

DOI: 10.3785/j.issn.1008-973X.2022.02.013

基于知识增强的图卷积神经网络的文本分类

王婷, 朱小飞, 唐顾
(重庆理工大学 计算机科学与工程学院, 重庆 400054)

摘 要: 针对文本分类问题, 提出新的基于知识增强的图卷积神经网络 (KEGCN) 分类模型. KEGCN 模型在整个文本集上构建了一个包含单词节点、文档节点、外部实体节点的文本图, 不同类型节点之间使用不同的相似性计算方法; 在文本图构建完成后将其输入到 2 层图卷积网络中学习节点的表示并进行分类. KEGCN 模型引入外部知识进行构图, 捕获长距离不连续的全局语义信息, 是第 1 个将知识信息引入图卷积网络进行分类任务的工作. 在 4 个大规模真实数据集 20NG、OHSUMED、R52、R8 上进行文本分类实验, 结果表明, KEGCN 模型分类准确率优于所有的基线模型. 将知识信息融入图卷积神经网络有利于学习到更精准的文本表示, 提高文本分类的准确率.

关键词: 知识嵌入; 图卷积网络; 神经网络; 文本分类; 自然语言处理

中图分类号: TP 391.1 文献标志码: A 文章编号: 1008-973X(2022)02-0322-07

Knowledge-enhanced graph convolutional neural networks for text classification

WANG Ting, ZHU Xiao-fei, TANG Gu

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: A new knowledge-enhanced graph convolutional neural network (KEGCN) classification model was proposed aiming at the problem of text classification. In the KEGCN model, firstly a text graph containing word nodes, document nodes, and external entity nodes was constructed on the entire text set. Different similarity calculation methods were used between different types of nodes. After the text graph was constructed, it was input into the two-layer graph convolutional network to learn the representation of the node and classified. The KEGCN model introduced external knowledge to compose the graph, and captured the long-distance discontinuous global semantic information, and was the first work to introduce knowledge information into the graph convolution network for classification tasks. Text classification experiments were conducted on four large-scale real data sets, 20NG, OHSUMED, R52 and R8, and results showed that the classification accuracy of the KEGCN network model was better than that of all baseline models. Results show that integrating knowledge information into the graph convolutional neural network is conducive to learning more accurate text representations and improving the accuracy of text classification.

Key words: knowledge embedding; graph convolutional network; neural network; text classification; natural language processing

文本分类是自然语言处理领域的基础性任务, 其目的在于将一段文本划分为预定义类别, 在情感分析、语言推理、主题分类、垃圾邮件检测、新闻过滤等领域有着广泛的应用. 目前, 绝

大多数的方法使用卷积神经网络结构 (convolutional neural networks, CNN) 和递归神经网络 (recurrent neural network, RNN) 进行文本分类, 但是现实生活中很多数据比如知识图谱、社交网络、交通

收稿日期: 2021-07-12. 网址: www.zjujournals.com/eng/article/2022/1008-973X/202202013.shtml

基金项目: 国家自然科学基金资助项目 (62141201); 重庆市技术创新与应用发展专项项目 (cstc2020jscx-dxwtBX0014); 重庆市教委语言文字科研资助项目 (yyk20103).

作者简介: 王婷 (1997—), 女, 硕士生, 从事文本分类研究. orcid.org/0000-0003-0318-141X. E-mail: ctingwang2021@163.com
通信联系人: 朱小飞, 男, 教授. orcid.org/0000-0001-8239-7176. E-mail: zxf@cqu.edu.cn

网络等并不是有规则的序列,因此图卷积神经网络(graph convolutional networks, GCN)应运而生.图卷积神经网络的本质是一个特征提取器,它将文本内容视为图结构数据,将文本中的每个单词看作一个节点,单词和单词之间存在关系即构边,进而构建成为一个图.图数据具有2个基本特性:1)每个节点都有自己的特征信息;2)节点和节点之间还存在结构信息.

虽然通过将文本构建成图的结构可以学到更多的信息,但是以往的研究工作^[1]只关注于捕获文本内部短距离连续的语义关系或者仅仅使用外部知识库来计算词与词之间的语义相似度,而忽略了外部知识在文本分类任务上更深层次的应用.其次,以往的研究工作仅考虑对词级上下文信息进行建模,而忽略了文档级和实体级的信息.

为了解决上述问题,提出新的用于文本分类的方法——知识增强的图卷积神经网络(knowledge-enhanced graph convolutional neural network, KEGCN).使用外部实体链接工具 Tagme 将文本中的实体链接至外部知识库中并得到其实体的表示,然后从整个语料库中构建一个包含3种节点(词级、文档级、实体级)、4种边(单词-单词、实体-单词、文档-单词、文档-实体)的大型图.本研究使用GCN对图进行建模,利用词共现构建词与词之间的边及词与实体之间的边,利用词频和词语的文档频率建立词节点和文档节点之间的边.因而,文本分类问题就转化为图节点分类问题.综上所述,主要创新点如下:1)提出新的应用于文本分类的知识增强图神经网络模型,使用外部知识库中的实体表示,作为知识层级表示,结合语义层面的单词和文档节点进行构图.2)所提出的KEGCN模型更好地利用了文本的分层结构信息,同时捕获了全局的语义信息和外部知识信息.

1 相关工作

1.1 文本分类任务

文本分类是自然语言处理中的经典问题,在现实中有广泛的应用.传统的文本分类模型如词袋模型^[2]、n-gram^[3]、支持向量机等都是使用特征工程和浅层算法.随着深度学习技术的发展,越来越多的深度学习模型被应用到文本分类中, Kim^[4]和 Liu等^[5]提出将CNN和RNN运用到文本分类任务中并取得了不错的表现.近年来随着GNN的

发展,越来越多的研究人员将图运用到分类任务中, Yao等^[1]提出的Text-GCN是目前最新的研究成果,本研究基于Text-GCN进行改进.

1.2 传统的文本分类方法

1) K最近邻算法(K-nearest neighbors, KNN)^[6]. KNN通过在 k 个最接近的有标记样本上找到样本最多的类别来对一个未标记样本进行分类. KNN是简单的分类器,但是使用KNN处理大规模数据集耗时较长.

2) 决策树(decision trees, DT). 决策树是有监督的树结构归纳分类算法,是用样本的属性作为节点,用属性的取值作为分支的树结构.常用的决策树算法有ID5^[7]、CART^[8]、C4.5^[9]等,基于CART、C4.5可以获取从属性到类的映射,从而有效地将未知的实体分类为新类别.此外,为了降低DT算法的计算代价, Vateekul等^[10]提出快速决策树(FDT),即预先选择一个特征集,并在不同的数据子集上训练多个DTs.通过数据融合技术将多个DTs的结果进行组合,解决类不平衡的问题.

1.3 基于深度学习的文本分类方法

1) 递归神经网络RNN^[11]. 基于RNN的模型通过获取序列信息来学习输入词之间的依赖性. Hochreiter等^[12]使用RNN的一种特定类型——LSTM来学习文本表示.这些方法虽然有效,但是仅考虑了局部的词序列信息.

2) 卷积神经网络CNN. 基于CNN的模型从卷积核中提取相关特征.为了将CNN用于文本分类任务, Kim^[4]提出TextCNN. Johnson等^[13]提出基于半监督学习的文本分类CNN模型, Zhao等^[14]在TextCNN的基础上提出ATT-CNN.此外, Xue等^[15]提出MTNA,该方法组合了BiLSTM和CNN层用于解决细粒度的文本分类任务.

1.4 图神经网络

图神经网络(graph neural network, GNN)^[16]近年来受到了广泛的关注.传统的神经网络^[17]只能对结构化的数据进行建模,而现实生活中很多数据比如社交网络^[18]、知识图^[19]都是树形或者图谱的结构.因此, GNN^[20]被提出用以处理非结构化数据, GNN是直接作用于图结构上的神经网络. GNN模型基于信息传播机制,每一个节点通过相互交换信息来更新自己的节点状态,直到达到某一个稳定值. GCN是引入卷积的GNN,通过提取空间特征来学习节点的表示,其典型的应用有节点分类、图分类、交通流量预测、用户偏好推荐等.实质上图中的每一个节点都与一个标签相关联,

GCN 通过其他已知标签信息的节点来预测未标记节点的标签, 从而进行分类。

2 KEGCN 模型

使用图神经网络最大的挑战在于须输入 GCN 可用的格式的数据. 须将文本数据转换为图的形式, 因此须巧妙地输入图形中的单词、文档及外部实体之间的交互进行建模. 本研究所提出的模型

如图 1 所示. 通过将单词表示为顶点, 词共现表示为边进行构图, 记为 $G=(V, E)$, 其中 V 表示节点, E 表示边. 图中有 3 类节点, 分别为单词节点、文档节点和实体节点, 分别以 W 、 T 、 E 作为符号表示. 在经过一个标准化的数据处理后 (包括对标点符号的处理及去除停用词), 使用 one-hot 对单词、文档及实体进行编码. 如图 1(a) 所示为模型输入时所构成的初始图, 如图 1(b) 所示为经过图卷积神经网络聚合节点特征信息后, 文本分类的结果。

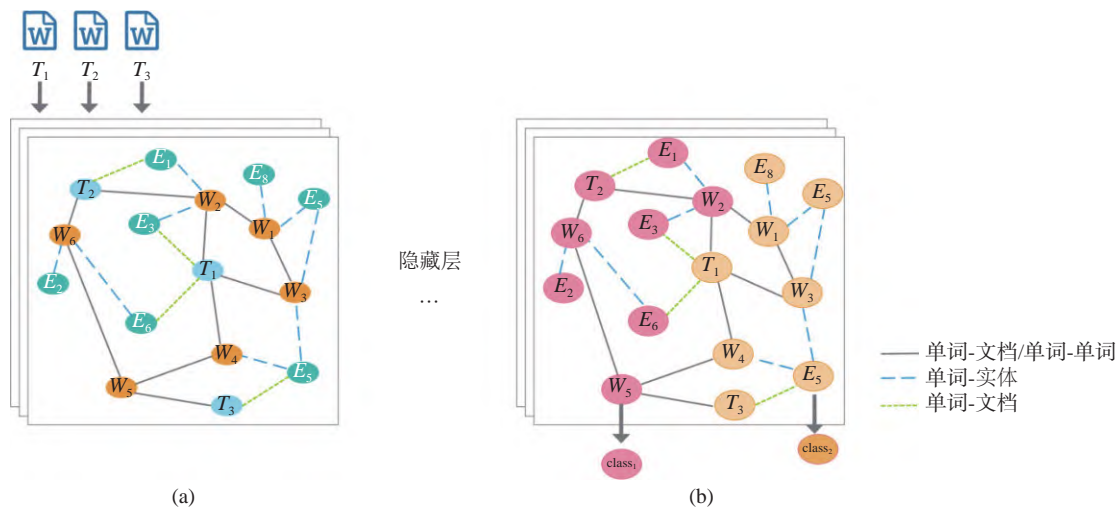


图 1 KEGCN 结构图

Fig.1 Illustration of construction of KEGCN

2.1 实体链接

实体链接 (entity linking) 就是将文档中出现的实体关联到外部知识库中的相关实体的任务. 详细过程如下: 首先提取文档中所有的实体, 其次使用实体链接工具 Tagme 将提取到的实体链接到外部知识库中, 找到候选实体并进行消歧, 最终得到与文档中实体对应的外部知识表示. 将所得的外部知识表示作为实体节点的嵌入表示。

2.2 构图

根据文档中的词共现 (文档-单词边), 整个语料库中的词共现 (单词-单词边) 及外部知识库中的词共现 (单词-实体边/文档-实体边) 在各节点之间构建边. 对于文档-单词和文档-实体的边的权重, 采用 BM25 算法进行计算. BM25 是一种用于计算词和文档相关性的算法, 计算公式如下:

$$\text{BM25} = \sum \ln \left(\frac{|D|}{|j: W_i \in D_j|} \right) \frac{f_i(K_1 + 1)}{f_i + K}, \quad (1)$$

$$K = K_1 (1 - b + b D_L / \bar{D}_L). \quad (2)$$

式中: $|D|$ 为文档总数, $|j: W_i \in D_j|$ 表示包含词 W_i 的

文档数; K_1 、 b 为调协因子, 分别设为 2、0.75; f_i 为词 W_i 在文档中出现的次数; D_L 为文档的长度; \bar{D}_L 为所有文档的平均长度。

为了利用全局词共现信息, 使用点互信息算法 (pointwise mutual information, PMI) 来计算单词-单词节点和实体-单词节点之间的权重. PMI 是一种常用的词关联度量方法, 其基本思想是统计 2 个词语在文本中同时出现的概率, 概率越大, 其相关性就越紧密, 关联度越高. 计算公式如下:

$$\text{PMI}(w_1, w_2) = \lg \frac{P(w_1 \& w_2)}{P(w_1)P(w_2)}, \quad (3)$$

$$P(w_1) = \text{win}(w_1) / \text{ws}, \quad (4)$$

$$P(w_1 \& w_2) = \text{win}(w_1, w_2) / \text{ws}. \quad (5)$$

式中: ws 表示滑动窗口总数, $\text{win}(w_1, w_2)$ 表示同时包含词 w_1 和词 w_2 的滑动窗口个数, $\text{win}(w_1)$ 表示仅包含词 w_1 的滑动窗口个数。

综上, 当节点 i 和节点 j 表示单词或知识实体时, 其边的权值使用 PMI 计算; 当计算文档-单词或文档-知识实体节点所构边的权值时, 使用 BM25

算法;当节点构成自环时,权值设置为1.0;其他情况下权值为0.边的权值表达式如下:

$$A_{ij} = \begin{cases} \text{PMI}(i, j), & i, j \text{ 均为单词, } \text{PMI}(i, j) > 0; \\ \text{BM25}(i, j), & i \text{ 为文档, } j \text{ 为实体或单词}; \\ 1, & i = j; \\ 0, & \text{其他}. \end{cases} \quad (6)$$

此时,输入的文本数据已经构成图.接下来按照 Kipf 等^[21]的方法,将构建好的图输入到一个2层的GCN中,通过 softmax 分类器进行预测.为了最小化损失,使用交叉熵函数进行计算.表达式如下:

$$\mathbf{H}^{(l+1)} = \text{ReLU}(\mathbf{D}^{1/2} \mathbf{A} \mathbf{D}^{-1/2} \mathbf{H}^{(l)} \mathbf{W}^{(l)}), \quad (7)$$

$$\hat{y}_i = \text{softmax}(\mathbf{H}^{(l+1)}), \quad (8)$$

$$L = - \sum y_i \lg(\hat{y}_i). \quad (9)$$

式中: l 取0和1分别得到2层GCN的节点表示, $\mathbf{W}^{(l)}$ 表示第 l 层的权重矩阵, \mathbf{A} 表示邻接矩阵, \mathbf{D} 为 \mathbf{A} 的度矩阵, $\mathbf{H}^{(l)}$ 表示特征向量矩阵, y_i 为第 i 个节点真实标签, L 为损失函数.假设一共有 N 个节点,每个节点都有自己的特征,这些特征组成一个 $N \times D$ 维的矩阵 \mathbf{X} ,对于输入层而言, $\mathbf{H}^{(0)} = \mathbf{X}$.

3 实验结果与分析

3.1 数据集

在4个广泛使用的真实数据集上进行实验.数据集包括20Newsgroups(20NG)、OHSUMED、R52和R8.对于所有的数据集,使用和文献[5]相同的方法划分训练集和测试集,再从训练集中随机选取10%的文本来构建验证集.数据集的统计结果如表1所示.

表1 情感分析数据集统计
Tab.1 Sentiment analysis dataset statistics

数据集	训练集文档数	测试集文档数	类别数	实体数	平均长度
20NG	11314	7532	20	26607	221.26
OHSUMED	3357	4043	23	9075	135.82
R52	6532	2568	52	7476	69.82
R8	5485	2189	8	6440	65.72

1)20Newsgroups. 20NG收集了18846个新闻文档,均匀分为20个不同主题的新闻组集合.训练集中共有11314个文档,测试集中共有7532个文档.每个文档平均长度为221.26个词.

2)OHSUMED.包含348566个文档.本研究的工作仅仅关注单标签文本分类,因此去除多类别文档,最后剩下7400个文档,其中训练集有3357个文档,测试集有4043个文档.

3)R52和R8. R8有8个类别,其中训练集有5485个文档,测试集有2189个测试文档. R52有52个类别,其中训练集有6532个文档,测试集有2568个文档.

3.2 基线模型

将本研究所提模型和8种不同的基线模型进行比较,8种模型如下.传统深度学习方法:CNN^[4]和LSTM^[11].使用词语上下文信息的分类方法:fastText^[22]和SWEM^[23].基于图的文本分类方法:Text-GCN^[8]和HETE-GCN^[24]. 1)CNN: CNN-rand表示使用随机初始化的词向量, CNN-non-static表示使用预训练的词向量进行初始化. 2)Bi-LSTM: 双向LSTM,使用训练好的词向量作为输入. 3)fast-Text: 简单高效的文本分类方法,将词向量的平均值输入到分类器中进行分类. 4)SWEM: 简单的词嵌入模型,对词向量进行池化操作. 5)Text-GCN: 将文本构成图,使用图卷积网络学习和预测图中节点(单词、文档)的特征向量表示. 6)HETE-GCN: 通过采样技术和图卷积运算学习高质量的节点嵌入.

3.3 参数设置和评估指标

对于所有的数据集,均划分出了训练集和测试集.将卷积层的嵌入大小设置为150,窗口大小设为20.根据经验,将Adam优化器^[25]的学习率设定为0.02, dropout设置为0.5, L2损失设置为0.为了进行公平的比较,其他基线模型设置相同的参数.对于使用预先训练好的词嵌入的基线模型,采用300维的Glove词嵌入^[26].

所采用的评价指标为正确率(Acc). Acc表示预测正确的样本数占样本总数的百分比.

3.4 实验结果

在4个公开数据集上进行8组实验以验证KEGCN模型的有效性,各个模型的性能对比结果如表2所示.图中,黑体加粗为最优结果.可以看出,KEGCN的表现最佳,显著优于所有基线模型.根据实验结果不难发现基于图的方法通常优于其他类型的模型,表明对文本进行构图有利于文本分类任务.其中,KEGCN在20NG数据集上的效果提升最明显,与以往最优的模型效果相比,提高了1.68%.本研究的模型效果更佳主要有以下

表 2 KEGCN 模型在 4 个数据集上的分类准确性对比

Tab.2 Comparison of classification accuracy of KEGCN model on four datasets

模型	Acc			
	20NG	OHSUMED	R52	R8
CNN-rand	0.769 3±0.006 1	0.438 7±0.010 0	0.853 7±0.004 7	0.940 2±0.005 7
CNN-non-static	0.821 5±0.005 2	0.584 4±0.010 6	0.875 9±0.004 8	0.957 1±0.005 2
LSTM(pretrain)	0.754 3±0.017 2	0.511 0±0.015 0	0.904 8±0.008 6	0.960 9±0.001 9
Bi-LSTM	0.731 8±0.018 5	0.492 7±0.010 7	0.905 4±0.009 1	0.963 1±0.003 3
fastText	0.793 8±0.003 0	0.577 0±0.004 9	0.928 1±0.000 9	0.961 3±0.002 1
SWEM	0.851 6±0.002 9	0.631 2±0.005 5	0.929 4±0.002 4	0.953 2±0.002 6
Text-GCN	0.863 4±0.000 9	0.683 6±0.005 6	0.935 6±0.001 8	0.970 7±0.005 1
HETE-GCN	0.871 5±0.001 5	0.681 1±0.007 0	0.943 5±0.002 5	0.972 4±0.001 0
KEGCN	0.882 2±0.004 5	0.697 1±0.005 9	0.945 1±0.001 8	0.974 1±0.002 5

2 点原因: 1) 将外部实体信息纳入图中作为词节点, 为大量的文本数据提供了额外的信息, 并且把词节点的语义信息和知识信息相结合, 获得了更丰富的表示. 2) 采用文本的分层结构, 将词节点作为桥梁连接文档节点和外部实体节点, 使得标签信息传播到整个图中.

为了进一步验证模型的有效性, 测试 4 种不同

的模型在 OHSUMED 和 R52 数据集上的 Loss 和 Acc, 结果如图 2 所示. 图中, Epoch 为迭代次数. 1 个 Epoch 表示将数据集全部数据都训练一次. 由图中曲线的整体走向可以看出, 相较于传统的深度学习方法, KEGCN 模型具有更好的收敛性和鲁棒性, 因为和 LSTM 的曲线相比, 由本研究提出的方法所生成的曲线更加平滑与稳定. 此外, 从实

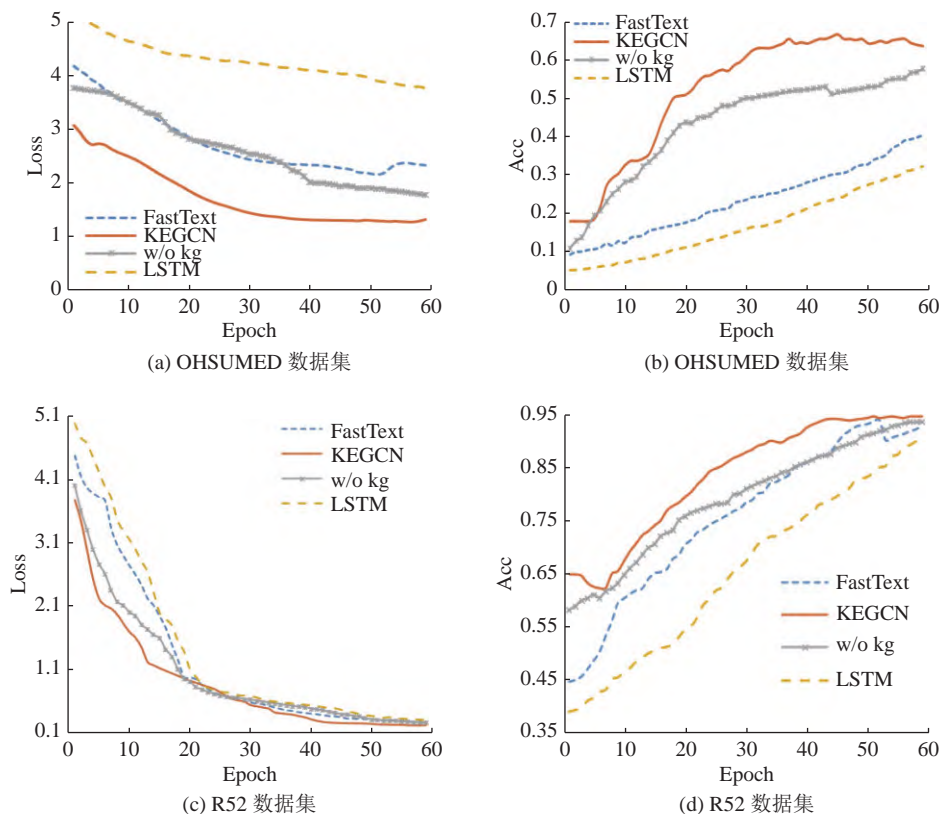


图 2 4 种模型在 OHSUMED 和 R52 数据集上损失和准确率的对比

Fig.2 Comparison of loss and accuracy of four models on OHSUMED and R52 datasets

验结果还可以直观地看出引入外部实体对于提升分类性能是有帮助的.

为了验证引入知识实体的有效性,将 KEGCN 模型和 KEGCN w/o knowledge 模型进行比较,在不同数据集上的准确率对比如图 3 所示.可以看出,KEGCN 除了在 R8 数据集上的提升相对较微弱以外,在其他数据集上的表现均不错.结合如表 1 所示的数据统计,认为表现不错的原因如下:R8 中的文档篇幅较短,数据较稀疏,导致构成的是低密度图,限制了标签消息在文档节点之间的传递.不过,词节点较少,导致链接到的实体数目也偏少,效果提升不明显.

除此之外,由于本研究使用和 Text-GCN 相同的数据处理方法,可以通过观察最终的实验结果,清楚地看出引入外部知识信息对于文本分类是有效果的.

为了分析 KEGCN 在不同特征向量维度 d 下的分类性能,在 OHSUMED 和 R8 数据集上进行对比实验,结果如图 4、5 所示.当特征向量维度较小时,模型并不能很好地将节点信息传播到整个图中,当嵌入维度过大时,性能会有一定程度下

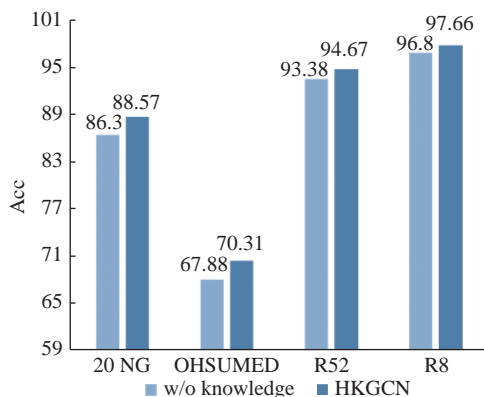


图 3 消融实验结果对比

Fig.3 Comparison of ablation experiment results

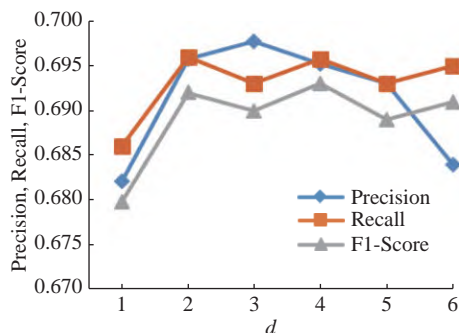


图 4 OHSUMED 数据集中不同向量维度下的性能对比

Fig.4 Performance comparison of different vector dimensions in OHSUMED dataset

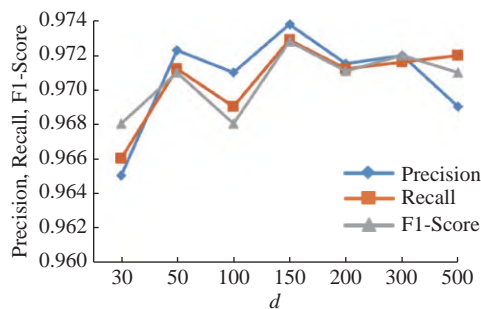


图 5 R8 数据集中不同特征向量维度下的性能对比

Fig.5 Performance comparison of different dimensions on R8 dataset

降,训练时间也会增加.可以看出,当特征向量维度设置为 150 时,效果最佳.

4 结 语

针对文本分类任务,为了提高分类结果的准确性,提出新的基于图的文本分类方法,该方法通过引入外部实体信息增强词节点表示并使用多层次结构进行构图.在不同的公开数据集上进行实验,结果表明,所提出的方法均取得了更优的分类结果,也验证了借助外部知识能够获得更为准确的文本表示.

虽然 KEGCN 模型在多个数据集上展现了良好的性能,但是本研究未对文本数据进行深层次的探究,同时未对图神经网络进行进一步的改进,未来考虑结合文本增强技术,设计带有注意力权重的图神经网络进行分类实验,并验证分析.

参考文献 (References):

- [1] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Honolulu: AAAI, 2019, 33(1): 7370-7377.
- [2] ZHANG Y, JIN R, ZHOU Z H. Understanding bag-of-words model: a statistical framework [J]. *International Journal of Machine Learning and Cybernetics*, 2010, 1: 43-52.
- [3] WANG S I, MANNING C D. Baselines and bigrams: simple, good sentiment and topic classification [C]// *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. Jeju Island: ACL, 2012: 90-94.
- [4] KIM Y. Convolutional neural networks for sentence classification [C]// *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha: EMNLP, 2014: 1746-1751.
- [5] 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**. New York: IJCAI, 2016: 2873–2879.
- [6] COVER T, HART P. Nearest neighbor pattern classification [J]. **IEEE Transactions on Information Theory**, 1967, 13(1): 21–27.
- [7] UTGOFF P E. ID5: an incremental ID3 [M]// **Machine Learning Proceedings 1988**. Ann Arbor: Morgan Kaufmann, 1988: 107 – 120.
- [8] LOH W Y. Classification and regression trees [J]. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, 2011, 1(1): 14–23.
- [9] QUINLAN J R. **C4.5: programs for machine learning** [M]. Massachusetts: Morgan Kaufmann Publishers, 1993.
- [10] VATEEKUL P, KUBAT M. Fast induction of multiple decision trees in text categorization from large scale, imbalanced, and multi-label data [C]// **2009 IEEE International Conference on Data Mining Workshops**. Miami, FL: IEEE, 2009: 320–325.
- [11] 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**. New York: IJCAI, 2016: 2873–2879.
- [12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. **Neural Computation**, 1997, 9(8): 1735–1780.
- [13] JOHNSON R, ZHANG T. Semi-supervised convolutional neural networks for text categorization via region embedding [J]. **Advances in Neural Information Processing Systems**, 2015, 28: 919–927.
- [14] ZHAO Z, WU Y. Attention-based convolutional neural networks for sentence classification [C]// **Proceeding of the 17th Annual Conference of the International Speech Communication Association**. San Francisco: INTER SPEECH, 2016: 705–709.
- [15] XUE W, ZHOU W, LI T, et al. MTNA: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews [C]// **Proceedings of the 8th International Joint Conference on Natural Language Processing**. Taipei: IJCNLP, 2017: 151–156.
- [16] ZHOU J, CUI G, HU S, et al. Graph neural networks: a review of methods and applications [J]. **AI Open**, 2020, 1: 57–81.
- [17] 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.
- [18] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs [C]// **Proceedings of the 31st International Conference on Neural Information Processing Systems**. Long Beach: NIPS, 2017: 1025–1035.
- [19] BI Z, ZHANG T, ZHOU P, et al. Knowledge transfer for out-of-knowledge-base entities: improving graph-neural-network-based embedding using convolutional layers [J]. **IEEE Access**, 2020, 8: 159039–159049.
- [20] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model [J]. **IEEE Transactions on Neural Networks**, 2008, 20(1): 61–80.
- [21] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks. [EB/OL]. (2017-02-22). <https://arxiv.org/abs/1609.02907>.
- [22] BOJANOWSKI P, GRAVE E, JOULIN A, et al. Bag of tricks for efficient text classification [C]// **Association for Computational Linguistics**. Valencia: ACL, 2017: 427–431.
- [23] SHEN D, WANG G, WANG W, et al. Baseline needs more love: on simple word-embedding-based models and associated pooling mechanisms [EB/OL]. (2018-05-24). <https://arxiv.org/abs/1805.09843>.
- [24] RAGESH R, SELLAMANICKAM S, IYER A, et al. Hetegen: heterogeneous graph convolutional networks for text classification [C]// **Proceedings of the 14th ACM International Conference on Web Search and Data Mining**. Queensland: WSDM, 2021: 860–868.
- [25] KINGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. (2014-12-22). <https://arxiv.org/abs/1412.6980v3>.
- [26] 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**. Doha: EMNLP, 2014: 1532–1543.