

基于情感的多头注意力卷积Transformer + CNN的假新闻检测

张亚立, 李征宇, 孙 平

沈阳建筑大学, 计算机科学与工程学院, 辽宁 沈阳

收稿日期: 2023年8月18日; 录用日期: 2023年9月18日; 发布日期: 2023年9月26日

摘 要

网络上虚假新闻的传播是我们社会的主要问题之一。虚假新闻是为了误导读者, 引发读者的强烈情绪, 并试图在社交网络上传播, 因此自动检测虚假新闻是一项艰巨的任务。尽管最近的研究已经探索了虚假新闻的不同语言模式, 但情感信号的作用尚未进行深入的探讨。本文研究了情感信号在假新闻检测中的作用, 利用了新闻发布者所传达的情感信息来对假新闻进行检测。我们提出了一个加入情感特征的多头注意力卷积Transformer + CNN模型(简称SMCT模型), 该模型不仅能捕获文本内容的局部和全局的依赖关系, 而且还能将文本内容与新闻发布者所传达的情感信息相融合来检测新闻的真假性。在两个真实的数据集上进行了大量的实验研究, 我们发现将情感特征引入模型中可以明显提升在区分真假新闻方面的准确率, 这一优化策略的应用有效地增强了模型在判别新闻真实性方面的能力。

关键词

假新闻检测, 注意力卷积, Transformer, CNN, 情感特征

Fake News Detection Based on Emotional Multi-Head Attention Convolution Transformer + CNN

Yali Zhang, Zhengyu Li, Ping Sun

School of Computer Science and Engineering, Shenyang Jianzhu University, Shenyang Liaoning

Received: Aug. 18th, 2023; accepted: Sep. 18th, 2023; published: Sep. 26th, 2023

Abstract

The spread of false news on the network is one of the main problems in our society. Fake news is

designed to mislead readers, arouse strong emotions among readers, and try to spread it on social networks, so the automatic detection of false news is a difficult task. Although recent studies have explored different linguistic patterns of fake news, the role of emotional signals has not been explored in depth. This paper studies the role of emotional signals in the detection of fake news, using the emotional information conveyed by news publishers to detect fake news. We propose a multi-head attention convolution Transformer + CNN model (SMCT model), which can not only capture the local and global dependence of the text content, but also integrate the text content with the emotional information conveyed by the news publisher to detect the truth and falsehood of news. After conducting a large number of experimental studies on two real datasets, we find that introducing emotional features into the model can significantly improve the accuracy in distinguishing true and false news. The application of this optimization strategy effectively enhances the ability of the model in judging the authenticity of news.

Keywords

Fake News Detection, Attention Convolution, Transformer, CNN, Emotional Features

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着互联网和移动智能设备的出现，人们在网络上发布和分享信息变得越来越方便和快捷。对于想要了解的新闻，可以不再需要用传统的电视台去接收新闻信息[1] [2]，现在人们只需要点击一下鼠标，就能够查找到想要查看的新闻。新闻媒体也从报纸、杂志、传统的电视台演变为了数字形式，如新闻平台、微博、人民日报和社交媒体等其他数字媒体形式[3]。然而，社交媒体平台是一把双刃剑，有利也有弊。好处是能够让用户讨论和分享自己的想法，并就民主和健康等问题进行探讨。弊端是这些平台很容易被某些实体以负面的视角去使用，有些实体为了获取经济利益[4] [5]，采用吸引人眼球的方式，在平台上发布虚假广告和政治言论。这些言论具有操纵性，旨在制造偏见、影响人们的心态，并传播讽刺或荒谬的信息，这些信息通常被称为假新闻。

虚假和误导性的新闻已经严重的影响了人们的生活，例如 2016 年美国总统大选[6]和英国脱欧公投事件使用了大量的虚假和误导性新闻，最终的结果对社会产生了许多方面的有害影响，包括社会、金融甚至政治领域。假新闻是虚假和真实信息的混合物，并且故意编写以迷惑读者并引发他们的强烈情绪。越来越多的事件已经表明，人们对后来被证明是虚假的消息做出了荒谬的反应[7] [8]。最近的一个案例就是新型冠状病毒的传播，其中关于病毒的起源、性质和行为的虚假报告在网络上急剧传播[9]。越来越多的人会在社交媒体上阅读到了虚假的新闻，情况也在进一步的恶化。因此，假新闻已经成为了一个全球性的问题，也是一个必须被发现和化解的重大的挑战。

为了解决这个问题，研究人员一直致力于开发自动检测假新闻的方法。据我们所知，目前的大部分研究都是利用新闻文本信息进行假新闻检测。许多研究者发现，假新闻会引发恐惧、厌恶和惊讶等情绪反应，而真实新闻则会引发喜悦、悲伤、信任和期待等情绪反应。然而，研究人员还尚未深入探索新闻发布者所传达情感在自动虚假信息检测中的有效性。因此，在进一步优化方法时，值得考虑将新闻发布者所要传达的情感信息纳入模型中，以提高假新闻检测的准确性和有效性。

在本文中,我们采用了我们提出的多头注意力卷积 Transformer + CNN (卷积神经网络)作为假新闻分类器。我们的研究受到 Zhang 等人[10]的启发,他们提出了 5 种情感特征,包括情感类别、情感词汇、情感强度、情感评分和其他辅助特征。我们在此基础上加入了一种新的情感特征,即情感词共现,将总共的情感特征扩展为 6 种,并将其应用到我们的分类器中。我们的研究还涉及不同领域的新闻,旨在进行真假新闻的检测。实验结果表明,本文的 Transformer + CNN + 6 种情感对于假新闻的检测表现出了很优异的效果,与现有的假新闻检测器相比,我们的模型提高了检测的准确性,并同时突显了情感特征能作为一种增强和补充功能,可以轻松地插入现有的假新闻检测分类器中。本文主要贡献总结如下:

1) 我们提出来的多头注意力卷积 Transformer + CNN 不仅能捕获新闻文本内容局部和全局的依赖关系,而且还能解决传统机器学习在数据不平衡时出现误分的问题。

2) 本文的 6 种情感特征可以很好地与我们的多头注意力卷积 Transformer + CNN 假新闻检测器兼容,有效提高假新闻检测性能。

3) 在两个真实的数据集上进行了大量的实验研究,实验结果表明: 1) 在对假新闻进行检测时,综合考虑新闻内容发布者的情感特征和新闻文本内容要比仅仅使用文本内容进行检测更为有效。2) 我们所提出的情感特征作为一种补充和增强功能,可以很容易的插入现有的假新闻检测器中,而且在识别虚假新闻方面提高了模型的性能。

2. 相关工作

假新闻检测又称谣言检测、假信息检测等,现有文献[11] [12] [13] [14]主要关注分析新闻的文本内容,包括标题和正文,以提取假新闻的特征。一些研究人员采用语言学方法来区分假新闻和真新闻,而其他人员则检查新闻标题和正文之间的一致性。例如, Rashkin 等人[11]专注于验证政治报道和检测假新闻。他们建议通过比较真实新闻和假新闻(如讽刺、恶作剧和宣传)的语言特征来识别错误信息。在文献[12]中, Bhatt 等人通过评估给定新闻文章的正文与其声明的一致性来进行假新闻检测,使用统计特征和外部特征构建了一个通用的假新闻分类器。Jin 等[13]认为在微博上,图像在新闻传播中发挥着重要作用。真假新闻图片的分布模式存在显著差异。因此,他们从视觉内容和统计两个角度提取图像特征。除文本特征外,新闻通常包含各种类型的数据。在早期信息可信度评价研究中, Castillo 等人[15]手动从新闻片段中提取了内容特征、发布者特征、主题特征和传播特征。研究发现,基于情感的特征(如情感词汇和感叹号的比例)对于评估信息的可信度是有效的。Ajao 等人[16]指出,新闻的真实性与发布文本的情绪之间存在关联,并引入了一个情感特征(消极词汇和积极词汇的数量之比),用于辅助纯文本假新闻检测器。Giachanou 等人[17]则从新闻内容中提取更丰富的情感特征,基于情感词汇进行假新闻检测,而不仅仅依赖单一特征。

然而,上述方法大多只考虑新闻文本中的单个情感特征,而忽视了从各个角度挖掘新闻文本的情感特征的重要性。因此,本文旨在融合新闻文本的内容与新闻发布者所传达的情感信息,以对假新闻进行检测。本文的 6 种情感综合考虑了新闻文本所表达的多方面情感特征,同时将该特征融入到多头注意力卷积 Transformer + CNN 模型中,进一步提高检测真假新闻的准确性。该模型采用多头注意力卷积机制,可以提取文本的关键特征、位置信息,捕捉全局和局部信息以及提高深度学习能力。通过结合局部特征,全局特征和位置信息,计算文本的整体注意力的输出,在模型中引入多头注意卷积机制,从多个子空间中学习更丰富的文本信息。

3. SMCT 模型

模型 SMCT 的框架结构如图 1 所示,主要由三部分组成:多头注意力卷积 Transformer、CNN 和情感特征嵌入。为了模拟用于假新闻检测的情感特征信号,我们利用了新闻发布者所传达出的情感信息。

具体而言, 情感嵌入部分展示了获取新闻发布者的情感特征的过程, 并将这些情感特征嵌入到提出的多头注意力卷积 Transformer + CNN 的检测器中, 以补充和增强新闻真假检测的功能。在本节中, 我们详细介绍了新闻发布者情感的特征提取以及多头注意力卷积 Transformer 的工作原理, 然后描述了将这些情感特征插入到多头注意卷积 Transformer + CNN 检测器的过程。SMTC 模型原理图如图所示:

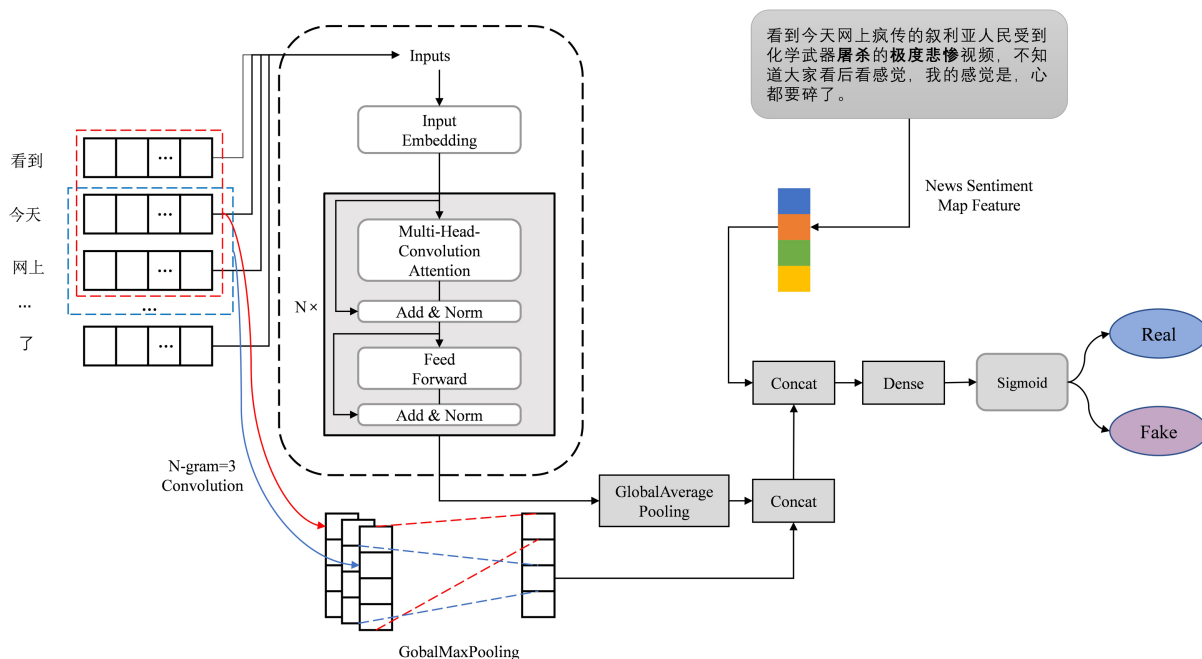


Figure 1. Schematic diagram of the SMTC model

图 1. SMTC 模型原理图

3.1. 新闻发布者情感

为了能将发布者的情感全面的表示出来, 本文采用了研究者 Zhang 等人[10]从新闻内容中提取到的多种特征, 包括情感种类, 情感词汇, 情感强度, 情感评分以及其他辅助特征。另外, 我们在此基础上提出一种新的情感特征——情感词共现关系。接下来, 本文将详细的介绍一下我们提出的情感特征, 并简要说明其他 5 种情感特征的相关内容。

对于给定的具有一定长度的新闻文本内容的输入序列 L , $T = [t_1, t_2, \dots, t_i, \dots, t_L]$, 其中 t_i 是新闻文本 T 中的词语, 目标是从新闻文本 T 中提取情感特征 emo_T 。

1) 情感词共现(emo_T^m): 情感词汇共现关系可以帮助我们了解不同情感词汇在文本中的相互作用, 因为文本中某些情感词的存在可以指示文本的整体情感。例如, 如果一篇新闻包含“快乐”、“开心”和“兴奋”等词, 并且这些情感词同时出现了多次, 那么该新闻的整体情绪很可能是积极的。另一方面, 如果新闻中包含“悲伤”、“沮丧”和“愤怒”等词, 那么新该闻的整体情绪很可能是负面的。通过分析新闻中情感词的共现, 我们可以提取用于新闻文本真假检测分类的特征。为了提取这些特征, 本文计算情感词汇之间的共现矩阵。给定情感词典 $\mathcal{E}_e$ 和新闻文本 T , 我们可以计算共现矩阵 M , 具体计算公式如(1)所示:

$$M[i][j] = count(v_{e,i}, v_{e,j}) \quad (1)$$

其中 $count(v_{e,i}, v_{e,j})$ 是情感词 $v_{e,i}$ 和情感词 $v_{e,j}$ 在新闻文本中共同出现的次数。

2) 情感类别(emo_T^{cate}): 情感类别用于描述文本所表达的情感类型, 例如喜悦、愤怒、悲伤等。通过将新闻文本映射到不同的情感类别, 可以更准确地理解文本中蕴含的情感。

3) 情感词汇(emo_T^{lex}): 通常, 一篇新闻文本会使用几个特定的词语来表达特定的情感, 适当使用情感词汇可以使新闻更具有感染力和说服力, 其中这些词大多会被包含在由专家注释的现有的情感词典中。例如, 在句子“我今天非常不开心”中, 我们计算单词“开心”在情感词典中“喜悦”上的分数。我们需要计算“开心”出现的频率, 以及其上下文的否定词“不”与程度词“非常”在现有情感词典中的分数。通过逐渐聚合每个单词和整个文本在所有情感上的分数, 以得到丰富的表征。

4) 情感强度(emo_T^{int}): 情感强度表示情感的强度或强烈程度。例如, 在表达高兴这一情感时, 词语“欣喜若狂”比“快乐”具有更高的强度。通过对情感词汇进行权重计算或使用一定的规则, 可以确定每个情感词所对应的情感强度, 从而更准确地理解文本中情感的强度。

5) 情感评分(emo_T^{sent}): 情感评分是对新闻文本整体情感的粗粒度得分, 它代表了整个新闻文本的正或消极性的程度。通过使用情感词典或公共工具包来计算, 用于量化新闻文本所表达的情感倾向。

6) 其他辅助特征(emo_T^{aux}): 引入了一组辅助特征来捕捉非词语元素背后的情感信号, 包括表情符号和标点符号, 并增加了情感词汇和人称代词的使用频率, 以增强用户对词汇用法的意识。

为了从内容中得到新闻文本 T 的新闻发布者情感, 我们将情感词共现 emo_T^m 、情感类别 emo_T^{cate} 、情感词汇 emo_T^{lex} 、情感强度 emo_T^{int} 、情感评分 emo_T^{sent} 和辅助特征 emo_T^{aux} 等进行拼接, 得到最终的情感特征表示 emo_T , 如式(2)所示, 式中 $\oplus$ 表示各情感特征拼接。

$$emo_T = emo_T^m \oplus emo_T^{cate} \oplus emo_T^{lex} \oplus emo_T^{int} \oplus emo_T^{sent} \oplus emo_T^{aux} \quad (2)$$

3.2. 多头注意力卷积 Transformer + CNN 模型

1) 多头注意力卷积 Transformer: 传统的 Transformer 的多头注意力机制利用了线性变换将输入向量映射为查询、键、值的组合, 然后计算注意力分数。本文提出的多头注意力卷积 Transformer 模型的多头注意力机制是将线性变换替换成卷积操作, 利用卷积层将输入向量映射为查询、键、值的组合。通过使用卷积滤波器, 将语义从复杂的词空间转换为信息更加丰富的卷积滤波器空间, 以捕获重要的 n-gram 信息。注意力卷积机制不仅能捕获局部和全局的依赖关系, 而且还能保留词语之间的序列关系或先后关系。具体的操作如下: 首先, 将输入的文本信息与训练好的词向量进行词嵌入 Embedding 表示, 然后对词向量进行卷积操作, 使用卷积滤波器捕获长度为 3 的 n-gram 特征。利用这些特征对卷积核进行加权, 得到文本特征映射空间表示, 其中包含了文本中各个 n-gram 局部表征。对局部表征进行最大池化, 找到文本与每个卷积滤波器的相关性。使用最大池化的特征映射来聚合卷积滤波器, 找到文本的整体语义, 得到全局表征。将局部表征, 全局表征和位置编码进行融合, 产生注意力计算的输入, 再将输入映射到一个多维子空间, 在该子空间中, 使用自注意机制捕获句子中单词的注意分布。接着, 对多个并行头的注意力分布应用全局注意力池化模块, 计算序列中每个注意力特征图的最大值, 从而产生一个固定长度的向量, 该向量概括了输入的最显著特征。最后, 在通过密集层生成最终输出之前, 将得到的向量重复并与多头卷积注意机制的输出连接, 以获得输入文本的最终特征表示, 使得模型更加专注于最重要的特征, 同时仍然捕获整体注意力的得分。

2) CNN 层: 能够有效的捕获文本中的局部特征。对文本检测的具体流程为: 首先, 将文本样本中的每个词或特殊字符转化为对应的向量表示。其次, 使用 300 个长度为 3 的卷积核对输入的向量进行滑动窗口卷积操作。对于每个卷积核, 将词嵌入 Embedding 与卷积核进行元素相乘, 并将结果相加。之后对特征映射的结果应用激活函数, 最后, 进行最大池化操作。

最后将多头注意力卷积 Transformer 的输出与 CNN 层的输出进行拼接，将两个模型进行组合的优势是利用注意力机制和 CNN 的特征提取能力，从输入文本中捕捉更丰富的语义、句法和局部信息，帮助模型理解新闻的上下文关系和逻辑，并更准确地判断新闻的真实性。

4. 实验分析

4.1. 数据集

本文使用了两个真实的中文数据集进行了实验，分别是 NAACL 上的中文的事实核查数据集 CHEF (论文作者来自清华大学，剑桥大学和北京师范大学 - 香港浸会大学联合国际学院的 PhilipYu 组)和北京市经济和信息化局&CCF 大数据专家委员会&中科院计算技术研究所提出的开源疫情期间互联网虚假新闻数据集。因为这两个数据集都共同拥有 5 个相同的领域，分别是医药健康、政治、教育考试、社会生活、科技。我们将这两个数据集的 5 个领域分别抽取出来，同时我们又自定义 Python 程序爬取微博平台辟谣网站中所属该 5 个领域的新闻信息，经过数据清洗工作来丰富现有的两个数据集，将每个抽出的领域组成了一个新的数据集。本文使用数据集简称来表示两个数据集，分别为互联网虚假新闻和事实核查新闻，统计结果如表 1。两个数据集均被划分成训练集、验证集和测试集，比率分别为 60%，20%，20%。假新闻的标记为 1，真新闻的标记为 0。

Table 1. Dataset statistics

表 1. 数据集统计

		政治	社会生活	文化	科技	医药健康
互联网虚假新闻数据集	真	905	3454	764	850	1960
	假	1566	4107	684	2858	1921
	总共	2471	7561	1448	3708	3881
事实核查新闻数据集	真	634	2786	1528	776	901
	假	408	2546	581	3076	2258
	总共	1042	5332	2109	3852	3159

4.2. 基线模型

本文使用现有的假新闻检测基线模型与我们提出的 SMCT 模型进行比较，在两个真实的数据集上进行实验，以此来验证 SMCT 模型的有效性。

1) BiGRU：双向门控循环单元(Bidirectional Gated Recurrent Unit, BiGRU)是基于循环神经网络(RNN)的模型，其由单向的、方向相反的 GRU 组成的神经网络模型，它们的输出共同决定了最终的输出结果。在每个时刻，输入会同时提供两个方向相反的 GRU，而输出则由这两个单向 GRU 共同决定。基于文本的模型，如 GRU [18]和 LSTM [19]，在文献[20] [21]中被证明对假新闻检测有效。同时，本文采用了 Chinese Word Vectors 作为中文词向量的表示方法。

2) BiGRU + Attention：BiGRU + Attention 模型采用六层结构，把文本词向量经输入层进入神经网络(BiGRU)层提取特征。接下来，引入注意力(Attention)机制突出词向量的重点信息。为了防止过拟合，引入了 Dropout 层。然后经过全连接层对特征进行处理。最后，将输出结果输入到 Softmax 层进行真假新闻检测。

3) Transformer: Transformer 的核心是多头自注意力机制,它能够对输入序列中的每个位置计算一个向量表示,该表示同时考虑到了所有其他位置的信息。通过进行多头自注意力机制,Transformer 能够捕捉到更长距离的依赖关系,从而更好地理解文本的语义。除了自注意力机制,Transformer 还引入了残差连接和层归一化等技术,有助于加快训练收敛速度并提高模型性能。

4.3. 实验设置

为了验证本文提出的方法的准确性和有效性,我们在参数设置如表 2 所示的实验环境下进行了一系列实验。

Table 2. Experimental environment and configuration
表 2. 实验环境与配置

实验环境	配置
操作系统	64 位 Windows11
CPU	Intel Core i5-12500H
内存/GB	16
GPU	RTX3050ti
编程语言	Python 3.9
深度学习框架	Keras 2.6.0

4.3.1. 情感资源

对于情感分类器,本文采用了被广泛认可的百度 AI 中文模型,该模型在提取新闻文本的情感方面具备极高的准确度。对于其他情感资源,采用情感词汇本体[22]提取情感词汇和情感强度特征;本文利用 HowNet [23]字典来计算情绪分数;对于表情符号,本文利用维基百科[24]的表情列表,将表情分为五种情绪,分别为快乐、愤怒、惊讶、悲伤和中性;对于情感词和程度词,本文使用知网[23]的双语情感词典。

4.3.2. 情感基准

在本次实验中,本文选择了两个基线情感特征来评估本文的 6 种情感特征的有效性。这次的评估我们使用了知网的情感词典来实现:

1) Emoratio: Ajao 等[16]提出了一种可以从新闻内容文本中提取的情感特征,称为 Emoratio。由消极情绪词汇数与积极情绪词汇数之比计算得出。

2) EmoCred: Giachanou 等[17]利用内容文本的情感词汇和强度特征。这些特征是根据词汇的出现频率计算出来的。

4.3.3. 参数设置

本文使用的预训练词向量为 300 d 的微博中文(Word + Character+ n-gram),在注意力卷积计算时使用 300 个核大小为 3 的卷积过滤器。多头注意力卷积 Transformer + CNN 模型的注意力卷积头使用 15 个,使用 3 层编码器,批处理为 16,学习率设置为 1e-3,Dropout 设置为 0.5,激活函数使用 GelUs,损失函数使用 Cross-Entropy,优化器使用 Adam。同时二分类的情感特征中的子特征的维数由选用的中文特定的情感资源决定。其中情感类别的输出设置为 8 维;情感词汇的大小设置为 29;情感共现矩阵的大小设置为 21,情感分数由知网计算得到,只保留 1 个维度;辅助特征的维度设置为 17。

4.3.4. 实验执行流程

1) 首先, 根据新闻文本的最大序列长度设置为 200, 定义输入层的张量大小。其次, 我们使用嵌入层将输入序列中的每个单词映射为密集向量表示。其中, 嵌入层用预先训练的嵌入矩阵初始化, 并且被设置为不可训练的。

2) 接下来, 注意力卷积 Transformer 模块的处理操作为: 从嵌入层获得的特征张量通过一个 3 层注意力卷积变换器。该层利用多头注意力卷积和前馈神经网络操作来处理输入序列, 并通过全局平均池捕获序列的上下文, 最终提供固定一个大小的输出。全局平均池化的作用在于对每个特征的所有时间步长取平均操作, 可以对序列进行重新映射, 将其简化为单个向量。

3) 与此同时, CNN 模块的处理操作为: 将嵌入层得到的特征张量通过 Conv1D 层提取原始新闻文本之间的局部特征, 其中卷积核数量设置为 300, 卷积核大小设置为 3。接着输入全局最大池化层, 提取新闻文本中最重要的特征信息。

4) 最后, 拼接注意力卷积 Transformer 的输出、卷积最大池化层的输出和情感特征向量, 作为最终的新闻文本特征表示, 将其输入到一个双层的 MLP 分类器中进行最终分类。使用交叉熵函数作为损失函数。在隐藏层和输出层分别使用 ReLu 与 Sigmoid 作为激活函数。Sigmoid 函数输出介于 0 到 1 之间的置信度。如果这个置信度大于等于 0.5 时, 本研究预测该条为假新闻, 否则, 预测为真新闻。

4.3.5. 评估指标

为了评估算法的性能, 我们使用了两种不同的度量标准, 准确率(Accuracy)和宏 F1 值(Macro F1 Score)。它们大多基于混淆矩阵。混淆矩阵是测试集上分类模型性能的表现, 它由四个参数组成: 真阳性(True Positive, TP)、假阳性(False Positive, FP)、真阴性(True Negative, TN)和假阴性(False Negative, FN)。

准确率: 准确率通常是最常用的指标, 是指被检测系统正确标记为假新闻或真新闻的文章的百分比。为了计算模型性能的准确率, 其计算方法如下式(3)所示:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

在大多数情况下, 高准确率代表一个好的模型, 但考虑到我们正在训练我们的案例中的分类模型, 一篇预测为真而实际上为假(假阳性)的文章可能会产生负面后果; 同样, 如果一篇新闻在包含事实数据的情况下被预测为虚假, 这可能会产生信任问题。因此, 我们使用了另外三个考虑到错误分类观察的指标, 即精确率、召回率和 F1 值。

召回率(Recall): 召回率表示超出真实类别的积极分类总数, 它表示在真实文章总数中预测为真实的文章数, 其计算方法如下式(4)所示:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

精确率(Precision): 精确率分数表示真阳性与所有预测为真的事件的比率。本文中, 精确率显示了所有正向预测的新闻中标记为 True 的新闻的数量, 计算公式如(5)所示:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

F1 值: F1 值代表精确性和召回之间的权衡, 计算两者之间的调和平均值, 同时考虑了假阳性和假阴性观察结果。F1 值计算公式如(6)所示:

$$F1\text{-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

4.4. 结果与分析

4.4.1. 假新闻检测器性能的评估

表 3 和表 4 展示了基线模型和本文提出的 SMCT 模型在互联网虚假新闻数据集上的实验结果, 评估指标包括 Macro F1 Score 和 Accuracy。模型包括传统的假新闻检测模型、注意力卷积 Transformer、注意力卷积 Transformer 接 CNN 以及与本文的 6 种情感特征拼接的模型。总体而言, 我们提出的 SMCT 模型在每个领域的真假新闻检测中相较于其他基线模型都有显著的提升。从结果中可以看出, 仅对新闻文本进行检测时, 我们提出的多头注意力卷积 Transformer + CNN 模型无论在 Macro F1 Score 还是 Accuracy 上相比于传统模型均提升了 1% 左右。这是因为多头注意力卷积 Transformer + CNN 模型结合了注意力卷积机制和卷积神经网络, 能够更好地捕捉到丰富的语义信息和局部特征。多头注意力卷积 Transformer 模型通过自注意力机制对整个句子进行全局建模, CNN 模型可以有效地提取局部特征。这种综合的模型结构使得我们的模型能够更好地理解文本语义并捕捉重要特征信息。相比之下, BiGRU + Attention 模型通过双向 GRU 和注意力机制来建模上下文信息, 但 GRU 存在梯度消失和梯度爆炸的问题, 对于较长的文本序列, 可能无法有效地捕捉到长距离的依赖关系。此外, 在加入 6 种情感特征后, 这两种模型的性能都得到了显著提升。这表明情感特征的引入对于假新闻检测非常有帮助, 并在提升模型性能方面发挥了关键作用。

Table 3. Macro F1 of each model on the internet fake news dataset (%)

表 3. 互联网虚假新闻数据集上各模型的宏观 F1 (%)

模型	政治	社会生活	文化	科技	医药健康
BiGRU	82.23	84.55	85.31	72.93	81.05
BiGRU + Att ¹	84.63	84.77	86.49	73.13	82.25
Transformer ^{*2}	84.20	85.94	86.93	75.34	84.04
Transformer [*] + CNN	84.09	86.14	87.70	76.02	84.62
BiGRU + Att + 6Sentiment ³	85.73	85.04	86.53	75.53	83.32
SMCT	86.20	87.28	88.49	77.29	84.83

¹Att 为 Attention 简写; ²加*为注意力卷积 Transformer; ³6Sentiment 为本文的 6 种情感特征。

Table 4. The accuracy of each model on the internet fake news dataset (%)

表 4. 互联网虚假新闻数据集上各模型的准确率(%)

模型	政治	社会生活	文化	科技	医药健康
BiGRU	84.04	84.60	85.32	82.28	81.10
BiGRU + Att	85.25	84.86	86.51	82.78	82.28
Transformer [*]	85.25	85.99	87.24	85.10	84.04
Transformer [*] + CNN	85.45	86.26	87.70	85.26	84.62
BiGRU + Att + 6Sentiment	86.67	85.19	86.90	84.27	83.33
SMCT	87.27	87.38	88.49	86.09	84.86

表 5 和表 6 展示了在事实核查新闻数据集的五个领域上所有模型的实验结果, 与其他模型相比, 本

文提出的 SMCT 模型在 Accuracy 和 Macro F1 Score 上均有 1%至 2%左右的提升。其中在文化领域, SMCT 的 Macro F1 Score 远高于基线模型, 高达 86.14%, 这表明该模型在识别真实新闻和假新闻方面具有相对较高的准确性和覆盖率, 同时在提取和学习对假新闻检测有用的特征方面更加有效。

Table 5. Macro F1 of each model on the fact-checking news dataset (%)

表 5. 事实核查新闻数据集上各模型的宏观 F1 (%)

模型	政治	社会生活	文化	科技	医药健康
BiGRU	84.71	81.74	79.47	78.67	75.55
BiGRU + Att	85.18	82.22	81.04	80.44	78.37
Transformer*	84.74	82.75	83.03	79.48	79.89
Transformer* + CNN	84.75	83.19	83.27	81.08	80.41
BiGRU + Att + 6Sentiment	84.87	82.63	81.29	80.23	80.19
SMCT	86.19	85.23	86.14	81.71	80.83

Table 6. The accuracy of each model on the fact-checking news dataset (%)

表 6. 事实核查新闻数据集上各模型的准确率(%)

模型	政治	社会生活	文化	科技	医药健康
BiGRU	85.62	83.44	82.97	82.11	82.75
BiGRU + Att	86.12	83.85	83.70	82.27	83.40
Transformer*	85.65	84.25	85.28	83.23	84.19
Transformer* + CNN	86.23	84.44	86.13	83.39	84.78
BiGRU + Att + 6Sentiment	86.67	85.19	86.90	84.27	83.33
SMCT	87.43	86.35	88.32	84.19	85.18

综合两个数据集的实验结果显示, 我们提出的 SMCT 模型在真假新闻检测任务中展现出更强大的特征提取能力, 明显优于传统的基线模型。这是因为我们的注意力卷积机制能够有效捕捉局部和全局特征之间的依赖关系, 并采用位置编码保留了文本之间的语序信息。此外, 采用多头多层结构来捕捉长距离依赖关系能够避免梯度消失问题的出现。同时, 引入情感特征可进一步提升原始模型的准确性。考虑情感词的使用背景以及与数据中其他特征的关联, 能够进一步增强模型对真假新闻的检测能力。

4.4.2. 对比实验

为了验证我们提出的模型以及加入 6 种情感特征的效果, 我们进行了对比实验, 在两个数据集的政治领域和社会生活领域进行了测试。尽管我们在这两个数据集的五个领域上都取得了最高精度的结果, 但这一步实验可以进一步证实我们的成果。实验结果分别展示在表 7 和表 8 中。通过观察这两个表格, 我们可以发现无论是传统的 Transformer 模型还是我们提出的注意力卷积 Transformer + CNN 模型, 在加入 6 种情感特征以及情感特征 EmoCred 和 Emoratio 后, 本文的 6 种情感特征相比于单一情感特征, 两个领域的实验结果指标 Macro F1 Score 和 Accuracy 指标均提升了约 2%左右, 这说明多维度的情感特征有助于真假新闻检测。特别是, 在互联网虚假新闻社会生活领域中, 将 6 种情感特征引入我们的注意

力卷积 Transformer + CNN 模型后, Macro F1 Score 达到了 87.28%, Accuracy 达到了 87.38%, 明显高于仅使用 EmoCred 和 Emoratio 情感特征。

同时观察实验结果发现, 在使用情感特征 Emoratio 或 EmoCred 时, 有时所选用的评估指标甚至会降低。研究表明, Emoratio 和 EmoCred 更容易被过度拟合, 因为它们都只关注内容, 而忽略了其他的新闻特征。这使得所考虑的新闻情感特征过于片面, 容易产生对抗特性, 并对最终的新闻检测产生负面影响。相反, 6 种情感特征可以在一定程度上避免这种情况的发生, 从而提高了模型的鲁棒性和准确性。

Table 7. Experimental results of adding different emotional features in the political field (%)

表 7. 政治领域加入不同情感特征的实验结果(%)

Model	互联网虚假新闻政治领域				事实核查新闻政治领域			
	Macro F1 Score	Accuracy	F1 score		Macro F1 Score	Accuracy	F1 score	
			Fake	Real			Fake	Real
Transformer	83.13	84.84	77.74	88.51	83.46	84.69	87.97	78.95
+Emoration	82.58	83.83	77.90	87.26	83.80	85.03	88.26	79.34
+EmoCred	81.38	83.23	75.52	87.25	82.96	84.43	87.96	77.97
+6Sentiment*	84.36	85.45	80.22	88.50	86.48	87.08	89.33	83.64
Transformer*	84.20	85.25	80.11	88.28	84.74	85.65	88.46	81.01
Transformer* + CNN	84.09	85.45	79.43	88.75	84.75	86.23	89.50	80.00
+Emoration	83.65	85.05	78.86	88.44	85.13	85.65	87.90	82.35
+EmoCred	84.25	85.66	79.54	88.96	84.39	85.63	88.78	80.00
SMCT	86.20	87.27	82.35	90.05	86.19	87.43	90.32	82.05

Table 8. Experimental results of adding different emotional features in the field of social life (%)

表 8. 社会生活领域加入不同情感特征的实验结果(%)

Model	互联网虚假新闻社会生活领域				事实核查新闻社会生活领域			
	Macro F1 Score	Accuracy	F1 score		Macro F1 Score	Accuracy	F1 score	
			Fake	Real			Fake	Real
Transformer	84.84	85.12	82.81	86.88	82.45	83.97	77.29	87.62
+Emoration	84.66	84.80	83.21	86.11	81.43	83.13	75.81	87.05
+EmoCred	84.54	84.66	83.19	85.89	81.07	82.81	75.34	86.81
+6Sentiment*	85.52	85.66	84.08	86.95	82.92	84.38	77.94	87.90
Transformer*	85.94	85.99	85.09	86.78	82.75	84.25	77.66	87.84
Transformer* + CNN	86.14	86.26	84.85	87.42	83.19	84.44	78.61	87.78
+Emoration	85.15	85.27	83.81	86.49	84.56	85.57	80.60	88.51
+EmoCred	86.38	86.45	85.39	87.37	83.64	84.69	79.50	87.78
SMCT	87.28	87.38	86.15	88.40	85.23	86.35	81.15	89.31

4.4.3. 消融实验

为了验证所提出的模型以及加入 6 种情感特征的效果, 本文进行了进一步的消融实验研究, 主要集中在两个数据集上的政治领域和社会生活领域。通过实验结果的分析, 我们得出了以下结论: 本文提出的 SMCT 模型在真假新闻检测任务中的表现出显著的提升效果。

消融研究 1:

经过分析表 7 和表 8 实验结果, 我们可以观察到, 在注意力卷积 Transformer 模型的基础上拼接了 CNN 模型之后, Macro F1 值和 Accuracy 指标均相较于之前提高了约 0.4% 左右。这表明采用多种模型结构组合可能会带来一定性能的提升。其中, 卷积层提取新闻文本特征的重要性在于其能够捕捉文本中的局部模式和关系。通过对输入文本的小部分应用过滤器, 卷积神经网络可以识别重要特征, 如单词组合和句子结构, 这些特征可以指示新闻文本的整体含义, 通过融合两种新闻文本语义的输出, 进一步增强了模型的特征提取能力, 更有效地提高了真假新闻模型检测的准确性。

消融研究 2:

通过分析表 7 和表 8, 我们可以看出, 在我们提出的注意力卷积 Transformer + CNN 模型中加入 6 种情感后, 其在 Macro F1 值和 Accuracy 方面的指标均提升了 1% 至 2%, 这说明了加入 6 种情感特征对检测假新闻的重要性。虚假类新闻一般会使用负面情感强烈的词语来突出假新闻表达的信息, 吸引人们的眼球。在社会生活和政治领域中, 假新闻经常会运用带有突出情感词的语言, 而这些情感词可以作为文章情感或语气的有力指标。通过识别这些突出情感词可以提高模型的分类能力。将这些情感词包括在训练数据中, 有助于模型学习将特定情绪词与新闻真假标签相关联的知识。通过这种方式, 模型可以更好地理解文章中的情感和语气, 并实现更准确的预测。研究表明, 将情绪分析纳入假新闻检测模型可以显著提高其性能。

5. 结论与未来工作

在本研究中, 我们引入了一种多头注意力卷积机制的 Transformer 架构, 并与 CNN 相结合, 同时加入了 6 种情感特征, 用于对新闻文本进行真假检测。通过在两个数据集上进行大量实验, 我们发现这 6 种情感特征能够轻松地嵌入到我们提出的多头注意卷积 Transformer + CNN 模型中。加入这些情感特征后, 经过对比其他传统的假新闻检测模型发现, 我们的模型表现出了最佳效果。在未来, 我们计划使用多模态扩展模型, 以提高模型的鲁棒性和泛化能力。

参考文献

- [1] Granik, M., and Mesyura, V. (2017) Fake News Detection Using Naive Bayes Classifier. 2017 *IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, Kyiv, 29 May-2 June 2017, 900-903. <https://doi.org/10.1109/UKRCON.2017.8100379>
- [2] Helmstetter, S. and Paulheim, H. (2018) Weakly Supervised Learning for Fake News Detection on Twitter. 2018 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, 28-31 August 2018, 274-277. <https://doi.org/10.1109/ASONAM.2018.8508520>
- [3] Douglas, A. (2006) News Consumption and the New Electronic Media. *The International Journal of Press/Politics*, **11**, 29-52. <https://doi.org/10.1177/1081180X05284317>
- [4] Lazer, D.M.J., Baum, M.A., Benkler, Y., et al. (2018) The Science of Fake News. *Science*, **359**, 1094-1096. <https://doi.org/10.1126/science.aao2998>
- [5] García, S.A., García, G.G., Prieto, M.S., Guerrero, A.J.M. and Jimenez, C.R. (2020) The Impact of Term Fake News on the Scientific Community. Scientific Performance and Mapping in Web of Science. *Social Sciences*, **9**, Article 73. <https://doi.org/10.3390/socsci9050073>
- [6] Holan, A.D. (2016) 2016 Lie of the Year: Fake News. Politifact, Washington DC.
- [7] Robb, A. (2017) Anatomy of a Fake News Scandal. *Rolling Stone*, **1301**, 28-33.

- [8] Soll, J. (2016) The Long and Brutal History of Fake News. *Politico Magazine*, **18**.
- [9] Hua, J. and Shaw, R. (2020) Corona Virus (COVID-19) “Infodemic” and Emerging Issues through a Data Lens: The Case of China. *International Journal of Environmental Research and Public Health*, **17**, Article 2309. <https://doi.org/10.3390/ijerph17072309>
- [10] Zhang, X., Cao, J., Li, X., et al. (2021) Mining Dual Emotion for Fake News Detection. *Proceedings of the Web Conference 2021*, Ljubljana, 19-23 April 2021, 3465-3476. <https://doi.org/10.1145/3442381.3450004>
- [11] Rashkin, H., Choi, E., Jang, J.Y., Volkova, S. and Choi, Y. (2017) Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, 7-11 September 2017, 2931-2937. <https://doi.org/10.18653/v1/D17-1317>
- [12] Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B. and Mittal, A. (2018) Combining Neural, Statistical and External Features for Fake News Stance Identification. *Companion Proceedings of the Web Conference 2018*, Lyon, 23-27 April 2018, 1353-1357. <https://doi.org/10.1145/3184558.3191577>
- [13] Jin, Z., Cao, J., Zhang, Y., Zhou, J. and Tian, Q. (2017) Novel Visual and Statistical Image Features for Microblogs News Verification. *IEEE Transactions on Multimedia*, **19**, 598-608. <https://doi.org/10.1109/TMM.2016.2617078>
- [14] Zhang, Q., Yilmaz, E. and Liang, S. (2018) Ranking-Based Method for News Stance Detection. *Companion of the Web Conference 2018*, Lyon, 23-27 April 2018, 41-42. <https://doi.org/10.1145/3184558.3186919>
- [15] Castillo, C., Mendoza, M. and Poblete, B. (2011) Information Credibility on Twitter. *Proceedings of the 20th International Conference on World Wide Web*, Hyderabad, 28 March-1 April 2011, 675-684. <https://doi.org/10.1145/1963405.1963500>
- [16] Ajao, O., Bhowmik, D. and Zargari, S. (2019) Sentiment Aware Fake News Detection on Online Social Networks. 2019 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, 12-17 May 2019, 2507-2511.
- [17] Giachanou, A., Rosso, P. and Crestani, F. (2019) Leveraging Emotional Signals for Credibility Detection. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, 21-25 July 2019, 877-880. <https://doi.org/10.1145/3331184.3331285>
- [18] Cho, K., van Merriënboer, B., Bahdanau, D. and Bengio, Y. (2014) On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, Doha, October 2014, 103-111.
- [19] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **8**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [20] Chen, T., Li, X., Yin, H.Z. and Zhang, J. (2018) Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection. In: Ganji, M., Rashidi, L., Fung, B. and Wang, C., Eds., *PAKDD 2018: Trends and Applications in Knowledge Discovery and Data Mining*, Springer, Cham, 40-52. https://doi.org/10.1007/978-3-030-04503-6_4
- [21] Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F. and Cha, M. (2016) Detecting Rumors from Microblogs with Recurrent Neural Networks. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI 2016)*, 3818-3824. https://ink.library.smu.edu.sg/sis_research/4630
- [22] Xu, L.H., Lin, H.F., Pan, Y., Ren, H. and Chen, J.M. (2008) Constructing the Affective Lexicon Ontology. *Journal of the China Society for Scientific*, **2**, 180-185. (In Chinese)
- [23] Dong, Z.D. and Dong, Q. (2003) HowNet: A Hybrid Language and Knowledge Resource. *International Conference on Natural Language Processing and Knowledge Engineering*, Beijing, 26-29 October 2003, 820-824.
- [24] Wikipedia (2020) List of Emoticons. https://en.wikipedia.org/wiki/List_of_emoticons