



计算机科学

COMPUTER SCIENCE

基于双图神经网络信息融合的文本分类方法

闫佳丹, 贾彩燕

引用本文

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 Jia-dan, JIA Cai-yan. [Text Classification Method Based on Information Fusion of Dual-graph Neural Network](#)[J]. Computer Science, 2022, 49(8): 230-236.

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)

[监督和半监督学习下的多标签分类综述](#)

Survey of Multi-label Classification Based on Supervised and Semi-supervised Learning

计算机科学, 2022, 49(8): 12-25. <https://doi.org/10.11896/jsjcx.210700111>

[基于多时间尺度时空图网络的交通流量预测模型](#)

Multi-time Scale Spatial-Temporal Graph Neural Network for Traffic Flow Prediction

计算机科学, 2022, 49(8): 40-48. <https://doi.org/10.11896/jsjcx.220100188>

[基于注意力机制的医学影像深度哈希检索算法](#)

Deep Hash Retrieval Algorithm for Medical Images Based on Attention Mechanism

计算机科学, 2022, 49(8): 113-119. <https://doi.org/10.11896/jsjcx.210700153>

[基于非局部注意力生成对抗网络的视频异常事件检测方法](#)

Non-local Attention Based Generative Adversarial Network for Video Abnormal Event Detection

计算机科学, 2022, 49(8): 172-177. <https://doi.org/10.11896/jsjcx.210600061>

[基于图卷积神经网络的文本分类方法研究综述](#)

Review of Text Classification Methods Based on Graph Convolutional Network

计算机科学, 2022, 49(8): 205-216. <https://doi.org/10.11896/jsjcx.210800064>

基于双图神经网络信息融合的文本分类方法

闫佳丹 贾彩燕

北京交通大学计算机与信息技术学院 北京 100044

交通数据分析与数据挖掘北京市重点实验室(北京交通大学) 北京 100044

(18120436@bjtu.edu.cn)

摘要 近年来,图神经网络在文本分类任务中得到了广泛应用。与图卷积网络相比,基于消息传递的文本级的图神经网络模型具有内存占用少和支持在线检测等优点。然而此类模型通常仅使用词共现信息为语料中的各个文本构建词汇图,导致获取到的信息缺少多样性。文中提出了一种基于双图神经网络信息融合的文本分类方法。该方法在保留原有词共现图的基础上,根据单词间的余弦相似度构建语义图,并通过阈值控制语义图的稀疏程度,更有效地利用了文本的多方位语义信息。此外,测试了直接融合和注意力机制融合两种方式对词汇图和语义图上学习到的文本表示融合的能力。实验使用 R8 和 R52 等 12 个文本分类领域常用的数据集来测试算法的精度,结果表明,与最新的 TextLevelGNN,TextING 和 MPAD 这 3 个文本级的图神经网络模型相比,双图模型能够有效提高文本分类的性能。

关键词: 文本分类;图神经网络;语义信息;信息融合;注意力机制;自然语言处理

中图法分类号 TP391

Text Classification Method Based on Information Fusion of Dual-graph Neural Network

YAN Jia-dan and JIA Cai-yan

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China

Abstract Graph neural networks are recently applied in text classification tasks. Compared with graph convolution network, the text level graph neural network model based on message passing(MP-GNN) has the advantages of low memory usage and supporting online testing. However,MP-GNN model only builds a lexical graph using the word co-occurrence information, and the obtained information lacks diversity. To address this problem,a text classification method based on information fusion of dual-graph neural network is proposed. Besides preserving the original lexical graph built in MP-GNN,this method constructs the semantic graph based on the cosine similarity between pairs of words,and controls the sparsity of the graph through a threshold, which makes more effective use of the multi-directional semantic information of the text. In addition,the ability of direct fusion and attention mechanism fusion are tested to fuse the text representation learned on lexical graph and semantic graph. Experimental results on 12 datasets(R8,R52 and other datasets commonly used for text classification) show that the proposed model demonstrates an obvious improvement on performance of text classification compared with the SOTA(state-of-the-art) methods TextLevelGNN,TextING and MPAD.

Keywords Text classification, Graph neural network, Semantic information, Information fusion, Attention mechanism, Natural language processing

1 引言

文本分类是自然语言处理中的基础任务,其目的是推断出给定文本的标签,在情感分析、新闻过滤、信息检索等方面有着广泛的应用^[1]。

文本表示学习是文本分类的关键步骤。现有的文本表示学习方法分为传统机器学习方法和深度学习方法。传统机器

学习方法通过特征工程学习文本表示,如词袋模型、n-grams 模型、TF-IDF 等。但此类方法依赖人工特征提取,因此效率较低。随着深度学习的发展,卷积神经网络(Convolutional Neural Networks,CNN)^[2]和递归神经网络(Recurrent Neural Networks,RNN)^[3]等序列学习模型可以自动地学习文本表示,被广泛应用于文本分类。序列学习模型优先考虑文本的局部性和时序性,在局部连续的词序列中能够有效地获得

到稿日期:2021-06-04 返修日期:2021-10-23

基金项目:中央高校基本科研业务费专项资金(2019JBZ110)

This work was supported by the Fundamental Research Funds for the Central Universities of Ministry of Education of China(2019JBZ110).

通信作者:贾彩燕(cyjia@bjtu.edu.cn)

语义信息和句法信息,但是无法获取文本的全局信息,因此缺乏获得非连续和长距离语义信息的能力。

近年来,图神经网络的提出成功解决了上述问题。与序列学习模型不同,图神经网络并非将文本看作一个序列,而是看作一组共现的单词,因此能够有效地处理复杂的结构化数据,并能在图节点表示中映射图的全局结构信息。文本分类任务使用两类图神经网络:图卷积网络(Graph Convolution Network,GCN)^[4]和基于消息传递(Message Passing,MP)^[5]的图神经网络。图卷积网络预先对整个语料库建图,引入文本节点将文本分类问题转化为节点分类问题,存在实际占用内存较大,且无法对新加入的文本进行分类的问题;而基于消息传递的图神经网络模型能够解决上述图卷积网络存在的问题。

基于消息传递的图神经网络能够对单个文本建图,通过邻居节点更新中心节点的词表示,从而生成文本表示对文本进行分类。在已有的研究工作中,此类模型仅使用词共现信息来构建词汇图,未考虑引入语义信息等其他类型的信息。采用不同的建图方式能够从不同角度获取文本的隐含信息,有助于提升文本分类的性能^[6]。

针对上述问题,本文提出了一种基于双图神经网络信息融合的文本分类方法,主要贡献包括以下3个方面:

(1)基于当前最新的3个基于消息传递的图神经网络模型,通过构建语义图来引入额外的语义信息,实现利用词汇图和语义图的双图更新机制。

(2)为融合双图信息提供了两种融合方法,即简单融合方法和注意力融合方法。

(3)在多个数据集上进行了充分实验,实验结果表明本文模型有效提升了文本分类的性能。

2 相关工作

近年来,鉴于图神经网络在处理复杂结构和保存全局信息方面表现出良好的性能,研究人员将图神经网络引入文本分类中,并取得了较好的效果。本文将基于图神经网络的文本分类模型分为图卷积网络模型和基于消息传递的图神经网络模型,图1给出了两类模型的整体框架。

图卷积网络模型基于全局词共现信息和文本-词的关系,首先构建包含词节点和文本节点的语料库级异构图,然后利用图卷积操作学得词表示和文本表示,最后将文本表示输入全连接分类器进行分类,主要架构如图1(a)所示。Yao等^[4]最先将图卷积应用到文本分类任务中,提出的TextGCN模型将词和文本作为节点,对整个数据集构建一个语料库级的图,使用标准的图卷积网络同时学习单词表示和文本表示。Wu等^[7]认为卷积层之间的非线性变换不是关键,因此去除了每层之间的非线性变换函数并且压缩了权重矩阵。相比TextGCN,该模型降低了复杂性,减少了冗余计算。Liu等^[6]提出了张量图神经网络模型用于协调和集成多图异构信息,构造文本图张量来描述语义、句法和顺序的上下文信息,然后在文本图张量上进行图内传播和图间传播。Hu等^[8]在异构图神经网络中引入了双层注意力结构,获得了不同粒度级别

的关键信息,降低了噪声信息的影响。以上图卷积网络模型对整个数据集构建语料库级图,导致模型占用大量内存,无法用于数据规模较大的数据集。同时,图的结构和参数依赖于数据集本身,无法对训练过程中没有出现过的新加入文本进行分类。

基于消息传递的图神经网络模型能够基于单个文本构建文本级图,通过消息传递机制来更新词节点表示,最终在读出层根据词表示求得文本表示用于分类,主要架构如图1(b)所示。Huang等^[9]提出的TextLevelGNN是首个用于文本分类的基于消息传递的图神经网络模型。该模型将文本中的单词作为图的节点,使用固定大小的滑动窗口来创建词间的连接,并将其作为图的边;为了利用全局词共现信息,TextLevelGNN计算全局的点互信息(Point-wise Mutual Information,PMI)矩阵并将其作为边的权重;在消息传递和读出阶段均使用了简单结构,使模型具有较低的复杂度。Zhang等^[10]提出的TextING构建了文本级的边权矩阵,在消息传递阶段使用门控图神经网络(Gated Graph Neural Network,GGNN)来更新词节点表示,取得了很好的效果。Nikolentzos等^[5]提出了MPAD,在构建文本级图时引入了一个文本节点,并将其与所有单词建立连接,获得全局统计信息;在读出阶段将词表示经过自注意力机制(Self-Attention)求得一个临时的文本表示,并与图中文本节点更新后的表示进行连接操作,作为最终的文本表示用于分类。与图卷积网络模型相比,基于消息传递的图神经网络模型使同一单词在不同上下文环境中具有不同的词表示,消除了单个文本与整个语料库之间的依赖关系,从而实现对新加入的文本进行分类,具有内存占用少、在线检测快等优点。但该类模型中缺少类似于图卷积模型^[6]中对多种语义图融合的相关研究,因此本文主要关注多图融合方法对文本级的图神经网络模型性能的改善。

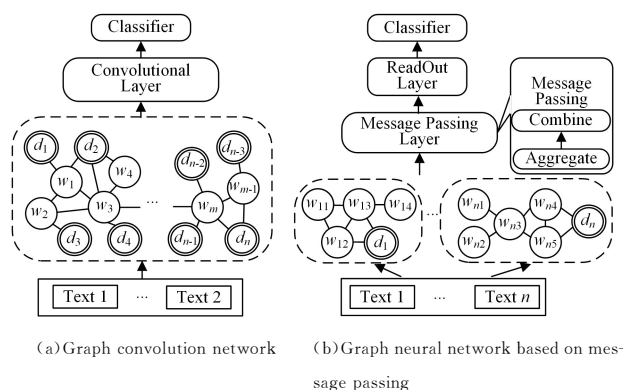


图1 基于图神经网络的文本分类模型的框架

Fig.1 Frameworks of text classification models based on graph neural network

3 基于消息传递的图神经网络模型

基于消息传递的图神经网络模型主要包括构建文本级图、消息传递和读出3个阶段。本节选取当前最新的TextLevelGNN,TextING和MPAD这3个模型作为本文的基础模型,围绕上述3个阶段进行详细介绍。

3.1 基本定义

假定长度为1的文本为 $T = \{h_1, \dots, h_i, \dots, h_l\}$, 任务目标是预测它的分类标签 y 。本文用 $G = (V, E)$ 表示图, 其中 $V (|V| = n)$ 为图中所有节点的集合, E 为图中所有边的集合, $v_i \in V$ 为其中的一个节点, $e_{ij} = (v_i, v_j) \in E$ 代表节点 v_i 到 v_j 的一条边。 $H \in \mathbb{R}^{n \times d}$ 代表节点特征矩阵(也称节点表示矩阵), $h_i \in \mathbb{R}^d$ 为节点 i 的特征向量(也称节点表示), d 为特征的维度。 $A \in \mathbb{R}^{n \times n}$ 为邻接矩阵, $D \in \mathbb{R}^{n \times n}$ 为度矩阵。

构建文本图 G 的过程即为确定图 G 的节点集、边集以及计算邻接矩阵的过程。下面通过文本词汇图的构建过程来给出词汇图的定义。

定义 1(词汇图, Lexical Graph) 将文本 T 中的单词作为节点来构建节点集。在构建边时, 使用固定大小为 p 的滑动窗口中位于窗口内的中心词与其相邻词来构建边。构建词汇图时需要满足本文对节点集和边集的定义 $V = \{v_i | i \in [1, l]\}$ 和 $E = \{e_{ij} | i \in [1, l], j \in [i-p, i+p]\}$ 。边 e_{ij} 的权重取自窗口滑动过程中统计的词共现信息, 即词对 $\langle v_i, v_j \rangle$ 共现的频率。

3.2 消息传递和读出

如图 1(b) 所示, 在基于消息传递的图神经网络模型中, 消息传递阶段为核心操作, 词节点根据其 n 阶邻居节点更新其自身节点表示。消息传递包含聚集和融合两个过程, 在图 G 中, 以节点 $v \in V$ 为例, 式(1)为聚集过程, 在第 $t+1$ 层根据 v 的邻居节点 $N(v)$ 的节点表示, 计算消息向量 $m_v^{t+1} \in \mathbb{R}^d$, 其表达式如下:

$$m_v^{t+1} = \text{AGGREGATE}^{t+1}(\{h_w^t | w \in N(v)\}) \quad (1)$$

融合过程如式(2)所示, 将节点 v 的当前节点表示 h_v^t 与消息向量 m_v^{t+1} 相融合, 计算出 v 的新特征表示 h_v^{t+1} , 其表达式如下:

$$h_v^{t+1} = \text{COMBINE}^{t+1}(h_v^t, m_v^{t+1}) \quad (2)$$

经过 T 次迭代, 节点 v 最终的特征表示 h_v^T 来自于以节点 v 为根节点且高度为 T 的子树中所有节点传递的消息向量, 它蕴含了节点 v 邻居节点的拓扑结构信息和邻域内节点表示分布信息。

由于在文本分类任务中对每个文本构建了一个图, 因此需要计算图级特征表示, 得到文本表示 h_G , 用于输入分类器, 即如式(3)所示的读出阶段, 其表达式如下:

$$h_G = \text{READOUT}(\{h_v^T | v \in V\}) \quad (3)$$

3.3 基础模型

在建图阶段, TextING 基于单个文本构建了词汇图。MPAD 不仅构建了词汇图, 而且增加了一个文本节点, 文本节点与所有词节点建立边, 边权设置为 1。TextLevelGNN 使用同样的方法构建了节点集和边集, 不同之处在于, TextLevelGNN 为了利用全局词共现信息, 基于整个数据集使用滑动窗口策略计算了节点间的点互信息 PMI, 构建了一个全局共享的权重矩阵, 词汇图中边的权值取自该权重矩阵。PMI 是一种常用的单词关联度量方法, PMI 为正值表示词对的语义相关性较高, PMI 为负值则表示词对的语义相关性较低或不存在相关性。因此, TextLevelGNN 模型只保留了 PMI 为正值词对间的边。PMI 的计算方法如式(4)所示。

$$\text{PMI}(i, j) = \log \frac{p(i, j)}{p(i)p(j)}, p(i, j) = \frac{\#W(i, j)}{\#W} \quad (4)$$

$$p(i) = \frac{\#W(i)}{\#W}$$

其中, $\#W(i)$ 是数据集中包含词 i 的窗口数量, $\#W(i, j)$ 是数据集中同时包含词 i 和 j 的窗口的数量, $\#W$ 是数据集中窗口的总数。

接下来分别介绍 3 个模型在消息传递和读出阶段的实现过程。

(1)TextLevelGNN 消息传递阶段。聚集时根据滑动窗口内所有邻居节点的边权和节点表示来计算消息向量; 融合时对中心词节点表示与消息向量进行加权求和, 以计算更新后的中心词节点表示。计算过程如式(5)、式(6)所示。

$$m_v^{t+1} = \max_{w \in N_v^p} e_{vw} h_w^t \quad (5)$$

$$h_v^{t+1} = (1 - \eta_v) m_v^{t+1} + \eta_v h_v^t \quad (6)$$

其中, $m_v^{t+1} \in \mathbb{R}^d$ 代表中心词节点 v 从其邻居节点收集的更新信息; \max 操作为取节点表示各维度上的最大值; N_v^p 为与节点 v 距离不大于 p 的节点的集合; e_{vw} 为节点 v 与 w 间的边权; h_v^t 和 h_v^{t+1} 分别为更新前后的节点表示, 由 η_v 决定保留多少原始信息。

最后, 读出阶段使用式(7)对迭代 T 次后的节点表示求和, 得到文本表示 h_G , 用于输入分类器。

$$h_G = \sum_{v \in N_t} h_v^T \quad (7)$$

(2)TextING 消息传递阶段。聚集过程使用式(8)所示的多层感知机 MLP 接收来自邻居节点的更新信息, 使用式(9)一式(12)所示的门控图神经网络对原有的中心词节点表示 h_v^t 和更新信息 m_v^{t+1} 进行融合, 本文将上述过程称为词交互。由于单层词交互仅能接收一阶邻居节点的信息, 因此 TextING 通过堆叠 T 层词交互的方式来获得高阶特征表示。

$$m_v^{t+1} = \text{MLP}^{t+1}(A h_v^t) \quad (8)$$

$$r^t = \sigma(W_r m_v^{t+1} + U_r h_v^t + b_r) \quad (9)$$

$$z^t = \sigma(W_z m_v^{t+1} + U_z h_v^t + b_z) \quad (10)$$

$$\tilde{h}_v^{t+1} = \tanh(W_h m_v^{t+1} + U_h (r^t \odot h_v^t) + b_h) \quad (11)$$

$$h_v^{t+1} = \tilde{h}_v^{t+1} \odot z^t + h_v^t \odot (1 - z^t) \quad (12)$$

然后, 使用式(13)所示的自注意力机制对迭代 T 次后的中心词节点表示 h_v^T 进行进一步转换。 f_1 和 f_2 均为多层感知机, $\sigma(f_1(h_v^T))$ 用于计算注意力权重, $\tanh(f_2(h_v^T))$ 对节点表示进行非线性特征转换。

$$h_v = \sigma(f_1(h_v^T)) \odot \tanh(f_2(h_v^T)) \quad (13)$$

最后的读出阶段如式(14)所示, 对转换后的文本表示矩阵 (h_1, \dots, h_n) 进行平均和最大池化, 得到两个临时的文本表示, 对二者进行求和, 得到最终的文本表示。

$$h_G = \frac{1}{|V|} \sum_{v \in V} h_v + \text{MaxPooling}(h_1, \dots, h_n) \quad (14)$$

(3)MPAD 消息传递阶段。如式(15)所示, 使用多层感知机对文本的节点表示矩阵 $H^t \in \mathbb{R}^{n \times d}$ 进行更新(其中除词节点以外还包含了一个特殊的文本节点)。聚集过程进行了归一化, 使 $D^{-1}A$ 各行元素之和分别为 1, 因此节点的更新信息实际上是其邻居节点表示的加权平均值。

$$M^{t+1} = \text{MLP}^{t+1}(D^{-1}A H^t) \quad (15)$$

融合过程使用门控神经网络对节点表示矩阵 \mathbf{H} 和更新信息 \mathbf{M}^{t+1} 进行融合,如式(16)–式(19)所示。

$$\mathbf{R}^{t+1} = \sigma(\mathbf{W}_R^t \mathbf{M}^{t+1} + \mathbf{U}_R^t \mathbf{H}^t) \quad (16)$$

$$\mathbf{Z}^{t+1} = \sigma(\mathbf{W}_Z^t \mathbf{M}^{t+1} + \mathbf{U}_Z^t \mathbf{H}^t) \quad (17)$$

$$\tilde{\mathbf{H}}^{t+1} = \tanh(\mathbf{W}^{t+1} \mathbf{M}^{t+1} + \mathbf{U}^{t+1} (\mathbf{R}^{t+1} \odot \mathbf{H}^t)) \quad (18)$$

$$\mathbf{H}^{t+1} = (1 - \mathbf{Z}^{t+1}) \odot \mathbf{H}^t + \mathbf{Z}^{t+1} \odot \tilde{\mathbf{H}}^{t+1} \quad (19)$$

经过 T 次迭代后,得到最终的节点表示矩阵 $\mathbf{H}^T \in \mathbb{R}^{n \times d}$ 和 $\hat{\mathbf{H}}^T \in \mathbb{R}^{(n-1) \times d}$,其中 $\hat{\mathbf{H}}^T$ 是只包含词节点而不包含文本节点的表示矩阵。

在读出阶段,对 $\hat{\mathbf{H}}^T$ 使用式(20)–式(22)所示的自注意力机制计算一个临时的文本表示 $\mathbf{u}^T \in \mathbb{R}^d$,并与消息传递阶段求得的 \mathbf{H}^T 中的文本节点的表示进行连接,得到新的文本表示 $\mathbf{h}_G^T \in \mathbb{R}^{2d}$ 。

$$\mathbf{Y}^T = \tanh(\hat{\mathbf{H}}^T \mathbf{W}_A^T) \quad (20)$$

$$\alpha_i^T = \frac{\exp(\mathbf{Y}_i^T \cdot \mathbf{v}^T)}{\sum_{j=1}^{n-1} \exp(\mathbf{Y}_j^T \cdot \mathbf{v}^T)} \quad (21)$$

$$\mathbf{u}^T = \sum_{i=1}^{n-1} \alpha_i^T \hat{\mathbf{H}}_i^T \quad (22)$$

MPAD 模型在每一层均执行读出函数,并将结果进行连接,得到最终的文本表示 $\mathbf{h}_G \in \mathbb{R}^{T \times 2d}$,用于输入分类器,上述连接过程如式(23)所示。

$$\mathbf{h}_G = \text{CONCAT}(\text{READOUT}(\mathbf{H}^t) | t=1, \dots, T) \quad (23)$$

4 双图神经网络信息融合模型

在基于消息传递的图神经网络模型已经存在词汇图的情况下,通过构建语义图为其引入额外的语义信息,并给出双图更新机制的整体框架以及双图信息融合方法。同样,我们根据语义图的构建过程给出语义图的定义。

定义 2(语义图, Semantic Graph) 在构建文本图时,通常使用预训练的词嵌入来初始化节点表示 \mathbf{h}_i ,由于其包含了语义信息,因此通过式(24)计算余弦相似度能够得到词对的语义相似度。构建节点集的过程与定义 1 中构建词汇图保持一致。构建边集时,首先在所有的节点间构建边,边的权重通过计算两个节点的初始化表示的余弦相似度(Cosine Similarity)来确定,仅保留权重大于或等于参数 $\theta \in [0, 1)$ 的边。

$$\text{CosineSimilarity} = \frac{\mathbf{h}_i \cdot \mathbf{h}_j}{\|\mathbf{h}_i\| \|\mathbf{h}_j\|} \quad (24)$$

在模型训练时,假设 \mathbf{h}_G^L 为通过单一的词汇图进行消息传递更新得到的文本表示, \mathbf{h}_G^S 为通过单一的语义图进行消息传递更新得到的文本表示。图 2 给出了双图更新机制的整体框架,首先按照定义 1 和定义 2 对文本构建词汇图和语义图;然后利用两个图分别进行消息传递和读出,从而得到两个文本表示 \mathbf{h}_G^L 和 \mathbf{h}_G^S ;最后对上述两个文本表示进行融合,得到最终的文本表示 \mathbf{h}_G 。

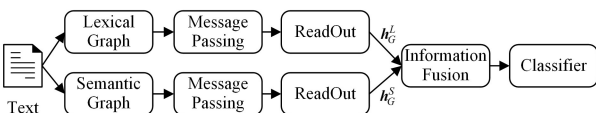


图 2 采用双图更新机制模型的整体框架

Fig. 2 Overall framework of model using both lexical graph and semantic graph

为融合双图信息,本文提出了简单融合和注意力融合两种融合方法。

(1)简单融合方法。该方法简单地认为通过词汇图和语义图更新得到的两个文本表示具有同等重要性,因此通过式(25)对两个文本表示取平均值,从而求得最终文本表示。

$$\mathbf{h}_G = \text{Avg}(\mathbf{h}_G^L, \mathbf{h}_G^S) \quad (25)$$

(2)基于注意力的融合方法。该方法假设通过词汇图和语义图更新得到的两个文本表示具有非均等的重要性,使用注意力机制对二者进行加权,这样能够对词汇图和语义图建立更细粒度的依赖,有利于提高文本分类的效果。使用注意力机制求得最终文本表示的过程如式(26)–式(29)所示。

$$\mathbf{H} = \mathbf{h}_G^L \oplus \mathbf{h}_G^S \quad (26)$$

$$\mathbf{Y} = \text{ReLU}(\mathbf{H}_G \mathbf{W}_A) \quad (27)$$

$$\alpha_i = \frac{\exp(\mathbf{Y}_i \cdot \mathbf{v}_A)}{\sum_{j=1}^{n-1} \exp(\mathbf{Y}_j \cdot \mathbf{v}_A)} \quad (28)$$

$$\mathbf{h}_G = \sum_{i=1}^{n-1} \alpha_i \mathbf{H}_i \quad (29)$$

其中, \oplus 为拼接操作, \mathbf{W}_A 和 \mathbf{v}_A 是模型要训练的参数矩阵和参数向量, α_i 是经过 Softmax 求得的权重。

在 3.3 节提到的 3 个基础模型之上,本文沿用其原有的消息传递层和分类器,使用简单融合方法和注意力融合方法,构建了 6 种基于双图神经网络信息融合的文本分类模型。

5 实验与结果

5.1 数据集

本文实验所使用的数据集可以归为新闻分类、主题分类和情感分类 3 类,表 1 列出了数据集的统计信息。CV 表示未划分训练集和测试集。使用 10 折交叉验证来评估算法性能。数据集的具体描述如下:

- (1)R8 和 R52 分别为包含 8 个类别的新闻文本和 52 个类别的新闻文本的 Reuters 新闻数据集;
- (2)BBCSport^[11] 包含来自 BBC 体育网站的 737 篇文章,这些文章为 2004–2005 年间 5 个主题领域的体育新闻文章;
- (3)Ohsumed 来自医学文献数据库 MEDLINE,每个文本都是医学摘要,并且属于 23 种心血管疾病类别中的一种;
- (4)Snippets 来自于 Google 搜索的片段,包含 8 个类别;
- (5)TREC^[12] 包含 6 类问题,其包含的文本均为疑问句;
- (6)MR^[13] 电影评论的每个评论只有一句话,包括正面情绪和负面情绪两类;
- (7)IMDB^[14] 电影评论的每个评论为一个段落,包含多个句子,包括正面情绪和负面情绪两类文本;
- (8)SST-1, SST-2^[15] 与 MR 相比, SST-1 具有更细粒度的类别(非常积极、积极、中性、消极、非常消极), SST-2 去除了 SST-1 中类别为中性的文本,并将剩余的文本重新划分为积极和消极两类;
- (9)MPQA^[16] 收集了各种新闻文章中的评论意见,包含正面和负面两类文本;
- (10)Subj^[17] 包含主观评价和客观评价两类情感文本。

表 1 数据集统计信息
Table 1 Statistics of datasets

Dataset	Task	# Docs	# Training	# Test	# Classes	Avg. Length	# Words
R8	News	7 674	5 485	2 189	8	64.53	2 923
R52	News	9 100	6 532	2 568	52	68.63	5 364
BBCSport	News	737	737	CV	5	337.61	13 253
Ohsumed	Topics	7 400	3 357	4 043	23	112.83	7 998
Snippets	Topics	12 340	10 060	2 280	8	18.07	29 251
TREC	Topics	5 952	5 452	500	6	8.52	8 739
MR	Sentiment	10 662	10 662	CV	2	18.69	18 322
IMDB	Sentiment	50 000	25 000	25 000	2	223.27	32 417
SST-1	Sentiment	11 855	9 645	2 210	5	18.18	17 830
SST-2	Sentiment	9 613	7 792	1 821	2	18.33	16 182
MPQA	Sentiment	10 606	10 606	CV	2	3.11	6 204
Subj	Sentiment	10 000	10 000	CV	2	21.19	20 892

5.2 基准算法

为评估本文模型的有效性,除了对比基础模型以外,还选取了文本分类领域中 5 个经典的文本分类算法与本文算法进行对比实验。

(1)TextCNN^[2]:基于卷积神经网络的模型,对词嵌入执行卷积和最大池化操作,得到文本表示。

(2)DAN^[18]:前馈神经网络模型,求词嵌入平均值再经过多个全连接层进行分类。

(3)TextRNN^[3]:基于递归神经网络的模型,该模型使用最后一个隐藏状态作为文本表示。

(4)HN-ATT^[19]:层次注意力模型,每层使用双向 GRU 进行编码。

(5)TextGCN^[4]:基于异构图的图卷积模型,将文本分类问题变为节点分类问题,使用图卷积求得文本表示。

5.3 实验结果与分析

本文从训练集中随机选取 10% 的文本作为验证集,用于

调试模型参数。实验使用 300 维的预训练词嵌入初始化词表示^[20],对未登录词则进行随机初始化。使用 Adam 优化器作为模型的优化方法,初始学习率设置为 0.001,L2 正则化参数设置为 0.0001,批训练大小设置为 32,根据使用的融合方法不同将阈值 θ 设置为 0.4 或 0.6,其余参数使用基础模型的默认设置。本文模型在 TITAN Xp 显卡上进行训练,实验使用准确率(Accuracy)作为评价指标。

实验结果如表 2 所列,第 1—第 4 行为基于序列学习的模型,第 5 行为图卷积网络模型,第 6—第 9 行为仅使用词汇图或仅使用语义图构建的两个 TextLevelGNN 模型,以及在 TextLevelGNN 基础上提出的两个双图神经网络信息融合模型(我们将简单融合模型标记为-S,将注意力融合模型标记为-A),同理第 10—第 13 行、第 14—第 17 行分别为 TextING 和 MPAD 单图模型及其改进的双图模型。分别标记每组数据的最优结果和次优结果,加粗的为最优结果,斜体加下划线的为次优结果。

表 2 本文模型与基础模型在文本分类任务上的性能对比

Table 2 Performance comparison of the proposed model and baselines model on text classification tasks

Algorithms	News Classification			Topics Classification			Sentiment Classification					
	R8	R52	BBCSport	Ohsumed	Snippets	TREC	MR	IMDB	SST1	SST2	MPQA	Subj
TextCNN	97.21	—	98.37	—	—	70.80	81.50	90.28	48.70	87.80	89.50	93.40
DAN	94.79	—	94.30	—	—	89.60	80.30	89.40	47.70	86.30	88.91	92.44
TextRNN	—	—	—	—	—	—	77.33	91.30	49.60	87.90	—	94.10
HN-ATT	97.25	—	96.73	—	—	90.80	80.78	90.06	49.00	86.71	89.08	92.92
TextGCN	97.07	93.56	—	68.36	79.90	89.76	76.74	88.52	39.57	81.13	—	—
TextLevelGNN(Lexical)	97.80	94.60	99.78	<u>69.40</u>	74.84	87.40	77.11	87.93	41.79	81.34	92.76	91.12
TextLevelGNN(Semantic)	<u>98.02</u>	<u>94.99</u>	99.41	69.26	73.20	88.27	<u>78.22</u>	<u>89.10</u>	43.27	82.39	92.39	<u>91.24</u>
TextLevelGNN-S	98.13	95.05	99.87	69.95	76.32	88.40	77.93	88.77	44.75	<u>83.09</u>	93.12	91.11
TextLevelGNN-A	97.58	94.67	99.87	68.91	<u>75.31</u>	90.00	78.58	89.70	<u>44.30</u>	84.13	<u>92.87</u>	91.78
TextING(Lexical)	97.42	94.62	99.40	70.20	86.14	98.68	79.50	OOM	48.10	85.38	89.08	91.64
TextING(Semantic)	97.38	93.69	99.62	68.66	86.14	98.60	80.22	OOM	47.38	84.51	90.06	92.26
TextING-S	98.13	<u>94.78</u>	99.99	71.06	<u>87.19</u>	99.00	81.93	OOM	<u>48.73</u>	<u>86.38</u>	90.28	<u>93.02</u>
TextING-A	<u>97.90</u>	95.13	<u>99.97</u>	<u>70.74</u>	87.50	99.00	<u>81.60</u>	OOM	48.91	86.71	<u>90.18</u>	93.04
MPAD (Lexical)	97.07	93.25	98.37	65.11	83.54	93.48	80.24	90.56	48.51	85.61	90.01	92.88
MPAD (Semantic)	97.27	93.03	<u>99.32</u>	64.61	82.85	92.00	79.55	88.90	47.83	85.23	89.91	92.59
MPAD-S	<u>97.62</u>	<u>93.85</u>	99.59	66.14	84.96	95.20	<u>80.32</u>	<u>90.88</u>	<u>49.64</u>	<u>86.60</u>	90.08	<u>93.07</u>
MPAD-A	97.72	93.89	99.19	<u>65.57</u>	<u>84.91</u>	<u>94.20</u>	80.68	90.99	50.05	87.20	<u>90.01</u>	93.16

实验结果表明,本文提出的双图模型在文本分类任务上的性能均优于各自的基础模型,而且在部分数据集上分类精度有显著的提升。具体地,与原始的 TextLevelGCN 模型相比,双图 TextLevelGCN 模型的准确率在 Snippets, MR 和 IMDB 数据集上均有 1.5% 左右的提升;在 TREC, SST-1 和

SST-2 数据集上均有高于 2.5% 的提升。与原始的 TextING 模型相比,双图 TextING 模型的准确率在 Snippets, TREC, SST-2, MPQA 和 Subj 数据集上均有 1.3% 左右的提升;在 MR 数据集上有 2.4% 的提升。与原始的 MPAD 模型相比,双图 MPAD 模型的准确率在 BBCSport, Ohsumed 和 Snip-

pets 数据集上均有高于 1% 的提升;在 TREC,SST-1 和 SST-2 数据集上均有高于 1.5% 的提升。

除此之外,本文提出的两种融合方法在不同类型数据集上的效果具有差异性。整体上,在新闻分类和主题分类数据集上,简单融合方法(-S)的效果优于注意力融合方法(-A);在情感分类数据集上,注意力融合方法(-A)的效果优于简单融合方法(-S)。具体地,TextLevelGCN-S 在新闻分类和主题分类数据集上的效果更好,TextLevelGCN-A 则在情感分类数据集上表现更优;TextING-S 在新闻分类和主题分类数据集上的效果更好,TextING-A 在情感分类数据集上取得了不错的效果;MPAD-S 在主题分类数据集上效果更好,MPAD-A 在新闻分类和情感分类数据集上表现更佳。

通过分析可知,本文提出的双图模型性能的提升源自以下两点:1)模型在保留词汇图信息的基础上,通过构建语义图

有效地引入了语义信息,使模型具备同步获取多方位文本信息的能力;2)两种信息融合方法各自适配不同类型的数据集,提升了模型对文本表示的编码能力。

5.4 参数影响

在本文模型中,参数 θ 会对文本分类的效果产生影响。参数 θ 是在构建语义图时定义两个节点的余弦相似度的阈值,模型仅保留相似度大于或等于 θ 的边,用于控制语义图的稀疏程度,从而控制引入的语义的信息量和精准性。图 3、图 4 分别给出了在不同的 θ 设置下,本文模型在 R8 和 MR 数据集上的准确率指标。可以看出,采用注意力融合方法时,将 θ 设置为 0.5 及以上可以达到最佳性能,这样可以保证引入适量的语义信息和较少的噪声。采用简单融合方法时,将 θ 设置在 0.5 以下可以达到最佳性能,虽然引入了更多的噪声,但同时引入了更多的语义信息,这些信息对提高分类性能具有帮助。

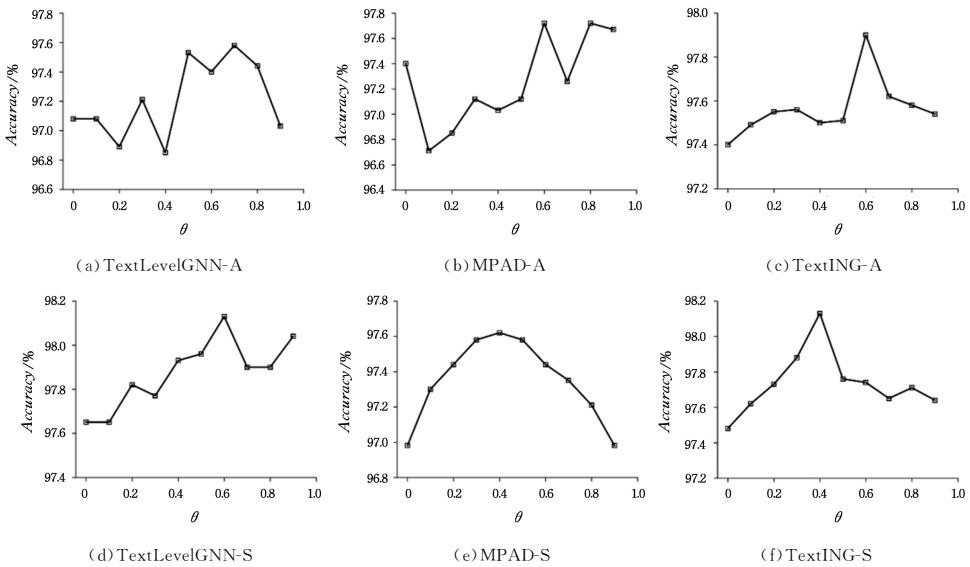


图 3 在 R8 数据集上参数 θ 对分类性能的影响

Fig. 3 Influence of parameter θ on classification performance on dataset R8

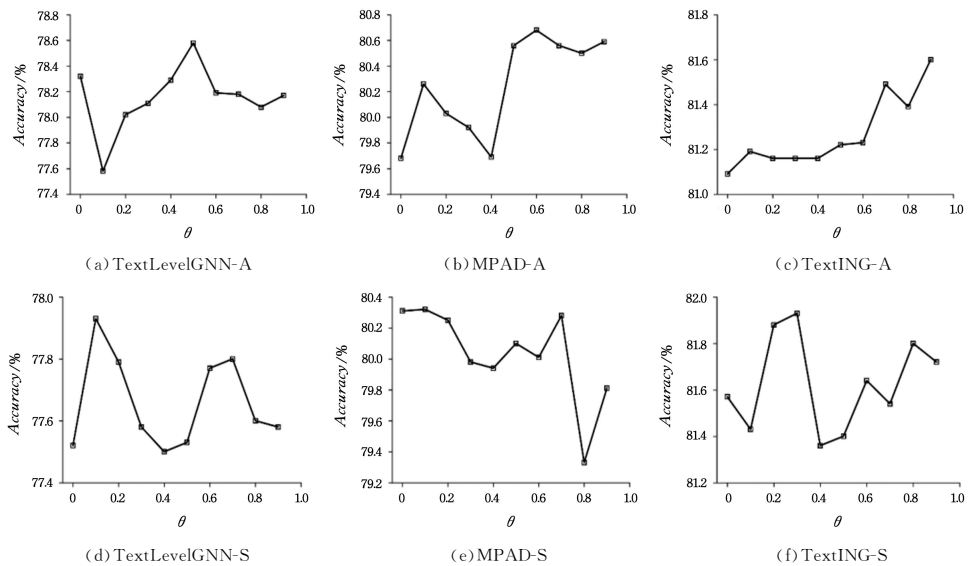


图 4 在 MR 数据集上参数 θ 对分类性能的影响

Fig. 4 Influence of parameter θ on classification performance on dataset MR

结束语 针对当前基于消息传递的图神经网络模型获取信息不足的问题,本文提出了基于双图神经网络信息融合的文本文本分类方法。该方法保留了原先使用词共现信息构建的词汇图,利用余弦相似度来度量单词间的语义关联程度从而构建语义图,通过调整图的稀疏程度来控制输入模型的语义信息量,简单融合方法和注意力融合方法则能够在不同粒度上对两个图的信息进行融合。在 R8, R52 等 12 个文本分类公开数据集上的实验结果表明,相比 TextLevelGNN, TextING 和 MPAD 这 3 个基础模型,本文模型性能有明显提高,具有较强的信息融合能力。但本文模型仅对两个图的文本表示进行融合,下一步将考虑在两个图之间建立联系,实现在模型训练过程中对双图信息进行交互。

参 考 文 献

- [1] SHERVIN M, NAL K, ERIK C, et al. Deep Learning Based Text Classification: A Comprehensive Review[EB/OL]. <https://arxiv.org/abs/2004.03705v1>.
- [2] KIM Y. Convolutional Neural Networks for Sentence Classification[C] // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1746-1751.
- [3] LIU P, QIU X, HUANG X. Recurrent neural network for text classification with multi-task learning[C] // Proceedings of the 25th International Joint Conference on Artificial Intelligence. 2016:2873-2879.
- [4] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019:7370-7377.
- [5] NIKOLENTZOS G, TIXIER A J, VAZIRGIANNIS M. Message Passing Attention Networks for Document Understanding[C] // Proceedings of the 34th AAAI Conference on Artificial Intelligence. 2020:8544-8551.
- [6] LIU X, YOU X, ZHANG X, et al. Tensor Graph Convolutional Networks for Text Classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020:8409-8416.
- [7] WU F, ZHANG T, YU T, et al. Simplifying Graph Convolutional Networks[C] // Proceedings of the 36th International Conference on Machine Learning. 2019:6861-6871.
- [8] HU L, YANG T, SHI C, et al. Heterogeneous graph attention networks for semi-supervised short 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). 2019:4823-4832.
- [9] HUANG L, MA D, LI S, et al. Text level graph neural network for text classification[C] // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. 2019: 3435-3441.
- [10] ZHANG Y, YU X, CUI Z, 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. 2020:334-339.
- [11] GREENE D, CUNNINGHAM P. Practical Solutions to the Problem of Diagonal Dominance in Kernel Document Clustering [C] // Proceedings of the 23rd International Conference on Machine Learning. 2006:377-384.
- [12] LI X, ROTH D. Learning question classifiers: the role of semantic information[C] // Proceedings of the 19th International Conference on Computational Linguistics. 2002:1-7.
- [13] PANG B, LEE L. Seeing Stars: Exploiting Class Relationships for Sentiment Categorization with Respect to Rating Scales [C] // Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. 2005:115-124.
- [14] MAAS A L, DALY R E, PHAM P T, et al. Learning Word Vectors for Sentiment Analysis[C] // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. 2011:142-150.
- [15] SOCHER R, PERELYGIN A, POTTS C, et al. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank[C] // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013:1631-1642.
- [16] WIEBE J, WILSON T, CARDIE C. Annotating Expressions of Opinions and Emotions in Language[J]. Language Resources and Evaluation, 2005, 39(2/3):165-210.
- [17] PANG B, LEE L. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts [C] // Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. 2004:271-278.
- [18] IYYER M, MANJUNATHA V, BOYD-GRABER J, et al. Deep Unordered Composition Rivals Syntactic Methods for Text Classification[C] // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015: 1681-1691.
- [19] 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.
- [20] 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). 2014:1532-1543.



YAN Jia-dan, born in 1995, postgraduate. Her main research interests include natural language processing and text classification.



JIA Cai-yan, born in 1976, Ph.D, professor, is a member of China Computer Federation. Her main research interests include machine learning, complex network analysis and social computing.