基于转折词的图卷积短语音 - 短文本模态转换的分类方法

徐克圣[1] 毛寅辉[1] 陈胜男[1]
XU Kesheng MAO Yinhui CHEN Shengnan

摘 要 提出了一种增加转折词后实体词注意力权重的短文本分类方法，旨在提高短文本分类的准确性和可靠性。所提出的方法结合了文本构图和图卷积网络技术，通过将文本数据转换为图形结构，利用图卷积神经网络对图形结构进行特征提取和模式识别，以捕捉文本数据的内在结构和语义关系。在训练过程中，使用转折词和置信度高的实体词作为关键信息，通过注意力机制强化这些信息在分类中的作用。通过多次循环训练，得到了一个高效的文本分类模型。实验结果表明，所提出的模型在短文本分类任务中具有较好的性能表现，能够有效提高分类的准确性和可靠性。为了验证模型的性能和泛化能力，选取了三个公开的短文本数据集 Ohsumed、AGNews 和 MR 数据集以及一个公开的短语音数据集 MELD 数据集。这些数据集具有不同的主题和领域，可以更好地评估模型的泛化能力。实验结果表明，所提出的模型在四个数据集上都取得了优于基线的分类效果，证明了模型的有效性和泛化能力。

关键词 图卷积网络；文本构图；注意力机制；短文本；语音数据

doi：10.3969/j.issn.1672-9528.2024.05.006

0 引言

随着移动互联网的迅猛发展和智能终端的普及，短视频软件已经成为人们日常生活中不可或缺的一部分。这些软件不仅提供了丰富的视频内容，还为人们提供了一个自由发表观点和交流思想的平台。然而，随着用户数量的不断增加，评论区中充满了海量的短文本信息和简短语音信息，使得网络言论的监管面临巨大的挑战[1]。为了更好地管理这些网络言论，对短文本和简短语音评论进行分类成为了一项迫切的任务。

传统的文本分类方法通常基于词袋模型或卷积神经网络，这些方法在处理长文本时表现较好，但在处理短文本和简短语音时，由于其信息量少、语义不完整等特点，这些方法无法很好地理解语义，导致分类精度不高[2]。此外，现有的短文本分类研究很少涉及对简短语音的分类，这使得对这一新型评论形式的监管和处理变得困难。

针对这些问题，本研究提出了一种基于图卷积神经网络的短文本和简短语音评论分类方法。与传统的文本分类方法相比，该方法能够更好地理解语义，并具有更强的泛化能力。此外，该方法还具有较好的泛化能力，可以应用于不同类型的短文本和简短语音评论的分类任务中。本研究不仅为短文本和简短语音评论分类提供了一种新的思路和方法，也为网络言论监管提供了有益的参考。

本文的主要贡献主要体现在以下两个方面。

（1）引入模态转换的思想，将不方便进行分类的语音模态信息转化为文本模态，利用发展较为成熟的文本分类技术解决语音分类的难题。

（2）提出转折词后实体语义增强的概念，在对短文本进行预处理时提取出转折词，使用命名实体技术识别出各类实体，对转折词后的实体添加不同的注意力权重，以提高分类任务时的准确率。

1 相关工作

早期，研究人员使用传统机器学习和手工特征进行短文本分类。后来，词嵌入技术和深度学习模型（如 CNN[3]、RNN、LSTM、GRU）的发展大幅度提升了分类性能。如今，预训练模型如 BERT、GPT 在大规模语料上预训练后，经微调就能显著提升短文本分类效果。

近些年来，图神经网络（GNN）发展飞速，图卷积神经网络（GCN[4]）能将图中的特征提取出来。基于这样的能力，短文本在 GCN 上的研究得到了广泛的关注，2019 年 Yao 等人[5]将整个语料库构建为一个图，并将文本分类的问题转化成了节点分类的问题，为如今的文本分类奠定了非常好的想基础。还有学者引入了主题和实体节点[6]来丰富短文本的语

1. 大连交通大学软件学院 辽宁大连 116028
[基金项目] 国产化公链基础软件研发与产业化（2022JH2/101300269）

义，并利用注意力机制来学习节点的重要性。又有许多学者在 TextGCN 模型的基础上进行了很多的改进，并衍生出许多更加优秀的模型，例如后来提出的 TensorGCN 模型 [7]，该模型在前人的基础上从语义、句法和序列三个角度构建了三张异构图，它有两种不同的传播方式，分别是图内的信息传播和图间的信息传播。Cui 等人 [8] 为了克服缺乏标记数据而造成的问题，提出了一种基于图卷积网络的自训练文本方法（ST-Text-GCN），这种方法没有引用额外的知识，而是计算每个单词的置信度，将高置信度的词加入训练样本中继续训练。这种方法主要是基于语料库构建一个文本图，并利用全局信息来优化文本的表示方式。然而，它仅仅依赖了文本内部的关联，并没有利用任何外部信息来进一步丰富文本信息，因此对文本的语义理解能力仍需继续提高。

基于此问题，本文提出将增加转折词后实体词注意力权重的方法应用于 ST-Text-GCN 模型，为短文本提供更丰富的特征表示，并使用实体链接技术引入了外部知识，从而提高分类性能。

2 模型描述

在深入了解基于图卷积网络的短文本分类之后，本文提出的基于转折词的图卷积短语音 - 短文本模态转换分类模型（transition-driven graph convolutional short audio-short text mode conversion classification model，TDGCM），如图 1 所示。

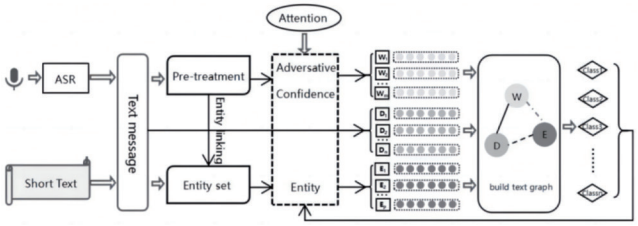

图 1 基于转折词的图卷积短语音 - 短文本模态转换分类模型

本模型通过自动语音识别技术（ASR）将短语音音频识别为短文本信息，同短文本数据一起进行数据清洗，比如将所有文本数据转换为小写格式，去除一些无用的停用词，去除定义的特殊符号、识别转折词等功能；然后对数据集进行短文本在单词 - 实体 - 文档节点之间建立边来构造文本图。词向量经过预处理之后输入到图卷积网络中执行训练过程。

TDGCM 的主要过程可以分为以下步骤：模态统一、转折词与实体识别、注意力机制与特征融合、文本构图以及自循环图卷积网络层。下面将详细解释模型的每个部分。

2.1 模态统一

本文将输入模型的语音数据集使用自动语音识别系统 [9]（ASR）进行处理将其转化为文本，ASR 系统主要包括 4 部分：特征提取、声学模型（AcousticModel）、语言模型（LanguageModel）和解码。本次实验先对语音数据集 MELD[10] 进行降噪、去噪音等预处理，使用长短时记忆网络模型对语音信号进行特征提取并分类，以实现将语音转换为文字的目的，再使用 n-gram 模型来提高语音识别的准确性，解码之后可得到相应的文本识别结果。模态转换部分的流程图如图 2 所示。

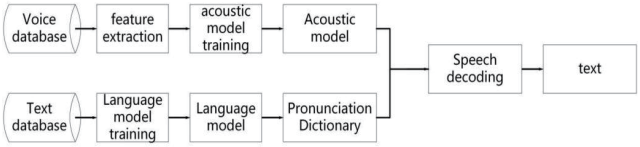

图 2 自动语音识别系统（ASR）训练过程

2.2 转折词与实体识别

首先对短文本数据和语音识别出的文本进行分词处理，使用词性标注工具 NLTK 对句子进行词性标注，通过依存句法分析技术来识别转折词。例如，“but”“though”等都是常见的转折词。然后运用命名实体识别技术，从短文本中精准识别出各类实体，使用 TagMe 这一实体链接工具，将文本中的各类实体单词精准地映射到维基百科的对应实体上，提取出与该实体相关的语义信息。最后将提取的语义信息与原始实体相结合，以丰富实体的语义表示。

2.3 注意力机制特征融合

根据实体的特征和上下文信息计算其置信度，选择置信度高于 0.9 的实体和转折词后的实体作为关键字。为选定的关键字计算注意力权重，权重值可以根据其置信度计算得出。

在自然语言处理方面，注意力机制对于准确识别句子含义起到了莫大的作用，通过对句子中不同的词添加不同的注意力权重，可以有效提高短文本分类的准确率。本文需要对已识别出的转折词之后的实体和置信度较高的单词添加注意力，并计算每个词相应的权重。首先在图卷积神经网络中，将每个单词的特征向量与其对应的注意力权重相乘，可得到加权后的特征向量。然后进行特征提取和分类，图卷积神经网络对这些加权特征进行进一步的处理和整合。最终输出分类结果。在训练过程中，使用梯度下降算法不断更新网络的参数，使得模型性能不断优化。

2.4 文本构图

本文构造单词 - 文档 - 实体异构图 $G=(v,\varepsilon)$，v 为节点集，ε 为边集。节点集 v 中分别包括单词节点 $W=\{w_1,w_2,\cdots,w_m\}$、文档节点 $D=\{d_1,d_2,\cdots,d_n\}$ 以及实体节点 $E=\{e_1,e_2,\cdots,e_p\}$，边集 ε 中的元素代表的是各节点之间的关系，使用 TF-IDF 乘以单词置信度可以确定文档节点和单词节点之间的边的权重。运用点互信息（PMI）方法来评估两个单词间的权重，从而决定它们在图网络中的连接强度。当建立文档节点与实体节

点之间的连接时，依据的是文档中单词映射到维基百科实体的精确性。对于图中的任意两个节点 m 和 n，其连接的权重可以用邻接矩阵 A 中的 $A_{m,n}$ 值来表示：

$$A_{m,n}=\begin{cases} TF\text{-}IDF\times \mathrm{Con}_{m,n} & m\text{是文档，}n\text{是单词} \\ PMI(m,n) & m\text{，}n\text{都是单词} \\ \mathrm{score}_{m,n} & m\text{是文档，}n\text{是实体} \\ 0 & \text{其他} \end{cases} \tag{1}$$

式中：PMI 描述了局部共现语言属性，计算出两个单词之间的关联程度，公式为（只考虑 PMI 为正的情况）：

$$PMI(m,n)=\frac{p(m,n)}{p(m)p(n)} \tag{2}$$

当 PMI 的数值增大时，两个单词之间的语义相关性也随之增强。为削弱歧义词可能带来的干扰，可进一步计算每个词的置信度。这一置信度的计算是基于带有标签的文档进行的，其中包括了带有预测标签的训练文本和测试文本。通过这种方法，能够更加准确地理解和使用每个词，从而提高整体文本处理的性能和效果。单词置信度计算公式为：

$$\mathrm{Con}_{m,n}=\begin{cases} \dfrac{Max(C_0,C_1,\ldots,C_k)}{\sum_{C=0}^{K}C_C} \\ \dfrac{1}{K} \end{cases} \tag{3}$$

式中：C_C 是包含单词 w_i 文档中类别 C 的文档数，K 是类别总数，Max 是最大函数。

本文使用的是一个简单的双层图卷积神经网络（two-layer GCN）构建成功的文本图输入其中，通过两层 GCN 网络对节点进行嵌入表示，然后将这些嵌入表示送入分类器进行分类。具体公式如下：

$$L^{(n+1)}=\sigma(\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{\frac{1}{2}}L^{(n)}\boldsymbol{W}_n) \tag{4}$$

式中：$\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{1/2}$ 是一个经过对称归一化的拉普拉斯矩阵，$\boldsymbol{A}$ 代表文本异构图的邻接矩阵，$\boldsymbol{D}$ 表示图中节点的度矩阵。σ 为 *softmax* 激活函数，$\boldsymbol{W}$ 为权重矩阵，j 为层数，$L^{(0)}=X$。

2.5 自循环的图卷积网络层

本实验构建了一个图卷积神经网络模型，该模型纳入了单词、文档和实体三种节点之间的关系。将实体、单词和文档视为节点，运用图卷积网络（GCN）进行构图和建模，通过利用词在文档中的出现（document-wordedges）和词在整个语料库中的共现（word-wordedges）以及实体链接技术，将文档中的单词与其对应的实体进行关联，并计算这种关联的置信度作为文档与实体之间边的权重来构建节点之间的边。自循环模块则将是关键字的单词作为伪标签数据添加到训练数据集中，经过多轮的训练和更新，最终输出所属类别。

3 实验与分析

3.1 数据集

为了确保评估的公正性和全面性，本文选择在三个公开的短文本数据集和一个公开的短语音数据集上进行了模型验证实验。三个短文本数据集分别是 Ohsumed、AGNews 和 MR 数据集，短语音数据集是 MELD 数据集。Ohsumed 数据集中包含了 13 929 篇独特的心血管疾病摘要集合，这个集合中的每个文档都有 23 个疾病类别中的一个或多个相关类别。还有 World、Sports、Business、Sci/Tec 四个类别的英文新闻数据集 AGNews，以及关于电影评论情感的 MR 数据集，其中 MR 数据集分为正向和负向两种情感类别。MELD 数据集是从电影老友记上摘取的片段，是一个多模态数据集，既有文本信息，也有与文本对应的音频和视频信息。MELD 中有超过 1400 组对话，总共 13 000 句。将 MELD 数据集的音频通过 ASR 系统转化为文本数据集，接下来为处理 Ohsumed、AGNews 数据集以及转化后的文本数据集，使用 NLTK 库去除了停用词和出现频率少于 5 次的单词。但由于 MR 数据集中的文本非常短，所以并未去除停用词或低频词，以此保证实验的准确性和可靠性。三个短文本数据集基本情况见表 1，短语音数据集基本情况见表 2。

表 1 Ohsumed、AGNews 和 MR 数据集基本情况

数据集	实例	train 数据	test 数据	类别数	平均长度
AGNews	1050	1000	950	4	38.87
Ohsumed	275	250	350	23	11.91
MR	35	30	45	2	21.02

表 2 MELD 数据集基本情况

MELD 数据集	train 数据	test 数据
独特单词数	10 643	4361
平均话语长度	8.03	8.28
最大话语长度	69	45
对话数	1039	280
演讲者人数	260	100

3.2 模型对比

为了全面评估本文模型的性能，本文选取了 7 种在文本分类领域表现卓越的基线模型作为对比对象。以下是这 7 种比较模型的详细信息。

（1）TF-IDF+LR：一种文本分类方法，它结合了 TF-IDF 特征提取技术和逻辑回归算法。

（2）CNN：卷积神经网络，研究了使用随机初始化单词嵌入的 CNN rand 和使用预训练单词嵌入的非静态 CNN。

（3）BiLSTM[11]：双向 LSTM，通常用于文本分类，将预先训练的单词嵌入输入到 Bi-LSTM 中。

（4）TextGCN：文本 GCN 将语料库构建成图，并将 GCN 应用于文本分类。

（5）TensorGCN：利用多维张量对文本进行建模，并使用图卷积操作对文本中的复杂关系进行捕捉和学习。

（6）ST-Text-GCN：基于图卷积神经网络（GCN）的自训练短文本分类模型。

（7）ETGCN[12]：融合了实体信息的图卷积模型。

3.3 实验结果与分析

3.3.1 实验环境

本实验使用 PyTorch 框架，在 NVIDIAGTX1080TiG-PUIntel(R)Core(TM)i5-11320H 上进行训练和测试。本次实验对基础数据进行预先设定，将词嵌入维度预设为 150，学习率设为 0.001，滑动窗口的大小预先设置成 20。为防止出现过拟合的现象，将损失率设为 0.5。本次实验采用了 adam 随机梯度优化算法对模型参数进行更新，并设定了最多进行 200 次迭代的实验计划。但是，如果在连续的 10 次迭代中，模型的性能没有显示出任何的改善，就会提前结束训练。

3.3.2 对比实验

模型性能通过分类准确率来评估，准确率愈高，模型表现愈佳，以下是 8 种模型在 4 个数据集上的实验结果，如表 3 所示。

表 3 8 种模型在 4 个数据集上的准确率

模型	Ohsumed /%	AGNews /%	MR /%	MELD /%	均值 /%
TF-IDF+LR	60.26	88.09	74.58	78.64	75.39
CNN	57.22	83.95	77.72	80.73	74.91
BiLSTM	58.76	83.77	77.64	81.07	75.31
TextGCN	62.31	86.80	76.78	83.49	77.35
TensorGCN	61.28	87.61	77.92	83.47	77.57
ST-Text-GCN	65.31	88.83	78.84	86.52	79.87
ETGCN	64.42	88.39	82.88	87.39	80.77
TDGCM	65.88	91.28	83.13	89.55	82.46

在表 3 中可明显看出，采用图卷积神经网络的模型在性能上超越了传统的神经网络模型。这得益于图结构的特点，即能够促进异类邻居节点间的信息传递，进而让节点能集成更多信息以丰富其特征表达。另外，相比传统模型局限于局部信息共享的缺陷，该模型利用单词间的词共现特征作为边的权重，更有利于全局信息共享，这也是其性能出色的一大原因。由于基于图的神经网络能够有助于挖掘单词和文档之间的关系信息，它可以通过捕获全局信息来提高分类的性能。从均值的准确率看，本次实验的模型比 ST-Text-GCN 和 Text-GCN 模型分别提高了 3.24% 和 6.61%，性能有了相当大的提升。TDGCM 模型在文本图中融入了加权的实体信息，增强了短文本的语义理解能力，使得本模型在短文本分类任务中，展现了出色的性能。

3.3.3 消融实验

接下来，本文使用消融实验来验证引入加强转折词后的实体词的注意力机制和链接实体信息对本模型的影响，由于原始模型没有加强转折词后实体词的注意力这一概念且没有引入外部知识，故而将转折词后的实体节点和链接实体信息作为变量，来验证本模型的有效性。如表 4 所示，结果显示，引入实体节点的文本图比仅由单词 - 文档之间的包含关系和词共现关系所构成的文本图效果要好。在引入外部信息之后，本模型的效果得到了有效的证实。

表 4 消融实验结果

	Ohsumed/%	AGNews/%	MR/%
原始模型	65.31	88.83	78.84
+ 注意力机制	69.44	91.97	83.23
+ 注意力机制 + 链接实体信息	70.88	93.28	85.13

3.3.4 参数对模型的影响实验

本次实验深入研究了滑动窗口尺寸和词嵌入维度对模型精度的影响。在确保其他参数一致的前提下，选取了滑动窗口大小为 5、10、15、20、25 和 30 的情况，并运用 TDGCM 模型在 Ohsumed 和 MR 数据集的测试集上进行了系统的实验，结果见图 3。

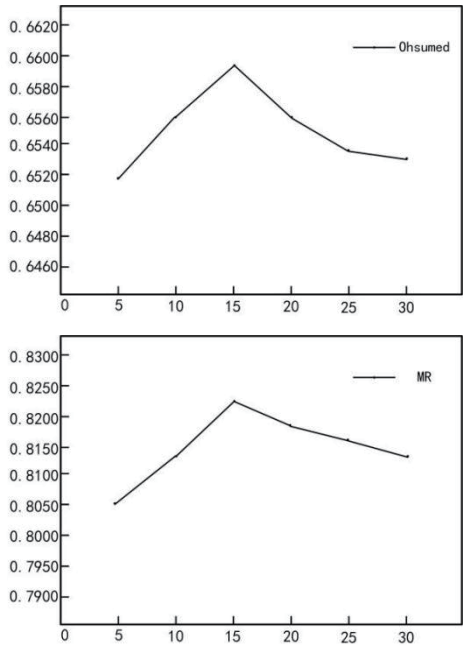

图 3 不同滑动窗口尺寸的 TDGCM 模型的 F_1 值

由图 3 可见，随着滑动窗口尺寸的增大，测试精度开始呈现上升趋势。然而，当窗口尺寸超过 15 时，平均精度便不再提高。这表明，一方面过小的窗口尺寸可能无法充分捕捉到全局的词共现信息，从而限制了模型性能的发挥；另一方面，过大的窗口尺寸可能会引入一些不太相关的节点之间

的连接，这可能会对模型的性能产生负面影响。保持其他参数不变，分别取词嵌入维度为 50、100、150、200、250 和 300，使用 TDGCM 模型在 MR 数据集的测试集上进行实验，结果见图 4。

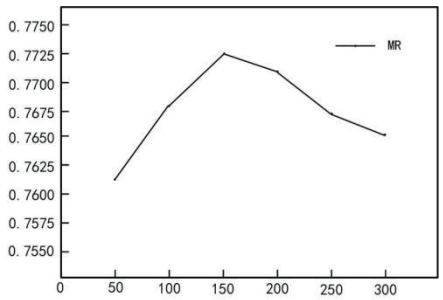

图 4 不同嵌入维度的 TDGCM 模型的 F_1 值

观察图 4 可知，当词嵌入维度为 150 时，F_1 值最高。同时也发现了与图 3 类似的趋势，过低的嵌入维度可能无法有效地将标签信息传递到整个图网络，而高维的嵌入无法提高分类性能，还可能会增加训练时间成本。

4 结论

本文介绍了一种创新的图卷积神经网络模型，该模型可将简短音频数据转换为短文本数据，实现了跨模态的文本分类方法，并且增加了转折词的后实体的注意力权重，使得在进行图神经网络的过程中能够更加准确地识别出短文本信息的含义。本模型结合了命名实体识别技术，构建了单词、文档和实体之间的异构文本图。这种模型可以有效地利用命名实体信息，提升文本处理的性能，还从外部知识库中引入实体信息，使模型可以捕捉到更加丰富的语义信息。而且使用 two-layerGCN 构建的网络结构，能进一步提高模型的表达能力和泛化能力。因此，在处理复杂的自然语言任务时，本模型取得了优于基线的表现。

参考文献：

[1] 祝利杰，罗迪凡，史彦丽 . 局部加权稀疏表示的文本分类算法研究 [J]. 信息技术与信息化 ,2023(8):24-27.

[2] 杨秀璋，武帅，张苗，等 . 基于 TextCNN 和 Attention 的微博舆情事件情感分析 [J]. 信息技术与信息化 ,2021(7):41-46.

[3]AHMED M R, SALEKUL I, SALEKUL A K M, et al. An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition[J].Expert systems with application, 2023, 218(5):1-21.

[4]KIPF T N，WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL].(2016-09-09)[2024-01-20]. https://arxiv.org/abs/1609.02907.

[5]YAO L, MAO C, LUO Y.Graph convolutional networks for text classification[C]//33rd AAAI Conference on Artificial Intelligence,v.9.AAAI Technical Tracks: Natural Language Processing; Planning, Routing, and Scheduling. Palo Alto:Association for the Advancement of Artificial Intelligence, 2019:7370-7377.

[6] HU L, YANG T, SHI C, et al. Heterogeneous graph attention networks for semi-supervised short text classification[C]// Conference on empirical methods in natural language processing 2019 and 9th international joint conference on natural language processing,vol. 8. Stroudsburg, PA: Association for Computational Linguistics, 2019:4820-4829.

[7]LIU X E, YOU X X, ZHANG X, et al. Tensor graph convolutional networks for text classification[J].Proceedings of the AAAI conference on artificial intelligence, 2020, 34(5): 8409-8416.

[8]CUI H, WANG G, LI Y, et al.Self-training method based on GCN for semi-supervised short text classification[J].Information sciences: an international journal, 2022,611:18-29.

[9] 高长丰，程高峰，张鹏远 . 面向鲁棒自动语音识别的一致性自监督学习方法 [J]. 声学学报 ,2023,48(3):578-587.

[10]PORIA S, HAZARIKA D, MAJUMDER N, et al. MELD: a multimodal multi-party dataset for emotion recognition in conversations [C]//57th annual meeting of the Association for Computational Linguistics,vol. 1.Stroudsburg:Association for Computational Linguistics, 2019:517-536.

[11] 汪辉，于璨 . 基于 BERT+BiLSTM+Attention 的对抗训练新闻文本分类模型 [J]. 西安文理学院学报 (自然科学版), 2023, 26(3):49-53.

[12] 王佳宇，李楹，马春梅，等 . 融合实体信息的图卷积神经网络的短文本分类模型 [J]. 天津师范大学学报 (自然科学版), 2023, 43(1):67-72.

【作者简介】

徐克圣（1965—），通信作者（email: Xks65@126.com），男，辽宁大连人，副教授，研究方向：自然语言处理、区块链、软件测试。

毛寅辉（1998—），女，山西朔州人，硕士研究生，研究方向：自然语言处理、区块链。

陈胜男（1998—），女，辽宁大石桥人，硕士研究生，研究方向：区块链分片技术、机器学习。

（收稿日期：2024-03-11）