引用格式：徐泽，曹三省．基于区块链和深度学习的全媒体虚假信息检测方法研究[J]．中国传媒大学学报（自然科学版），2024，31(02)：60-70.
文章编号：1673-4793(2024)02-0060-11

基于区块链和深度学习的全媒体虚假信息检测方法研究

徐泽[1,2]，曹三省[1,3*]
（1.中国传媒大学媒体融合与传播国家重点实验室，北京 100024；2.湖北工程学院，孝感 432000；
3.中国传媒大学信息科学与技术学部，北京 100024）

摘要：在当前的全媒体数字信息环境下，虚假文本，包括假新闻、篡改内容和各类虚假信息，越来越受到关注。为了应对这一挑战，本文引入了一种创新方法，将区块链技术的强大功能与深度学习技术相结合，用于检测虚假文本。该方法构建了一个可信的信息传播网络，保证了信息的安全和透明共享；利用集成的词向量表示来有效地保存内容之间的细粒度关系；采用GRU-CNN模型对输入数据进行训练，训练结果反馈给集成媒体区块链网络。最后的实验结果表明，本文所提出方法在全媒体环境下以短视频为代表的内容任务中检测虚假评论是可行和有效的。
关键词：区块链；虚假信息；深度学习；信息传播；全媒体
中图分类号：G202　**文献标识码**：A

A study of the method of all-media fake information detection based on blockchain and deep learning

XU Ze[1,2], CAO Sanxing [1,3*]
（1.State Key Lab of Media Convergence and Communication, Communication University of China, Beijing 100024, China; 2.Hubei Engineering University, Xiaogan 432000, China; 3. Faculty of Information Sciences and Technologies, Communication University of China, Beijing 100024, China）

Abstract: Within the environment of all-media digital information at present, fake texts, including fake news, falsified content and all types of fake information, are more and more widely focused. Replying to this challenge, in this paper a new method for the detection of fake text was introduced by the combination of the power of the blockchain technology with deep learning. With the proposed method, a trustworthy information communication network that ensured the security and transparent sharing of information was built; by using the integrated presentation of word vectors, the fine-grained relations within the contents was stored; via the training of input data with GRU-CNN, the results were fed to the integrated media blockchain network. The final experiment results indicate that the proposed method is feasible and effective in the detection of fake comments in the typical content tasks of short video in the environment of all-media.
Keywords: blockchain; fake information; deep learning; information communication; all-media

1　引言

近年来，以抖音、快手等为代表的短视频平台迅速发展，已成为影响人们生活、工作、娱乐的重要全媒体产品之一。许多媒体机构或自媒体人借助短视频平台短、平、快的优势，整合碎片化资源，吸引大量粉丝好友

基金项目：中国传媒大学国家级重大项目培育专项项目（CUC22GP004）
作者简介（*为通讯作者）：徐泽（1992-），男，博士，讲师，主要从事互联网信息与网络信息工程方面研究。Email：xuze201609@163.com；曹三省（1977-），男，博士，教授，博士生导师，主要从事互联网信息、媒体融合与人工智能方面研究。Email：sxcao@cuc.edu.cn

并赚取流量，这种新闻创作和交流的个性化，在便利性和相关性方面创造了优势，但也增加了诸如假新闻、虚假宣传和阴谋论等形式的错误信息的风险，对社会造成不利后果，严重影响舆论稳定与社会和谐。

文本作为虚假信息的一种重要载体，其表现形式多样，包括新闻文章、社交媒体帖子、评论回复或聊天信息等。它能够通过数字传播渠道快速便捷地传播，并进而对个人、组织甚至整个社会产生重大影响。文本错误检测作为虚假信息检测的重要一环，它是指识别基于文本内容中的虚假或不准确信息的过程。这个过程包括分析文本的内容，以确定它所呈现的信息是真实和准确的，还是包含虚假、误导或有偏见的信息。通常使用自然语言处理（NLP）技术、情感分析及机器学习算法来分析大量文本，并识别与虚假或误导性信息相关的模式和特征。但由于生成信息的数量和速度、基于文本的内容中使用的语言的多样性和复杂性以及用于生成错误信息的技术越来越复杂，虚假信息检测的难度也越来越高。虚假信息可以被故意设计成可信的，甚至可能包含一些真实的元素，以至于现有成熟的技术方法也无法完全将其与准确的信息区分开来。因此，在前人研究的基础上，本文使用区块链技术创建一个具有去中心化技术属性的、来源可信的平台来共享和验证信息，通过区块链成员管理确保全域融媒区块链网络中节点身份的真实可信，借助激励和共识机制能够提高信息传播环节的内容质量和效果，并使用深度学习算法来分析和识别数据，形成虚假信息判别的语言模式。通过结合这些技术，为抑制全媒体环境下的虚假信息传播提供一种更为有效的方式。

2 相关工作综述

2.1 虚假信息检测研究现状

近年来，在全媒体场景下，针对虚假消息检测问题，许多专家学者已经做了大量的研究工作，主要分为基于传统机器学习的方法和基于深度学习的方法，这些方法在不同场景下考虑的侧重点各不相同。Gilda等人[1]探索了自然语言处理技术在检测假新闻中的应用。而Granik等人的另一篇论文则使用朴素贝叶斯分类器识别假新闻[2]。Vedova等人[3]提出了一种新的机器学习假新闻定位技术，该技术将新闻内容和社会背景结合在一起，在Facebook Messenger聊天机器人中执行了所提策略，并应用该方法获得了81.7%的假新闻发现精度。Tashnim等人将人的因素与人工智能方法结合起来，利用假新闻和恶搞发现，提出了一种杂交的机器群方法，用以区分可能具有误导性新闻的策略[4]。Bajaj提出了一种新颖的设计[5]，将一种类似注意力的机制集成到卷积网络中，可以根据新闻内容预测一条新闻是否是假的，并比较几种不同自然语言处理模型的结果。这些研究大多仅仅是对虚假信息本身处理而无法找到其爆发的根源，也无法找到消除其传播的通用且有效的途径。因此，需要结合新技术研究虚假信息传播机制，从根源上、作用点上尽量减少损失。以往对于信息传播机制的研究涉及物理学、社会学和心理学等多学科的努力。Vega-Oliveros等人[6]在研究中发现社交网络上存在更有影响力的传播者，进而提出的模型为他们分配了更高的传播信息的概率。在新型社交媒体和网络（如微信、微博、新闻客户端、抖音等）的背景下，Zhao等人[7]提出一种名为“易感—感染—抵愈”（SIR）谣言传播模型，在该模型中，将传播过程分为易感、感染和抵愈三个阶段。该项工作是基于这样一个假设：无知的人很容易受到散布谣言者的影响，并且存在根据现实情况将改变谣言散布者转变为谣言中止者的可能性。Hamid等人[8]对新冠疫情期间虚假信息问题展开研究，以MediaEval 2020任务的解决方案为背景，分析与新冠肺炎和5G阴谋理论相关的推文，利用图神经网络和自然语言处理技术来检测误报传播者。然而，鉴于信息传播动力学的复杂性、社会网络的多样性和信息传播媒体的出现，这些研究大多找不到虚假信息爆炸的根源，也找不到消除虚假信息传播的普遍有效的方法。已有研究[9-10]表明，谣言的停止或爆发主要与特定网络的阻塞和遗忘机制有关。

区块链以其独特的技术特性正好契合信息传播过程中的安全需求，通过创建一个可用于验证信息准确性的可信来源数据库，使用去中心化合约来激励信任网络进行安全信息交换，从而帮助减少虚假信息对在线平台的影响。Fraga-Lamas等人[11]深入探索分布式账本技术和区块链技术对抗数字欺骗的潜力，回顾目前正在研究的举措，并阐述当前数字传播面临的主要挑战。Arquam等人[12]提出一种基于区块链的安全可信的网络社交信息传播框架，在该框架下网络节点根据其可信度将信息传播到对等节点。Agrawal等人[13]提出一种基于深度学习的混合模型，并使用区块链检测假新闻，它将区块链的优势与智能深度学习模型相结合，以增强对抗假新闻障碍的鲁棒性和准确性。Jaroucheh等人[14]提出命名为TRUSTD的方法，利用该方法达到打击虚假内容的目的，它是一个基于区块链和集体签名的生态系统，可帮助内容创建者使其内容得到社区的支持，并帮助用户来判断这些内容的正确性和可信

度。但是,区块链的性能瓶颈一直是阻碍其应用发展的重要因素,伴随全媒体机构的不断加入和服务用户量的增多,区块链系统中节点数目也越来越多,带来的交易延迟问题和耗能问题也是非同小可的。

在全媒体传播体系和主流网络舆论环境的建设道路上,打击虚假信息任重而道远,充分利用信息技术解决虚假信息检测是第一步,也是最关键的一步,区块链技术与人工智能技术的结合有望为该问题的解决提供新方案。

2.2 相关技术概况

2.2.1 超级账本概述

超级账本(Hyperledger Fabric)是Linux基金会在2015年赞助启动的全球合作项目Hyperledger的子项目之一,是分布式账本技术的一种实现,为企业构建联盟链提供了模块化区块链架构[15]。与公有链不同,Fabric具有私有和许可的特性,它通过成员管理服务提供商(Management Service Provider,MSP)注册所有成员,并通过智能合约(即链码)读写账本数据。此外,Fabric通过划分通道来维护分类帐数据。未加入通道的成员无法查看账本数据,不同通道之间的数据无法共享,实现了通道之间的数据隔离。Fabric作为企业级联合体区块链系统,在设计过程中充分考虑了实际应用环境对安全性、私密性和性能的需求。如图1所示,Fabric系统架构主要包括两层:应用层和核心层。

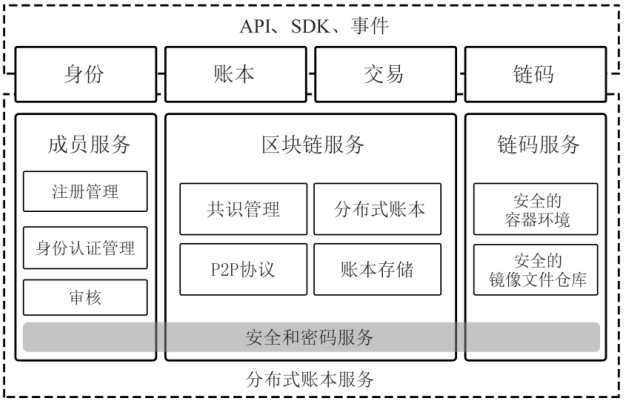

图1 Fabric系统架构

上层作为与应用程序交互的中介,包括身份管理、账本管理、事务处理和链码服务。身份管理根据联盟区块链的业务特点和安全隐私要求,通过PKI系统和CA模块实现成员、权限和证书的管理。账本管理将所有交易信息记录在账本中,授权用户可以记录和读取账本数据。此外,区块链数据还可以通过块号、事务散列等多种方式查询。事务处理功能允许客户端通过发起事务请求来写入账本数据。链码服务实现了用于更新账本数据的各种业务逻辑,使用Docker容器管理链码,同时还实现了用户数据隔离和快速环境部署。下层是区块链的核心实现,包括会员服务、区块链共识服务、链码服务、安全加密服务。成员服务通过加密机制生成身份证书,并使用MSP组件基于身份证书对用户身份进行身份验证。共识服务通过背书者模拟交易提议并签署背书结果,通过订购者和共识插件对交易进行订购和打包,通过提交者验证交易并将交易数据写入账本。安全和加密服务提供密钥生成、消息签名和验证、加密和解密、获取哈希函数等功能。它们具有可插拔的组件特性,并可以扩展定制的加密安全服务算法。

2.2.2 深度学习算法

研究发现,深度学习算法可以自动学习表达语言的结构,使计算机能够理解自然语言,可以用于分析和分类文本数据,是检测文本中虚假信息的有用工具。在虚假信息检测的情况下,深度学习算法可以在经过验证的真实和虚假信息的大型数据集上进行训练,以学习区分它们的模式和特征。

卷积神经网络(Convolutional Neural Network,CNN)作为目前应用较为成熟的一种深度学习算法[16],通常用于图像和视频分析,也可以应用于自然语言处理任务,如虚假信息检测。在虚假中文文本检测任务中,CNN用于分析文本数据,并学习有助于区分真假文本信息的复杂特征,通常包括以下步骤:1)对文本数据进行预处理,这可能包括标记化、词干化和删除停止词等任务;2)使用诸如单词嵌入或一次热编码的技术将预处理的文本数据编码为数字向量;3)在编码的文本数据上训练CNN模型,以识别文本中单词和短语之间的模式和关系。它们可以自动从输入数据中学习特征,而不需要手工制作特征或预先了解假新闻的特征。同时,它们对不同类型的输入数据也有很强的适应性,从而能够适合检测不同形式的假文本。相较于长短记忆网络(Long Short-Term Memory,LSTM)等其他深度学习模型,CNN模型对于短文本特征提取有较好效果,且结构简单易实现,训练速度通常更快,在处理大量数据时更有效。除此之外,CNN模型还能通过分析账户活动和内容的模式来识别机器人和虚假社交媒体账户。

门控循环单元(Gate Recurrent Unit,GRU)是2014年首次引入的一种循环神经网络[17],它旨在解决

传统循环神经网络中常见的梯度消失问题，现已广泛应用于自然语言处理任务中，如机器翻译、情感分析和文本生成[18]等。与LSTM模型相比，GRU可以捕获序列数据中的长期依赖关系，且具有更简单的结构和更少的参数，计算效率更高。它通过使用门控机制来控制网络中的信息流，允许它有选择地记住或忘记之前时间步骤中的信息。其基本结构由更新门和重置门组成，这两个门决定了有多少以前的状态应该被遗忘，有多少当前输入应该被添加到新的状态。它还有一个候选隐藏状态，表示要存储在新状态中的信息。更新门决定保留多少以前的隐藏状态，添加多少新信息，而重置门决定忘记多少以前的隐藏状态。GRU应用于虚假新闻检测的特点和优势在于，其能够有效提取新闻文本丰富的语义特征，为语义分析和多模态融合虚假新闻检测提供有效支持。

双向Transformer编码器（Bidirectional Encoder Representations from Transformers，BERT）是一种用于自然语言处理任务的预训练深度学习模型，由Google AI Language的研究人员于2018年开发，现已成为该领域最受欢迎和广泛使用的模型之一[19]。该模型由输入、编码器和预训练任务三部分组成。其中，输入是一系列令牌嵌入，主要包括词向量、段向量、位置向量。编码器基于Transformers体系结构，由多头自注意机制和全连接前馈神经网络两个子层组成，自注意机制允许模型关注输入序列的不同部分，而前馈网络则单独且相同地处理每个位置。通过编码器可以使模型能够捕获输入序列的局部和全局上下文，同时还能够捕获输入序列及其表示之间的复杂非线性关系。另外，BERT模型还使用掩码语言模型（Masked Language Modeling，MLM）和下一句预测（Next Sentence Prediction，NSP）两个任务进行预训练，来学习语言的深度表示，从而实现对广泛的下游NLP任务进行微调，例如文本分类、命名实体识别、问题回答和自然语言推断。MLM任务通过在给定的句子中随机屏蔽一定比例的标记，然后根据周围的上下文训练BERT模型来预测被屏蔽的标记。而NSP任务则是预测原文中两个句子是否连续。与传统语言模型单向处理文本不同，BERT模型使用双向训练来处理文本，以捕获句子中每个单词的完整上下文，且能够生成高质量的句子嵌入或文本的向量表示。在虚假新闻检测问题中，BERT模型的双向文本处理机制能够更为有效地进行深度语义和细节语义理解，使得针对自然语言环境下的隐喻、反讽、暗示等通常存在于全媒体新闻信息传播生态中的、同虚假信息传播密切相关的语义特征能够更为有效地得以侦测和识别。

3 全媒体虚假信息检测混合方法研究

3.1 总体框架设计

全媒体传播体系是当前的深度媒体融合在创新发展上的总体趋势和主要方向，结合全媒体传播体系建设的落地场景和人机协作智能相关理论与方法，本文提出了结合不同模态下的内容生产机制与分发体系的全域融媒（All-Generated Content，AGC）框架。在全域融媒信息传播环境中，针对虚假文本检测问题，本文考虑利用区块链和深度学习的技术优势来解决传统虚假信息检测方法的局限性。一方面，深度学习算法能够处理大量数据并识别复杂的模式，使其能够有效地检测虚假信息。但它对训练样本的数量和质量要求比较高，计算效率低，且容易受到对抗性攻击，存在恶意行为者故意引入虚假信息来逃避检测的可能。另一方面，区块链技术提供了一个防篡改和透明公开的分类账本，可用于安全地存储和跟踪数据。这使得它非常适合检测和防止虚假信息的传播，任何改变数据的恶意行为都难以实现。因此，本文将两者结合，提出一种新的虚假信息检测方法，其总体框架如图2所示。

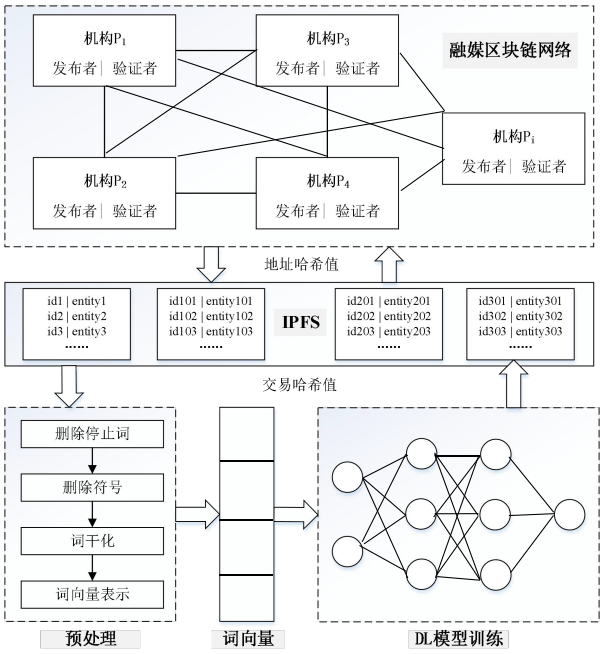

图2 混合方法总体框架

该方法主要由基于区块链的全域融媒体信息传播网络和基于深度学习的分类模型两部分组成，其基本思想是利用区块链技术进行数据验证和溯源，再利用深度学习技术进行特征提取和分类。在信息传播生态中，各

媒体相关机构共同组成全域融媒区块链网络,网络中各节点同时具备信息发布者和信息验证者双重身份,且节点间相互信任,不需要中介机构或中心授权机构作背书,其身份信息经过全网验证并存储在区块链上。同时,节点之间的交易记录在所有参与者之间共享并存储在分布式账本IPFS(Inter Planetary File System)上。每笔交易都由网络中的多个节点使用加密技术进行验证,以确保交易是有效的,不能被篡改。将来自全域融媒区块链网络的数据输入深度学习模型进行预处理、特征提取及训练,并通过调参提升检测速度和准确度。

3.2 融媒区块链网络搭建

区块链作为一种去中心化、安全透明的基础设施,可以为信息传播提供可信来源和完整性保障。本文针对全媒体环境下虚假信息检测问题构建基于Hyperledger Fabric的全域融媒区块链网络,基本结构如图3所示。该网络主要完成以下三方面的内容:1)定义网络结构,设置组织、节点规则,编写相应配置文件;2)创建网络通信通道并初始化;3)测试与部署网络。

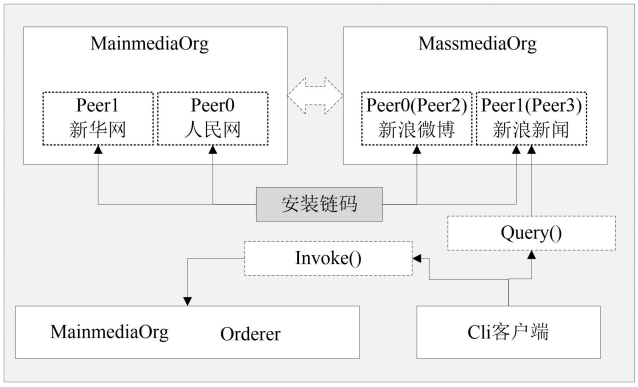

图3 全域融媒区块链网络结构图

在区块链网络中,每个组织和节点都具有独一无二的标识,可以通过公共密钥加密来验证身份。节点使用IP地址作为其网络地址,并通过点对点通信实现与其它节点的交互,从而确保多个参与者之间的无障碍通信。为了实现全域融媒区块链网络的搭建,本文使用一个虚拟机系统,并在其上搭载具有通用意义的Ubuntu操作系统。通过使用命名虚拟IP地址,使得该虚拟机可以模拟多个不同的节点,这样可以有效地减少硬件资源的占用,并且更加灵活方便。相比于传统的一节点一主机的方式,这种方法不仅能够减少物理机的数量,也能够提高整个网络的安全性和稳定性,更加适用于多主体协同场景下的应用。由于Hyperledger Fabric是一种允许机构加入并参与管理节点的联盟链[20],可供机构访问,并依靠组织管理节点来保持其完整性。另外,为避免网络中的一些节点通过故意发送不正确的信息或破坏节点之间的通信来恶意操作的情况,Fabric区块链系统也引入了实用拜占庭容错共识算法(Practical Byzantine Fault Tolerance,PBFT),为节点网络就账本状态达成共识提供了一种更有效和可扩展的方式。这些与全媒体场景下的信息传播需求十分契合,在媒体融合场景下,信息传播变得去中心化、随机化、个性化,任何人或机构都可以利用各种渠道和平台与受众分享新闻、观点等信息,通过创建虚假社交媒体账户,或使用机器人来传播虚假信息,放大分裂或有争议的信息,或操纵引导公众舆论。基于此,本文采用支持PBFT共识机制的Hyperledger Fabric网络架构进行创建,整个网络由Infrastructure、Application、Chaincode三部分构成。按照表1(全域融媒区块链组织信息和节点信息配置策略如表1所示)给出的组织信息和节点信息进行配置,编写相应配置文件。其中主要包括三类组织,即主流媒体机构组织、大众媒体机构组织和排序组织。其中主流媒体机构组织、大众媒体机构组织是网络结构根据业务逻辑设计添加的,其身份是通过映射现有媒体组织来建立的。为了实现这一点,MainmediaOrg和MassmediaOrg在联盟链中充当组织的结构和标识。组织的域名对于映射其管理节点的地址至关重要,并且与计算机网络中的地址域名相同。排序组织是网络结构必需的,不具备独特身份,虽然由其他媒体机构组织共同建立和管理,但它拥有自己的域名,并使用可信的分布式存储系统存储交易和上行链路数据。由于区块链网络节点通信依赖通道实现,待网络组织、节点信息配置完毕后,下一步需要定义各种参数,包括通道容量、组织结构、排序节点配置、通道配置及网络入口。网络入口充当设置Genesis块、通道及其关联组织配置名称的指南。为了简化相关Genesis块和通道文件的创建,通过在网络入口中指定它们的配置名称,使得配置文件能更容易生成和引用。同时,通过组织机构配置定义组织实体,有助于后续配置的实现,因为它提供了特定的详细信息,包括组织应用ID名称、MSP相关文件的位置以及与组织关联的锚节点。由于基于一台虚拟机构建的区块链网络需要所有节点都使用自己的IP地址启动和操作,通过将这些IP地址映射到网络上可访问的Docker容器,并在Docker容器中部署相应的智能合约,使客户端在容器中执行初始化操作。

表1 全域融媒区块链网络结构配置信息

组织名	组织标识	组织ID	域名	节点信息 （节点名、IP地址、端口号）
主流媒体机构组织	MainmediaOrg	MainmediaMSP	Mainmedia.com	人民网 peer0.mainmedia.cuc.com "7051:7051"/ "7053:7053" 新华网 peer1.mainmedia.cuc.com "17051:7051"/ "17053:7053"
大众媒体机构组织	MassmediaOrg	MassmediaMSP	Massmedia.com	新浪微博 peer0.massmedia.cuc.com "27051:7051"/ "27053:7053" 新浪新闻 peer1.massmedia.cuc.com "37051:7051"/ "37053:7053"
排序组织	ControlOrg	ControlMSP	Control.com	排序节点 orderer.control.cuc.com "7050:7050"

3.3 GRU-CNN模型训练

在全域融媒区块链网络中，媒体信息以去中心化、去信任的方式得以迅速传播，这为虚假信息的检测带来了机遇和挑战。一方面，信息的链式存储和不可篡改为虚假信息的溯源提供有力支持，同时，以智能合约的形式部署深度学习检测模型能够有效避免对抗性攻击，使得检测结果客观、真实。另一方面，网络节点身份无需向权威信任机构请求背书，当恶意或扮演“网络水军”角色的节点数量达到一定规模后，虚假信息的产生和传播就变得十分容易。另外，媒体融合场景下，数据大都以非结构化的形式呈现。考虑到短视频平台的评论具有短小精悍、个性化、情绪化等特点，单一词向量特征对评论的真实意思表达存在欠缺，需要利用上下文来增大词向量矩阵以求获得更多的特征信息。

基于此，本文从节点身份和信息内容两方面着手，通过在信息源头和传播过程两方面检测识别虚假信息，从而提高虚假信息检测的有效性和准确率。采用GRU-CNN模型对来自全域融媒区块链网络的采编内容利用深度学习模型进行检测，融合账户身份信息及评论信息进行特征提取。该模型将卷积结构引入到GRU中进行多尺度特征提取，增强了高水平特征学习能力。然后使用softmax分类器来压缩学习到的特征，以提高类别的紧凑性，该模型执行流程如图4所示，包括模型输入、模型训练和模型输出三个阶段。在模型输入阶段，先将账号昵称、粉丝数、关注数等身份信息以及评论文本信息采用Word2Vec方法获得词向量，再利用BERT预训练模型得到更准确和更有意义的单词表示，将获得的字向量与Word2Vec词向量进行融合形成模型词嵌入。在模型训练阶段，采用GRU-CNN模型先局部捕捉身份信息特征，再结合其他文本信息捕捉全局上下文特征，并适时调整模型参数以达到最优效果。最后根据模型训练的上下文语义及节点身份信息的类别标签（真、假）进行预测分类。

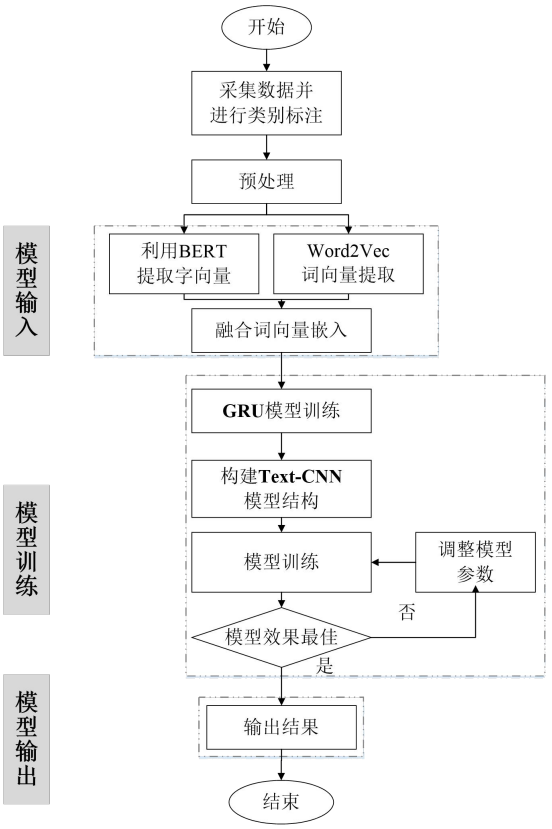

图4 模型训练流程图

（1）模型输入

传统的Word2Vec方法通常涉及两种训练模式[21]，

即CBOW和Skip-gram。CBOW模型是一种基于窗口的词嵌入神经网络架构,它以单词序列作为输入,并根据给定窗口内的上下文单词预测目标单词。该模型经过训练,可以根据目标单词周围的上下文最大化预测目标单词的可能性。与Skip-Gram模型相比,CBOW需要更少的训练迭代,故而它的训练速度更快。Skip-Gram模型是NLP中用于学习高质量词嵌入的另一种流行算法。它的目标是预测目标单词的周围单词,并对模型进行训练,以最大程度地减少预测单词和实际单词之间的差异。经过训练后,Skip-Gram会为语料库中每个单词生成密集的向量表示,从而捕获其上下文和含义,它能够捕获单词之间更细粒度的关系,例如同义词、反义词和类比。此外,它通过负抽样进行增强,从而加快训练速度并提高学习嵌入的质量。因此,本文针对账号昵称、粉丝数、关注数等身份信息采用CBOW模型提取词向量,对评论文本信息采用Skip-Gram模型提取词向量,再将它们与BERT预训练模型提取的字向量融合作为GRU-CNN模型的输入。在使用Word2Vec方法得到的词向量均为100维,经BERT预训练模型得到的字向量是768维,则融合向量是968维。

(2)模型训练

GRU-CNN模型利用各种大小的卷积核来捕获输入文本的基本和复杂的特征。这些特征随后通过池化层进行组合,以生成有利于分类任务的信息表示。因此,本文构建GRU和Text-CNN模型进行特征提取,先从全局上下文确定各特征关系,再从局部分析账号身份信息与评论文本信息间的密切关系。两模型的结构如图5、6所示,其中,GRU模型使用32个门控循环单元,与一句话字符数相同,模型输入是由Word2Vec词向量与BERT字向量构成的融合向量,每个门控循环单元输出维数与输入融合向量维数相同,为968维。GRU模型的最终输出是32个968维的向量。Text-CNN模型由词嵌入、卷积层、池化层、全连接和输出层组成,其输入来自于GRU模型的输出,接受一个维度为32*968的向量输入,表示每个段落中的32个字符以及每个单词的单词向量的长度。为了考虑由不同字长组成的中文汉字具有不同特征,Text-CNN模型采用不同大小(2,3,4)的多个卷积核来提取不同粒度的信息。每个内核被设置为128,生成的特征被连接到卷积层的单个输出中。最大池化后得到3*128个标准化特征值,来自池化层的特征组合在全连接层创建384*1向量。

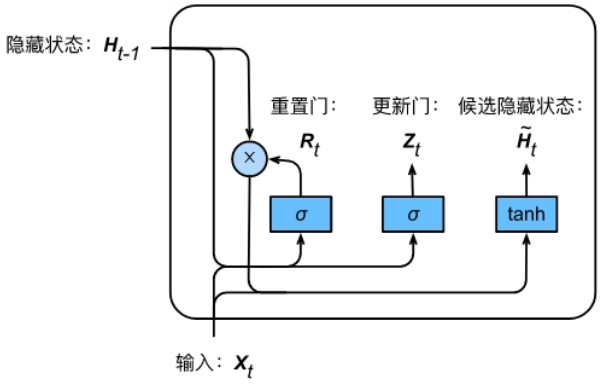

图5 GRU模型结构图

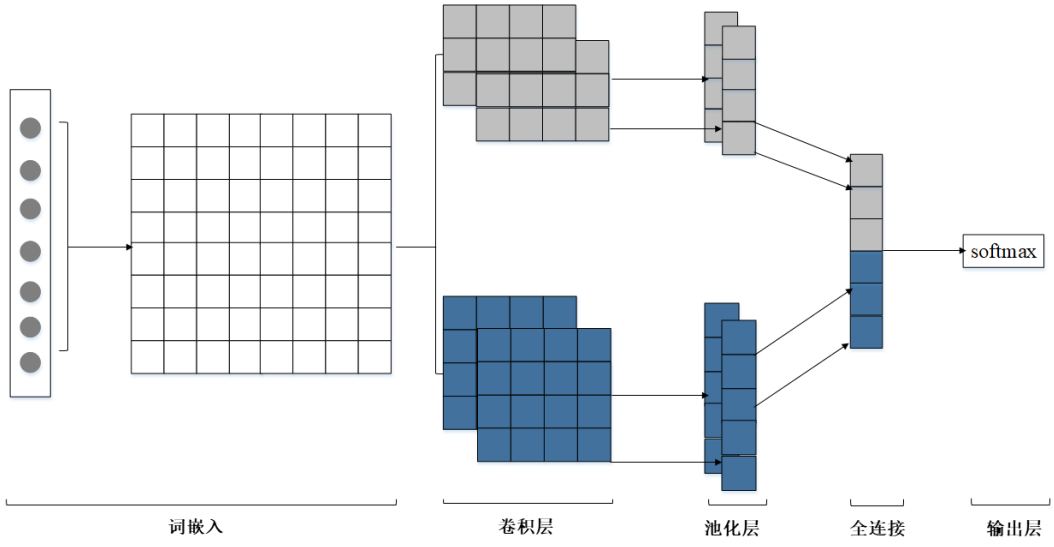

图6 Text-CNN模型结构图

（3）模型输出

输入文本经过词嵌入和训练后，模型得到一个特征向量$\boldsymbol{Z}$，用于输入类分类。情感分类采用全连接层，结合softmax激活函数，将特征划分为多个类。这种方法使模型能够有效地解决多类分类问题，利用全连接层的维数来表征特征向量。

3.4 反馈操作

将深度学习模型与全域融媒区块链网络联动，是提高检测准确率的有效方式之一。传统的方法都是针对深度学习算法模型的超参数进行调优，操作随机性太强且缺乏可解释性。本文主要将深度学习模型以智能合约的形式部署在区块链网络上，根据模型训练结果结合共识和激励机制既可以抑制虚假账号生产内容，又可以促进优质内容创作。其工作思路是：先收集全域融媒区块链上采编的媒体信息，借助IPFS形成模型训练数据集；然后将数据集按7:1:2划分为训练集、验证集和测试集进行模型训练，将训练结果存储至IPFS；最后，将训练好的模型部署在区块链网络中，对节点身份启动惩罚机制，防止虚假信息的产生和传播。

4 实验与分析

4.1 实验数据集与模型评价标准

由于现有全媒体平台采用区块链技术运营还在起步阶段，在真实的融媒系统中获取大量文本信息难以实现，而且缺乏全媒体场景下公开可用的数据集。本文主要关注抖音、快手等短视频平台的评论内容，但这些平台为了保护用户隐私采用了严格的加密算法和反爬虫机制。因此，本文使用的研究数据主要是通过实验室人工手段收集，数据集由抖音、快手应用程序上各种新闻媒体账户围绕“抗击新冠疫情”为主题发布的视频评论组成，主要包含抖音数据30125个条目，快手数据9875个条目，总计40000条数据。

随着全媒体环境下信息情绪化特征愈来愈明显，虚假信息检测问题变得越来越复杂。然而，虚假信息检测和虚假账号识别实质上都是二分类问题，在现有研究基础上[22-24]，本文对采集的数据进行预处理，将其初步分为真实信息、虚假信息和重复信息三类，表2列出了数据集的详细统计数据。

表2 数据集统计信息

数据来源	标签（情感特征）	数目	总计
抖音	真实信息（正向、中性）	19037	30125
	虚假信息（负面、疑惑、不文明）	9814	
	重复信息	1274	
快手	真实信息（正向、中性）	6661	9875
	虚假信息（负面、疑惑、不文明）	3042	
	重复信息	172	

4.2 实验配置

本文实验主要分为两部分：全域融媒区块链网络测试、GRU-CNN分类模型。前者的环境配置大都在上节网络构建中已详细说明，需要补充的是在Ubuntu18.04环境中选用支持PBFT共识机制的Hyperledger/fabric v1.4.4作为底层区块链网络，相关工具明细见表3示。基于GRU-CNN模型训练实验使用PyTorch深度学习框架实现[25]。这个开源库是由FaceBook人工智能研究所于2017年开发的，具有内置引导的自动神经网络和强大的CPU加速张量计算，其强大的功能和灵活性使其成为本研究中实现和评估深度学习模型的理想选择。表4概述了所使用的具体配置和环境细节。

表3 相关工具明细

工具/软件	版本号
Linux	Ubuntu 18.04
Fabric	1.4.4
Go	1.15.14
Docker	19.03.5
Docker-Compose	1.25.4
VSCode	1.72.0
Node	9.9.0

表4 模型训练实验环境配置情况

名称	值
操作系统	Ubuntu 18.04
处理器	CPU E5-2640 v4 @ 2.40GHz
GPU	NVIDIA Quadro P4000
RAM	16GB
深度学习框架	PyTorch
编程语言	Python
编程工具	Visual Studio Code

实验采用Adam Optimizer优化器来训练网络。在训练过程中,针对模型训练效果进行不断调优,最后设定模型超参数如表5所示。

表5 模型超参数设置

参数	值
Word2Vec词向量维度	100
BERT字向量维度	768
GRU门控单元数	32
卷积核大小	2、3、4
每组Filter个数	256
激活函数	softmax
池化策略	最大池化
学习率	0.001
Dropout比率	0.5
Num_Epochs,Batch Size	16,64

4.3 实验分析

鉴于目前将区块链技术与深度学习技术结合应用于解决虚假信息检测的研究有限,本文仅从计算性能和网络性能两方面验证本文所提方法的有效性。

(1)计算性能

从方法概述中发现,与区块链网络相比,基于深度学习的分类模型是评估所提方法计算性能的重要模块,考虑到将深度学习模型部署在智能合约上需要大量的计算、存储资源,本文在链下针对GRU-CNN模型依据Precision、F1-Score和计算时间等各项指标对训练效果进行评估。使用所采集的数据集,本文设计了FastText、TextRNN、CNN-LSTM、Att-TextRNN和BERT共5个实验来对比不同输入下模型的性能,其比较结果如表6所示。

表6 对比实验结果

	抖音			快手		
	Accuracy	Macro-F1	Time	Accuracy	Macro-F1	Time
FastText	0.831	0.839	706s	0.744	0.755	216s
TextRNN	0823	0.844	552s	0.769	0.771	241s
CNN-LSTM	0.955	0.975	1023s	0.847	0.863	324s
Att-TextRNN	0.873	0.884	995s	0.821	0.806	298s
BERT	0.951	0.967	1140s	0.849	0.852	378s
GRU-CNN	**0.947**	**0.963**	**453s**	**0.851**	**0.846**	**192s**

从上述结果可知,在数据样本充足的情况下,基于深度学习模型的文本分类能取得较好的效果。不同模型的结构及超参数设置都不同,造成在检测准确度上稍有差别。结果表明,与GRU-CNN模型相比,CNN-LSTM模型和BERT模型的分类效果更好些,但它们结构复杂,需要的计算资源相对较多,耗时长。然而,在区块链网络上部署深度学习模型尽量考虑轻量级的,GRU-CNN模型结构相对简单且准确度较高,适合部署。

(2)网络性能

区块链网络性能瓶颈是影响其广泛应用的重要因素。在本文所提方法中,针对全域融媒区块链信息传播网络,本文以Fabric-Samples v1.4.4的First-Network实例为标准,利用网络仿真软件模拟方案实施,测试方案运行100次、200次、300次、400次、500次、600次的通信带宽消耗情况,具体测试结果见图7所示。

从图7中可以看出,随着运算规模的增大,网络带宽消耗也在不断加大,全域融媒区块链网络结构相比原生Fabric更复杂,因此消耗的网络资源更多。同

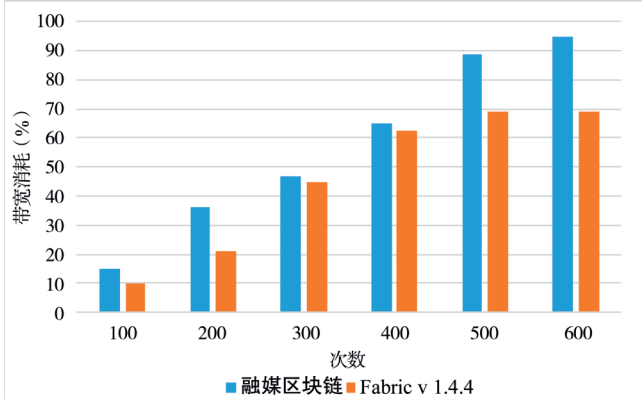

图7 不同规模下BYFN交易通信带宽消耗比较

时,根据结果发现当交易规模达到临界状态500次时,原生Fabric网络带宽消耗也达到饱和状态,此时说明网络负载已超出范围,交易计算可能无法完成。而针对全域融媒区块链,测试实验表明还未到达其交易饱和点,可在原有的基础上继续增加交易数量。这也从侧面体现了Fabric可插拔、分布式和并行处理的优势。因此,在算力、存储等硬件条件允许下,在区块链网络上部署深度学习模型是可行的,这样可以有效防

止对抗性攻击，确保用于虚假信息检测的模型和数据是可信的、防篡改的。

5 结语

本文介绍了一种基于区块链和深度学习的虚假信息检测方法。该方法结合区块链和深度学习的技术优势，在融合媒体环境下对短视频评论进行虚假检测。首先，构建了基于区块链的融合媒体信息传播网络，保证了数据的安全性和可靠性，为深度学习模型训练提供了值得信赖的数据源。然后，考虑文本分类的特点，将账号身份信息和评论文本信息向量化，采用GRU-CNN模型进行特征训练。最后，将培训结果反馈给区块链网络，使用共识和激励机制来控制内容生产和传播。在真实数据集上进行了一系列对比实验，以评估所提出的混合方法对假信息检测的有效性。实验结果表明，该方法能有效提高假信息检测任务的准确性和效率。

然而，本文所提出的混合方法的可扩展性还有待提高。一方面，随着网络规模或数据量的增加，介质区块链的性能会显著下降；另一方面，CNN模型训练依赖于大量的短视频评论数据，需要较高的计算能力和存储硬件。此外，本文仅通过实验验证了该方法的有效性，并未对计算代价进行详细分析。因此，下一步工作是考虑设计用于特征提取的轻量级深度学习模型，并将所提方法的性能与其他最先进的假信息检测方法进行比较。

参考文献（References）：

[1] Gilda S. Evaluating machine learning algorithms for fake news detection[C]// 2017 IEEE 15th Student Conference on Research and Development (SCOReD), 2017: 110-115.

[2] Granik M, Mesyura V. Fake news detection using naive Bayes classifier [C]// 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2017: 900-903.

[3] Vedova M L D, Tacchini E, Moret S, et al. Automatic online fake news detection combining content and social signals[C]// 2018 22nd Conference of Open Innovations Association (FRUCT), 2018: 272-279.

[4] Tashnim A, Nowshin S, Akter F, et al. Interactive interface design for learning numeracy and calculation for children with autism[C]// 2017 9th International Conference on Information Technology and Electrical Engineering (ICITEE), 2017:1-6.

[5] Bajaj S. The pope has a new baby! fake news detection using deep learning[R/OL]. http://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2710385.pdf,2017.

[6] Vega-Oliveros D A, Costa L F, Rodrigues F A. Rumor propagation with heterogeneous transmission in social networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2017, 2017(2): 023401.

[7] Zhao L, Cui H, Qiu X, et al. Sir rumor spreading model in the new media age[J]. Physica A: Statistical Mechanics and its Applications, 2013, 392(4):995 - 1003.

[8] Hamid A, Sheikh N, Said N, et al. Fake news detection in social media using graph neural networks and NLP techniques: a COVID-19 use-case[DB/OL]. arXiv:2012.07517, 2020.

[9] Daley D J, Kendall D G. Epidemics and rumours[J]. Nature, 1964, 204(4963): 1118 - 1118.

[10] Nekovee M, Moreno Y, Bianconi G, et.al. Theory of rumour spreading in complex social networks[J]. Physica A: Statistical Mechanics and its Applications, 2007, 374(1):457-470.

[11] Fraga-Lamas P, Fernandez-Carames T M. Fake news, disinformation, and deepfakes: leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality[J]. IT Professional,2020, 222:53-59.

[12] Arquam M, Singh A, Sharma R. A blockchain based secured and trusted framework for information propagation on online social networks[J]. Social Network Analysis and Mining, 2021, 11(1):1-16.

[13] Agrawal P, Anjana P S, Peri S. DeHiDe: deep learning-based hybrid model to detect fake news using blockchain[C]// Proceedings of the 22nd International Conference on Distributed Computing and Networking, 2021: 245-246.

[14] Jaroucheh Z, Alissa M, Buchanan W J, et.al. TRUSTD: combat fake content using blockchain and collective signature technologies [C]// 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC), 2020: 1235-1240.

[15] Androulaki E, Manevich Y, Muralidharan S, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains[C]// Proceedings of the Thirteenth EuroSys Conference, 2018.

[16] Sharma P, Berwal Y, Ghai W. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation[J]. Information Processing in Agriculture, 2020, 7(4): 566-574.

[17] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[DB/OL]. arXiv: 1412.3555, 2014.

[18] Qu G, Qiu T, Si Y, et al. Remaining useful life prediction for aero-engine based on hybrid CNN-GRU model[C]// 2022 IEEE International Conference on Unmanned Systems (ICUS), 2022:1523-1528.

[19] Devlin J, Chang M W, 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, 2019, 1 : 4171-4186.

[20] Zhang S, Hua S, Pi B, et al. Performance diagnosis and optimization for Hyperledger Fabric[C]// 2020 2nd Conference on Blockchain Research & Applications for Innovative Networks and Services (BRAINS), 2020: 210-211.

[21] Onishi T, Shiina H. Distributed representation computation using CBOW model and Skip-gram model[C]// 2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI), 2020: 845-846.

[22] Zhang H, Jin W, Zhang J, et al. YNU-HPCC at SemEval 2017 Task 4: using a multi-channel CNN-LSTM model for sentiment classification[C]// Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), 2017: 796-801.

[23] Wang Y, Sun A, Han J, et al. Sentiment analysis by capsules[C]// Proceedings of the 2018 World Wide Web Conference, 2018: 1165-1174.

[24] Du J, Gui L, He Y, et al. Convolution-based neural attention with applications to sentiment classification[J]. IEEE Access, 2019, 7: 27983-27992.

[25] Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019: 8026-8037.

编辑:赵志军