Modern Engineering and Application 现代工程与应用, 2026, 4(3), 60-70
Published Online August 2026.
https://ufascience.com/mea/index

基于加权微调 BERT 的稀疏 RNN 虚假新闻检测方法

何文博

青岛大学经济学院，青岛 266071，山东，中国

摘要：虚假新闻指以图像、文章或视频形式传播、伪装成真实新闻以试图操纵公众观点的错误信息或不实报道。然而，由于语言风格、语境与信息来源的多样性，现有检测系统难以捕捉虚假新闻的多样特征，导致识别不准确。为此，本文提出一种用于虚假新闻检测的深度学习（DL）方法——面向 Transformer 的加权微调双向编码器表示与稀疏循环神经网络相结合的模型（WFT-BERT-SRNN）。首先，从 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 数据集获取数据以评估 WFT-BERT-SRNN；随后通过停用词去除、分词与词干提取进行预处理，剔除无用短语或词汇；接着用 WFT-BERT 提取特征；最后用 SRNN 检测并把新闻分类为真实或虚假。本文把 WFT-BERT-SRNN 与深度神经网络虚假新闻检测（DeepFake）、联合学习 BERT、面向虚假新闻检测的多 EDU 结构（EDU4FD）、基于图像描述的方法以及细粒度多模态融合网络（FMFN）等现有技术进行了比较。相较上述方法，WFT-BERT-SRNN 在 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 数据集上分别取得 0.9847、0.9724、0.9624 与 0.9725 的更优准确率。

关键词：深度学习；虚假新闻；自然语言处理；稀疏循环神经网络；面向 Transformer 的加权微调双向表示

Weighted Fine-tuned BERT-based Sparse RNN for Fake News Detection

WenBo He

School of Economics, Qingdao University, Qingdao 266071, Shandong, China

Abstract: Fake news refers to misinformation or false reports shared in the form of images, articles, or videos that are disguised as real news to try to manipulate people's opinions. However, detection systems fail to capture diverse features of fake news due to variability in linguistic styles, contexts, and sources, which lead to inaccurate identification. For this purpose, a weighted fine-tuned-bidirectional encoder representation for transformer-based sparse recurrent neural network (WFT-BERT-SRNN) is proposed for fake news detection using deep learning (DL). Initially, data is acquired from Buzzfeed, PolitiFact, Fakeddit, and Weibo datasets to evaluate WFT-BERT-SRNN. Pre-processing is established using stopword removal, tokenization, and stemming to eliminate unwanted phrases or words. Then, WFT-BERT is employed to extract features. Finally, SRNN is employed to detect and classify fake news as real or fake. Existing techniques like deep neural networks for fake news detection (DeepFake), BERT with joint learning, multi-EDU structure for fake news detection (EDU4FD), image caption-based technique, and fine-grained multimodal fusion network (FMFN) are compared with WFT-BERT-SRNN. The WFT-BERT-SRNN achieves a better accuracy of

文章引用: 何文博. 基于加权微调 BERT 的稀疏 RNN 虚假新闻检测方法[J]. 现代工程与应用, 2026, 4(3): 60-70.
DOI: https://doi.org/10.61784/mea2032

0.9847, 0.9724, 0.9624, and 0.9725 for Buzzfeed, PolitiFact, Fakeddit, and Weibo datasets compared to existing techniques like DeepFake, BERT-joint framework, EDU4FD, image caption-based technique, and FMFN.

Keywords: Deep learning; Fake news; Natural language processing; Sparse recurrent neural network; Weighted fine-tuned-bidirectional representation for transformers

一、引言

虚假新闻是对新闻业与全球商业构成严重威胁的主要因素之一，并伴随相当程度的附带损害 [1]。社交网络平台为用户构建、获取与分享各类信息提供了重要渠道；随着众多用户在相应时段接收并查找最新动态，社交媒体网络的使用规模不断扩大。然而，社交媒体也为向用户传播海量虚假与误导性信息提供了机会，而这类信息会对社会产生破坏性后果 [2]。虚假新闻的扩散明显比真相更深更快，对社会与个人都造成严重负面影响 [3]。与推特、谷歌以及 Gmail、Yahoo 等网络邮箱相比，脸书是传播虚假新闻最常用的媒体平台之一 [4]。随着人们花在社交媒体上的时间不断增加、并把它当作唯一的主要新闻来源，由此引发的担忧也随之扩大 [5]。互联网缺乏管束加之操作简便，使虚假新闻得以广泛扩散 [6]。在 2016 年美国大选期间，据称社交媒体上出现了大量虚假新闻，其中包括总统选举期间印度新任空军元帅提名的消息 [7]。与他人评估并交流知识的有效性使在线社交网络颇具吸引力；此外，以极小代价高速散布即时信息也使虚假新闻等不实信息得以广泛扩散 [8]。假信息可分为蓄意造假与错误信息两类：蓄意造假是为误导公众而发布的不实信息，具有经济与社会影响。此外，当虚假新闻已被大规模、多次转发之后，要加以终止十分困难。因此，虚假新闻的传播对社会能力、与个人生活都产生了负面影响，虚假新闻检测已成为社交媒体的重要议题之一。为此，本研究提出一种高效的检测方法，以规避这些恶意意图并帮助人们不受其害。近期借助深度学习检测虚假新闻的研究，通过利用用户特征、文本数据与用户反馈等多种社交媒体新闻特征，已取得可观成效。研究表明，人在没有专门辅助的情况下识别不实信息的能力仅为 54%，因此需要计算机化的真假新闻检测。

然而，由于语言风格、语境与来源的多样性，检测系统难以捕捉虚假新闻的多样特征，导致识别不准确。KALIYAR 等张量分解实现了 DeepFakE：收集用户的新闻互动并与用户社群数据整合，构建语境、内容与用户社群的三模态张量；为获取新闻文章的潜在表示，执行了耦合矩阵—张量分解；随后用 DeepFake 与 XGBoost 完成分类任务。该方法通过整合语境与内容技术取得了较好表现，但由于其静态特性，DeepFake 在应对演变模式与捕捉虚假新闻动态方面存在困难。CHE 等提出用于虚假新闻检测的稀疏与图正则化 CANDECOMP/PARAFAC（SGCP）张量分解学习：用 CP 张量分解建立新闻因子矩阵，以反映用户与新闻之间的复杂关联；该方法用 Buzzfeed 与 PolitiFact News 两个数据集进行评估，在保持新闻因子矩阵稀疏性的同时保留了原空间的流形结构。但当大量样本缺失时，该方法的检测性能下降。

PALANI 等开发了用于虚假新闻检测的 CB-Fake：用 BERT 提取保留词间语义关系的文本特征，用 CapsNet 捕捉图像中的视觉特征，再把二者整合以获得更丰富的数据表示，从而判定新闻真伪。CB-Fake 方法在虚假新闻检测中高效且可扩展，但它仅从新闻文章中生成视觉与文本特征，未评估用户档案数据与行为特征。SHISHAH 提出结合命名实体识别（NER）与关系特征分类（RFC）的联合学习 BERT 用于虚假新闻检测：其 SPR 编码器修改了 BERT 中第 k 层的动态注意力范围，借助所提供预训练方法中的先验知识构建上下文向量。该 BERT 联合方法的独到之处在于为特征赋予有意义的权重，从而取得更好表现；但该方法在区分不同信息类型方面存在困难，影响了分类效果。KALIYAR 等实现了兼用语境与内容数据、基于回音室效应的深度学习方法（EchoFakeD）：把新闻内容与张量融合以获得社交语境与新闻内容数据的潜在表示，再用带最优超参数的深度神经网络（DNN）加以归类。在该方法中采用张量分解带来了较好表现，但把内容与语境数据整合在一起的 EchoFakeD 在泛化方

面存在问题。

WANG 等提出用于增强文本表示以检测虚假新闻的多 EDU 结构方法 EDU4FD：先提取修辞关系以构建 EDU 依存图，再用关系图注意力网络（RGAT）获得基于图的 EDU 表示；最后用带全局注意力的门控递归单元把两种表示组合为增强的文本表示以检测虚假新闻。但 EDU4FD 同样受语言风格、语境与来源多样性的困扰。LIU 等提出一种基于图像描述的方法，以改善用于虚假新闻检测的图像语义信息：把图像描述信息融入文本，弥合图文之间的语义鸿沟；随后用 Transformer 融合多模态内容；把图像的目标特征与全局特征结合，提高了图像利用率并改善了图文之间的语义交互。但该方法依赖预定义词表，难以准确描述复杂的视觉概念。WANG 等提出细粒度多模态融合网络（FMFN），融合文本与视觉特征以检测虚假新闻：用深度卷积神经网络（CNN）提取图像的多种视觉特征，用缩放点积注意力增强并融合视觉与文本特征，最后把融合特征送入二分类器进行检测。缩放点积不仅考虑视觉特征之间的相关性，还捕捉文本与视觉特征之间的依赖关系；但 FMFN 因融合过程复杂而存在过拟合问题，导致表示出现偏差并降低了在未见数据上的泛化性能。综合来看，现有技术存在泛化困难，以及受语言风格、语境与来源多样性影响导致虚假新闻识别不准确等局限。为解决这一问题，本文提出 WFT-BERT-SRNN 以准确检测虚假新闻。

本研究的基本贡献是：（1）用 WFT-BERT 提取特征，使模型能够在预训练阶段优先关注特定领域或任务，提升其捕捉领域专有信息的能力；（2）用 SRNN 检测并把新闻分类为真实或虚假，它具备捕捉语义数据中长程依赖的潜力；（3）用 BuzzFeed、PolitiFact、Fakeddit 与 Weibo 四个基准数据集评估 WFT-BERT-SRNN 方法。

本文其余部分结构如下：第二部分讨论所提出的方法；第三部分详述面向 Transformer 的加权微调双向编码器表示与稀疏循环神经网络；第四部分给出结果与讨论；第五部分为结论。

二、所提出的方法

本研究提出 WFT-BERT-SRNN 用于虚假新闻检测。首先，从 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 基准数据集获取数据以评估该方法；随后通过停用词去除、分词与词干提取进行数据预处理，剔除无用短语或词汇；用 WFT-BERT 提取特征，用 SRNN 检测并把新闻分类为真实或虚假。所提出方法的框图见图 1。

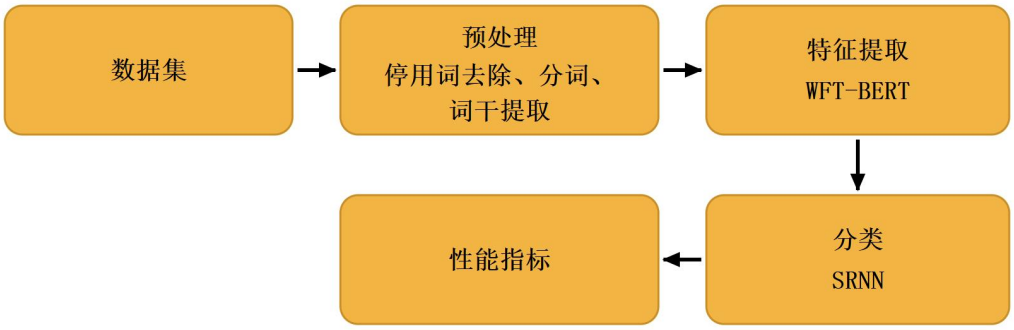

Figure 1 Block Diagram for the Proposed Approach

图 1 所提出方法的框图

（一）数据集

本文用 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 四个基准数据集评估虚假新闻检测。Buzzfeed 包含真实与虚假两类新闻集，取自有关 2016 年美国总统大选的不实新闻报道，共含从脸书采集的 1700 篇新闻文章；该数

据集所选用的词项有 nation、country、party、political、democrat、bill 等。PolitiFact 数据集包含由 3634 位作者提供的 14 055 篇文章，人均 3.86 篇；这些文章覆盖 52 个主题，且每篇文章涉及一个以上主题。所获数据被送入预处理阶段。

Fakeddit 数据集采自流行社交媒体 Reddit，包含超过 100 万条样本，并生成涵盖元数据、文本、评论与图像的多粒度标签。Weibo 数据集来自中国流行的社交媒体平台，每条新闻均含对应的图像、标签与文本。

（二）预处理

获取数据后，预处理通过停用词去除、分词与词干提取完成，具体讨论如下。分词从文本数据中去除标点，停用词去除剔除介词、代词与连词，词干提取则处理副词、动词等语法性词汇。

分词：即把原始文本切分为更小片段的过程，这些片段称为词元。此外，该方法还可去除文本数据中的标点；用数字过滤器剔除特定句子中的数字词项；用大小写转换器把文本数据转换为大写或小写；最后用 N 字符过滤器剔除字符数较少的词。

停用词去除：停用词并不恰当也不关键，但在关联表达与句子成形中被频繁使用；它们在每个句子中十分普遍且常见，却不承载信息。此外，英语中约有 500 个停用词，包括连词、介词与代词，被视为少量“停用词”，例如 a、when、on、what、am、an、under 等。因此，去除停用词可节省处理时间与存储空间。

词干提取：词干提取的基本目标是获得不同词形所共有、含义相同的基本词形。该过程把形容词、副词、动词与名词等众多语法性词汇转换为原形，例如“consulting”与“consultants”都源自“consult”。因此，把词还原为规范基本形式是一种高效做法。至此，文本、数字与停用词等冗余与字符类词项在预处理阶段被过滤掉，处理结果送入特征提取环节。

三、面向 Transformer 的加权微调双向编码器表示与稀疏循环神经网络（WFT-BERT-SRNN）

（一）特征提取

预处理后的数据作为输入送入 WFT-BERT，为虚假新闻检测提取特征。它能够通过突出重要词或短语来提升模型的表示能力，从而提高检测准确率；同时它在特定数据上微调时利用预训练知识，保证更好的上下文理解。文本序列记为 A = {a1, …, aL}，其中 al（1 ⩽ l ⩽ L）表示句子，L 表示文本序列长度。借助双向预训练方法，文本序列 A = {a1, …, aL}被编码为定长句向量 h，作为输入源元素。句向量 sl 由式（1）给出。

$$sl = \mathrm{BERTsent}(al) \tag{1}$$

式中，BERTsent(•)表示把句子编码为句向量。转换后句向量 sl 的隐藏向量表示 ul 借助多层感知机（MLP）获得，其数学表达见式（2）。

$$ul = \tanh(W1\ sl + b1) \tag{2}$$

式中，W1 与 b1 表示权重与偏置参数。常用的文本表示方法丢弃了句子之间的交互信息，导致部分语义损失。这里把全部源元素视为上下文信息，以获得语义更丰富的文本表示。例如，捕捉某一源元素 hk 与所有源元素(s1, s2, …, sL)之间的相关性；al 的语义权重由 ak 分配，源元素记为 α kl，见式（3）。

$$\alpha\ kl = \exp(ul^{\top}\ uk) / \ \Sigma l{=}1..L\ \exp(ul^{\top}\ uk) \tag{3}$$

ik 由式（4）给出。

$$ik = \ \Sigma l{=}1..L\ \ \alpha\ kl\ sl \tag{4}$$

单个源的每个元素都与所有源发生交互，并获得所有源元素与单个源元素之间的交互，如式（5）所示。该交互用于非等长文本的最终表示，并加入注意力层以支持类似分类的数据交互。s 表示对 I 的相容性得分权重，I 表示交互表示。在联合词嵌入过程中产生整段文本的相容性得分，因此最终文本 T 由式（6）给出。

$$I = (i1, i2, \cdots, iL) \tag{5}$$

$$T = sI \tag{6}$$

每个句子记为 al = {wd1, wd2, ⋯, wdn}，其中 wd1 表示句中的每个词。预训练 BERT 把所有句子 al = {wd1, wd2, ⋯, wdn}编码为相应的词嵌入形式{E1, ⋯, En}，见式（7）；其中 BERTtoken(al)表示把词编码为词向量。整段文本 A = {a1, ⋯, aL}的词嵌入表示记为 E = {E1, E2, ⋯, EL} = {{e1, ⋯, en}, {e1, ⋯, en}, ⋯, {e1, ⋯, en}}，其中 n 表示词的总数。此外，b 表示与文本序列 A 相关联的标签，由 BERT 编码为标签嵌入 F = {f1, f2, ⋯, fk}，其中 K 表示类别数。F = {f1, f2, ⋯, fk}由式（8）给出。

$$Vl = BERTtoken(al) \tag{7}$$

$$fk = BERTtoken(b) \tag{8}$$

标签与词被嵌入到同一联合空间。计算标签—词对之间相容性 G 的简单方法是余弦相似度，见式（9）。矩阵或向量运算中的逐元素相除记为⊘；Ĝ表示规模为 K × L 的归一化矩阵，其元素表示为ĝ = ‖ck‖ ‖el‖，其中‖ • ‖表示 2 范数；el 与 fk 分别表示第 l 个词与第 k 个标签的嵌入。借助非线性函数，在求取标签—词对相容性时计算相邻词之间的空间相对信息。式（10）刻画了全部标签与第 q 个短语之间的高层相容性。

$$G = (F^{\top} E) \oslash \hat{G} \tag{9}$$

$$eq = ReLU(Gq-i:q+i\ WRD2 + b2) \tag{10}$$

式中，Gq−i:q+i 表示标签到词元的相容性，WRD2 表示权重，b2 表示偏置。最大池化操作依据全部标签取第 q 个短语的最大相容性值，见式（11）；整段文本序列的相容性得分见式（12）；sq 表示 Softmax 的第 q 个元素，见式（13）。

$$mq = \max(eq) \tag{11}$$

$$s = Softmax(m) \tag{12}$$

$$sq = \exp(mq) / \Sigma q{=}1..L \exp(mq) \tag{13}$$

式中，L 与 m 分别表示长度与向量。整段文本序列的相容性得分 s 由所学词嵌入与标签嵌入计算得到。s 用于捕捉大量交互信息，并对最终交互文本表示 I = (i1, i2, ⋯, iL)加权。此外，由于标签学习到大量文本信息，分类器可有效利用这些加权标签进行分类；s 还用于对最终标签向量 Fk 加权，见式（14）与式（15）。

$$T = \Sigma q\ sq\ iq \tag{14}$$

$$\dot{F} = \Sigma q\ sq\ Fk \tag{15}$$

式中，T 与$\dot{F}$分别表示最终文本表示与标签表示。此外，在分析阶段还提取了不同陈述的相似度得分。训练好的分类方法记为 C: S → Y。这里考虑软标签设定，即攻击者查询分类器以由所生成输入获取输出概率，而模型参数与训练数据不可访问。例如，加权样例记为 S_weight，需要为给定输入对生成，并满足条件 C(S_weight) ≠ y；S_weight 在语法上须正确并保持语义 S。为生成加权样例 S_weight，本文设定了两类词元级扰动：（1）在满足 a ∈ S 的条件下把某词元替换为另一词元；（2）向 S 中插入新词元 a。

相较其他词元，输入中较少的词元对经 C 得到的最终检测结果贡献更大。词元替换或新词元插入对更新分类器的检测结果影响显著。对每个“a”，把“a”从 S 中移除并降低正确标签（y）检测概率的计算，即可得到词元重要性 Ii。本文用预训练 BERT 检测相似词元，这类相似词元与文本语法及语境契合良好。在替换词元时，若出现多个词元均导致 C 对 S 误分类，则依据相似度得分选取使 S_weight 与原始 S 最接近的词元；若未发生误分类，则选取另一个能降低检测概率的词元。词元扰动被反复施加，直至 C(S_weight) ≠ S 或 S 中所有词元均被扰动。由于上下文长度固定，BERT 建模长程依赖的能力有限，因此 BERT 仅用作提取过程而不用于分类。特征提取之后，用具备捕捉语义数据长程依赖潜力的 SRNN 完成分类。

（二）分类

提取特征之后，用 SRNN 有效地检测并分类虚假新闻。把 WFT-BERT 提取的特征由稠密 BERT 嵌入转换为稀疏表示后输入稀疏 RNN，使 SRNN 能够生成所提取特征中蕴含的序列数据，并依据所学的时序依赖进行检测与分类。SRNN 把新颖的稀疏连通性与标准 RNN 架构相结合，以有效捕捉长期依赖；该方法通过保留关键连接、剔除冗余连接来保存重要信息，从而保证检测的准确性。RNN 是自然语言处理（NLP）中常用的一类神经网络，因其具备记忆先前估计的能力：它把此前所有计算纳入当前层结果的评估。这与所有输入彼此独立的传统神经网络不同——RNN 的每个输入都前向传递至后续层，每层的结果都依赖于前面各层；之所以称为“循环”，是因为它对所有组件执行相同的函数。然而，RNN 用于分类时训练过程缓慢且复杂。为克服这一问题，本文采用 SRNN：它通过学习恰当的模式与连接提升泛化能力，并在推理与训练两方面获得效率与效果的提升。这里引入的 RNN 稀疏训练是 SRNN 的一个新类别。

1. 稀疏拓扑初始化

在稀疏拓扑初始化中，网络 y 由式（16）表示，其中 θ ∈ R 表示网络的稠密参数。该方法不从稠密参数出发，而使网络从 θ s 开始。由于对稀疏连接的支持有限，这里用掩码来强制稀疏结构。网络的基本初始化见式（17）。

$$y = f(x;\ \theta) \tag{16}$$

$$\theta s = \theta * M \tag{17}$$

式中，M 表示含非零分量的二值掩码，由 Erdos-Renyi 分布或随机分布初始化。按 Erdos-Renyi 方式，神经元 hj(k−1)与 hi(k)之间的 Mij(k)以式（18）所示概率存在。

$$P(Mij(k)) = \varepsilon (nk + nk-1) / (nk\ nk-1) \tag{18}$$

式中，nk、nk−1 表示 hk 与 hk−1 的神经元数量，ε 表示按稀疏度 s 评估得到的参数。按 Erdos-Renyi 拓扑初始化时，权重规模较大的层比规模较小的层具有更高的稀疏度。稀疏初始化是另一种方法，它按与总体稀疏度 s 相同的稀疏度初始化每一层。

2. 剪枝策略

在稀疏演化训练（SET）中，与基于幅值的剪枝不同——后者在每轮训练后剔除每层中最小正权重与最大负权重的ς部分——本文另选按绝对值最小的方式剪枝。对每个 θ s(i)，其重要性定义为其绝对值，见式（19）。S(θ S)的第 p 百分位由某一剪枝率 p 按升序 γ 确定，新掩码见式（20）。此外，训练过程中剪枝率 p 被迭代衰减至 0，使稀疏拓扑收敛至最优。

$$S(\theta s(i)) = |\theta s(i)| \tag{19}$$

$$M = S(\theta S) > \gamma \tag{20}$$

3. 再生长策略

本文利用非零参数的信息随机再生新权重，从而在前向与后向过程中都维持纯粹的稀疏结构。这正是 ST-RNN 与 SNFS（从头开始的稀疏网络）、RigL（操纵的彩票）等基于梯度的稀疏训练方法的主要差别：基于梯度的再生长高度依赖全部梯度参数，且每 ΔT 次迭代至少仍需一次稠密前向传播；而 SRNN 明确保持后向稀疏传播，所需浮点运算（FLOPs）更少。随机再生长见式（21）。

$$M = y + R \tag{21}$$

式中，R 表示二值张量，其非零分量随机分布。新激活的连接总数与被剔除的连接数相同，以维持相同的稀疏水平。此外，每层的稀疏度保持固定，训练模型所需的 FLOPs 与其稠密对应物成比例。由于 SRNN 以可变

时间步生成序列数据，模型得以有效捕捉长程依赖；它处理稀疏输入序列、恰当的上下文并检测虚假新闻模式。此外，通过分类，它依据所学的序列模式区分真实与误导性信息，从而提升准确率并有效检测虚假新闻。表 1 给出符号说明。

Table 1 Notation Description

表 1 符号说明

符号	说明
A	文本序列
BERTsent(·)	把句子编码为句向量
ul	隐藏向量表示
W1 与 b1	权重与偏置参数
s	相容性得分
T	最终文本
BERTtoken(al)	把词编码为词向量
n	词的总数
K	类别数
⊘	矩阵或向量运算
$\hat{G}$	归一化矩阵
fk	第 l 个词与第 k 个标签的嵌入
eq	全部标签与第 q 个短语之间的高层相容性刻画
Gq−i:q+i	标签到词元的相容性
WRD2	权重
b2	偏置
sq	Softmax 的第 q 个元素
L 与 m	长度与向量
T 与$\dot{F}$	最终文本表示与标签表示
S_weight	加权样例
y	网络
θ ∈ R	网络的稠密参数
M	二值掩码
nk, nk−1	hk 与 hk−1 的神经元数量
ε	按稀疏度 s 评估得到的参数
p	剪枝率
γ	升序
R	二值张量

四、结果与讨论

本文在配备 16 GB 内存、Intel Core i5 处理器与 Windows 10 操作系统的 Python 3.8 环境中仿真 WFT-BERT-SRNN。用于评估所提出方法的指标为准确率、召回率、F1 分数与精确率，其数学表达见式（22）至式（25）。式中，FP 为假阳性，TP 为真阳性，FN 为假阴性，TN 为真阴性。

$$Accuracy = (TP + TN) / (TN + TP + FN + FP) \times 100 \quad (22)$$

$$Precision = TP / (TP + FP) \quad (23)$$

$$Recall = TP / (TP + FN) \quad (24)$$

$$F1\text{-}Score = 2 \times TP / (2 \times TP + FP + FN) \times 100 \quad (25)$$

（一）性能分析

本节给出 WFT-BERT-SRNN 的定性与定量评估，见表 2 至表 5。表 2 给出在 Buzzfeed 数据集上不同特征提取方法的结果，把词袋（BoW）、Word2Vec、词频—逆文档频率（TF-IDF）与 BERT 等现有技术与 WFT-BERT 相比较；与上述现有技术相比，WFT-BERT 取得 0.9847 的更优准确率，原因在于 WFT-BERT 能够捕捉上下文细微差别与词间语义关系，提供更丰富的表示；这种更强的理解能力提升了虚假新闻的特征提取与检测准确率。

Table 2 Different Feature Extraction Methods Utilizing Buzzfeed Dataset

表 2 Buzzfeed 数据集上不同特征提取方法的比较

性能指标	BoW	Word2Vec	TF-IDF	BERT	WFT-BERT
准确率	0.8532	0.8614	0.8736	0.8771	0.9847
召回率	0.8425	0.8356	0.8547	0.8625	0.9704
精确率	0.8520	0.8453	0.8347	0.8598	0.9712
F1 分数	0.8435	0.8374	0.8215	0.8603	0.9674

表 3 给出在 Buzzfeed 数据集上的分类性能，把 DNN、CNN、LSTM 与 RNN 等现有技术与 SRNN 相比较；与上述现有技术相比，SRNN 取得 0.9847 的高准确率，原因在于所提出方法利用稀疏连通性提升了模型效率并减轻了过拟合，使其能更好地捕捉文本数据中的细微模式。

Table 3 Classification Performance Utilizing Buzzfeed Dataset

表 3 Buzzfeed 数据集上的分类性能

性能指标	DNN	CNN	LSTM	RNN	SRNN
准确率	0.9425	0.9547	0.9586	0.9635	0.9847
召回率	0.9357	0.9468	0.9515	0.9596	0.9704
精确率	0.9426	0.9536	0.9615	0.9654	0.9712
F1 分数	0.9357	0.9235	0.9567	0.9588	0.9674

表 4 给出在 PolitiFact 数据集上不同特征提取方法的结果，把 BoW、Word2Vec、TF-IDF 与 BERT 等现有技术与 WFT-BERT 相比较；与上述现有技术相比，WFT-BERT 取得 0.9724 的更优准确率：它通过为不同片段或词元赋予不同的重要性，使模型能更多地关注文本中的相关部分，从而提升上下文理解与虚假新闻检测的准确率。

Table 4 Different Feature Extraction Methods Utilizing PolitiFact Dataset

表 4 PolitiFact 数据集上不同特征提取方法的比较

性能指标	BoW	Word2Vec	TF-IDF	BERT	WFT-BERT
准确率	0.8336	0.8426	0.8563	0.8625	0.9724

召回率	0.8416	0.8320	0.8436	0.8536	0.9612
精确率	0.8425	0.8303	0.8361	0.8584	0.9547
F1 分数	0.8507	0.8436	0.8635	0.8502	0.9309

表 5 给出在 PolitiFact 数据集上的分类性能，把 DNN、CNN、LSTM 与 RNN 等现有技术与 SRNN 相比较；SRNN 取得 0.9724 的优异准确率，原因在于与 DNN、CNN、LSTM 与 RNN 等现有方法相比，所提出的方法借助稀疏连通性提升了模型效率。

Table 5 Classification Performance Utilizing PolitiFact Dataset

表 5 PolitiFact 数据集上的分类性能

性能指标	DNN	CNN	LSTM	RNN	SRNN
准确率	0.9412	0.9456	0.9548	0.9625	0.9724
召回率	0.9354	0.9375	0.9456	0.9520	0.9612
精确率	0.9432	0.9357	0.9435	0.9468	0.9547
F1 分数	0.9135	0.9257	0.9215	0.9265	0.9309

（二）对比分析

表 6 至表 8 给出在 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 数据集上，现有技术与所提出方法之间的对比分析。在 Buzzfeed 数据集上，用文献[17]、[22]与 WFT-BERT-SRNN 比较；在 PolitiFact 上，用文献[17]、[20]、[22]与 WFT-BERT-SRNN 比较；在 Fakeddit 与 Weibo 数据集上，用文献[23]、[24]与所提出方法比较。与这些现有技术相比，所提出的 WFT-BERT-SRNN 借助 BERT 的上下文嵌入获得丰富的文本表示，并借助 SRNN 的序列建模捕捉新闻文章中的模式，在 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 数据集上分别取得 0.9847、0.9724、0.9624 与 0.9725 的更优准确率。这一组合通过有效处理时序依赖与复杂语言特征，提高了区分虚假与真实新闻的准确率。

Table 6 Comparative Analysis with Existing Techniques Utilizing Buzzfeed

表 6 Buzzfeed 数据集上与现有技术的对比分析

性能指标	DeepFake[17]	EDU4FD[22]	本文 WFT-BERT-SRNN
准确率	0.8649	0.7488	0.9847
召回率	0.8696	0.7486	0.9704
精确率	0.8333	0.7519	0.9712
F1 分数	0.8511	0.7475	0.9674

Table 7 Comparative Analysis with Existing Techniques Utilizing PolitiFact

表 7 PolitiFact 数据集上与现有技术的对比分析

性能指标	DeepFake[17]	BERT 联合框架[20]	EDU4FD[22]	本文 WFT-BERT-SRNN
准确率	0.8864	0.84	0.7162	0.9724

召回率	0.8460	N/A	0.7111	0.9612
精确率	0.8210	N/A	0.7155	0.9547
F1 分数	0.8404	0.87	0.7110	0.9309

Table 8 Comparative Analysis with Existing Techniques Using Fakeddit and Weibo Datasets

表 8 Fakeddit 与 Weibo 数据集上与现有技术的对比分析

性能指标	Fakeddit：基于图像描述的方法[23]	Fakeddit：本文 WFT-BERT-SRNN	Weibo：基于图像描述的方法[23]	Weibo：FMFN[24]	Weibo：本文 WFT-BERT-SRNN
准确率	0.9251	0.9624	0.8886	0.885	0.9725
召回率	0.9374	0.9515	0.9201	N/A	0.9615
精确率	0.9383	0.9520	0.8692	N/A	0.9536
F1 分数	0.9379	0.9588	0.8939	N/A	0.9621

（三）讨论

本研究展示了用 WFT-BERT-SRNN 进行虚假新闻检测。与现有方法相比，该方法在 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 数据集上把检测准确率显著提升至 0.9847、0.9724、0.9624 与 0.9725，这归因于该模型能够利用微调后的上下文嵌入与稀疏循环处理更好地捕捉虚假新闻中的细微模式。本节讨论所提出的 WFT-BERT-SRNN 的优势与现有技术的局限。现有技术的局限包括：DeepFake 因其静态特性，在应对演变模式与捕捉虚假新闻动态方面存在困难；SGCP+SVM 在大量样本缺失时检测性能下降；BERT 联合框架在区分不同信息类型方面存在困难，影响分类；EDU4FD 可能损失上下文连贯性，导致误解；基于图像描述的方法依赖预定义词表，难以准确描述复杂视觉概念。所提出的方法克服了上述现有技术的局限。

在 WFT-BERT 中，深度双向学习有助于在特定任务上取得与传统架构相当甚至更好的结果；SRNN 则通过学习恰当的模式与连接提升泛化能力，并在推理与训练两方面获得效率与效果的提升。本文的主要目的是借助自然语言处理提高虚假新闻检测系统的准确性与稳健性。通过把微调 BERT 嵌入与 SRNN 结合，所提出的方法增强了模型辨识与错误信息相关的细微语言线索与上下文信息的能力，并借助二者优势的整合弥补了现有方法的不足。本研究的意义在于其提升虚假新闻检测系统可靠性与准确性的潜力。

五、结论

由于错误信息对公众舆论与社会稳定具有广泛影响，虚假新闻检测是一个关键研究领域。本研究提出 WFT-BERT-SRNN 以准确检测虚假新闻。WFT-BERT 增强了模型在训练中优先关注特定领域或任务的能力，使其能更有效地捕捉专有信息，从而提高稳健性与准确率；SRNN 的优势在于能够聚焦最相关的模式与连接，从而高效地处理与学习大量文本数据，提升泛化能力并使模型能准确地把新闻分类为真实或虚假。与 DeepFake、BERT 联合框架、EDU4FD、基于图像描述的方法与 FMFN 等现有方法相比，WFT-BERT-SRNN 在 Buzzfeed、PolitiFact、Fakeddit 与 Weibo 数据集上分别取得 0.9847、0.9724、0.9624 与 0.9725 的更优准确率。这些结果标志着虚假新闻检测领域的实质性进展：高准确率表明 WFT-BERT-SRNN 可作为新闻机构识别并遏制错误信息传播的合适工具；该方法提供了稳健的框架，提升了信息的整体可信度。未来将考虑采用混合深度学习检测方法以进一步提升模型性能。

参考文献

[1] 钟将, 高晋鹏, 黄敬旺, 等. 基于证据增强和局部语义交互的多模态虚假新闻检测. 计算机学报, 2025, 48(3): 556-574.

[2] 杨思佳, 李显勇, 杜亚军. 基于在线社交网络的虚假新闻检测方法综述. 西华大学学报（自然科学版）, 2025(3): 37-46.

[3] Saleh H, Alharbi A, Alsamhi S H. OPCNN-FAKE: optimized convolutional neural network for fake news detection. IEEE Access, 2021, 9: 129471-129489.

[4] Dixit D K, Bhagat A, Dangi D. Automating fake news detection using PPCA and levy flight-based LSTM. Soft Computing, 2022, 26(22): 12545-12557.

[5] Jing J, Wu H, Sun J, et al. Multimodal fake news detection via progressive fusion networks. Information Processing and Management, 2023, 60(1): 103120.

[6] Sousa-Silva R. Fighting the fake: a forensic linguistic analysis to fake news detection. International Journal for the Semiotics of Law, 2022, 35(6): 2409-2433.

[7] 李润东, 汪旦丁, 王正佳, 等. 基于词抽象化的虚假新闻检测方法. 信息传播研究, 2025(1): 2-12.

[8] 顾亦然, 顾立强, 黄丽亚. 基于新闻环境的可解释虚假新闻检测方法. 小型微型计算机系统, 2025, 46(9): 2058-2065.