文章编号:1008-1542(2024)01-0052-07

基于文本图神经网络的小样本文本分类技术研究

安相成,刘保柱,甘精伟

(中国电子科技集团公司第五十四研究所,河北石家庄　050051)

摘　要:为了解决文本图神经网络小样本文本分类精度较差的问题,设计了基于文本图神经网络的原型网络,采用预训练语言模型,利用文本级图神经网络为每个输入文本构建图并共享全局参数,将文本图神经网络的结果作为原型网络的输入,对未标注文本进行分类,并验证新模型在多个文本分类数据集上的有效性。实验结果表明,与需要大量标注文档的监督学习方法相比,所采用的方法未标注文本的分类精度提高了 1%～3%,在多个文本分类数据集上验证了新模型性能先进,内存占用更少。研究结果可为解决小样本文本分类问题提供参考。

关键词:自然语言处理;小样本文本分类;预训练模型;图神经网络;原型网络

中图分类号:TP183　　**文献标识码**:A　　**DOI**:10.7535/hbkd.2024yx01006

Research on few-shot text classification techniques based on text-level-graph neural networks

AN Xiangcheng, LIU Baozhu, GAN Jingwei

(The 54th Research Institute of CETC, Shijiazhuang, Hebei 050051, China)

Abstract: In order to solve the problem of poor accuracy of text classification in text graph neural network with small samples, a text level graph neural network-prototypical (LGNN-Proto) was designed. An advanced pre-training language model was adopted, and the text graph neural network was used to construct the graph for each input text, then the global parameters were shared. The result of the text graph neural network was used as the input of the prototype network to classify the unlabeled text, and the validity of the new model on multiple text classification data sets was verified. The results show that the accuracy of unlabeled text classification is improved by 1% ～ 3% compared with that of supervised learning, which requires a large number of labeled documents, and the new model is validated on multiple text classification data sets with advanced performance and lower memory consumption. The research results can provide reference for solving the problem of text classification with small sample size.

Keywords: natural language processing; few-shot text classification; pre-trained model; graph neural network; prototype network

收稿日期:2023-11-23;修回日期:2024-01-08;责任编辑:张士莹

基金项目:河北省智能化信息感知与处理重点实验室发展基金(SXX22138X002);LZH 联合 QB 数据融合与共享服务项目

第一作者简介:安相成(1997—),男,河北石家庄人,硕士,主要从事知识图谱方面的研究。

通信作者:甘精伟,研究员。E-mail:fiberChina@163.com

安相成,刘保柱,甘精伟.基于文本图神经网络的小样本文本分类技术研究[J].河北科技大学学报,2024,45(1):52-58.

AN Xiangcheng, LIU Baozhu, GAN Jingwei. Research on few-shot text classification techniques based on text-level-graph neural networks[J]. Journal of Hebei University of Science and Technology, 2024, 45(1):52-58.

文本分类是自然语言处理(NLP)的一个核心问题，在垃圾邮件检测[1]和新闻过滤[2]等领域应用发挥着重要作用。其中文本表示学习作为文本分类中的关键步骤，是目前文本分类研究的重点和难点。深度学习是一种“数据饥饿技术”，需要大量标记数据来训练模型参数，否则就会出现过拟合、模型鲁棒性和泛化性能较差的问题[3-5]。然而，在现实生活中，分类数据很难得到，文本分类费时费力。

小样本学习[6](few-shot learning，FSL)使深度学习在样本稀有案例上的部署成为可能，其主要目标是通过极少量训练样本来学习新类别的分类器。在这个领域的开创性工作可以追溯到 21 世纪初[7]。近期的研究主要分为 2 类：基于相似度的方法[6-9]和基于优化的方法[10-11]。此外，研究人员还在各种自然语言处理任务中探索了 FSL 的应用[12-14]。小样本学习通过从极少的标记样本中识别新的类别来解决数据不足的问题。只有一个或很少样本的局限性挑战了深度学习中的标准微调方法。早期研究采用数据增广和正则化技术来缓解由数据稀疏性导致的过拟合问题，但效果有限。相反，受人类学习的启发，有学者探索了元学习[15]。现代小样本学习方法通常按照测试和训练条件必须匹配的原则，将训练过程分解为一个辅助元学习阶段，包括许多元任务，通过将元任务从一个小批处理切换到另一个小批处理来提取一些可转移的知识。因此，小样本模型可以用一个小的标签支持集来分类新类别。随着深度学习的快速发展，利用卷积神经网络(convolutional neural network，CNN)[16]和循环神经网络(recurrent neural network，RNN)[17]等进行文本表示的研究方法已趋于成熟，有效提高了文本分类的精度和时效性。近年来，作为一种新型神经网络，图神经网络(graph neural network，GNN)因具有能有效利用文本数据结构信息、非线性映射能力、适用不同类型的文本数据等优势，引起了学者的广泛关注[18-31]，并已成功应用于文本分类。然而，以往的小样本文本分类方法普遍存在 2 个缺陷。首先，都采用 GloVe 词嵌入结合 CNN 或 RNN 结构进行文本嵌入，而没有采用近年来提出的更高级的预训练语言模型，如 ELMo[32]，GPT[33]和 BERT[34]。其次，在解决语料库不固定的小样本文本分类问题上无法获得较高的精度。

针对上述问题，本文提出基于文本图神经网络的原型网络(text level graph neural network-prototypical，LGNN-Proto)进行小样本文本分类。采用文本图神经网络模型，为每个输入文本构建图并共享全局参数，而不是为整个语料库构建单个图。该方法消除了单个文本与支持在线测试的整个语料库之间的依赖关系，但仍然保留了全局信息。此外，在文本中使用较小的窗口构建图，不仅能提取更多的局部特征，而且大大减少了边缘数和内存消耗。

1　模型介绍

如图 1 所示，所建模型由 3 个主要组件组成，即文本嵌入组件、文本级图神经网络组件和原型网络组件。文本嵌入组件主要用于从文本中提取语义特征，并将原始文本序列转化为文本嵌入。通过文本级图神经网络对文本的类内相似度和类间不相似度进行建模，度量文本样本与文本样本之间的潜在关系，最后利用一个原型网络对查询集中的样本进行分类。

在小样本文本分类任务中，定义标记样本集为支持集 S，未标记样本集为查询集 Q。采用 N-way，k-shot 设置，其中支持集 S 包含每个 k 标记的 N 个类的样本，值得注意的是在小样本学习任务中，k 通常较小。研究目标是进行元学习训练，提取可转移的知识，能够在支持集 S 上更好地进行小样本学习，并对来自查询集 Q 的样本进行尽可能准确的分类。首先，使用编码器(encoder)如 BERT 对输入序列进行编码。通过 BERT 等编码器获得上下文有关的词嵌入(embedding)后，将其与静态的词嵌入，与字符级别的嵌入拼接起来，送入文本图神经网络获得聚合的词表示。将文本图神经网络中的整个语料库设置成一个大的矩阵，每 2 个单词之间通过共现数据设置一个关系值，作为边之间的权重。然后，将出现次数$<k$ 的边统一映射到一个统一的表示，将低频次的边删去，可以使模型更注重高频的关系，参数得到充分训练。

图 1 呈现了一个针对双向 2-shot 小样本文本分类任务的工作流。不同颜色的文档图标代表来自不同类别的文本，深色组成的图标表示查询文本，浅色表示文本嵌入。为了更清晰地展示，在图 1 中将节点“proud”的参数设置为 $p=2$(节点和边以红色表示)，而将其他节点的参数设置为 $p=1$(以蓝色表示)。实际情况中，会话期间的 p 值是唯一确定的。图 1 中的所有参数都源自全局共享表示矩阵，如底部所示；左下方是特征矩阵；右下方为全局共享表示矩阵，这个矩阵为每个单词之间都生成一个关系，因此是 $V\times V$。

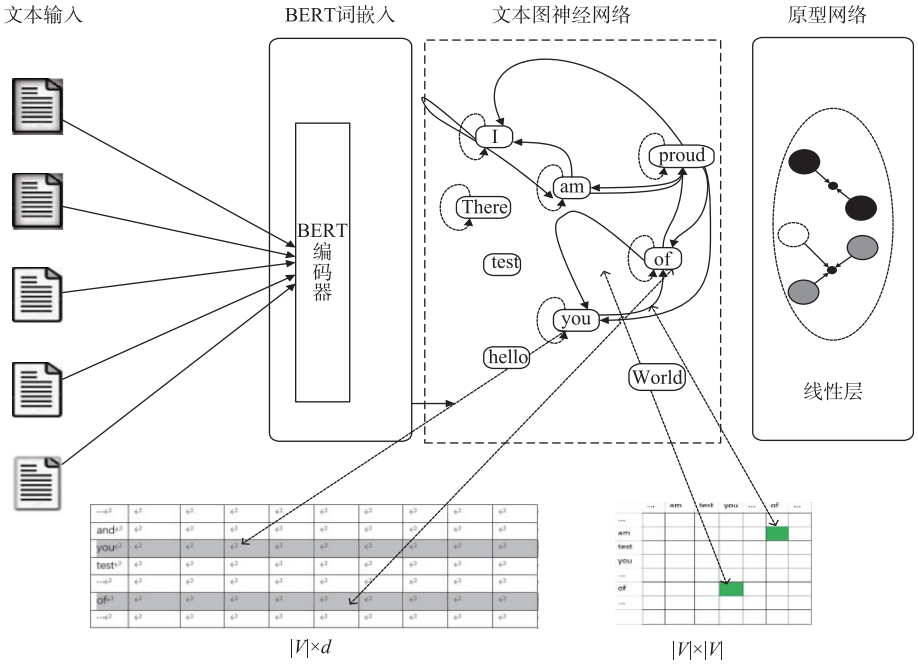

图 1 LGNN 模型

Fig. 1 Model of LGNN

2 模块设计

2.1 文本嵌入

一项有限的文本分类任务中，只能利用有限的注释数据来训练分类器。鉴于此，决定采用预训练的语言模型，以更好地提取文本特征。具体而言，选择了 BERT 变压器编码器。该编码器堆栈由 12 个编码器层组成，每个编码器层包括双向自注意力层和全连接层。在堆栈输入中，首先将每个令牌嵌入到一个已学习的 d 维嵌入中，然后在每次遍历 BERT 编码器层时逐步转换。从理论上讲，每个编码层的词嵌入输出都包含了整个文本的特征。采用标记文本开头的特殊符号[CLS]，从倒数第 2 层开始的词嵌入作为整个文本 t 的文本嵌入，表示为

$$e = f_{\text{emb}}(t \mid \theta_{\text{emb}}) \text{ 。} \tag{1}$$

其中 θ_{emb} 表示文本嵌入分量中的参数，在训练时进行微调。

2.2 文本级图神经网络

为更好地解决小样本文本分类无法支持在线测试的问题，引用 HUANG 等[28]提出的文本级图神经网络(Text-Level-GNN)方法，这是一种基于图神经网络(GNN)的文本分类方法，在图结构、在线测试和内存消耗等方面表现出显著的优越性，并被证实在文本分类领域中能够有效发挥作用。

考虑到任务中所有样本的文本嵌入，首先构造一个文本级图，其中每个节点代表每个样本，每个边代表连通节点之间的关系。该方法具体描述为为每个输入文本生成一个文本级别的图。在文本级图中，采用一种较小窗口的方式连接单词节点，而非直接连接所有单词节点。将带有 l 个单词的文本标记为 $T=\{r_1, r_2, \cdots, r_l\}$，其中 r_i 表示第 i 个单词的表示。R_i 是一个由 d 维词嵌入初始化的向量，可以通过训练进行更新。对给定的文本构建图，将文本中出现的所有单词视为图的节点。每条边都从文本中的一个单词开始，并以其相邻的单词结束。具体来说，文本 T 的图形定义为

$$N=\{r_i \mid i \in [1,l]\}, \tag{2}$$

$$E=\{e_{ij} \mid i \in [1,l]; j[i-p, i+p]\} \text{ 。} \tag{3}$$

式中：N 和 E 是图的节点集和边集，N 中的词表示和 E 中的边权值取自全局共享矩阵；p 表示图中与每个单词相连的相邻单词数。此外，将训练集中出现次数小于 k 次的边均匀映射到“公共”边上，使参数得到充分训练。

在此设计中，相同节点的表示和边的权值是全局共享的，可以通过消息传递机制在文本级别的图中更新。在该机制中，节点通过获取相邻节点的信息来更新其表示。本文采用一种名为信息传递机制(MPM)的

非谱方法进行卷积。MPM 使节点的表示方式受邻域的影响，即表示方式可以从上下文中获取信息。因此，即使是多义词，其在上下文中的确切含义也可以通过来自邻居的加权信息的影响来确定。此外，文本级图的参数取自全局共享矩阵，这意味着文本级图的表示也能像其他基于图的模型一样带来全局信息。

通过文本嵌入初始化节点特征，即 e，并根据连接节点是否属于同一类初始化边缘特征，交替更新直至收敛，得到更新后的节点特征见式(4)，之后对 e 进行进一步处理。

$$e_f = f_{e\mathrm{GNN}}(e \mid \theta_{e\mathrm{GNN}}) \ 。 \tag{4}$$

通过如此设计，构建出图的结构只依赖于每个文本，边的权重则来自于全局关系，使得图的大小得到缩减，并且适用于新来的文本。文本级图消除了单个输入文本和整个语料库之间的依赖负担，支持在线测试。此外，文本级图的优势在于在一个较小的上下文窗口内连接单词，从而减少内存消耗。这是因为它排除了文本中与当前单词关系不大、距离较远的许多单词，大幅减少了边的数量。消息传递机制使得图中节点能够感知周围的信息，从而在特定的上下文中更准确地获取含义。

2.3　原型网络

本文与 HUANG 等[28]的方法不同，没有直接通过边缘特征和节点特征来计算查询节点的预测概率。相反，引入了原型网络，可以从更一般的角度对样本进行分类。原型关系网络包括 2 个模块：Prototype 模块生成每个类样本的原型，Relation 模块计算查询集中的每个样本与每个类每个原型之间的关系。

将样本集 S 中的样本 x_i 和基于文本构建的查询集 Q 中的样本 x_j 送入嵌入网络，通过嵌入网络得到它们的特征向量 $\boldsymbol{f}_{\mathrm{linear}}$，计算样本集中每个类样本的原型。每个样本特征 e_f 通过一个线性层进行变换，得到新的特征表示

$$e_l = \boldsymbol{f}_{\mathrm{linear}}(e_f \mid \theta_{\mathrm{linear}}) \ 。 \tag{5}$$

式中 θ_{linear} 表示原型网络的参数，可在训练时微调。

计算从查询集 Q 中的每个样本到每个类原型的欧氏距离。在训练过程中，通过最小化同类样本之间的距离和最大化不同类样本之间的距离来优化模型参数。在测试过程中，将分类问题转化为最近邻问题，输出与测试样本最接近的原型类别，也就是通过比较所有原型向量与查询向量之间的距离，对查询样本进行分类。

3　实验部分

采用 BERT-base 作为文本编码器，将文本嵌入后的样本用一个 768 维的向量表示。

3.1　数据集

HUANG 等[28]采用的 R8，R52 数据集是路透社数据集 Reuters-21578 的子集派生的其他数据集。R8 有 8 个类别，分为 2 189 个测试样本和 5 485 个训练样本。R52 有 52 个类别，分为 6 532 个训练样本和 2 568 个测试样本。因为 R8，R52，Ohsumed 数据集无法获得小样本条件，所以没有像文献[28]一样在这些数据集上进行指标比较。在 2 个公开的小样本任务分类数据集上进行评估，分别为亚马逊评论情感分类数据集(Amazon review sentiment classification, ARSC)[35]和小样本关系分类数据集(few-shot relation classification, FewRel)[36]。ARSC 数据集包含了亚马逊上 23 种产品的英文评论。

在小样本分类任务中，一般有 3 个数据集：训练集、支持集和测试集。训练集有自己的标签空间，与支持集、测试集是不相交的，支持集和测试集具有相同的标签空间。创建一个包括 12 个类别和 5 个样本的文本分类任务。FewRel 是一个大规模的监督数据集，包含来自维基百科的 100 个关系上的 70 000 个实例。在 5 类别和 5 个样本以及 10 类别和 5 个样本的设置上对所建方法进行了评估，删除带有 2 个或更多标签的文本。对于以上所有的数据集，从训练集中随机选取 10%的文本构建验证集。

3.2　基线模型

对于在 ARSC 数据集中的实验，本文将所提 LGNN-Proto 模型与 Matching Networks 模型、Prototypical Networks 模型、Graph Network 模型、Relation Networks 模型、Text-GCN 模型、ROBUSTTC-FSL 模型、Induction Networks 模型和 EGNN-Proto 模型进行比较。对于在 FewRel 数据集中的实验，本文将所建 LGNN-Proto 模型与 Finetune 模型、KNN 模型、MetaN 模型、GNN 模型、Text-GCN 模型、Proto 模型、PHATT 模型、HAPN 模型和 EGNN-Proto 模型进行比较。

3.3　结果和分析

ARSC 和 FewRel 实验结果如表 1 和表 2 所示。表 1 和表 2 的结果显示，所提出的 LGNN-Proto 模型在

表1 ARSC数据集分类精度

Tab.1 Classification accuracy of ARSC data set

模型	分类精度/%
Matching Networks	65.73
Prototypical Networks	68.17
Graph Network	82.61
Relation Networks	83.07
Text-GCN	82.57
ROBUSTTC-FSL	83.12
Induction Networks	85.62
EGNN-Proto	88.26
LGNN-Proto	89.13

表2 FewRel数据集分类精度

Tab.2 Classification accuracy of FewRel data set

模型	分类精度/%	
	5 way 5 shot	10 way 5 shot
Finetune	68.66±0.41	55.04±0.31
KNN	68.77±0.41	55.87±0.31
MetaN	80.57±0.48	69.23±0.52
GNN	81.28±0.62	64.02±0.77
Text-GCN	79.40±0.22	68.33±0.25
Proto	89.05±0.09	81.46±0.13
PHATT	90.12±0.04	83.05±0.05
HAPN	91.02±0.11	84.16±0.19
EGNN-Proto	92.29±0.23	86.09±0.31
LGNN-Proto	92.82±0.49	86.53±0.51

ARSC数据集和FewRel数据集上算法精度均有很大提高。

由表1和表2可知:1)基于图的模型结果优于传统的基线模型,如匹配网络、原型网络和图网络。这可能是由于图结构的特性。图结构允许存在不同数量的邻居节点,使得单词节点可以通过不同搭配学习更准确表示;2)单词之间的关系可以记录在边缘权值中,并在全局共享。这些对于传统模型来说是不可实现的。此外,所建模型相对于基于图的模型(如Graph Network)表现更为出色。Graph Network使用单词袋模型来表示文档,与本文所提方法相似,但它们在一个没有加权边缘的大窗口内连接单词节点,无法区分不同单词的重要性。相比之下,本文所建模型采用可训练的边缘权值,使得单词能够在不同搭配中以不同方式表达自己;3)这些权值是全局共享的,意味着它们可以通过整个语料库中包含相同搭配的所有文本进行训练,相对于之前的EGNN-Proto模型表现更佳。这可能归因于更多的边的表达以及表征学习的差异。EGNN通过语料库级共现来学习单词表示,而本文所建模型则在上下文窗口内进行训练,类似于传统的单词嵌入,能够从预训练的单词嵌入中受益,并取得更好的性能。

EGNN-Proto和本文所建模型在内存消耗方面的比较结果如表3所示。由表3可以看出,所建模型具有显著优势,其单词仅与文本中相邻的单词相连,而EGNN-Proto基于语料库级别图,在一个较大的窗口内连接节点。由于EGNN-Proto使用共现信息作为固定权值,为了获取更准确的共现权值,必须增大窗口。因此,本文所建模型产生的边缘权值矩阵比EGNN-Prot更为稀疏。此外,由于文本表示是由文本中的单词节点表示的总和计算的,因此模型中不存在文本节点,进一步降低了内存消耗。

表3 内存消耗比较

Tab.3 Comparison of memory consumption

数据集	内存消耗/MB	
	EGNN-Proto	LGNN-Proto
FewRel	18 810	9 535
ARSC	26 642	16 506

为了深入了解不同连接窗口的影响,比较了具有不同 p 值数据集的性能,如图2所示。

由图2可以看出,随着 p 的增加,该算法的精度也在增加,并且当连接到大约3个邻域时,该算法的精度达到最佳。随着 p 值的增加,精度的波动率降低。这表明,当节点仅使用最近的邻域连接时,节点无法理解跨越多个单词的依赖关系;而当节点使用距离较远的邻域(p 较大)连接时,图变得越来越类似于忽略局部特征的全连通图。此外,边的减少也导致内存消耗的降低。相较于之前的方法,所提模型具有更少的边,进一步显示了模型的优越性。

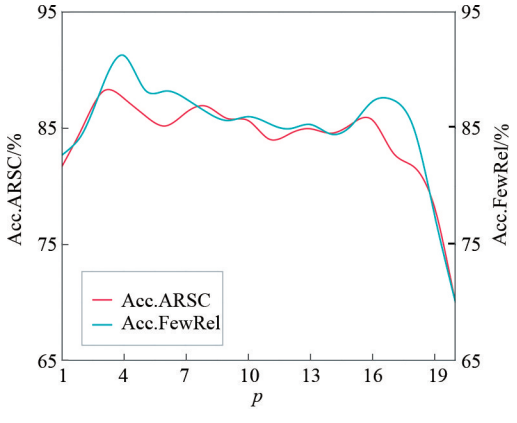

图2 不同 p 值的ARSC和FewRel数据集的精度变化

Fig.2 Accuracy variation of ARSC and FewRel data sets with different p values

3.4 消融实验

LGNN-Proto模型主要有3个重要组件:原型网络、文本图神经网络模块和Bert解码器。本文将Bert

解码器设置为 A，文本图神经网络模块设置为 B，原型网络组件设置为 C。为了验证这 3 个组件对小样本命名实体识别模型 EGNN-Proto 的影响程度，分别去掉或替换其中一个组件，在 ARSC 和 FewRel 数据集上进行实验，并与原模型进行比较。消融实验结果如表 4 所示。

表 4　消融实验结果

Tab. 4　Ablation results

实验设置	准确度/%	
	ARSC	FewRel
B	68.71	66.35
A+B	71.16	68.33
B+C	85.58	86.15
A+B+C	89.13	92.828

从表 4 可以看出，文本图神经网络模块表现并不是很好，在样本较少的数据集上并不能够很好地优化模型，可能会造成标签的随机预测，这表明传统的机器学习方法和迁移学习方法不能够适应样本较少的环境。BERT 依靠不经微调的预训练语言模型 BERT 进行编码，通过词嵌入的余弦相似性预测标签类型，效果得到了一定的提升，表明基于度量的方法在小样本环境下比传统的机器学习模型更具优势。其中，原型网络影响最大，将原型网络组件去掉，简单采用文本图神经网络原文中所采用的分类方法后，实验结果的准确率大幅降低。通过对消融实验的分析可知，原型网络组件、Bert 解码器都能够增强 EGNN-Proto 模型在小样本命名实体任务中的识别效果。

4　结　语

1）本文提出的模型解决了小样本文本分类问题。通过借助预训练语言模型 BERT，成功提取了文本的语义特征。引入文本图神经网络，采用整个语料库范围的文本级图，利用一个原型网络对查询文本进行分类。文本级图消除了单个输入文本和整个语料库之间的依赖负担，减小了内存消耗。

2）所提方法在 ARSC 和 FewRel 数据集上均表现出先进性能，可以考虑将该方法推广到其他不完善领域的自然语言处理问题的研究中。例如在信息化作战条件下，行动与任务是相互关联、紧密配合的。通常一个使命会具体到多个任务来达成，同时任务又包含不同的行动和目的。而我方在大多数情况下只能搜集到对方有关行动的部分零散情报信息，仅对单个行动或者军事实体进行情报分析可能难以获得有效信息，也很难快速分析出对手的行动全貌、关联行动、军事实体以及对手正在实施的任务，需要通过对较少样本数量的情报信息进行分析推理。

未来工作中，将考虑结合元学习的小样本方法，提升军事情报小样本知识图谱补全领域的识别效果。

参考文献/References：

[1] JINDAL N，LIU Bing. Review spam detection[C]//Proceedings of the 16th International Conference on World Wide Web. Canada：Association for Computing Machinery，2007：1189-1190.

[2] AGGARWAL C C，ZHAI Chengxiang. A Survey of Text Classification Algorithms[M]. Boston：Springer，2012.

[3] RUSSAKOVSKY O，DENG Jia，SU Hao，et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision，2015，115(3)：211-252.

[4] LIN T Y，MAIRE M，BELONGIE S，et al. Microsoft COCO：Common objects in context[C]//Computer Vision-ECCV 2014. Zurich：Springer，2014：740-755.

[5] WANG Yaqing，YAO Quanming，KWOK J T，et al. Generalizing from a few examples：A survey on few-shot learning[J]. ACM Computing Surveys，2020，53(3)：1-34.

[6] SNELL J，SWERSKY K，ZEMEL R. Prototypical networks for few-shot learning[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach：Curran Associates Inc，2017：4080-4090.

[7] LI Feifei，FERGUS R，PERONA P. One-shot learning of object categories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2006，28(4)：594-611.

[8] SUNG F，YANG Y X，ZHANG L，et al. Learning to compare：Relation network for few-shot learning[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City：IEEE，2018：1199-1208.

[9] VINYALS O，BLUNDELL C，LILLICRAP T，et al. Matching networks for one shot learning[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona：Curran Associates Inc，2016：3637-3645.

[10] RAVI S，LAROCHELLE H. Optimization as a model for few-shot learning[C]//5th International Conference on Learning Representations. Toulon：ICLR，2017：1-11.

[11] PRECUP D，THE Y W. Proceedings of the 34th International Conference on Machine Learning-Volume 70[R]. Sydney：JMLR，2017.

[12] YU Mo，GUO Xiaoxiao，YI Jinfei，et al. Diverse few-shot text classification with multiple metrics[C]//Proceedings of the 2018 Conference

of the North American Chapter of the Association for Computational Linguistics. New Orleans: Association for Computational Linguistics, 2018: 1206-1215.

[13] GU Jiatao, WANG Yong, CHEN Yun, et al. Meta-learning for low-resource neural machine translation[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels: Association for Computational Linguistics, 2018: 3622-3631.

[14] GAO Tianyu, HAN Xu, LIU Zhiyuan, et al. Hybrid attention-based prototypical networks for noisy few-shot relation classification[C]// The Thirty-third AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019: 6407-6414.

[15] LIU Xiaoqian, ZHOU Fengyu, LIU Jin, et al. Meta-learning based prototype-relation network for few-shot classification[J]. Neurocomputing, 2020, 383: 224-234.

[16] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing(EMNLP). Doha: Association for Computational Linguistics, 2014: 1746-1751.

[17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[18] 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. Barcelona: Curran Associates Inc, 2016: 3844-3852.

[19] GORI M, MONFARDINI G, SCARSELLI F. A new model for learning in graph domains[C]//2005 IEEE International Joint Conference on Neural Networks. Montreal: IEEE, 2005: 729-734.

[20] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20 (1): 61-80.

[21] LI Yujia, TARLOW D, BROCKSCHMIDT M, et al. Gated Graph Sequence Neural Networks[EB/OL]. https://arxiv.org/abs/1511.05493, 2015-11-17.

[22] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral Networks and Locally Connected Networks on Graphs[EB/OL]. https://arxiv.org/abs/1312.6203, 2013-12-21.

[23] HENAFF M, BRUNA J, LECUN Y. Deep Convolutional Networks on Graph-Structured Data[EB/OL]. https://arxiv.org/abs/1506.05163, 2015-06-16.

[24] GARCIA V, BRUNA J. Few-shot Learning with Graph Neural Networks[EB/OL]. https://arxiv.org/abs/1711.04043, 2017-11-10.

[25] KIM J, KIM T, KIM S, et al. Edge-labeling graph neural network for few-shot learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach: IEEE, 2019: 11-20.

[26] GIDARIS S, KOMODAKIS N. Generating classification weights with GNN denoising autoencoders for few-shot learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach: IEEE, 2019: 21-30.

[27] YAO Liang, MAO Chengsheng, LUO Yuan. Graph convolutional networks for text classification[C]//The Thirty-third AAAI Conference on Artificial Intelligence(AAAI-19). Palo Alto: AAAI Press, 2019: 7370-7377.

[28] HUANG Lianzhe, MA Dehong, LI Sujian, 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. Hong Kong: Association for Computational Linguistics, 2019: 3444-3450.

[29] YAN Leiming, ZHENG Yuhui, CAO Jie. Few-shot learning for short text classification[J]. Multimedia Tools and Applications, 2018, 77 (22): 29799-29810.

[30] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1(Long Papers). Louisiana: Association for Computational Linguistics, 2018: 2227-2237.

[31] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving Language Understanding by Generative Pre-Training[EB/OL]. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf, 2015-06-16.

[32] DEVLIN J, CHANG Mingwei, 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, Volume 1(Long and Short Papers). Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.

[33] BLITZER J, DREDZE M, PEREIRA F. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification [C]//Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics. Prague: Association for Computational Linguistics, 2007: 440-447.

[34] HAN Xu, ZHU Hao, YU Pengfei, et al. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018: 4803-4809.