

融入句子中远距离词语依赖的图卷积短文本分类方法

张虎 柏萍

山西大学计算机与信息技术学院 太原 030006

摘要 随着图神经网络技术在自然语言处理领域中的广泛应用,基于图神经网络的文本分类研究受到了越来越多的关注,文本构图是图神经网络应用到文本分类中的一项重要研究任务,已有方法在构图时通常不能有效捕获句子中远距离词语的依赖关系。短文本分类是待分类文本中普遍较短的一类特殊文本分类任务,传统的文本表示通常比较稀疏且缺乏丰富的语义信息。基于此,文中提出了一种融入远距离词语依赖关系进行构图的图卷积短文本分类方法。首先结合词语共现关系、文档和词语之间的包含关系、远距离词语依赖关系为整个文本语料库构建一个文本图;然后将文本图输入到图卷积神经网络,通过2层卷积后,对每个文档节点进行类别预测。在 on_line_shopping_10_cats、中文论文摘要和酒店评论3个数据集上的实验结果表明,所提方法相比已有基线模型取得了更好的效果。

关键词:短文本分类;图卷积神经网络;句法关系;文本构图;自然语言处理

中图法分类号 TP391

Graph Convolutional Networks with Long-distance Words Dependency in Sentences for Short Text Classification

ZHANG Hu and BAI Ping

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

Abstract With the wide application of graph neural network technology in the field of natural language processing, the research of text classification based on graph neural networks has received more and more attention, Building graph for text is an important research task in the application of graph neural networks for text classification. Existing methods cannot effectively capture the dependency of long-distance words in sentences when building graph. Short text classification is a special type of text classification task in which the classified text is generally short, so the traditional text representation is usually sparse and lacks rich semantic information. Based on this, in this paper we propose a short text classification method based on graph convolutional neural networks incorporating long-distance words dependency. Firstly, by using the co-occurrence relationship of words, the containment relationship between documents and words, and the long-distance words dependency in sentences, a text graph is constructed for the entire text corpus. Then, the text graph is input into the graph convolutional neural networks, and the category label prediction is made for each document node after 2-layer convolution. The experimental results on the three datasets of online_shopping_10_cats, summaries of Chinese papers and hotel reviews show that the proposed method achieves better results than the existing baselines.

Keywords Short text classification, Graph convolutional neural network, Syntactic relationship, Building graph for text, Natural language processing

1 引言

文本分类是自然语言处理领域中的一项目基础任务,旨在使用文本中的特定信息将文本划分为不同的类别。目前文本分类技术在舆情分析、新闻分类、垃圾邮件过滤、抽取式问答

等方面已得到深入应用。

文本表示是文本分类中的一项重要研究内容,典型的文本表示模型有基于向量空间的模型、基于矩阵的模型和基于神经网络的模型。向量空间模型将每个文本表示成一个向量,向量中的元素是文本的特征,整个文档集合表示为高维

到稿日期:2020-12-07 返修日期:2021-05-08

基金项目:国家重点研发计划项目(2018YFB1005103);国家自然科学基金(62176145);山西省自然科学基金(201901D111028)

This work was supported by the National Key Research and Development Program of China(2018YFB1005103), National Natural Science Foundation of China(62176145) and Natural Science Foundation of Shanxi Province, China(201901D111028).

通信作者:张虎(zhanghu@sxu.edu.cn)

稀疏矩阵,此类方法的文本分类精度相对较差。基于矩阵的文本深层表示以“词-上下文”矩阵为核心,一般需构建一个“词-上下文”矩阵,并从矩阵中获取词的表示,但该方法无法有效解决文本中的一词多义问题,并且计算出的特征向量没有对应的物理解释。

随着深度学习技术的发展,基于神经网络的文本表示方法受到了广泛的关注,基于卷积神经网络(Convolutional Neural Networks, CNN)、循环神经网络(Recurrent Neural Networks, RNN)和图神经网络(Graph Neural Networks, GNN)的文本表示模型被相继提出。Kim^[1]提出了一个包含两个单词向量通道的简单CNN,允许同时使用动态更新和静态单词嵌入实现句子级分类任务,然而CNN一般只能表示局部窗口中上下文的语义信息。Zhou等^[2]尝试利用CNN提取短语表示,并将其输入到长短期记忆网络中获取句子表示,提出的模型能捕获短语的局部特征以及句子的局部语义。为了进一步提高模型表示的灵活性,Wang等^[3]将注意力机制引入文本分类中,但注意力机制主要关注局部连续词序列,不能显式地使用语料库中的全局词共现信息。

近年来,图神经网络^[4-7]的研究受到越来越多的关注,Defferrard等^[8]将卷积神经网络由低维规则网格归纳到高维不规则域,首次将图卷积神经网络应用于文本分类任务。Zhang等^[9]提出了一种用来编码句子的长短期记忆网络结构(Long Short-Term Memory, LSTM),该结构由每个单词的并行状态组成,其中的递归步骤用于在单词之间同时执行局部和全局信息交换。Wang等^[10]提出了改进算法模型RHS-CNN(Regularization Hierarchical Softmax CNN),采用正则化的方法,对H-Softmax的节点参数进行约束,避免过拟合,增强了模型的泛化能力。这些研究有效弥补了基于向量空间模型和矩阵模型的不足,但在进行文本表示时仍缺乏对句子中远距离词语的关系表示。短文本通常由几个到十几个词组成,其生成的文本表示会更稀疏,因此短文本的表示对句子中远距离词语依赖关系的刻画就更为重要。

针对上述问题,本文提出了融入句子中远距离词语依赖的图卷积短文本分类模型。首先以单词和文档作为图的节点,结合词语共现关系、文档和词语之间的包含关系以及远距离词语依赖关系,为整个文本语料库构建一个文本图。然后将文本图输入到图卷积神经网络,通过2层卷积后对文档节点进行类别预测,实现文本分类。

本文的主要贡献体现在以下两点:

(1)在已有基于图神经网络文本分类模型的基础上,引入依存句法关系到文本图的构建过程中,提出融入句子中远距离词语依赖的图卷积短文本分类模型(Syntactic Graph Convolution Network, SGCN)。

(2)同时在中文论文摘要数据集(Summaries of Chinese Papers, SCP)、online_shopping_10_cats(OS)、酒店评论数据集(Hotel Review Dataset, HR)上进行了多项对比实验,验证

了所提模型的有效性和通用性。

2 相关工作

传统的文本分类方法研究主要集中在特征工程和分类算法上,目前常见的用于提取文本特征的方法有one-hot和TF-IDF^[11]等,分类算法通常采用支持向量机(Support Vector Machine, SVM)^[12-13]、朴素贝叶斯^[14]等传统机器学习算法。传统的文本分类方法在处理短文本时,会造成更严重的特征稀疏,为了解决这个问题,国内外学者通常利用外部知识库或使用内部语义提取关联规则来扩展短文本特征。对于每个分类任务,Phan等^[15]都会收集一个大规模的外部数据集,然后在(少量)带标签的训练数据集和从该数据集中发现的丰富的隐藏主题集上建立分类器。然而,已有的特征扩展方法一般耗时较长,且不能捕获短文本的完整语义。

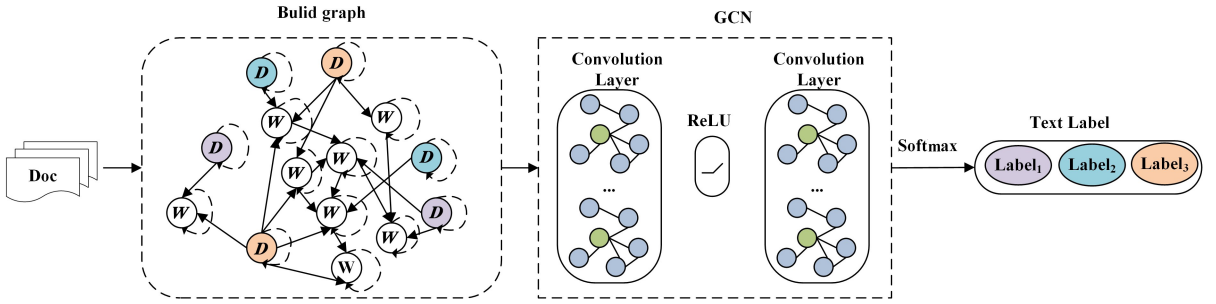
近年来,基于深度学习的文本分类方法在自然语言处理任务上的应用越来越广泛,深度学习方法相比传统文本分类方法最重要的一点就是能够解决文本表示问题,并且可以很好地捕获局部连续词序列中的语义,实现端到端地解决问题。Zhang等^[16]提出利用特征级卷积网络对中文新闻语料进行文本分类,通过挖掘文本中不同层次的信息来缓解稀疏性。Wang等^[17]使用一个词嵌入聚类卷积神经网络扩展语义,在一定程度上解决了对上下文语义的敏感性,使分类结果得到了一定程度的提高。尽管基于深度学习的文本分类方法相比传统的文本分类方法表现较好,但仍会忽略句子中非连续和长距离词语的依赖关系。

随着图神经网络的发展,基于图神经网络的方法也被用来解决文本分类问题。Yao等^[18]将文本分类问题转化为节点分类问题,首先根据整个文本语料库构建文档和单词之间具有全局关系的文本图,然后利用图卷积网络对文档节点分类。Huang等^[19]为每个文档单独建图,通过在文档上的滑动窗口来连接单词节点,更好地捕捉了局部特征并且大大减少了内存资源的消耗。Hu等^[20]提出了一种基于两层注意力机制的异构图注意力网络,用于短文本分类任务,异构信息网络可以集成各种类型的附加信息以及它们之间的关系。现有基于图神经网络的文本分类方法可以捕获文本中非连续实体的关系,但未充分考虑句子中远距离词语的依赖关系。

显然,已有的方法主要考虑了文本的局部特征和句子中连续词语的关系,对于句子中的远距离词语依赖关系表示不足。因此,探索融入句子中远距离词语的依赖关系的图卷积短文本分类方法具有重要的研究意义。

3 SGCN 模型描述

本文方法的模型结构如图1所示。首先对完整的文本数据集构图,然后将构建的图输入到GCN,在2层卷积后对文档节点进行类别预测。



注:以“D”开头的是文档节点,以“W”开头的是单词节点,虚线表示节点自连接,实线表示节点之间的边

图1 SGCN整体架构图

Fig.1 Overall architecture of SGCN

3.1 文本构图

本文将语料库中的单词和文档作为节点,并通过3种特定关系在特定节点之间建立连边来构建一个大型异构文本图 $G=(V,E)$,其中, $V(|V|=n)$ 和 E 分别是节点和边的集合。设 $X \in R^{n \times m}$ 为特征矩阵,其中, n 指的是节点的个数, m 为特征向量的维数, $x_v \in R^m$ 指的是节点 v 的特征向量,将特征矩阵设置为单位矩阵。边的3种关系分别为:词语共现关系以及文档与词语之间的包含关系、远距离词语依赖关系。

3.1.1 词语共现关系

本文对语料库中的所有文档使用固定大小的滑动窗口来统计词语共现信息,对存在共现关系的词语之间建立一条连边,并使用点互信息(Pointwise Mutual Information, PMI)计算两个单词之间的权重。因为当 PMI 值为正时,语料库中词语之间的语义相关性较高,当 PMI 值为负时,语料库中词语之间的语义相关性较低或不存在,所以本文只在 PMI 值为正的词语之间建立边。关于单词 a 和单词 b 之间权重值 PMI 的计算式如下:

$$PMI(a,b) = \log \frac{p(a,b)}{p(a)p(b)} \quad (1)$$

$$p(a,b) = \frac{N(a,b)}{N} \quad (2)$$

$$p(a) = \frac{N(a)}{N} \quad (3)$$

其中, $N(a,b)$ 是既包含单词 a 也包含单词 b 的滑动窗口的数量, $N(a)$ 是包含单词 a 的滑动窗口的数量, N 是整个语料库中包含的滑动窗口的总数。

3.1.2 文档与词语之间的包含关系

基于文档和词语之间的包含关系来建立文档节点和词语节点之间的边。文档节点和词语节点之间的边的权重用词语在文档中的词频-逆文档频率指数(Term Frequency-Inverse Document Frequency, TF-IDF)计算,词频指词语在文档中出现的次数,逆文档频率指总文档数除以包含该词语的文档的数目。

$$TF-IDF = TF * IDF \quad (4)$$

$$tf_{ij} = \frac{c_{i,j}}{\sum_k c_{k,j}} \quad (5)$$

$$idf_i = \log \frac{|D|}{1 + |\{j: t_i \in d_j\}|} \quad (6)$$

其中, $c_{i,j}$ 是词语 t_i 在文档 d_j 中出现的次数, $\sum_k c_{k,j}$ 是文档 d_j 中所有词汇出现的次数总和, $|D|$ 是语料库中的文档总数, $|\{j: t_i \in d_j\}|$ 表示包含词语 t_i 的文档数目。如某个词语未在

语料库中出现,就会导致分母为零,因此一般情况下会使用 $1 + |\{j: t_i \in d_j\}|$ 表示包含词语 t_i 的文档数目。

3.1.3 远距离词语依赖关系

本文使用哈尔滨工业大学的语言技术平台(Language Technology Platform, LTP)^[21]对文档进行依存关系分析,选取句子中存在远距离依赖关系的词语来扩展文本图,这将有利于丰富所构图的语义信息,可进一步提升模型的性能。LTP 共包括 24 种依存关系,例句 1“我刚刚吃了一个苹果,而他吃了一个苹果和一个芒果。”和例句 2“他吃了一个面包并且喝了一瓶牛奶”,其对应的依存结构如图 2 和图 3 所示。

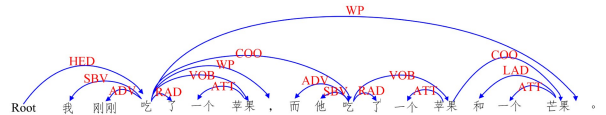


图2 例句1的依存图结构

Fig.2 Dependency graph structure

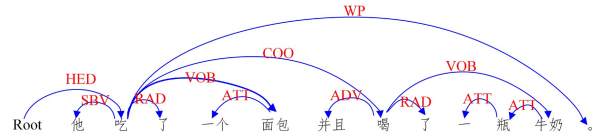


图3 例句2的依存图结构

Fig.3 Dependency graph structure

通过对上述句子的依存结构进行分析,可以得到多种远距离词语依赖关系,如 SBV 关系“我/吃”、VOB 关系“吃/苹果”、ATT 关系“一个/苹果”、ADV 关系“刚刚/吃”和 COO 关系“吃/喝”等。本文分析了不同依存关系的含义,选择了 5 种依存关系作为模型的远距离词语依赖关系,具体示例如表 1 所列。

表1 远距离依赖关系

Table 1 Long-distance dependency

依存关系	示例
SBV(主谓关系)	“我……吃”
VOB(动宾关系)	“吃……苹果”
ADV(状中结构)	“为了……生活”
COO(并列关系)	“…吃…,…喝…”
ATT(定中关系)	“一个..苹果”

基于远距离依赖关系建立的边的权重通过 PMI 计算,并只保留 PMI 为正值时的边^[22]。设 w_1 和 w_2 是具有远距离依赖关系 rel 的词语,边的权重计算式如下:

$$I_{rel}(w_1, w_2) = \lg \frac{p(w_1, w_2 | rel)}{p(w_1 | rel)p(w_2 | rel)} \quad (7)$$

其中:

$$p(\omega_1, \omega_2 | rel) = \frac{p(\omega_1, rel, \omega_2)}{p(rel)} \quad (8)$$

其中的概率用极大似然估计的方法计算。

$$p(\omega_1, rel, \omega_2) = \frac{Count(\omega_1, rel, \omega_2)}{Count(*, *, *)} \quad (9)$$

$$p(\omega_1 | rel) = \frac{Count(\omega_1, rel, *)}{Count(*, rel, *)} \quad (10)$$

$$p(\omega_2 | rel) = \frac{Count(*, rel, \omega_2)}{Count(*, rel, *)} \quad (11)$$

$$p(rel) = \frac{Count(*, rel, *)}{Count(*, *, *)} \quad (12)$$

其中, * 表示可能存在的词语或者依存关系, 最后能够推导得到式(13):

$$I_{rel}(\omega_1, \omega_2) = \lg \frac{Count(\omega_1, rel, \omega_2)Count(*, rel, *)}{Count(\omega_1, rel, *)Count(*, rel, \omega_2)} \quad (13)$$

3.2 图卷积神经网络

在构建文本图之后, 将该图输入到图卷积神经网络中。其由两层 GCN 构成, 第一层使用 ReLU 作为激活函数, 在这一层中每个节点接收邻域信息进行融合, 以更新其表示, 通过 ReLU 操作将节点的所有特征信息做非线性变换。然后将通过 ReLU 操作得到的表示输入到第二层, 使用 Softmax 函数作分类得到最后的分类结果。设 \mathbf{A} 为邻接矩阵, \mathbf{D} 为度矩阵, 其中 $D_{ii} = \sum_j A_{ij}$ 。假设每个节点都与自身相连, 因此 \mathbf{A} 的对角元素都设置为 1。SGCN 模型的表达式如下:

$$Z = \text{Softmax}(\tilde{\mathbf{A}}\text{ReLU}(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}_0)\mathbf{W}_1) \quad (14)$$

其中,

$$\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}} \quad (15)$$

其中, \mathbf{W}_0 和 \mathbf{W}_1 为权重参数, 通过梯度下降训练得到。

4 实验

4.1 数据集

实验采用了 3 个公开的文本分类数据集进行模型验证, 包括 SCP, HR 和 OS。

(1)SCP。此数据集包含 25900 个样本, 共涉及到 7 个类别。由于样本较长, 作者将每个样本只截取了前 70 个字符^[23]。本文实验从每个类别中随机选择 900 个样本作为训练集, 选择 300 个样本作为测试集。

(2)OS。该数据集共有 10 个类别, 包含正、负向各 3 万条评论数据。为了便于比较, 本文实验从中整理出 6 个类别的数据, 分别是书籍、水果、洗发水、衣服、酒店和计算机, 各个类别的数据规模如表 2 所列。

表 2 OS 数据集统计

Table 2 Summary statistics of OS datasets

Category	Train	Test	Category	Train	Test
Book	2567	1284	Clothes	4000	2000
Fruits	4000	2000	Hotel	4000	2000
Shampoo	4000	2000	Computer	2662	1330

(3)HR。该数据集的语料规模为 10000 篇^[24]。语料被整理成 4 个子集, 本文实验使用了第 3 个子集, 其选取的正负类样本各 3000 篇。

本文实验所用的 3 个数据集的训练样本数、测试样本数、类别数以及语料包含的词语数量统计参数如表 3 所列。

表 3 数据集统计

Table 3 Summary statistics of datasets

Dataset	Num-classes	Train	Test	Words
SCP	7	6300	2100	22700
OS	6	21229	10614	93973
HR	2	3000	3000	24063

4.2 基线模型

为了评估 SGCN 模型的性能, 实验采用了 5 个基线模型。

(1)CNN。CNN 把每个句子作为模型的输入, 利用多个不同大小的卷积核来提取句子的关键信息, 以便于更好地捕捉局部相关性。

(2)LSTM。LSTM 能够保留句子中词语的时序关系, 在一定程度上解决梯度消失问题, 同时还能获取词语间的依赖关系。

(3)Bi-LSTM。Bi-LSTM 是由正向 LSTM 和反向 LSTM 叠加而成的, Bi-LSTM 相比单向的 LSTM, 其能更好地捕获句子中的上下文信息。

(4)Graph-CNN-C^[25]。该模型利用切比雪夫多项式近似拟合卷积核, 降低 GCN 的卷积维度和计算复杂度。

(5)Text_GCN。该模型将文档分类转化成节点分类问题, 为语料构建文本图, 并利用图卷积网络进行文本分类。

4.3 参数设置

本文实验随机选择 10% 的训练集作为验证集, 实验用到的主要参数与参数值如表 4 所列。

表 4 实验参数设置

Table 4 Experimental parameter setting

Parameter	Parameter value
Learning_rate	1×10^{-3}
Dropout	0.5
L2 penalty	0
Sliding window	20
Epochs	200
GCN	2

同时, 为了分析 GCN 的层数对结果的影响, 本文分别在 3 个数据集上进行了多项对比实验, 结果如图 4 所示。由于图卷积神经网络包含了聚合邻居节点特征的操作, 当网络堆叠多层后使节点之间的特征过于平滑, 造成过拟合现象, 因此, F1 值在 3 个数据集上都呈现先增大后减小的趋势, 在堆叠层数为 2 时, 该网络就已经能达到较好的效果。

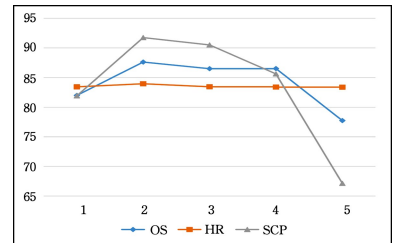


图 4 不同 GCN 层数对应的 F1 值

Fig. 4 F1 with different GCN layers

4.4 实验结果与分析

本文采用 F1 值来评价模型的性能, 并将 5 次运行结果

的平均值作为最终结果,具体结果如表 5 和表 6 所列。表 5 和表 6 的实验结果表明,本文方法在多个数据集上相比已有的基线模型取得了更好的效果。CNN 模型在各个数据集的表现相对较差,其原因在于 CNN 的池化层在处理上下文信息时会丢失大量有价值的信息,忽略了局部与整体之间的关联性,使训练结果收敛于局部最小值。LSTM 和 Bi-LSTM 模型在解决序列文本的远距离依赖问题方面存在一定优势,使它们在各个数据集上的表现均优于 CNN。然而,由于它们本质上都是基于序列的模型,因此其在捕获远距离依赖特征方面仍存在一定不足。Graph-CNN-C 在 HR 上的表现最好,但在 SCP 上的表现却最差,其原因在于 HR 属于情感分类数据

集,该模型能保留情感分类中较为重要的词序信息。Text-GCN 与其他模型相比,能捕捉高阶邻域信息,能更好地整合语料的全局信息,因此其分类结果有进一步的提升。不同模型在不同数据集上的实验结果的变化趋势如图 5 所示。

表 5 模型在 SCP 和 HR 数据集上的 F1 值

Table 5 Model F1 values on SCP and HR datasets

Model	SCP	HR
CNN	86.94	82.90
LSTM	87.78	83.59
Bi-LSTM	86.56	85.50
Graph-CNN-C	83.57	85.83
Text_GCN	91.67	83.55
SGCN	91.75	83.96

表 6 模型在 OS 数据集上的 F1 值

Table 6 Model F1 values on OS dataset

Model	Book	Fruits	Shampoo	Clothes	Computer	Hotel	Average
CNN	77.04	89.26	83.96	87.95	77.76	82.38	83.06
LSTM	78.55	89.85	86.38	88.12	79.14	82.39	84.07
Bi-LSTM	79.76	91.16	87.07	89.44	80.62	83.65	85.28
Graph-CNN-C	83.84	83.80	86.33	86.33	82.45	84.83	84.60
Text_GCN	81.06	89.08	92.16	91.16	84.57	86.45	87.41
SGCN	81.85	89.24	92.26	91.41	84.45	86.53	87.62

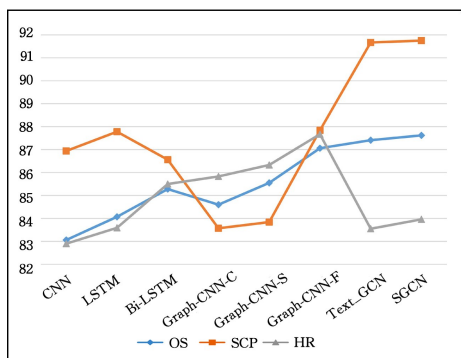


图 5 不同数据集上实验结果的变化趋势

Fig. 5 Variation trend of results in different datasets

如图 5 所示,尽管 SGCN 模型在 HR 数据集上的 F1 值低于 Graph-CNN-C 模型,但其相比 Text_GCN 模型,在 3 个数据集上都有不同程度的提升。因此,实验结果表明,本文提出的 SGCN 模型同时使用了词语共现关系、文档与词语之间的包含关系以及句子中远距离词语的依赖关系进行构图,能够捕获更丰富的语义信息。SGCN 在 HR 数据集上的表现不如部分模型的原因是,GCN 忽略了文本的词序信息,而词序信息对于情感分类数据是较为重要的。

4.5 消融实验

为了验证远距离词语依赖关系、词语共现关系、文档和词语之间的包含关系等特征对所提模型的影响,本文在 2 个数据集上对 3 类关系进行了消融实验,实验结果如表 7 所列。

(1)DW, 仅用文档与词语之间的包含关系构图。

(2)DW-WC, 同时使用文档与词语之间的包含关系和词共现关系构图。

(3)DW-WC-LD, 同时利用远距离词语依赖关系、词语共现关系、文档和词语之间的包含关系构图。

表 7 消融实验结果

Table 7 Results of ablation experiments

	SCP	HR
DW	90.96	80.40
DW-WC	91.67	83.55
DW-WC-LD	91.75	83.96

4.6 参数对模型的影响

为了验证滑动窗口大小和嵌入维度对实验结果的影响,本文在其他参数不变的前提下,分别分析了滑动窗口大小对 OS 数据集中 Clothes 和 Shampoo 的分类结果的影响,对比了嵌入维度对 HR 数据集的分类结果的影响,不同趋势变化如图 6 和图 7 所示。

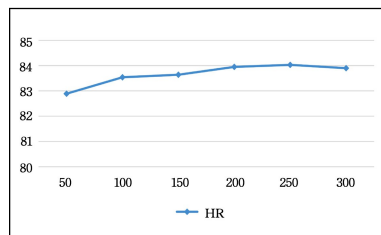


图 6 不同嵌入维度对应的 F1 值

Fig. 6 F1 with different embedding dimension

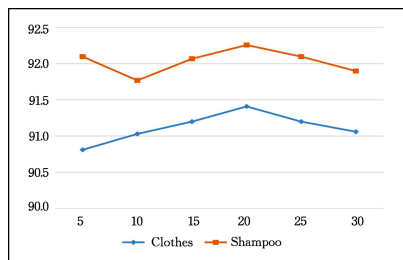


图 7 不同滑动窗口大小对应的 F1 值

Fig. 7 F1 with different sliding window sizes

如图 6 所示, $F1$ 值随着嵌入维度的变化而变化, 当嵌入维度为 250 时, $F1$ 值达到最大。这表明过低的嵌入维度不能较好地 将标签信息传播到整个图, 而高嵌入维度也不一定 能提高分类性能。

图 7 给出了 $F1$ 值会随着滑动窗口大小的增大而增大, 当滑动窗口增加到 20 时, $F1$ 值达到最大。这表 明, 太小的滑动窗口不能生成足够的全局词协同信息, 而太大的窗口可能会使联系不密切的节点之间形成边。

结束语 本文将具有远距离依赖关系的词语引入到文本 构图过程中, 同时利用词语共现关系、文档和词语之间的包含 关系、远距离词语依赖关系为整个文本语料库构建文本图, 提 出了融入句子中远距离词语依赖的图卷积短文本分类方法。 尽管 3 个数据集上的实验结果表明了本文方法的有效性, 但 短文本的表示仍存在特征稀疏和语义信息缺乏的问题, 因此 接下来我们将探索外部知识增强的短文本分类方法。

参 考 文 献

- [1] 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.
- [2] ZHOU C, SUN C, LIU Z, et al. A C-LSTM Neural Network for Text Classification[J]. *Computer*, 2015, 1(4): 39-44.
- [3] WANG Y, HUANG M, ZHU X, et al. Attention-based LSTM for Aspect-level Sentiment Classification[C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016: 606-615.
- [4] ZHOU J, CUI G, ZHANG Z, et al. Graph neural networks: A review of methods and applications[J]. *arXiv: 1812. 08434*, 2018.
- [5] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. *arXiv: 1609. 02907*, 2016.
- [6] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C] // Advances in Neural Information Processing Systems. 2017: 1024-1034.
- [7] CHEN J, MA T, XIAO C. FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling[C] // International Conference on Learning Representations. 2018.
- [8] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C] // Advances in Neural Information Processing Systems. 2016: 3844-3852.
- [9] ZHANG Y, LIU Q, SONG L. Sentence-State LSTM for Text Representation[C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018: 317-327.
- [10] 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 (Natural Science)*, 2020, 34(5): 187-195.
- [11] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3: 993-1022.
- [12] JOACHIMS T. Text categorization with support vector machines; Learning with many relevant features[C] // European Conference on Machine Learning. Berlin; Springer, 1998: 137-142.
- [13] WANG X, CHEN R, JIA Y, et al. Short text classification using

wikipedia concept based document representation[C] // 2013 International Conference on Information Technology and Applications. IEEE, 2013: 471-474.

- [14] MCCALLUM A, NIGAM K. A comparison of event models for naive bayes text classification [C] // AAAI-98 Workshop on Learning for Text Categorization. 1998: 41-48.
- [15] PHAN X H, NGUYEN L M, HORIGUCHI S. Learning to classify short and sparse text & web with hidden topics from large-scale data collections[C] // Proceedings of the 17th International Conference on World Wide Web. 2008: 91-100.
- [16] ZHANG X, ZHAO J, LECUN Y. Character-level convolutional networks for text classification[C] // Advances in Neural Information Processing Systems. 2015: 649-657.
- [17] WANG P, XU B, XU J, et al. Semantic expansion using word embedding clustering and convolutional neural network for improving short text classification[J]. *Neurocomputing*, 2016, 174: 806-814.
- [18] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 7370-7377.
- [19] 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 and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019: 3435-3441.
- [20] HU L M, YANG T C, 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.
- [21] CHE W X, LI Z H, LIU T. Ltp: A chinese language technology platform[C] // Coling 2010: Demonstrations. 2010: 13-16.
- [22] YUAN L C. A statistical model of syntactic analysis based on dependence relations[J]. *Journal of Central South University: Natural Science Edition*, 2009(6): 1630-1635.
- [23] HAO M, XU B, LIANG J Y, et al. Chinese Short Text Classification with Mutual-Attention Convolutional Neural Networks [J]. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 2020, 19(5): 1-13.
- [24] WU X H, CHEN L, WEI T T, et al. Sentiment analysis of Chinese short text based on Self-Attention and Bi-LSTM[J]. *Journal of Chinese Information Processing*, 2019, 33(6): 100-107.
- [25] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. *Advances in Neural Information Processing Systems*, 2016, 29: 3844-3852.



ZHANG Hu, born in 1979, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include natural language processing and representation learning.