

融合 BiLSTM 的双图神经网络文本分类模型

宋婷婷¹, 吴赛君¹, 裴颂文^{1,2,3}

(1. 上海理工大学 光电信息与计算机工程学院, 上海 200093; 2. 中国科学院 计算技术研究所 计算机体系结构国家重点实验室, 北京 100190; 3. 软硬件协同设计技术与应用教育部工程研究中心(华东师范大学), 上海 200062)

摘要: 采用图神经网络模型为整个语料库构建异构图处理文本分类任务时, 存在难以泛化到新样本和词序信息缺失的问题。针对上述问题, 提出了一种融合双图特征和上下文语义信息的文本分类模型。首先, 为每个文本独立构建共现图和句法依存图, 从而实现对新样本的归纳式学习, 从双图角度捕获文本特征, 解决忽略单词间依存关系的问题; 其次, 利用双向长短期记忆网络(bi-directional long short-term memory, BiLSTM)编码文本, 解决忽略词序特征和难以捕捉上下文语义信息的问题; 最后, 融合双图特征, 增强图神经网络模型的分类性能。在 MR, Ohsumed, R8, R52 数据集上的实验结果表明, 相较于经典的文本分类模型, 该模型能够提取更丰富的文本特征, 在准确率上平均提高了 2.17%, 5.38%, 0.61%, 2.48%。

关键词: 文本分类; 图神经网络; 双向长短期记忆网络; 句法依存图; 共现图

中图分类号: TP 391 文献标志码: A

Dual graph neural networks with BiLSTM for text classification

SONG Tingting¹, WU Saijun¹, PEI Songwen^{1,2,3}

(1. School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 2. State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; 3. Engineering Research Center of Software/Hardware Co-design Technology and Application, Ministry of Education (East China Normal University), Shanghai 200062, China)

Abstract: When a heterogeneous graph was constructed for graph neural network model for the whole corpus to text classification, it was difficult to generalize to new samples and word order information was missed. To solve these problems, a text classification that incorporated dual graph features and contextual semantic information was proposed. Firstly, a syntactic dependency graph and a co-occurrence graph were constructed for individual text to inductive learning for new samples. Text features were captured from two perspectives, which solved the problem of ignoring dependency information between words. Secondly, bi-directional long short-term memory network (BiLSTM) was used to encode text, which solved the problem of ignoring word order features and having difficulty in

收稿日期: 2023-02-08

基金项目: 国家自然科学基金资助项目(61975124); 上海市自然科学基金资助项目(20ZR1438500); 计算机体系结构国家重点实验室开放课题(CARCHA202111); 软硬件协同设计技术与应用教育部工程研究中心开放课题(OP202202)

第一作者: 宋婷婷(1998-), 女, 硕士研究生。研究方向: 自然语言处理。E-mail: sstmiracle@163.com

通信作者: 裴颂文(1981-), 男, 教授。研究方向: 计算机体系结构、智能计算等。E-mail: swpei@usst.edu.cn

capturing contextual semantic information. Finally, dual graph features were fused to enhance the classification performance of graph neural network model. The experimental results on MR, Ohsumed, R8 and R52 datasets show that the model can extract richer text features and have an average increase of 2.17%, 5.38%, 0.61% and 2.48% in accuracy compared with classical text classification models.

Keywords: *text classification; graph neural networks; bi-directional long short-term memory network; syntactic dependency graph; co-occurrence graph*

文本分类是自然语言处理的重要内容，旨在将无标签的文本分类到预先定义类别中，被广泛应用于情感分析、意图识别等领域。传统的文本分类方法使用词袋模型、词频-逆文档频率指数表示文本，忽略了单词间的语义联系，同时存在数据稀疏的问题。随着深度学习的发展，神经网络被广泛应用于文本分类中，如卷积神经网络(CNN)、循环神经网络(RNN)、长短期记忆网络(LSTM)等。CNN和RNN能够捕捉文本局部信息，但由于模型的局部性和序列性特点，难以有效地捕捉非连续和远距离语义信息。进而衍生出一些变体如双向长短期记忆网络(BiLSTM)、DPCNN模型^[1]、双通道RNN模型^[2]等。Devlin等^[3]在2018年提出双向模型BERT，其核心组成是Transformer编码器^[4]。BERT采用多任务预训练技术在大型语料库上训练。上述模型在文本分类中表现良好，但无法对单词的全局共现关系建模。

图神经网络技术在图结构数据上具有强大的处理能力，吸引了众多研究者将其应用于文本分类^[5]。2019年Yao等^[6]根据词共现关系和单词文本关系为整个语料库构建一张异构图，之后使用图卷积神经网络(graph convolutional network, GCN)^[7]提取图特征。为了丰富图信息，Hu等^[8]在构图时引入主题节点，Xin等^[9]引入了标签信息。2020年Liu等^[10]提出构建词共现图、语义图、句法图，从多个角度捕捉文本信息。但以上研究是转导式学习，构图时使用了测试样本。当有新样本加入时，就要重新构建图和训练模型，这会浪费时间且不适用于现实情况^[11]。因此，归纳式文本分类模型不断涌现，其能够解决上述图神经网络存在的问题。InducT-GCN文本分类模型^[12]基于训练样本构建图，在测试样本上执行一维图卷积。一些学者提出为每个文本构建图。例如：Huang等^[13]使用边共享矩阵捕捉全局信息，采用消息传递机制捕获文本特征；Zhang等^[14]提出为每个文

本构建单词共现图，并使用门控图神经网络(gated graph neural networks, GGNN)^[15]传播信息。上述模型仅使用词共现信息构建文本图，未考虑语义等其他类型信息。Li等^[16]提出一种使用语义图和句法图的方法，用于方面级情感分析。Dai等^[11]使用共现信息和预训练词嵌入构建4种不同类型的文本图，在不重构整个文本图的情况下，学习系统也能够对新文本进行推理。为弥补单图信息不足的缺陷，本文探索多种构图方式，从多维度捕获文本特征，提升文本分类的性能。同时图神经网络文本分类模型仅关注文本的词交互，忽略了文本的词序信息^[14,17]，不能充分捕捉上下文语义特征。针对上述问题，本文提出了融合BiLSTM的双图神经网络(dual graph neural networks with BiLSTM, 简称DGNN-B)模型，通过结合不同的语义信息实现归纳式文本分类。

本文通过DGNN-B模型为每个文本独立构图，将文本分类问题转换为图分类问题，实现了对新文本的归纳式学习。针对模型获取信息不足的问题，为文本构建共现图、句法依存图，从双图角度提取文本特征^[18]。引入双向长短期记忆网络(BiLSTM)提高模型捕捉上下文语义信息的能力。在4个基准数据集上的实验结果表明，所提方法优于经典模型，在文本分类任务中具有较高的分类精度。

1 DGNN-B 文本分类模型

针对为整个语料库构建异形图的文本分类模型无法轻易泛化到新样本的问题，本文提出了DGNN-B模型，模型架构如图1所示。通过DGNN-B模型为每个文本构建词图，可以对新文本进行归纳式分类，无需从头训练。为丰富构图质量，基于单词共现和依存关系构建文本图，充分捕捉共现信息和句法依赖特征，提高文本分类

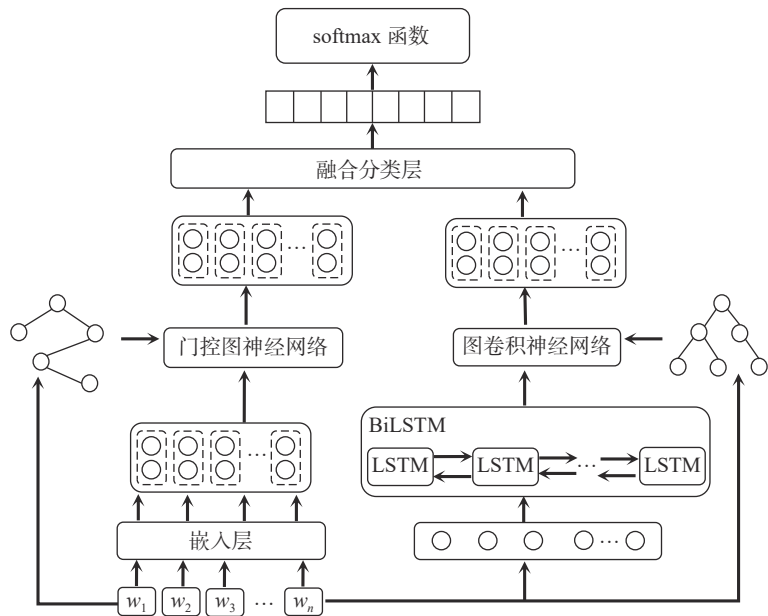


图 1 融合 BiLSTM 的双图神经网络 (DGNN-B)

Fig.1 Dual graph neural networks with BiLSTM (DGNN-B)

性能。本文模型还利用 BiLSTM 捕捉文本的词序信息。

DGNN-B 模型的关键部分是：构建文本图、嵌入层、BiLSTM 深层特征提取模块、图特征提取模块、融合分类层。

1.1 构建文本图

本文为每个文本独立构建共现图和句法依存图，专注于文本自身结构，当有新样本加入时，模型不需要重新训练，同时有效地利用了文本的多维度语义信息。现介绍如何基于单词间的关系将文本构建成图结构。拥有 n 个单词的文本 $T = \{w_1, \dots, w_i, \dots, w_n\}$, w_i 是文本中的第 i 个单词。通过嵌入层，使用预训练的 Glove 词向量将单词映射为 d 维向量。对于文本 T ，以单词为节点，单词间的关系为边，为其构建文本图 $G_t = (V_t, E_t)$, V_t, E_t 为文本图的节点集和边集。文本图的特征矩阵 $X_t \in \mathbb{R}^{n_t \times d}$ ，第 i 个节点的向量表示记作 $x_i \in \mathbb{R}^d$, n_t 是文本图 G_t 的节点数， $n_t = |V_t|$ 。文本图的邻接矩阵 $A_t \in \mathbb{R}^{n_t \times n_t}$ 。本文为每个文本构建 2 种不同类型的图：共现图和句法依存图。现介绍这 2 种文本图的构建过程。

1.1.1 共现图

单词共现关系描述单词局部共现的语言特征，被多位学者用于文本构图，是一种经典的构图方法。共现图 $G_1 = (V_1, E_1)$ ，将文本中的单词视作节点，单词之间的共现关系视作边。使用固定

尺寸的窗口在文本序列上从左至右滑动，在同一窗口内出现的单词存在共现关系，如图 2 所示。

局部滑动窗口可以在图构建期间有效地捕捉词与词的局部共现特征。共现图构建完成后，本文使用 GGNN 网络传播和融合词特征。

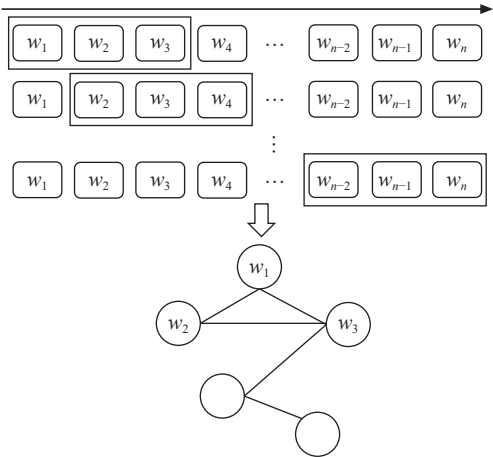


图 2 共现图构建

Fig.2 Co-occurrence graph construction

1.1.2 句法依存图

基于共现关系构建的文本图可以捕捉局部关联，但缺乏丰富的语义信息。通过对文本构建句法依存图，从双图角度提取文本特征。依存分析^[19-20]深入理解句子内部结构，分析句子成分，提取单词间的依存关系如主谓关系、定状关系等，帮助理解文本语言结构和含义。首先，对于语料库中的文本，使用 StanfordNLP 工具包^[21]提

取单词间的依赖关系。对于给定的文本通过解析器得到的分析结果如图3所示。虽然提取的依赖关系是定向的,但本文将其视作无向关系,以便后续文本图的构建和特征提取。对于一个文本,其依存关系解析结果为: $DP = \{r_{ij} | i \neq j; i, j < n_2\}$, r_{ij} 表示文本中单词 w_i 和 w_j 间的依存关系。依存关系提取完成之后,以词为节点、依存关系为边,构建句法依存图,记作 $G_2 = (V_2, E_2)$, 其中, $E_2 = \{e_{ij} | r_{ij} \in DP\}$, e_{ij} 为单词 w_i 和 w_j 之间的边。由此构建的句法依存图包含丰富的语义和句法特征。

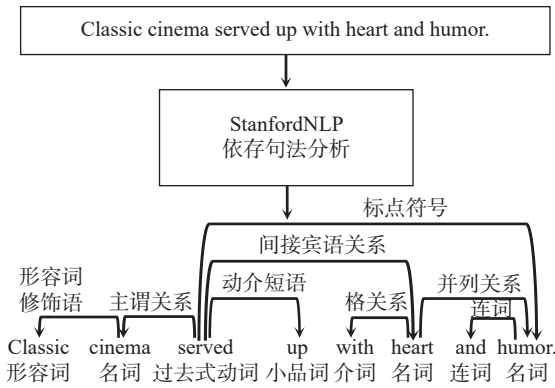


图3 依存句法分析示例

Fig.3 An example of dependency parsing

将每个文本转换为独立图结构进行文本分类时,模型遇到新样本不需要从头训练,能够对文本进行归纳式学习。与构建单一文本图相比,构建2种不同类型的文本图,可以捕捉到更加丰富的特征,在图上进行信息传播和卷积操作时,可以从2个方面提取特征,相互补充。

1.2 BiLSTM 深层特征提取模块

将文本构建成图结构,聚合邻居可以有效地捕捉远距离词节点,但忽略了文本的词序信息。因此,利用 BiLSTM 捕捉文本的双向上下文语义信息并保留文本的词序特征。

通过 BiLSTM 对文本序列进行双向编码表示,然后利用 GCN 网络进一步优化。不同于共现图使用 Glove 词向量技术初始化节点的嵌入表示,句法依存图通过 BiLSTM 捕捉上下文信息,提取更深层次的文本特征。一方面,文本是一种非欧几里德结构的数据, BiLSTM 可以保留文本的位置信息,捕获文本的词序特征;另一方面, BiLSTM 的双向机制通过充分考虑上下文,保证每个单词获得丰富的语义信息。

1.3 图特征提取模块

为了捕获单词之间的依赖关系,采用图卷积

神经网络提取句法依存图特征。GCN 可以出色地捕捉节点间关系,被广泛应用于自然语言处理。聚合邻居时为了选择重要的节点信息、遗忘不重要的信息,并捕捉较远距离的节点,使用门控图神经网络提取共现图的特征。该网络采用门控循环单元选择性地聚合或丢失邻居信息,从而增强了网络的长期记忆能力。

对于构建的句法依存图,其邻接矩阵记为 A_2 , BiLSTM 捕捉到的节点表示作为句法依存图的特征矩阵。在图上使用经典的 GCN,具体的卷积过程见文献 [7]。不同的数据集设置相应的卷积层数,最终得到句法依存图的嵌入表示,记为 $H_{dp} \in \mathbb{R}^{n_2 d_2}$, d_2 为句法依存图嵌入表示的维度。

在构建的共现图上,使用 GGNN 网络更新节点特征。在消息传递过程中,一个节点接受来自邻域节点的信息后,与上一时间点的表示融合起来,更新节点自身的隐藏表示。GGNN 具有长期记忆的能力,并且不再需要通过约束参数的方式来确保模型的收敛。在 t 时刻模型的传播过程如下:

$$N^t = A_1 H^{t-1} W_N + b_N \quad (1)$$

$$U^t = \sigma(W_U N^t + M_U H^{t-1}) \quad (2)$$

$$R^t = \sigma(W_R N^t + M_R H^{t-1}) \quad (3)$$

$$\tilde{H}^t = \tanh(W_H N^t + M_H (R^t \odot H^{t-1})) \quad (4)$$

$$H^t = H^{t-1} \odot (1 - U^t) + \tilde{H}^t \odot U^t \quad (5)$$

式中: $A_1 \in \mathbb{R}^{n_1 n_1}$ 为共现图的邻接矩阵; N 为聚合邻居信息; W, M 为可训练的权重; b 为偏差,不同时刻不共享; $\sigma, \tanh(k)$ 为激活函数, $\sigma(k) = 1/(1 + e^{-k})$, $\tanh(k) = (e^k - e^{-k})/(e^k + e^{-k})$; U, R 为更新门和重置门,分别控制遗忘信息和新产生的信息; R^t 决定从哪些过去的信息中产生新信息; $1 - U^t$ 选择遗忘不重要的信息, U^t 选择记住新产生的信息; $\odot$ 为逐元素乘积; $\tilde{H}^t$ 为新产生的信息; H^t 为 t 时刻节点的嵌入表示。

设置 $H^0 = X_1$, X_1 为共现图的初始化嵌入矩阵。最终共现图的嵌入表示记作 $H_{co} \in \mathbb{R}^{n_1 d_1}$, d_1 为共现图嵌入表示的维度。

1.4 融合分类层

词节点信息被传播后,聚合生成图表示。然后通过拼接操作融合共现图和句法依存图的图表示,从双图角度提取文本特征。

首先,使用软注意力、最大池化及平均池化聚合句法依存图、共现图的节点生成图级表示为

$\mathbf{H}_{G-dp}, \mathbf{H}_{G-co}$ 。以共现文本图为例, 聚合前文通过 GGNN 学习的共现图的词嵌入表示 $\mathbf{H}_{co}$, 得到全局表示 $\mathbf{H}_{G-co}$ 。

$$a_v = \sigma(o(\mathbf{H}_{v,co})) \tag{6}$$

$$f_v = \tanh(p(\mathbf{H}_{v,co})) \tag{7}$$

$$\mathbf{H}_v = a_v \odot f_v \tag{8}$$

$$\mathbf{H}_{G-co} = m(\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_{|V|}) + \frac{1}{|V|} \sum_{v \in V} \mathbf{H}_v \tag{9}$$

式中: o, p 为 2 个多层感知机; m 为最大池化操作; a_v 为节点 v 的注意力权重; f_v 为通过非线性特征变换后的节点 v 的嵌入表示; $\mathbf{H}_v$ 为节点 v 的加权词向量。

式(6)阐述了采用软注意力机制衡量节点 $\mathbf{H}_{v,co}$ 对当前文本的重要性。在计算文本特征时, 使用加权平均词特征, 而不是简单的平均词特征, 能够关注对文本重要的单词, 学习更精准的文本表示。此外使用最大池化操作能够捕捉到节点集的不同方面。最终得到共现图的图表示 $\mathbf{H}_{G-co}$ 。同理, 可求得句法依存图的图表示 $\mathbf{H}_{G-dp}$ 。

其次, 采用拼接方式融合双图文本特征, 文本的最终表示为

$$\mathbf{H}_G = \mathbf{H}_{G-co} \parallel \mathbf{H}_{G-dp} \tag{10}$$

最后, 通过全连接层和softmax函数对文本表

示 $\mathbf{H}_G$ 进行分类。损失函数用于量化模型预测标签和真实标签之间的差异^[22]。本文使用交叉熵函数作为损失函数。通过反向传播算法对模型进行训练以最小化损失函数。

$$\hat{y} = \text{softmax}(\mathbf{W}_G \mathbf{H}_G + \mathbf{b}_G) \tag{11}$$

$$L = - \sum_i g_i \log \hat{y}_i \tag{12}$$

式中: $\mathbf{W}_G, \mathbf{b}_G$ 为权重和偏差; $\hat{y}$ 为预测的标签; g_i 为第 i 个文本真实标签的 one-hot 表示; L 为交叉熵损失。

2 实验结果和分析

为了验证上述模型的有效性, 在 4 个基准数据集上评估 DGNN-B 模型, 并通过消融实验验证各模块对文本分类的作用。

2.1 数据集

为了保持一致性, 本研究所使用的数据集包括 MR, Ohsumed, R8, R52^[6]。MR 是典型的情感二分类任务, 将电影评论分为积极或消极; Ohsumed 选自 MEDLINE 数据库, 本文仅关注单标签文本分类, 因此, 只使用其中 7400 篇摘要, 进行 23 种疾病分类; R8 和 R52 是 Reuters 数据集的子集, 分别有 8 个类别和 52 个类别。数据集的信息如表 1 所示。

表 1 数据集信息汇总

Tab.1 Summary of datasets information

数据集	类别	训练集	测试集	文本数	平均长度
MR	2	7 108	3 554	10 662	20.39
Ohsumed	23	3 357	4 043	7 400	135.82
R8	8	5 485	2 189	7 674	65.72
R52	52	6 532	2 568	9 100	69.82

2.2 基准方法

为了评估本文模型的有效性, 选取多个经典文本分类模型与本文模型进行对比分析。基于开源代码重现 TextING 的实验结果, 其他模型的结果取自文献 [6, 10, 11, 13]。现介绍本文选取的经典文本分类模型。

a. CNN: 在文本序列上执行卷积操作和最大池化操作, 得到文本的嵌入表示。

b. BiLSTM: 由前向 LSTM 与后向 LSTM 组合而成。

c. TextGCN: 将文档和单词作为节点, 构建一

张异构图, 利用图卷积网络捕捉文本特征^[6]。

d. TensorGCN: 提出张量图卷积网络, 构建基于语义的图、基于句法的图和基于序列的图。使用图内传播、图间传播策略协调和整合多个图的异构信息^[10]。

e. Text-level: 为文本独立构建图, 图中参数取自全局共享矩阵, 使用消息传递机制学习文本的图表示^[13]。

f. TextING: 提出基于 GGNN 的文本分类模型, 为每个文本构建单词共现图, 实现新单词的归纳学习^[14]。

g. GFN: 提出 4 种不同文本图的构建方式, 然

后利用多头融合模块集成文本表示^[11]。

2.3 实验参数

针对上述提及的数据集，以9:1的比例将训练集划分为真实的训练数据和验证数据。实验中，通过 Glove 初始化词向量时，将嵌入维度设置为 300。BiLSTM 的嵌入维度设置为 100。采用 Adam 优化器，学习率设置为 0.01。为防止过拟合，将随机激活 dropout 方法的节点保留率设置为 0.5。

表 2 多模型准确率对比
Tab.2 Multi-model accuracy comparison

模型	MR	Ohsumed	R8	R52
CNN	0.777 5	0.584 4	0.957 1	0.875 9
BiLSTM	0.776 8	0.492 7	0.963 1	0.905 4
TEXTGCN	0.767 4	0.694 0	0.970 7	0.935 6
TensorGCN	0.779 1	0.701 1	0.980 4	0.950 5
Text-level	0.761 3	0.694 0	0.978 0	0.946 0
TextING	0.785 3	0.702 9	0.975 0	0.953 3
GFN	0.780 4	0.702 0	0.982 2	0.953 1
DGNN-B	0.797 1	0.706 8	0.978 5	0.956 2

相较于 TextGCN 和 TensorGCN，本文利用 DGNN-B 为每个文本独立构图，实现对新样本的归纳式分类。因此，当有新样本加入时，无需重新训练整个模型。DGNN-B 在 4 个数据集上的准确率相较于 TextGCN 分别提升了 2.97%，0.87%，0.63%，1.6%。这是因为 TextGCN 关注文本的全局特征，忽略了文本中的重要信息如细粒度词交互及上下文语义特征，而 DGNN-B 为文本独立构图，关注文本自身的结构特征，同时利用 BiLSTM 学习上下文语义特征，保留了词序信息。在 MR 数据集上，DGNN-B 相较于 TextGCN 模型性能提高幅度最大。由于 MR 的文本长度较短，导致 TextGCN 构建的图密度较低，这限制了文本节点之间的标签信息传递，与之不同的是本文的文本图专注于自身结构，不依赖于这种标签信息传递机制。

与使用单个文本图方法相比，DGNN-B 的准确率有所提升。这是因为该模型构建双图，从多角度捕捉文本信息；同时使用 BiLSTM 捕捉文本的双向上下文语义信息，弥补图神经网络忽略文本词序特征的不足。实验结果表明，双图特征和语义特征对文本分类具有重要作用，这与 GFN^[11]的研究一致。相较于 GFN，DGNN-B 缺少了对全

2.4 实验结果和分析

在 4 个数据集上，对比分析 DGNN-B 模型与基准模型的分类性能，结果如表 2 所示。

与传统的 CNN 和 BiLSTM 模型相比，本文模型性能显著。实验结果表明，将文本构建成图结构，能较好地挖掘出文本中隐藏的结构信息。依据单词的共现关系建立边，通过共同邻居实现信息传播，可以捕捉远距离词特征。

局特征的捕捉，但亦取得了不错的性能。综合上述分析，相较于其他模型，DGNN-B 模型在文本分类中具有优越性。

2.5 消融实验

现通过消融实验验证双图机制和 BiLSTM 对文本分类的有效性。DGNN-B 以及移除不同组件的准确率如表 3 所示。

- a. CoGraph：基于共现图的文本分类模型。
- b. DpGraph：基于句法依存图的文本分类模型。
- c. BiLSTM(w/o)：融合共现图和句法依存图的文本分类模型。
- d. CoGraph(w/o)：融合句法依存图和 BiLSTM 的文本分类模型。
- e. DpGraph(w/o)：融合共现图和 BiLSTM 的文本分类模型。

为了验证双图机制在多个维度上捕捉文本信息的能力，构建了单图模型 CoGraph，DpGraph，CoGraph(w/o)，DpGraph(w/o) 和双图神经网络融合模型 BiLSTM(w/o)，DGNN-B。如表 3 所示，双图模型在文本分类任务上优于各自的基础模型。DGNN-B 模型与单图模型 CoGraph(w/o)，DpGraph(w/o) 相比，在 4 个数据集上的准确率平均提高了 0.27%，

表 3 消融实验结果

Tab.3 Results of ablation experiments

模型	MR	Ohsumed	R8	R52
CoGraph	0.785 3	0.702 9	0.975 0	0.953 3
DpGraph	0.789 2	0.643 8	0.971 7	0.945 9
BiLSTM(w/o)	0.792 0	0.703 9	0.975 6	0.954 7
CoGraph(w/o)	0.795 8	0.647 8	0.976 7	0.949 3
DpGraph(w/o)	0.793 1	0.705 4	0.975 9	0.954 5
DGNN-B	0.797 1	0.706 8	0.978 5	0.956 2

3.02%, 0.22%, 0.43%。实验结果表明, 共现图特征和句法依存图特征相互补充, 丰富文本信息。

如表 3 所示, 移除共现图模块或句法依存图模块时, DGNN-B 模型的性能不同。根据数据分析得出结论, 2 个模块在不同的数据集上扮演的角色不同。共现图模块在 Ohsumed, R52 数据集上表现优于句法依存图模块, 而在 MR, R8 数据集上相反。因为, MR 数据集文本长度较短且句式简单、句法结构明显, 句法依存图可以有效地捕捉文本句法特征。Ohsumed 文本较长, 通过单词共现图可以将不同句子的单词通过共同邻居连接, 使用 GGNN 网络可以捕捉跨句上下文概念, 获得更加丰富的表示。CoGraph(w/o) 模型在 Ohsumed 数据集上的性能与其他模型差异较大。通过实验验证, CoGraph(w/o) 模型表现不佳的原因与其使用的图特征提取器 GCN 有关。因此, 可以进一步从可解释性方面探索该模型。

为了验证 BiLSTM 对文本词序信息的学习能力, 对比移除 BiLSTM 的模型 CoGraph, DpGraph, BiLSTM(w/o) 与未移除 BiLSTM 的模型 DpGraph(w/o), CoGraph(w/o), DGNN-B。如表 3 所示, 使用 BiLSTM 模型在文本分类任务上优于各自的基础模型。与 BiLSTM(w/o) 模型相比, DGNN-B 在

MR, Ohsumed, R8, R52 数据集上的准确率分别提升 0.51%, 0.29%, 0.29%, 0.15%。因为, 文本理解的关键是语义表示, 在一定“作用域”中的词语对表达同一主题具有一定的共性, BiLSTM 可以有效地捕捉文本词序和双向上下文信息。MR 数据集是一种典型的短文本情感数据集, 文本词序不同, 表达的含义就会相差甚远, 从而影响文本情感分析的准确度。Ohsumed, R8, R52 的文本较长, BiLSTM 可以有效地捕捉双向上下文语义。实验结果表明, 通过 BiLSTM 使得 DGNN-B 模型更好地保留文本词序, 捕捉上下文语义信息。

除了将准确率作为评价指标之外, 在 MR, R8 数据集上进一步将精确率、召回率、调和平均值作为衡量模型优劣的评价指标。模型的多指标性能对比如图 4 所示。实验结果表明, DGNN-B 模型在多个指标上都取得了较优的效果。在 MR 数据集上, DGNN-B 相较于其他模型, 精确率、召回率和调和平均值平均提高了 0.59%, 0.56%, 0.56%。调和平均值是精确率和召回率的调和平均。调和平均值越高, 表示模型的质量越好。DGNN-B 的调和平均值相较于其他模型偏大, 进一步验证了 DGNN-B 模型的有效性。

经过分析可知, DGNN-B 模型的优势主要表

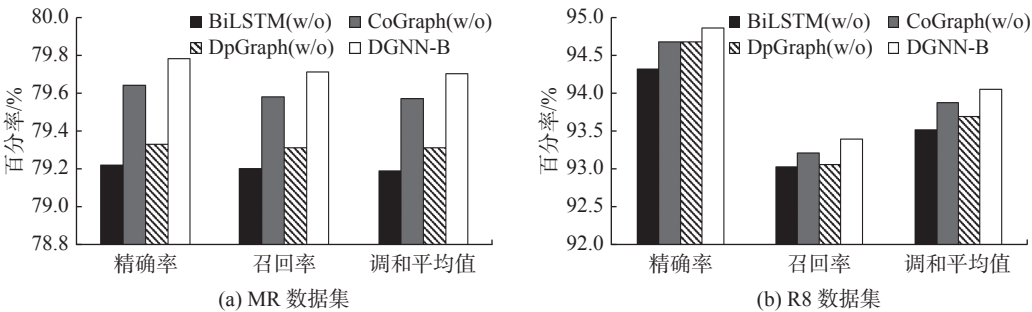


图 4 多指标实验性能评价

Fig.4 Multi-index experimental performance evaluation

现在以下两个方面：通过构建双图，该模型能够捕获多维度文本信息；利用 BiLSTM 保留文本词序，提升模型捕捉上下文信息的能力。

2.6 参数分析

在 4 个数据集上，GGNN 层数、GCN 层数对模型性能的影响如图 5 和图 6 所示。通过堆叠图神经网络层数，模型能够接收高阶邻居的特征信息，进而获得更精确的表示。堆叠层数过多会导致过平滑现象，即不同节点的特征趋于同质化，使得节点难以区分，从而影响模型性能。因此，

在 4 个数据集上，随着层数的增加，模型的准确率呈现先增大而后减小的趋势。由于 R8, R52 数据集文本分类较为简单，因此，GCN 层数设置为 1 便能较好地捕获文本特征。

滑动窗口的尺寸对模型性能的影响如图 7 所示。太小的滑动窗口不能反映单词间的联系，太大的滑动窗口导致不必要的节点连接。适当地选择窗口尺寸可以有效地传播局部语义信息。实验结果表明，在 MR, Ohsumed, R8, R52 数据集上，最优的滑动窗口大小为 3, 3, 7, 7。

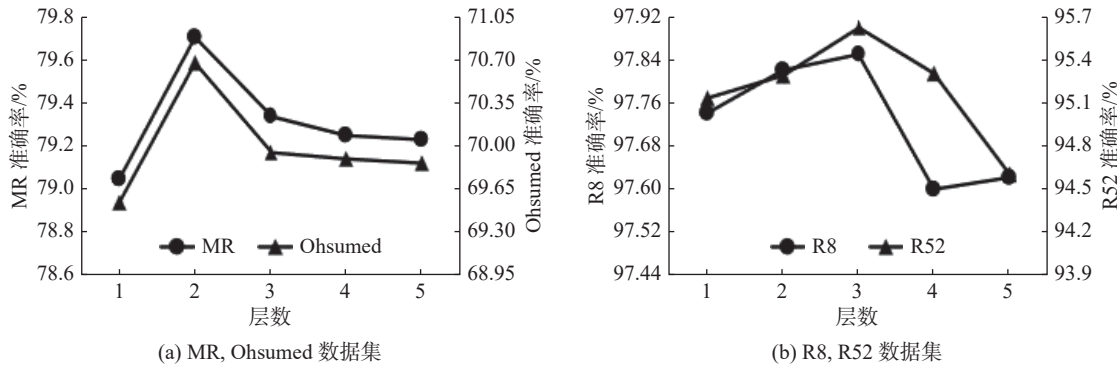


图 5 不同 GGNN 层数对应的准确率
Fig.5 Accuracy with different GGNN layers

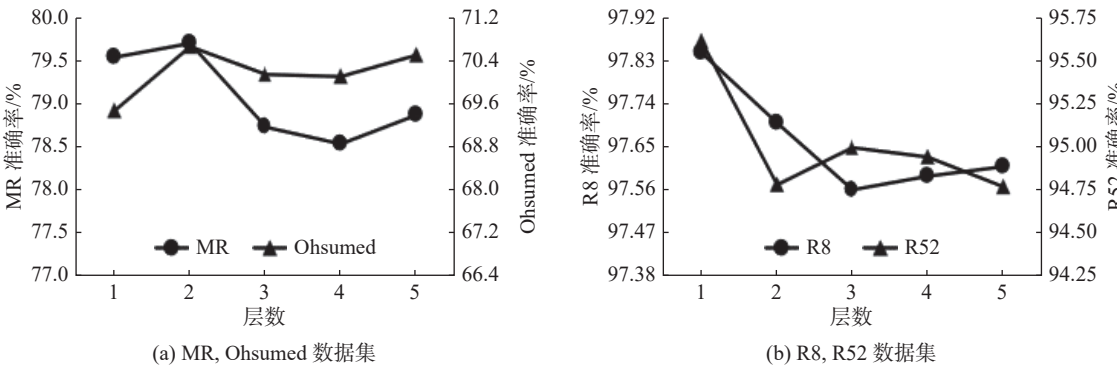


图 6 不同 GCN 层数对应的准确率
Fig.6 Accuracy with different GCN layers

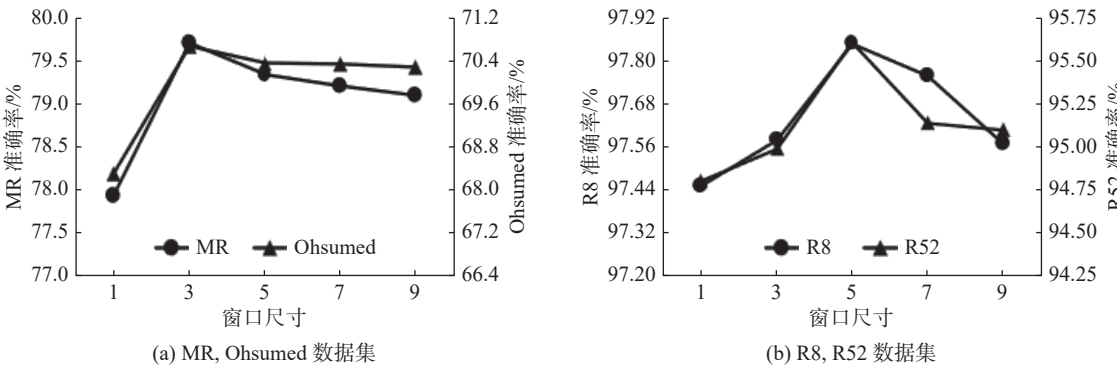


图 7 不同滑动窗口尺寸对应的准确率
Fig.7 Accuracy with different sliding window sizes

3 结束语

提出了融合 BiLSTM 的双图神经网络模型 (DGNN-B) 用于文本分类。该模型引入 BiLSTM 捕捉上下文语义信息并保留文本词序特征。同时为文本构建共现图、句法依存图, 有效地利用文本的多维度信息。在 4 个数据集上的实验结果表明, DGNN-B 的性能优于许多先进方法。同时通过消融实验, 验证不同的文本图是互补的, 以及 BiLSTM 捕捉的语义信息可以进一步提高模型性能。本文不足之处在于构建的是静态文本图, 未来可以探索构建动态文本图, 灵活地捕捉文本特征, 以及进一步研究在整个语料库上构建文本图并实现归纳式文本分类。

参考文献:

- [1] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: Association for Computational Linguistics, 2017: 562–570.
- [2] XU C, HUANG W R, WANG H W, et al. Modeling local dependence in natural language with multi-channel recurrent neural networks[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence and 31st Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019: 677.
- [3] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019: 4171–4186.
- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing System. Long Beach: Curran Associates Inc., 2017: 6000–6010.
- [5] SI Y H, ZHOU Y C. Deep graph neural networks for text classification task[C]//Proceedings of the 7th International Conference on Cyber Security and Information Engineering. Brisbane: Association for Computing Machinery, 2022: 272–275.
- [6] 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: AAAI Press, 2019: 7370–7377.
- [7] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview. net, 2017: 266–285.
- [8] HU L M, YANG T C, SHI C, et al. Heterogeneous graph attention networks for semi-supervised short text classification[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics, 2019: 4821–4830.
- [9] XIN Y, XU L L, GUO J L, et al. Label incorporated graph neural networks for text classification[C]//Proceedings of the 25th International Conference on Pattern Recognition. Milan: IEEE, 2021: 8892–8898.
- [10] LIU X E, YOU X X, ZHANG X, et al. Tensor graph convolutional networks for text classification[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020, 34: 8409–8416.
- [11] DAI Y, SHOU L J, GONG M, et al. Graph fusion network for text classification[J]. [Knowledge-Based Systems](#), 2022, 236: 107659.
- [12] WANG K Z, HAN S C, POON J. InducT-GCN: Inductive graph convolutional networks for text classification[C]//Proceedings of 2022 26th International Conference on Pattern Recognition. Montreal: IEEE, 2022: 1243–1249.
- [13] HUANG L Z, MA D H, LI S J, et al. Text level graph neural network for text classification[C]//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: Association for Computational Linguistics, 2019: 3444–3450.
- [14] 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: Association for Computational Linguistics, 2020: 334–339.
- [15] BECK D, HAFFARI G, COHN T. Graph-to-sequence learning using gated graph neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne: Association for Computational Linguistics, 2018: 273–283.

- machine learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc. , 2017: 4151–4161.
- [14] DOZAT T. Incorporating nesterov momentum into Adam[C]//The 4th International Conference on Learning Representations. San Juan: Open Review, 2016: 1–4.
- [15] ILYA L, FRANK H. Fixing weight decay regularization in Adam[C]//The 6th International Conference on Learning Representations. Vancouver: Open Review, 2018: 1–14.
- [16] LOSHCILOV I, HUTTER F. Decoupled weight decay regularization[C]//The 7th International Conference on Learning Representation. New Orleans: ICLR, 2019: 1–4.
- [17] QURESHI M N, UMAR M S, SHAHAB S. A transfer-learning-based novel convolution neural network for melanoma classification[J]. *Computers*, 2022, 11(5): 64.
- [18] KESKAR N S, SOCHER R. Improving generalization performance by switching from Adam to SGD[DB/OL]. [2017-12-20]. <https://arXiv preprint arXiv: 1712.07628>.
- [19] ROBBINS H, MONRO S. A stochastic approximation method[J]. *The Annals of Mathematical Statistics*, 1951, 22(3): 400–407.
- [20] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning[C]//Proceedings of the 30th International Conference on International Conference on Machine Learning. Atlanta: JMLR. org, 2013: 1139–1147.
- [21] MARTENS J. Deep learning via hessian-free optimization [C]//Proceedings of the 27th International Conference on Machine Learning. Haifa: ICML, 2010: 735–742.
- (编辑: 石 瑛)

(编辑: 石 瑛)

[上接第 111 页]

- [16] LI R F, CHEN H, FENG F X, et al. Dual graph convolutional networks for aspect-based sentiment analysis[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Stroudsburg: ACL, 2021: 6319–6329.
- [17] WEI X D, HUANG H, MA L X, et al. Recurrent graph neural networks for text classification[C]//Proceedings of the IEEE 11th International Conference on Software Engineering and Service Science. New York: IEEE, 2020: 91–97.
- [18] 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.
- [19] 范国凤, 刘璟, 姚绍文, 等. 基于语义依存分析的图网络文本分类模型 [J]. 计算机应用研究, 2020, 37(12): 3594–3598.
- [20] 邵党国, 张潮, 黄初升, 等. 结合 ONLSTM-GCN 和注意力机制的中文评论分类模型 [J]. 小型微型计算机系统, 2021, 42(7): 1377–1381.
- [21] QI P, DOZAT T, ZHANG Y H, et al. Universal dependency parsing from scratch[C]//Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. Brussels: Association for Computational Linguistics, 2018: 160–170.
- [22] 闫育铭, 李峰, 罗德名, 等. 基于深度迁移学习的糖尿病视网膜病变的检测 [J]. 光学仪器, 2020, 42(5): 33–42.
- (编辑: 石 瑛)

(编辑: 石 瑛)