计算机科学
COMPUTER SCIENCE

基于不变图卷积神经网络的文本分类

黄瑞，徐计

引用本文

黄瑞，徐计．基于不变图卷积神经网络的文本分类[J]．计算机科学，2024，51(6A)：230900018-5.

HUANG Rui, XU Ji. Text Classification Based on Invariant Graph Convolutional Neural Networks[J]. Computer Science, 2024, 51(6A): 230900018-5.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230900018

基于不变图卷积神经网络的文本分类

黄 瑞[1] 徐 计[2]
1 贵州大学计算机科学与技术学院 贵阳 550025
2 贵州大学省部共建公共大数据国家重点实验室 贵阳 550025
(gs.huangr21@gzu.edu.cn)

摘 要 文本分类是自然语言处理中一个基本而又重要的任务,近年来,图神经网络被越来越多地应用于文本分类中。然而,使用图神经网络的图表示学习在涉及文本分类的任务中不能很好地满足新词的归纳学习,其一般假设训练和测试数据来自相同的分布,但现实中这个假设经常不成立。为了克服这些问题,文中提出了 Invariant-GCN,用于通过 GCN 进行归纳文本分类。首先为每个文档构建单个图,使用 GCN 根据其局部结构学习细粒度的单词表示,这可以有效地为新文档中没见过的单词生成嵌入进而将单词节点作为文档嵌入合并;然后提取最大限度地保留不变类内信息的期望子图,使用这些子图进行学习不受分布变化的影响;最后通过图分类方法完成文本分类。在 4 个基准数据集上与 5 种分类方法进行了比较,实验结果表明 Invariant-GCN 具有良好的文本分类效果。

关键词:文本分类;图卷积神经网络;因果学习;文本图构建
中图分类号 TP391

Text Classification Based on Invariant Graph Convolutional Neural Networks

HUANG Rui[1] and XU Ji[2]
1 College of Computer Science and Technology,Guizhou University,Guiyang 550025,China
2 State Key Laboratory of Public Big Data,Guizhou University,Guiyang 550025,China

Abstract Text classification is a basic and important task in natural language processing,and graph neural networks have been applied to this task in recent years. However,the graph representation learning using graph neural networks can not well satisfy the generalization learning of new words in the task involving text classification. It is generally assumed that training and testing data come from the same distribution,which is often invalid in reality. To overcome these problems,this paper puts forward the Invariant-GCN,which is used for text categorization by GCN reported. First,to build a single figure for each document,use GCN to learn fine-grained word representation according to its local structure,which can effectivelygenerate embeddings for words not seen in the new document and then merge the word nodes as document embeddings. And then extract the maximum limit retained within the same class information expectations subgraph,use the graph to study is not affected by the distribution change. Finally,the text classification is completed by graph classification method. In four benchmark datasets,the the Invariant-GCN is compared with five classification methods,and the experimental results show that it has a good effect of text categorization.

Keywords Text classification,Graph convolutional neural network,Casual learning,Text graph construction

1 引言

文本分类是自然语言处理的一项基本任务,其目的是将标签分配给文本单元。文本分类的应用场景广泛,如垃圾邮件检测[1]、情感分析[2]、意图识别[3]等。

文本表示是文本分类问题中必不可少的一步,文本表示的研究可概括为两种:特征工程和特征学习。特征工程是指一段文本用手工制作的特征表示,例如使用词频(TF)、词频-逆文档频率指数(TF-IDF)等作为特征。特征学习是通过机器学习从原始文本中自动提取特征表示。根据学习模型的不同,特征学习又分为基于序列的学习模型以及基于图的学习模型。最常用的基于序列的学习模型包括卷积神经网络(CNN)[4]和循环神经网络(RNN)[5],它们促进了从局部连续单词序列中捕获文本特征,但可能会忽略携带非连续和长距离语义的语料库中的全局单词共现。基于图的学习方法被用于解决这个问题,该方法不将文本视为序列,而是将其视为一组共现词。并且基于图的学习模型如图神经网络(GNN)[6]可以直接处理复杂的结构化数据,并优先考虑利用全局特征。

近年来,使用 GNN 的图表示学习在涉及文本分类的任务中取得了巨大成功。然而,现有的基于 GNN 的图表示学习一般假设训练和测试数据来自相同的分布,现实中这个假设经常不成立。训练和测试分布之间通常不匹配,即存在分

基金项目:国家自然科学基金(61966005,62366008)
This work was supported by the National Natural Science Foundation of China(61966005,62366008).
通信作者:徐计(jixu@gzu.edu.cn)

布偏移。分布外泛化失败成为图表示学习应用于实际的主要障碍，因此，在常规欧几里德数据上实现分布外泛化受到了越来越多的关注，研究学者提出了几种解决方案，其中因果关系的不变性原理是这些解决方案的核心。该原理利用独立因果机制(Independent Causal Mechanisms，ICM)假设，表明只关注标签原因的模型预测可以保持对大类分布偏移的不变性[7]。

虽然不变性原理在欧几里德数据上取得了成功，但是图上的分布变化更加复杂，因此不能直接采用该原则。为了解决这个问题，本文提出了 Invariant-GCN 来捕获因果不变图结构，以实现在不同分布变化下保证分布外泛化。首先使用滑动窗口为每个文档构建单个图，其次在图生成过程中对不变特征和混淆特征的两种相互作用进行建模，然后使用不变性原理识别出不变子图，最后用不变子图去预测标签。

本文的主要贡献如下：

1)为每个文档构建图结构，将文本数据构建成文本共现图，使用滑动窗口学习文档的局部结构，考虑了更多的上下文信息。

2)提出了一种用于识别文本拓扑结构中不变子图的方法(Invariant-GCN)，解决分布偏移问题。

3)在多个基准数据集上进行了广泛的实验，说明了 Invariant-GCN 对文本分类的有效性。

2 相关工作

2.1 基于传统机器学习的文本分类

传统机器学习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 TF-IDF 和词袋模型，具有代表性的传统机器学习的分类方法有朴素贝叶斯(Naive Bayes，NB)[8]、支持向量机(Support Vector Machines，SVMs)[9]、K 最近邻算法(K-Nearest Neighbors，KNN)[10]等。然而传统机器学习表示的分类效果不佳，这是因为传统机器学习是浅层次的特征提取，对文本背后的语义、结构、序列和上下文理解不充分，模型的表征能力有限。

2.2 基于深度学习的文本分类

基于深度学习的文本分类方法有自动获取特征和进行端到端学习的能力，并且深度学习模型能够学习到文本深层的语义信息。Kim[11]提出使用文本卷积神经网络(TextCNN)进行文本分类，利用多个不同尺寸的滤波器来捕获文本的局部特征信息。Liu 等[12]提出使用循环神经网络(TextRNN)进行文本分类，循环神经网络能捕获序列的历史信息，使得其在文本分类任务中能捕获文本的上下文语义信息。Wang 等[13]使用一个词嵌入聚类和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。虽然基于深度学习的文本分类方法表现较好，但会忽略句子中非连续和长距离词语的依赖关系。

2.3 基于图神经网络的文本分类

现有的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。Yao 等提出了 TextGCN 模型[14]，使用 GCN 将文本分类问题转化为节点分类问题。Huang 等[15]通过引入消息传递机制和减少内存消耗来改进 TextGCN。最近，Zhang 等[16]通过学习文本级的单词交互，使用 GNN 进行归纳文本分类，Liu 等[17]构造了一个文本图张量来描述语义、句法和顺序上下文信息，在文本图张量上进行图内传播和图间传播。

2.4 分布偏移

分布偏移指的是模型在训练和测试阶段面对不同数据分布时性能下降的现象。在实际场景中，由于数据来源不同、数据采集时间不同等原因，训练数据和测试数据的分布可能会发生变化，进而导致模型在测试集上的性能不如在训练集上的性能。

分布偏移包含 3 种类型：协变量偏移、标签偏移、概念偏移。协变量偏移指训练数据和测试数据的输入特征分布不一致，即 $P(x_{\text{train}})\neq P(x_{\text{test}})$；标签偏移指训练数据和测试数据的输出分布不一致，即 $P(y_{\text{train}}\mid x_{\text{train}})\neq P(y_{\text{test}}\mid x_{\text{test}})$；概念偏移指训练数据和测试数据随着时间、环境、人群等因素的变化而发生变化即 $P(x_{\text{train}},y_{\text{train}})\neq P(x_{\text{test}},y_{\text{test}})$。

3 研究方法

目前已有大量关于文本分类的算法，而本文提出的新方法从因果关系角度出发，和已有工作存在明显不同。本文提出的模型将文本构建成图数据结构，识别图数据结构中的不变子图，将文本分类转化为图分类问题，从而解决了分布偏移中协变量偏移的问题，提升了文本分类效果。

如图 1 所示，Invariant-GCN 包含两个关键组件：构建文本图结构和构建基于图的学习框架。其中基于图的学习框架包括识别不变子图模块以及分类模块。

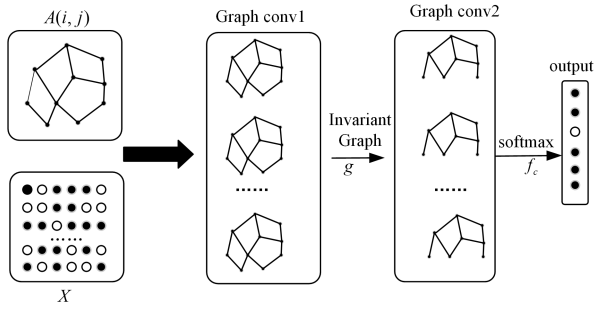


图 1 Invariant Graph-GCN 总体框架图

Fig. 1 Overall framework of Invariant Graph-GCN

3.1 构建文本图结构

首先以标准方式对文本进行预处理，包括标点符号和停用词的删除。顶点的嵌入用单词特征初始化，表示为 $\boldsymbol{h}\in\boldsymbol{R}^{|v|\times d}$，其中 d 为嵌入维度。使用 $G=(V,S)$ 表示为图结构，其中 $\boldsymbol{V}$ 是顶点集，$\boldsymbol{S}$ 是边集。图的构建过程是把每个单词作为顶点，根据单词之间的某种规则或者联系来确定两个单词之间是否存在连边。本文利用单词共现的方式确定单词节点之间的连边，构建文本图网络。

假定一段长度为 l 的文本序列，用 $\boldsymbol{T}=\{t_1,t_2,\cdots,t_n\}$ 表示，设定滑动窗口大小为 s，共现文本图的构建步骤如下：

1)从给定文本序列中提取单词集合，此单词集合就是共现图的节点集合，每一个单词代表一个节点。

2)滑动窗口沿着文本序列从左向右滑动，窗口初始单词

是 t_i，如果 t_j 和 t_i 在一个窗口内，就构建 t_j 和 t_i 单词节点之间的连边。

以“Jack is a basketball star who has won many awards”这个句子为例，图 2 给出了将一个文本转换成文本共现图的例子。

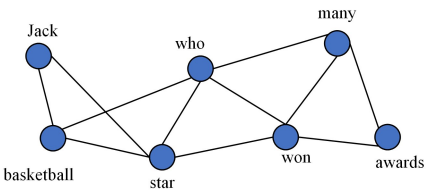
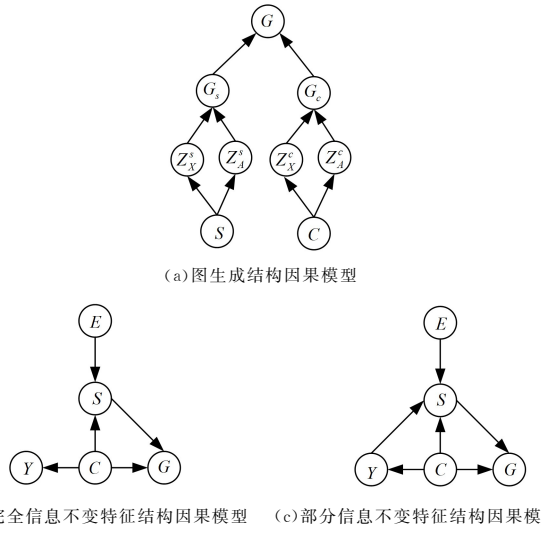


图 2　文本共现图的构建

Fig. 2　Construction of text co-occurrence graph

在进行文本分类前，首先对文本进行预处理，去掉一些停用词，因此在图 2 的例子中，只选取了“Jack”“basketball”“star”“who”“won”“many”“awards”作为图中的节点，同时设定滑动窗口大小为 3。因此“Jack”和“basketball”两个单词节点间存在一条无向边。



(a)图生成结构因果模型

(b)完全信息不变特征结构因果模型　(c)部分信息不变特征结构因果模型

图 3　图分布偏移下的结构因果模型

Fig. 3　SCMS on graph distribution shifts

3.2　构建结构因果模型

在图生成过程中，假设图是通过映射 $f_{\text{gen}}: z \to G$ 生成的，其中 $z \subseteq R^n$ 表示潜在空间，$G = \bigcup_{N=1}^{\infty} \{0,1\}^N \times R^{N \times d}$ 表示图空间，$\boldsymbol{X}$ 表示顶点特征，$\boldsymbol{A}$ 表示邻接矩阵，E 表示环境。我们将潜在变量划分为不变部分 $C \in \boldsymbol{R}^{n_c}$ 和变化部分 $S \in \boldsymbol{R}^{n_s}$，$n = n_c + n_s$。我们在假设 1 和图 3 中详细阐述了图生成过程的结构因果模型(Structural Causal Model，SCM)，为了简单起见省略了结构方程中的噪声。

假设 1　图生成结构因果模型：$G_c = f_{\text{gen}}^{G_c}(C)$，$G_s = f_{\text{gen}}^{G_s}(S)$，$G = f_{\text{gen}}^{G}(G_c, G_s)$，其中 f_{gen}^{G} 分解为 $f_{\text{gen}}^{G_c}$ 与 $f_{\text{gen}}^{G_s}$，其分别控制 G_c 与 G_s 的生成。G_c 继承了 C 的不变信息，其不受 $\boldsymbol{E}$ 的影响，G_s 的生成受到 E 的影响，因此 G_s 继承了关于 Y 的虚假特征。

将 C 和 S 之间的交互分为部分信息不变特征(Partially Informative Invariant Feature，PIIF)和完全信息不变特征(Fully Informative Invariant Feature，FIIF)，两者的区别在于不变部分是否完全提供关于标签 Y 的信息，相应的 SCM 的格式定义如下：

假设 2　FIIF Structural Casual Model：

$Y = f_{\text{inv}}(C)$，$S = f_{\text{spu}}(C, E)$，$G = f_{\text{gen}}(C, S)$

假设 3　PIIF Structural Causal Model：

$Y = f_{\text{inv}}(C)$，$S = f_{\text{spu}}(Y, E)$，$G = f_{\text{gen}}(C, S)$

在上述两个 SCM 中，f_{gen} 对应假设 1 中的图生成过程，f_{spu} 描述 S 如何在潜在空间中受到 C 和 E 的影响。根据定义，FIIF 中 S 由 C 直接控制，PIIF 中 S 由 C 通过 Y 间接控制。$f_{\text{inv}}: C \to Y$ 表示标签过程，仅根据 C 为相应的 G 分配标签 Y。分类任务的必要分离假设是当给定 Y 时，C 比 S 更好聚类，即假设 4。

假设 4　聚类性质：$H(C|Y) \leqslant H(S|Y)$，其中 H(·)为信息熵。

3.3　构建基于图的学习框架

为了扩展 SCM 下图的不变性原理，需要在 PIIF 和 FIIF 下识别一组与 Y 具有稳定因果关系的变量。根据 ICM 假设，$C \to Y$ 不受其他过程的影响，即对于环境变量 E 的干预，条件分布 $P(Y|C)$保持不变。

Invariant-GCN 将 GCN 分解为两个子组件，1)特征器 g：$G \to \hat{G}_c$；2)分类器 $f_c: \hat{G}_c \to Y$。g 和 f_c 的学习目标为：

$$\max_{f_c, g} I(\hat{G}_c; Y) \quad \text{s.t. } \hat{G}_c \perp\!\!\!\perp E, \hat{G}_c = g(G) \tag{1}$$

其中，最大化 $I(\hat{G}_c; Y)$等价于最小化 $R(f_c(\hat{G}_c))$边界。G_s 与 Y 共享部分信息，若只最大化 $I(\hat{G}_c; Y)$可能会导致 $\hat{G}_c$ 包含 G_s 子图，同时 E 的不可用性使得识别 G_c 更具有挑战性。

为了缓解上述问题，需要将 G_c 的属性转换为可微且满足 G_c 与 $\boldsymbol{E}$ 条件独立的目标。假设不变子图 G_c 具有相同大小的 s_c，当最大化 $I(\hat{G}_c; Y)$时，FIIF 和 PIIF 都能将 G_s 引入 $\hat{G}_c$。因此，新的目标需要从 $\hat{G}_c$ 中消除 G_s 的部分子图，使得估计的 $\hat{G}_c$ 只能包含 G_c。

在 FIIF 和 PIIF 的 SCM 下，当两个环境 e_1 和 e_2 具有相同因果因素 c 时，$\hat{G}_c^{e_1}$ 和 $\hat{G}_c^{e_2}$ 将有很高的互信息，即$(G_c^{e_1}, G_c^{e_2}) \in \arg\max I(\hat{G}_c^{e_1}; \hat{G}_c^{e_2})$。但在同一环境 e_1 下，当识别出不同的因果因素 c 与 c_1 时，若 $\hat{G}_c^{e_1}$ 和 $\hat{G}_{c_1}^{e_1}$ 都包含 $G_c^{e_1}$ 和 $\hat{G}_s^{e_1}$ 的子图，则会增加两者的互信息，因此应让$(G_c^{e_1}, \hat{G}_{c_1}^{e_1}) \in \arg\min I(\hat{G}_c^{e_1}; \hat{G}_{c_1}^{e_1})$。$G_c$ 有一个重要属性：

$$G e_{1_c} \in \arg\max_{\hat{G}_C^{e_1}} I(\hat{G}_c^{e_1}; \hat{G}_c^{e_2} | C=c) - I(\hat{G}_c^{e_1}; \hat{G}_{c_1}^{e_1} | C=c, c_1 \neq c) \tag{2}$$

在 FIIF 和 PIIF 的 SCM 下，C 和 Y 都具有稳定的因果关系，Y 可以替换式(2)中的 C。此外，当 $I(\hat{G}_c^{e_1}; \hat{G}_c^{e_2} | C=c)$和 $I(\hat{G}_c; Y)$最大时，$I(\hat{G}_c^{e_1}; \hat{G}_{c_1}^{e_1} | C=c, c_1 \neq c)$自动最小。式(2)适用于任何环境，因此可去除环境上标。用式(2)替换式(1)中的条件独立，即：

$$\max_{f_c, g} I(\hat{G}_c; Y) \quad \text{s.t. } \hat{G}_c \in \arg\max_{\hat{G}_c = g(G), |\hat{G}_c| \leqslant s_c} I(\hat{G}_c; \tilde{G}_c | Y) \tag{3}$$

其中，$\widetilde{G}_c=g(\widetilde{G})$，$\widetilde{G}\sim P(G|Y)$，$\widetilde{G}$ 是从与 G 共享相同标签 Y 的训练图中采样而来。本文使用有监督采样的对比学习来计算 $I(\hat{G}_c;\widetilde{G}_c|Y)$。

$$I(\hat{G}_c;\widetilde{G}_c|Y)\approx -L_{\hat{G}_c,\widetilde{G}_c,G_c^i} \tag{4}$$

$$L_{\hat{G}_c,\widetilde{G}_c,G_c^i}=-E_{\substack{\hat{G}_c,\widetilde{G}_c\sim P_g(G|y=Y)\\ \{G_c^i\}_{i=1}^M\sim P_g(G|y\neq Y)}}\log\frac{e^{\phi(h_{\hat{G}_c},h_{\widetilde{G}_c})}}{e^{\phi(h_{\hat{G}_c},h_{\widetilde{G}_c})}+\sum_{i=1}^{M}e^{\phi(h_{\hat{G}_c},h_{G_c^i})}} \tag{5}$$

其中，$(\hat{G}_c,\widetilde{G}_c)$是提取与 G 共享相同标签的子图，G_c^i 是提取与 G 不同标签的子图；$h_{\hat{G}_c}$，$h_{\widetilde{G}_c}$ 以及 $h_{G_c^i}$ 是估计子图的图表示；ϕ 是图表示的相似度量。最终将识别出的 $\hat{G}_c$ 输入到后续主干 GNN 中，对不变子图做图卷积操作。将卷积的结果再经过 softmax 函数进行映射得到分类预测结果。

3.4 模型实施步骤

本算法的伪代码如算法 1 所示，算法输入为训练集的文本数据 D 以及对应标签 y，输出是文本分类的主干网络以及分类器的可学习参数 θ 与 ω，其中 window_size 为构建图结构过程中滑动窗口大小。

算法 1 Invariant-GCN

Input: D, y
Input Parameters: window_size
Initialize: backbone g(.;θ), classifier f(.;ω)
1. D ← preprocess(D) {文本预处理}
2. X ← get(D) {得到预训练词向量}
3. G_{full} ← construct(X, window_size) {构建文本共现图}
4. for M epochs do
5. G_{sub} ← subgraph(X, G_{full}) {获取不变子图}
6. for N epochs do
7. G ← f(X, G_{sub};θ)
8. $\hat{y}$ ← CLS(G;ω)
9. θ, ω∈ argmin L(y, $\hat{y}$)
10. end
11. Output: backbone g(.;θ), classifier f(.;ω)

4 实验

为了验证本文模型的有效性，在 4 个公开数据集上做了文本分类实验，使用正确率作为评价模型性能的指标。

4.1 实验器材

本文实验环境和机器配置如表 1 所列。

表 1 实验配置表

Table 1 Experiment configurations

实验环境	环境配置
操作系统	Linux
GPU	NVIDIA RTX 3090
编辑语言	Python 3.9
深度学习框架	PyTorch, PyTorch Geometric

4.2 数据集

MR 是一个电影评论数据集，该数据集将电影评论分为积极或消极评论。Ohsumed 数据集来自医学文献数据库 MEDLINE，其将医学摘要分类为 23 种心血管疾病。R8 和 R52 分别是包含 8 个类别和 52 个类别的新闻文本的 Reuters 新闻数据集。数据集统计信息如表 2 所列，其中 Prop. NW 表示在测试中新词的比例。

表 2 数据集统计信息

Table 2 Statistics of datasets

Dataset	# Training	# Test	# Classes	Prop. NW/%
MR	7108	3554	2	30.07
Ohsumed	3357	4043	23	8.46
R8	5485	2189	2568	2.60
R52	6532	8	52	2.64

4.3 基准算法

为评估本文模型的有效性，我们考虑了 3 种类型的模型作为基线：1）传统的深度学习方法，包括 TextCNN[11] 和 TextRNN[12]；2）基于单词特征的简单但有效的策略，包括 fastText[18] 和 SWEM[19]；3）基于图的文本分类方法 TextGCN[14]。

TextCNN：该模型使用预训练的 GloVe 词向量和 CNN 网络对输入文本进行嵌入和卷积操作，最后通过全连接层进行分类。

TextRNN：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。

fastText：一种快速的文本分类方法，该模型只有 3 个网络层，模型简单，训练速度快。

SWEM：一种简单快速的文本分类方法，把预训练词向量简单池化后，直接进行文本分类。

TextGCN：使用紧耦合的方式将整个语料库构建成一张异构图，然后使用图卷积神经网络学习文本的嵌入表示。

4.4 实验参数设置

所有数据集给出了训练集和测试集，将训练集以 9∶1 的比例划分为训练集和验证集，超参数根据验证集的性能进行调整。在构建文本图结构的过程中，滑动窗口大小设置为 20，滑动窗口的步长设置为 1。图中最大节点数设置为 350，如果超过 350，则只截取前 350 个。用 Glove 词向量方法表示单词向量，其嵌入维度设为 300，对词汇表外的词进行随机初始化。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，dropout 率设为 0.5，批处理大小设置为 64。

4.5 实验结果

为减小各模型的对比误差，本文将每个模型分别在相应的数据集上运行 10 次，然后采用均值加减标准差的形式获得实验结果。各模型在 Ohsumed，R8，R52 和 MR 数据集上的分类准确率如表 3 所列，其中粗体部分表示各数据集上取得的最优分类结果。表 3 中 TextGCN 和本文方法均为图网络模型。

表 3 不同方法的分类准确率对比

Table 3 Classification accuracy comparison of different methods

Model	MR	Ohsumed	R8	R52
TextCNN[11]	77.75±0.72	58.44±1.06	95.71±0.52	87.59±0.48
TextRNN[12]	77.68±0.86	49.27±1.07	96.31±0.33	90.54±0.91
fastText[18]	75.14±0.20	57.70±0.49	96.13±0.21	92.81±0.09
SWEM[19]	76.65±0.63	63.12±0.55	95.32±0.26	92.94±0.24
TextGCN[14]	76.74±0.20	68.36±0.56	97.37±0.13	93.56±0.18
ours	**78.61±0.15**	**69.83±0.51**	**97.07±0.10**	**94.17±0.20**

由表 3 可以看出，基于图网络的模型明显比其他类型的模型在性能上更加出色，有利于文本处理。此外，本文方法在

4个数据集上取得了最优结果，这表明识别文本图结构的不变子图进行文本分类是有效的。

结束语　本文提出了一种新颖的基于图神经网络的文本分类方法，每个文本有自己的结构图，可以学习文本级的单词交互，将传统文本分类问题转换为图分类问题。首先为每个文档构建单个图，使用GCN根据其局部结构学习细粒度的单词表示；然后提取最大限度地保留不变类内信息的期望子图，使用这些子图进行学习不受分布变化的影响；最后通过图分类方法完成文本分类。实验结果表明，从因果角度考虑将不变性原理应用到文本中提升了文本分类效果。未来我们将考虑通过反事实解释生成反事实数据进行数据增强，进一步提升文本分类效果。

参 考 文 献

[1] ROY P K, SINGH J P, BANERJEE S. Deep learning to filter SMS Spam[J]. Future Generation Computer Systems, 2020, 102: 524-533.

[2] ZHANG L, WANG S, LIU B. Deep learning for sentiment analysis: A survey[J]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018, 8(4): e1253.

[3] ZENG Z, HE K, YAN Y, et al. Modeling Discriminative Representations for Out-of-Domain Detection with Supervised Contrastive Learning[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). 2021: 870-878.

[4] 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 (Volume 1: Long Papers). 2017: 562-570.

[5] 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.

[6] WAN S, PAN S, YANG J, et al. Contrastive and generative graph convolutional networks for graph-based semi-supervised learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(11): 10049-10057.

[7] AHUJA K, CABALLERO E, ZHANG D, et al. Invariance principle meets information bottleneck for out-of-distribution generalization[J]. Advances in Neural Information Processing Systems, 2021, 34: 3438-3450.

[8] 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.

[9] JOACHIMS T. Text categorization with support vector machines: Learning with many relevant features[C]// European Conference on Machine Learning. Berlin, Springer, 1998: 137-142.

[10] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.

[11] KIM Y. Convolutional Neural Networks for Sentence Classification[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2014.

[12] LIU P, QIU X, HUANG X. Recurrent neural network for text classification with multi-task learning[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016: 2873-2879.

[13] 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.

[14] YAO L, MAO C, LUO Y. Graph convolutional networks for text classification[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 7370-7377.

[15] 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: 3444-3450.

[16] 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.

[17] 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.

[18] JOULIN A, GRAVE É, BOJANOWSKI P, et al. Bag of Tricks for Efficient Text Classification[C]// Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers. 2017: 427-431.

[19] SHEN D, WANG G, WANG W, et al. Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2018.

HUANG Rui, born in 1999, postgraduate, is a member of CCF (No. N8705G). Her main research interests include machine learning and natural language processing.

XU Ji, born in 1979, Ph.D, professor, is a member of CCF (No. 12919M). His main research interests include data mining, granular computing and machine learning.