interests include machine vision, embedded system, software programming and IoT design.

(责任编辑:柯颖)