CN 43-1258/TP
ISSN 1007-130X

计算机工程与科学
Computer Engineering & Science

第 46 卷第 4 期 2024 年 4 月
Vol. 46, No. 4, Apr. 2024

文章编号:1007-130X(2024)04-0752-09

基于用户权威度和多特征融合的微博谣言检测模型*

许莉芬[1],曹霑懋[1],郑明杰[1],肖博健[2]
(1.华南师范大学计算机学院,广东 广州 510631;2.华南师范大学人工智能学院,广东 佛山 528200)

摘　要:网络谣言的广泛传播及其对社会的负面影响急切需要高效的谣言检测模型。由于数据集的文本缺乏语义信息和严格的句法结构,结合用户特征和语境特征来丰富语义信息显得很有意义。对此,提出一种基于用户权威度和多特征融合的微博谣言检测模型 MRUAMF。首先,抽取出用户信息完整度、用户活跃度、用户交际广度和用户平台认证指数 4 项指标构建用户权威度定量计算模型,通过级联用户权威度及其构成指标,并使用 2 层全连接网络融合特征,有效量化用户特征。其次,考虑到语境对谣言理解的有效性,提取相关语境特征。最后,使用 BERT 预训练模型提取文本特征,并结合多模态适应门 MAG 融合用户特征、语境特征与文本特征。在微博数据集上进行的实验表明,相比基线模型,MRUAMF 模型的检测性能更优,准确率达 0.941。

关键词:谣言检测;BERT;MAG;用户权威度;层次分析法
中图分类号:TP391　　**文献标志码**:A
doi:10.3969/j.issn.1007-130X.2024.04.020

A microblog rumor detection model based on user authority and multi-feature fusion

XU Li-fen[1], CAO Zhan-mao[1], ZHENG Ming-jie[1], XIAO Bo-jian[2]
(1. School of Computer Science, South China Normal University, Guangzhou 510631;
2. School of Artificial Intelligence, South China Normal University, Foshan 528200, China)

Abstract: The widespread dissemination of online rumors and their negative impact on society urgently require efficient rumor detection. Due to the lack of semantic information and strict syntactic structure in the text of the dataset, it is meaningful to combine user characteristics and contextual features to enrich semantic information. In this regard, MRUAMF is proposed. Firstly, four indicators including user information completeness, user activity, user communication span, and user platform authentication index are extracted to construct a quantitative calculation model for user authority. By cascading user authority and its constituent indicators, and using a two-layer fully connected network to fuse features, user characteristics are effectively quantified. Secondly, considering the effectiveness of context in understanding rumors, relevant contextual features are extracted. Finally, the BERT pre-training model is used to extract text features, which are then combined with the Multimodal Adaptation Gate (MAG) to fuse user features, contextual features, and text features. Experiments on the microblog dataset show that compared with the baseline model, the MRUAMF model has better detection performance with an accuracy rate of 0.941.

Key words: rumor detection; bidirectional encoder representations from transformers(BERT); multimodal adaption gate(MAG); user authority; analytic hierarchy process

* **收稿日期**:2023-09-04;**修回日期**:2023-10-22
通信作者:曹霑懋(caozhanmao@scnu.edu.cn)
通信地址:510631 广东省广州市华南师范大学计算机学院
Address: School of Computer Science, South China Normal University, Guangzhou 510631, Guangdong, P. R. China

1 引言

社交媒体是新闻传播、政治活动、科学发现或产品广告的流行媒介。由于社交媒体可以快速广泛地传播信息,总有人为了特定的利益,在网络上发布谣言信息,严重影响了网络的良性发展,甚至影响社会稳定。有效地检测谣言有利于净化网络空间和维护社会稳定,具有十分重要的现实意义。

为了进行谣言检测,一些研究对社交网络中的用户特征进行建模,从用户群体中提取特征,如用户信息[1]和用户历史行为[2,3]等,并结合机器学习和深度学习模型实现谣言检测。

Rieh[4]发现,信息的可信度由其来源的权威度所支配。微博用户作为信息发布和传播的主体,其权威度是传播影响因素中的一个重要评价指标。研究用户权威度,对于预测帖子的真实性具有积极的意义。

研究表明[5],信息的语言风格和语气有助于区分虚假陈述和真实陈述。说谎者在讲述虚假的故事时倾向于更频繁地使用负面情绪词,作为内疚的标志,同时也倾向于在文本中频繁使用感叹号和问号,表达发布者激动的心情。Li 等人[6]验证了将发布的文本中的语境特征(如情绪和标点符号的使用特征等)视为传播谣言有关的线索是合乎逻辑的。

在谣言检测领域,研究人员在 3 个方面打下了坚实的理论基础:用户、语境和文本特征表示。但是,传统的用户特征强调用户属性和用户关联的关系,在谣言检测任务中还没有对社交网络中的用户权威度进行精确的定量表示,这使得量化用户权威度并将其与谣言识别相结合很难。

针对上述问题,本文提出基于用户权威度和多特征融合的谣言检测模型 MRUAMF(Microblog Rumor detection model based on User Authority and Multi-feature Fusion),主要工作如下:

(1)提出了一种用户权威度的计算方法并将其用于谣言检测。

(2)提出了一种基于用户特征、语境特征和文本特征融合的特征表示策略。使用 BERT(Bidirectional Encoder Representations from Transformers)模型获得文本特征,并结合多模态适应门 MAG(Multimodal Adaptation Gate)[7]将用户特征、语境特征与文本特征进行融合。

(3)本文在微博谣言数据集上进行了实验,以检验 MRUAMF 模型的有效性。

2 相关工作

现有的谣言检测方法大致可以分为 2 类:手工特征提取的方法和基于深度学习的方法。

传统的基于手工特征提取的方法主要从 3 个方面来设计谣言检测模型:(1)关注统计文本内容和用户信息的特征构造检测模型。如 Liang 等人[8]利用用户行为的手工特征进行建模检测,Castillo 等人[9]提取用户信息和发表的微博内容特征来构造谣言检测模型。(2)关注传播路径和传播节点等特征的传播结构检测模型。如 Kwon 等人[10]利用谣言传播结构的特征设计谣言检测模型。(3)关注文本信息随时间变化的统计特征的时间序列检测模型。如 Ma 等人[11]认为谣言文本和非谣言文本在时间序列上变化的模式不同,并抓取多种社会上下文特征随时间流逝的变化,以此设计谣言检测模型。然而,基于手工特征提取的方法依赖特征的选取,且缺乏一种标准和系统的方法来设计跨平台的通用特征和处理不同类型的谣言。

深度学习的迅速发展催生了许多基于深度学习的谣言检测模型。由于增强了自动表示学习的能力,基于深度学习的谣言检测模型的检测性能要优于传统手工特征提取的谣言检测模型。大多数现有的基于深度学习的方法主要侧重于从文本内容、用户评论和图像中提取文本特征和视觉特征。Ma 等人[12]首先提出用循环神经网络 RNN(Recursive Neural Network)进行谣言识别,他们基于谣言传播过程中的转发时间序列数据,使用门控循环单元 GRU(Gate Recurrent Unit)学习时间和文本特征。Shu 等人[13]提出了一种共同关注网络,利用新闻内容和用户评论进行谣言检测。Jin 等人[14]提出了一种用于提取视觉、文本和社会背景特征的模型。此外,一些研究还采用了其它深度学习技术,如多任务学习[15]和对抗学习[16],来学习更丰富的内容感知特征。然而,有些谣言是故意通过模仿真实的新闻来撰写的。由于缺乏必要的领域知识,仅从内容特征方面构建谣言检测模型很难进一步提高检测性能。

一些研究人员意识到用户在谣言传播中发挥着重要作用。例如,Zhang 等人[17]提出了一种名为 Fake Detector 的自动假新闻可信度推理模型。作者分析了多种属性,如用户个人资料特征、用户与假新闻创建者之间的联系,并使用深度扩散神经

模型来学习文章内容、创作者和主题的特征。Dong等人[18]提出了一种用于谣言检测的具有用户和情感信息的分层注意网络。Chen等人[19]提出了用于谣言检测的具有注意力的用户方面多视图学习模型,有效地学习传播帖子的用户的个人资料视图、结构视图和时间视图表示,将学习到的用户特征与内容特征连接起来,用于检测任务。

尽管上述研究中提到的融合用户特征的方法在社交网络环境中检测谣言是有效的,但是也有一定的局限性。这些方法在用户级别上强调用户资料和用户关联的关系,但没有更进一步地挖掘用户资料。本文借助用户特征、语境特征与文本特征,并从用户权威度角度构建用户特征,通过深度学习来分析信息的真实性。

3 相关定义与模型准备

本节介绍用户权威度和语境的相关定义和计算方法,所涉及的重要符号如表1所示。

Table 1 Description of symbols

表1 符号描述

符号	描述
$I(u)$	用户 u 的信息完整度
$P(u)$	用户 u 的平台认证指数
$A(u)$	用户 u 的活跃度
$C(u)$	用户 u 的交际广度
$Au(u)$	用户 u 的权威度
$Mfans(u)$	用户 u 的互粉数
$Pfans(u)$	用户 u 的纯粉丝数
$Att(u)$	用户 u 的关注数

3.1 用户权威度

用户权威度[20]指用户在社交网络关系中具有的影响力与公众对其信服的程度。为了量化用户权威度,本文通过用户的交际广度(Commuication Span)来衡量用户的影响力,并从平台认证指数(Platform Authentication Index)、用户信息完整度(Information Integrity)和用户活跃度(Activity Degree)方面度量公众对其信服的程度。基于这些指标构建用户权威度。首先,考虑了用户信息完整度,完整的用户信息能够让公众更好地了解用户的背景和专业领域,从而更愿意信任用户的观点和意见。其次,考虑了用户的平台认证指数,认证身份能够提高用户在社交网络中的信誉度和影响力。第三,考虑了用户在微博上的活跃度,高度活跃的用户能够吸引更多的关注和互动,从而提高其影响力和信服程度。最后,鉴于广泛的社交网络关系能够提高用户的影响力和信服程度,本文考虑了用户的交际广度。

3.1.1 用户信息完整度

用户的信息能够反映用户的真实性,具有完整的、真实信息的用户会具有更高的权威度。用户信息包含昵称、性别、简介和地址信息。本文构建向量$\boldsymbol{V}=(v_1,v_2,\cdots,v_n)$用以表示用户基本信息的填写情况。其中,$v_i$表示序号为$i$的标签是否包含信息。当$v_i=1$时,表示第$i$号的标签存在有效信息;当$v_i=0$时,表示第$i$号的标签不存在有效信息。定义用户$u$的信息完整度$I(u)$为用户愿意公开的信息占所有信息标签的比例,如式(1)所示:

$$I(u)=\frac{1}{n}\sum_{i=1}^{n}v_i \tag{1}$$

其中,n表示向量$\boldsymbol{V}$的维度。

3.1.2 用户平台认证指数

定义用户u的平台认证指数$P(u)$为平台对用户给出的认证评价。平台认证是微博官方认证平台对用户进行审查认证。$P(u)$的计算方法如式(2)所示:

$$P(u)=\begin{cases}1,\text{用户 } u \text{ 为平台认证的用户}\\0,\text{否则}\end{cases} \tag{2}$$

3.1.3 用户活跃度

定义用户u的活跃度$A(u)$为用户在一定时间内发布帖子的频率。用户活跃度计算方法如式(3)所示:

$$A(u)=\frac{2}{\pi}\times\arctan\frac{Num(u)}{t} \tag{3}$$

其中,$Num(u)$表示用户在时间段t内发布的帖子数量,这里的t指从用户注册起到获取数据的这段时间的天数。

3.1.4 用户交际广度

在微博社交网络中,用户通过关注成为对方的粉丝。粉丝表明他人对用户的关注,是对用户微博评论转发的潜在用户。用户拥有粉丝越多,与粉丝的交互程度越高,用户的影响力越大,权威度越高。用户u的交际广度$C(u)$的定义如式(4)所示:

$$C(u)=\frac{\omega_1\times Pfans(u)+\omega_2\times Mfans(u)}{Fans(u)+Att(u)} \tag{4}$$

其中,$Pfans(u)$表示用户纯粉丝数;$Mfans(u)$表示用户互粉数;$Fans(u)$表示用户粉丝数,$Fans(u)=Mfans(u)+Pfans(u)$;$Att(u)$表示用

户关注数;ω_1 表示用户纯粉丝数的权重系数,ω_2 表示用户互粉数的权重系数,初始值 $\omega_1=0.7$,$\omega_2=0.3$。

3.1.5 用户权威度模型的构建

基于上述 4 个指标,本文使用层次分析法确定权重系数,并利用特征向量法计算用户权威度评价特征的权值,求解步骤如下:

步骤 1 设用户权威度评价的判断矩阵为 $\boldsymbol{B}$,其中的元素 $b_{i,j}$ 表示特征 i 相比特征 j 对评价结果影响的重要程度的倍数。本文认为,最能体现用户权威度的首要因素是用户交际广度,其次是用户平台认证指数,第 3 是用户活跃度,第 4 是用户信息完整度。之所以将用户信息完整度置于第 4,是因为用户信息中可能含有虚假的成分。按照四元组 $(I(u),A(u),P(u),C(u))$ 为判断矩阵 $\boldsymbol{B}$ 赋值,如式(5)所示:

$$\boldsymbol{B}=\begin{bmatrix}1 & 1/3 & 1/7 & 1/9\\ 3 & 1 & 1/5 & 1/7\\ 7 & 5 & 1 & 1/2\\ 9 & 7 & 2 & 1\end{bmatrix} \tag{5}$$

步骤 2 求解判断矩阵 $\boldsymbol{B}$ 的最大特征值的特征向量并进行归一化处理,得到各个指标的权重 $\boldsymbol{\mu}$,如式(6)所示:

$$\boldsymbol{\mu}=(0.0439,0.0885,0.3286,0.5390) \tag{6}$$

本文构建的用户权威度定量计算模型的量化计算方法如式(7)所示。

$$Au(u)=(I(u),A(u),P(u),C(u))\times\boldsymbol{\mu}^{\mathrm{T}} \tag{7}$$

3.2 语境

语境对信息的传播有强大的影响力。情绪是语境的重要组成部分。谣言往往具有消极的情绪,具体来说,谣言的语气总是很强烈,感叹号的使用比较频繁[6],尤其是当情感是怀疑或惊讶时,问号和感叹号的出现次数较多,因此可以将情绪、问号与感叹号的数量作为区分谣言的特征。

3.2.1 情绪

本文提取用户发布的帖子和对应转发的帖子的文本进行情绪分析。调用中文情感分析库 cnsenti,将情绪划分为好、乐、哀、怒、惧、恶、惊 7 种情绪,并统计 7 种情绪词的频数。本文选择其中频数最大的情绪词作为帖子的情绪代表。

3.2.2 问号和感叹号数量

本文统计源帖及其转发帖子中问号和感叹号的数量,并将其作为区分谣言的特征之一。

4 MRUAMF 模型

4.1 模型结构

MRUAMF 模型主要包括 5 个部分:用户特征提取、语境特征提取、文本特征提取、多特征融合和预测分类。在用户特征挖掘上,将用户权威度及其 4 项构成指标作为用户特征。在语境特征挖掘上,进行情绪统计分析,并统计符号数量。图 1 显示了 MRUAMF 模型的总体结构。

4.2 输入

按照第 3 节的描述收集用户信息,包括:(1)昵称;(2)简介;(3)性别;(4)粉丝数;(5)互粉数;(6)关注数;(7)是否经过微博认证;(8)地址信息;(9)历史发布的帖子数量。每条源帖下包含了其他用户转发的帖子,本文对源贴及其转发帖子的文本内容进行统计分析,获取文本情绪标记、问号和感叹号的数量。

4.3 特征提取与融合

4.3.1 用户特征提取

本文使用 2 层全连接网络提取与用户权威度相关的指标作为用户特征,包括:(1)平台认证指数;(2)活跃度;(3)交际广度;(4)信息完整度;(5)权威度。

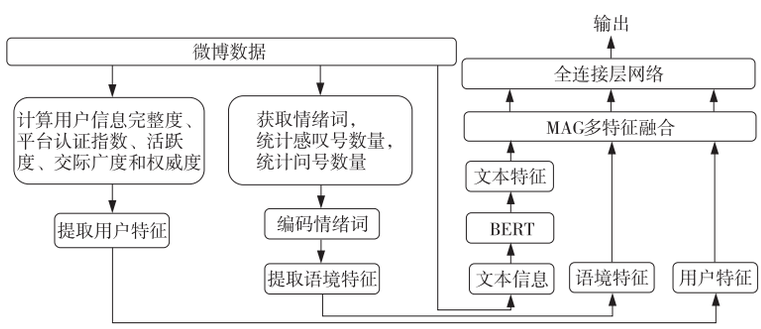

Figure 1 Structure of MRUAMF model

图 1 MRUAMF 模型结构

将用户特征用向量 $\boldsymbol{V}_u=(v_p,v_a,v_c,v_i,v_r)$ 表示,向量中的每个值分别对应上述 5 种特征,即平台认证指数、活跃度、交际广度、信息完整度和权威度的数值。本文使用 2 层全连接网络对用户特征进行学习和融合,2 层全连接网络的计算过程和输出如式(8)和式(9)所示:

$$\boldsymbol{U}'=ReLU(\boldsymbol{W}_{u_1}\boldsymbol{V}_u+\boldsymbol{b}_{u_1}) \tag{8}$$

$$\boldsymbol{U}=ReLU(\boldsymbol{W}_{u_2}\boldsymbol{U}'+\boldsymbol{b}_{u_2}) \tag{9}$$

其中,$\boldsymbol{W}_{u_1}$ 和 $\boldsymbol{W}_{u_2}$ 分别表示第 1 层和第 2 层全连接网络的权重矩阵,$\boldsymbol{b}_{u_1}$ 和 $\boldsymbol{b}_{u_2}$ 表示相应的偏移项。$\boldsymbol{U}'$ 表示相应的第 1 层全连接网络输出的中间向量,$\boldsymbol{U}$ 表示相应的第 2 层全连接网络输出的用户融合特征向量,$\boldsymbol{U}\in\mathbf{R}^{768}$。本文使用修正线性单元 ReLU(Rectified Linear Unit)函数作为激活函数。

4.3.2 语境特征提取

根据第 3 节的描述,本文选取以下语境特征:(1)情绪;(2)问号的数量;(3)感叹号的数量。

对情绪标签进行预处理,对情绪的分类标签使用数值代替,将 7 种情绪映射到 0～6 的整数。用向量 $\boldsymbol{V}_c=(v_s,v_q,v_e)$ 表示语境特征,向量中的每个值分别对应上述语境特征的数值。本文使用 2 层全连接网络对语境特征进行学习和融合,2 层全连接网络的计算方法和输出公式如式(10)和式(11)所示:

$$\boldsymbol{C}'=ReLU\ (\boldsymbol{W}_{c_1}\boldsymbol{V}_c+\boldsymbol{b}_{c_1}) \tag{10}$$

$$\boldsymbol{C}=ReLU\ (\boldsymbol{W}_{c_2}\boldsymbol{C}'+\boldsymbol{b}_{c_2}) \tag{11}$$

其中,$\boldsymbol{W}_{c_1}$ 和 $\boldsymbol{W}_{c_2}$ 分别表示第 1 层和第 2 层全连接网络的权重矩阵,$\boldsymbol{b}_{c_1}$ 和 $\boldsymbol{b}_{c_2}$ 分别表示相应的偏移项。$\boldsymbol{C}'$ 是第 1 层全连接网络输出的中间向量,$\boldsymbol{C}$ 是第 2 层全连接网络输出的语境融合特征向量,$\boldsymbol{C}\in\mathbf{R}^{768}$。

4.3.3 文本特征提取

本文使用 BERT 提取文本特征。BERT 模型能够联系上下文语义进行学习,结合自注意力机制考虑每个词语对其他词语的重要程度,预训练出来的向量表示效果比 word2vec 的更好。BERT 模型主要包括 2 个阶段:编码阶段和生成向量表示阶段。使用 BERT 模型在数据集上进行微调,获得文本特征 $\boldsymbol{Z}\in\mathbf{R}^{768}$。

4.3.4 特征融合

受多模态适应门(MAG)的启发,本文将 3 种特征即文本特征、用户特征和语境特征通过 MAG 进行特征融合,如图 2 所示。MAG 单元接收 3 个特征作为输入。令三元组 $(\boldsymbol{Z},\boldsymbol{U},\boldsymbol{C})$ 表示文本特征、用户特征和语境特征输入。$\boldsymbol{Z}$ 表示文本特征,$\boldsymbol{U}$ 表示用户特征,$\boldsymbol{C}$ 表示语境特征,其维度均为 768。将文本特征分别与用户特征和语境特征拼接得到 $[\boldsymbol{Z};\boldsymbol{U}]$ 和 $[\boldsymbol{Z};\boldsymbol{C}]$,并利用它们生成 2 个注意力门控向量 $\boldsymbol{g}^u$ 和 $\boldsymbol{g}^c$:

$$\boldsymbol{g}^u=ReLU\ (\boldsymbol{W}_{gu}[\boldsymbol{Z};\boldsymbol{U}]+\boldsymbol{b}_u) \tag{12}$$

$$\boldsymbol{g}^c=ReLU\ (\boldsymbol{W}_{gc}[\boldsymbol{Z};\boldsymbol{C}]+\boldsymbol{b}_c) \tag{13}$$

其中,$\boldsymbol{W}_{gu}$ 和 $\boldsymbol{W}_{gc}$ 分别表示用户和语境的权重矩阵,$\boldsymbol{b}_u$ 和 $\boldsymbol{b}_c$ 表示相应的偏移向量。

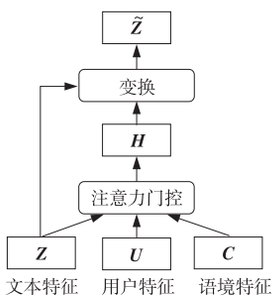

Figure 2 Structure of MAG combining user feature and context feature

图 2 结合用户特征与语境特征的 MAG 结构

然后,分别将 $\boldsymbol{U}$ 和 $\boldsymbol{C}$ 与各自的注意力门控向量相乘获得向量 $\boldsymbol{H}$,如式(14)所示:

$$\boldsymbol{H}=\boldsymbol{g}^u\cdot(\boldsymbol{W}_u\boldsymbol{U})+\boldsymbol{g}^c\cdot(\boldsymbol{W}_s\boldsymbol{C})+\boldsymbol{b}_H \tag{14}$$

其中,$\boldsymbol{W}_u$ 和 $\boldsymbol{W}_s$ 分别表示用户信息和语境信息的权重矩阵,$\boldsymbol{b}_H$ 表示偏移向量。

将式(14)获得的向量 $\boldsymbol{H}$ 与 $\boldsymbol{Z}$ 加权求和得到融合特征 $\tilde{\boldsymbol{Z}}$:

$$\tilde{\boldsymbol{Z}}=\boldsymbol{Z}+\alpha\boldsymbol{H} \tag{15}$$

$$\alpha=\min\left(\frac{\|\boldsymbol{Z}\|_2}{\|\boldsymbol{H}\|_2}\beta,1\right) \tag{16}$$

其中,β 表示通过交叉验证过程选择的超参数,$\|\boldsymbol{Z}\|_2$ 和 $\|\boldsymbol{H}\|_2$ 分别表示 $\boldsymbol{Z}$ 和 $\boldsymbol{H}$ 的 L_2 范数。

4.4 预测分类

本文使用 Softmax 分类器进行谣言检测,预测结果如式(17)所示:

$$Prediction=Softmax\ (\boldsymbol{W}_z\tilde{\boldsymbol{Z}}+\boldsymbol{b}_z) \tag{17}$$

其中,$\boldsymbol{W}_z$ 表示权值矩阵,$\boldsymbol{b}_z$ 表示偏置向量,$Softmax(\cdot)$ 激活函数用于判断目标是否为谣言。

5 实验与结果分析

为了验证 MRUAMF 模型的有效性,本节在同一个微博数据集上将其与其它基线模型进行比较,并设计一系列实验验证 MRUAMF 模型的合理性。

5.1 实验设置

5.1.1 实验数据集

数据集来源于 Ma 等人[21]构建的微博谣言数据集,该数据集包含 4 664 个源帖,每条源帖下有若干条转发的帖子,同时还包含了用户的信息。

本文分别选取 60%的微博数据集组成训练集,30%组成测试集,10%组成验证集。数据集的统计信息如表 2 所示。

Table 2 Statistic of dataset

表 2 数据集统计信息

统计项	条数
源贴数量	4 664
谣言帖子数量	2 313
非谣言帖子数量	2 351
最大转发数量	59 318
最小转发数量	10
平均转发数量	816

5.1.2 实验参数设置

本文使用 PyTorch 深度学习框架编码模型,使用 Adam 优化器进行训练,学习率为 5e−5;使用交叉熵损失函数;设置 BERT 网络的随机失活率为 0.1,批大小为 16。

5.1.3 评价指标

本文采用准确率(*ACC*)、精确度(*P*)、召回率(*R*)和 F1 分数(*F*1)作为评估指标来衡量模型的性能。

5.1.4 基线模型

在微博数据集上,本文提出的 MRUAMF 模型与下列基线模型进行对比实验:

(1)DTC(Decision Tree Classifier)[22]:基于用户行为特征,使用决策树分类器 DTC 进行谣言识别的模型。

(2)SVM-RBF(Support Vector Machine using Radial Basis Function)[23]:使用径向基函数 RBF 作为核函数的支持向量机 SVM 模型。

(3)GRU[14]:基于 RNN 的谣言识别模型,用于捕获谣言识别输入的上下文信息。

(4)TD-RvNN(Top-Down tree-structured Recur- sive Neural Network)[24]:一种基于 RNN 的树结构模型,该模型在树结构中嵌入隐藏的指示信号,并探索帖子内容对谣言检测的重要性。

(5)PLAN(Post-Level Attention Network)[25]:一种用于谣言检测的分层令牌和后级注意力模型。

(6)GCAN(Graph-aware Co-Attention Network)[26]:一种图感知共同关注网络,利用源帖的内容及其基于传播的用户来检测信息的真实性。

(7)UMLARD(User-aspect Multi-view Learning with Attention for Rumor Detection)[19]:一种用于谣言检测的用户端注意力多视角学习模型,该模型学习传播帖子用户的不同视图表示,并将学到的用户方面特征与内容特征连接起来。

所有模型在测试集上的实验结果如表 3 所示。

Table 3 Experimental results of different models (R:rumor,N:non rumor)

表 3 不同模型的实验结果(R:谣言,N:非谣言)

模型	类别	*ACC*	*P*	*R*	*F*1
DTC	R	0.731	0.747	0.715	0.731
	N		0.715	0.747	0.742
SVM-RBF	R	0.741	0.745	0.735	0.740
	N		0.738	0.747	0.742
GRU	R	0.762	0.728	0.809	0.767
	N		0.803	0.715	0.757
TD-RvNN	R	0.832	0.821	0.861	0.841
	N		0.832	0.812	0.821
PLAN	R	0.857	0.893	0.805	0.835
	N		0.829	0.904	0.857
GCAN	R	0.880	0.866	0.929	0.896
	N		0.911	0.861	0.885
UMLARD	R	0.928	0.894	0.944	0.928
	N		0.942	0.965	0.924
MRUAMF	R	0.941	0.931	0.948	0.939
	N		0.949	0.933	0.941

5.2 实验分析

5.2.1 与基线模型的比较与分析

表 3 的实验结果表明,传统的基于手工特征提取的模型如 DTR、SVM-RBF 的效果均不佳。这些模型基于帖子的统计数据使用手工提取的特征,不足以捕捉文本的可概括性特征,无法形成文本特征间的高级交互。

基于深度学习的模型优于基于手工特征提取的模型。GRU 和 TD-RvNN 均是基于 RNN 的模型,PLAN 使用自注意力机制用于模拟帖子之间的交互,其效果比 GRU 和 TD-RvNN 的要优。但是,上述 3 种模型主要关注文本信息而忽略了其他类型的特征。

GCAN 从用户相似度建模,提取传播帖子的

用户特征和源帖文本特征。UMLARD通过对用户的信息进行建模丰富了谣言检测的工作主体。GCAN和UMLARD的检测效果比PLAN的性能更佳,这个结果也验证了用户特征在传播谣言的过程中起着重要的作用,用户是错误信息的主要传播者。

MRUAMF模型在微博数据集上优于所有对比基线模型。与最佳基线模型UMLARD相比,MRUAMF模型通过用户权威度构造用户特征,并融合了语境特征,从而提高了检测效果。

5.2.2 不同输入特征对预测结果的影响

本文在特征提取部分介绍了文本、用户和语境融合特征。本节介绍基于不同特征组合的消融实验。设置了4种特征组合,第1组为文本特征,第2组为文本特征+用户特征,第3组为文本特征+语境特征,第4组为文本特征+用户特征+语境特征。消融实验结果如图3所示。

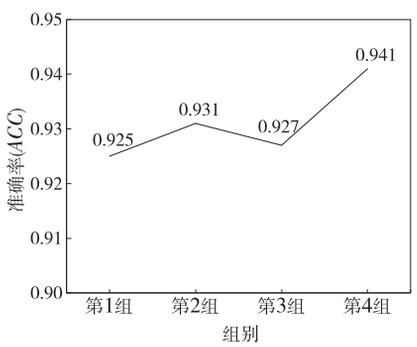

Figure 3 Results of ablation experiment

图3 消融实验结果

图3表明,相较于第1组,其余3组在该数据集上表现更好,这说明增加用户特征或语境特征作为辅助特征可以丰富谣言语义信息,从而提高检测效果。在*ACC*指标上,第4组的测试结果相比第2组和第3组均有所提高,这证实了本文使用文本、用户和语境3种类型的特征对提高谣言检测能力有积极的影响。通过从用户信息和语境信息这2个不同的角度提取特征,引入多特征融合,使模型能够更好地实现分类预测。

5.2.3 不同特征融合方法对预测结果的影响

本文使用MAG处理特征融合,本节将其与其它常用的特征融合方法进行对比,以探索它们对性能的影响,结果如表4所示。

表4中的融合方法包括:(1)加法:$\boldsymbol{Z}+\boldsymbol{U}+\boldsymbol{C}$;(2)串联:$[\boldsymbol{Z};\boldsymbol{U};\boldsymbol{C}]$;(3)MLP:$\boldsymbol{Z}+\tanh(\boldsymbol{W}_1\boldsymbol{U}+\boldsymbol{b}_1)+\tanh(\boldsymbol{W}_2\boldsymbol{C}+\boldsymbol{b}_2)$,其中,$\boldsymbol{W}_1$和$\boldsymbol{W}_2$表示权重矩阵,$\boldsymbol{b}_1$和$\boldsymbol{b}_2$表示偏移项,$\tanh(\cdot)$是激活函数。

Table 4 Results of different feature fusion methods

表4 不同特征融合方法结果

特征融合方法	*ACC*
加法	0.927
串联	0.929
MLP	0.934
MAG	0.941

从表4可以看出,使用MAG的融合方法在*ACC*指标上均优于另外3种方法,这表明了使用MAG进行多特征融合的有效性。

5.2.4 用户权威度分析

令标签UR和UN分别表示发布谣言的用户和发布非谣言的用户。图4和图5分别是间隔为0.1的权威度区间上UR和UN的数量分布图。

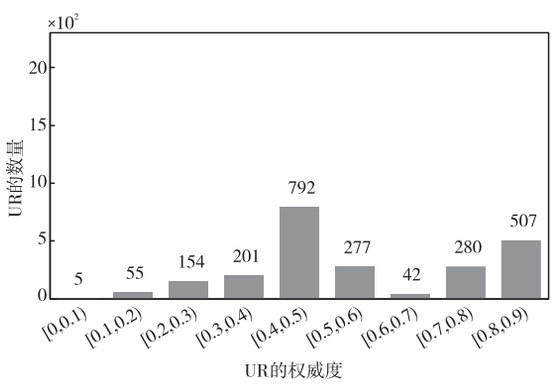

Figure 4 Distribution of UR based on authority interval

图4 基于权威度区间的UR数量分布

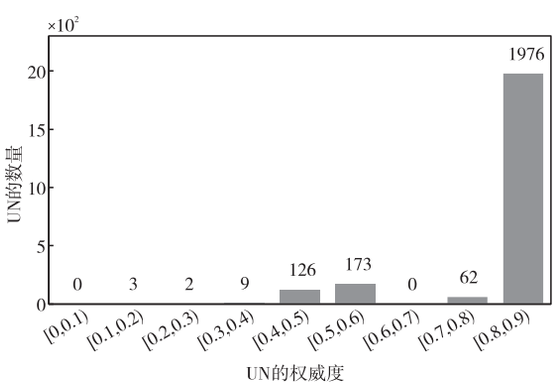

Figure 5 Distribution of UN based on authority interval

图5 基于权威度区间的UN数量分布

图4表明发布谣言的用户,权威度主要集中在[0.4,0.5)和[0.8,0.9)。图5表明发布非谣言的用户,权威度主要集中在[0.8,0.9)。由此可见,总体上发布谣言的用户权威度较发布非谣言的用户权威度低,这也证实了权威度是谣言检测的一个重要的评价指标。通常来说,权威度较低的用户更容易散布谣言。

6 结束语

本文提出的 MRUAMF 模型实现了对微博帖子的谣言识别。首先,考虑到用户权威度对谣言检测的积极作用,本文通过级联用户权威度及其相关指标,使用 2 层全局连接层网络提取特征,有效量化和压缩用户特征。其次,对帖子中的语境信息进行挖掘,这有助于提高谣言的识别能力。使用 BERT 预训练文本获得文本特征表示,并结合多模态适应门将用户特征、语境特征与文本特征融合。实验结果表明,本文提出的模型能有效地实现谣言检测。未来,将探索更有效的谣言检测模型,如利用社交网络的拓扑结构来提升谣言分类器的性能。并将在更具有差异性的数据集上进行深入探索。

参考文献:

[1] Cheng Y Y, Huo L A, Zhao L J. Dynamical behaviors and control measures of rumor-spreading model in consideration of the infected media and time delay[J]. Information Sciences, 2021, 564: 237-253.

[2] Xiao Y P, Li W, Qiang S, et al. A rumor & anti-rumor propagation model based on data enhancement and evolutionary game[J]. IEEE Transactions on Emerging Topics in Computing, 2022, 10(2): 690-703.

[3] Hosni A I E, Li K, Ahmad S. Minimizing rumor influence in multiplex online social networks based on human individual and social behaviors[J]. Information Sciences, 2020, 512: 1458-1480.

[4] Rieh S Y. Judgment of information quality and cognitive authority in the web[J]. Journal of the American Society for Information Science and Technology, 2002, 53(2): 145-161.

[5] Newman M L, Pennebaker J W, Berry D S, et al. Lying words: Predicting deception from linguistic styles[J]. Personality and Social Psychology Bulletin, 2003, 29(5): 665-675.

[6] Li Z M, Zhao Y, Duan T, et al. Configurational patterns for COVID-19 related social media rumor refutation effectiveness enhancement based on machine learning and fsQCA[J]. Information Processing & Management, 2023, 60(3): 103303.

[7] Rahman W, Hasan M K, Lee S, et al. Integrating multimodal information in large pretrained transformers[C]//Proc of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 2359-2369.

[8] Liang G, He W B, Xu C, et al. Rumor identification in microblogging systems based on users' behavior[J]. IEEE Transactions on Computational Social Systems, 2015, 2(3): 99-108.

[9] Castillo C, Mendoza M, Poblete B. Information credibility on Twitter[C]//Proc of the 20th International Conference on World Wide Web, 2011: 675-684.

[10] Kwon S, Cha M, Jung K, et al. Prominent features of rumor propagation in online social media[C]//Proc of 2013 IEEE 13th International Conference on Data Mining, 2013: 1103-1108.

[11] Ma J, Gao W, Wei Z Y, et al. Detect rumors using time series of social context information on microblogging websites[C]//Proc of the 24th ACM International Conference on Information and Knowledge Management, 2015: 1751-1754.

[12] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks[C]//Proc of the 25th International Joint Conference on Artificial Intelligence, 2016: 3818-3824.

[13] Shu K, Cui L M, Wang S H, et al. DEFEND: Explainable fake news detection[C]//Proc of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 395-405.

[14] Jin Z W, Cao J, Guo H, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs[C]//Proc of the 25th ACM International Conference on Multimedia, 2017: 795-816.

[15] Ma J, Gao W, Wong K F. Detect rumor and stance jointly by neural multi-task learning[C]//Proc of the 24th International Conference on Artificial Intelligence, 2018: 585-593.

[16] Ma J, Gao W, Wong K F. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning[C]//Proc of the World Wide Web Conference, 2019: 3049-3055.

[17] Zhang J, Cui L, Fu Y, et al. Fake news detection with deep diffusive network model[J]. arXiv: 1805.08751, 2018.

[18] Dong S J, Qian Z, Li P F, et al. Rumor detection on hierarchical attention network with user and sentiment information[C]//Proc of the 9th CCF International Conference on Natural Language Processing and Chinese Computing, 2020: 366-377.

[19] Chen X, Zhou F, Trajcevski G, et al. Multi-view learning with distinguishable feature fusion for rumor detection[J]. Knowledge-Based Systems, 2022, 240: 108085.

[20] 张仰森,郑佳,唐安杰. 基于多特征融合的微博用户权威度定量评价方法[J]. 电子学报, 2017, 45(11): 2800-2809.
Zhang Yang-sen, Zheng Jia, Tang An-jie. A quantitative evaluation method of microblog user authority based on multi-feature fusion [J]. Acta Electronica Sinica, 2017, 45(11): 2800-2809.

[21] Ma J, Gao W, Wong K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]//Proc of the 55th Annual Meeting of the Association for Computational Linguistics, 2017: 708-717.

[22] Zhao Z, Resnick P, Mei Q Z. Early detection of rumors in social media from enquiry posts[C]//Proc of the 24th International Conference on World Wide Web, 2015: 1395-1405.

[23] Yang F, Yu X H, Liu Y, et al. Automatic detection of rumor on Sina weibo[C]//Proc of the ACM SIGKDD Workshop on

Mining Data Semantics,2012:1-7.

[24] Ma J,Gao W,Wong K. Rumor detection on Twitter with tree-structured recursive neural networks[C]//Proc of the 56th Annual Meeting of the Association for Computational Linguistics,2018:1980-1989.

[25] Khoo L M S,Chieu H L,Qian Z, et al. Interpretable rumor detection in microblogs by attending to user interactions[C]//Proc of the AAAI Conference on Artificial Intelligence,2020:8783-8790.

[26] Lu Y J,Li C T. GCAN:Graph-aware co-attention networks for explainable fake news detection on social media[C]//Proc of the 58th Annual Meeting of the Association for Computational Linguistics,2020:505-514.

作者简介:

许莉芬(1999 -),女,广东汕头人,硕士生,研究方向为深度学习和自然语言处理。**E-mail**:xulfen@163.com

XU Li-fen, born in 1999, MS candidate, her research interests include deep learning and natural language processing.

曹霑懋(1967 -),男,甘肃靖远人,博士,副教授,研究方向为无线网状网算法开发和人工智能。**E-mail**:caozhanmao@scnu.edu.cn

CAO Zhan-mao, born in 1967, PhD, associate professor, his research interests include wireless mesh network algorithm development and artificial intelligence.

郑明杰(1997 -),男,广东揭阳人,硕士生,研究方向为资源匹配算法。**E-mail**:1372187793@qq.com

ZHENG Ming-jie, born in 1997, MS candidate, his research interest includes resource matching algorithm.

肖博健(1999 -),男,湖南娄底人,硕士生,研究方向为深度学习和自然语言处理。**E-mail**:xbj5080@163.com

XIAO Bo-jian, born in 1999, MS candidate, his research interests include deep learning and natural language processing.

第 39 届中国计算机应用大会 CCF NCCA 2024
征 稿 通 知

随着人工智能、大数据、区块链、物联网、云计算、5G、数字孪生等为代表的计算机应用技术快速发展,并与各行各业融合程度进一步加深,支持各类工业设备、信息系统、业务流程、企业产品与服务、人员之间的互操作技术也愈加复杂。构建一个更高效、更安全、更智能的互操作技术体系是各行业领域当前面临的重要挑战。

由中国计算机学会(CCF)主办,CCF 计算机应用专业委员会、东北林业大学、黑龙江省计算机学会联合承办的 CCF 第 39 届中国计算机应用大会(CCF NCCA 2024)将围绕“数字经济赋能向北开放新高地,助推产研应用厚植新质生产力”主题于 2024 年 7 月 15—18 日在黑龙江·哈尔滨举办。大会将邀请 10 余位中国科学院院士、中国工程院院士及 50 余位优青、杰青、长江学者等计算领域及其行业应用领域的国家级人才、顶级专家学者、企业家共同探讨人工智能+应用,尤其是在大规模预训练语言模型赋能千行百业的应用方面,将共享产学研创新合作新成就,共谋经济社会应用新前景。中国计算机应用大会已经发展成为 CCF 的重要学术交流会议之一,大会连续多年被评为中国科学技术协会重要学术会议指南推荐会议之一。

本届大会还设有专题讲习班、分领域论坛、企业成果展示、学术论文交流、不同赛道的竞赛等形式多样的活动。会议现面向学术界与产业界进行征文,大会将评选一定数量的优秀论文推荐期刊发表,并在颁奖大会上颁发证书予以表彰。

征稿截止日期:2024 年 4 月 15 日

录用通知日期:2024 年 5 月 15 日

大会召开日期:2024 年 7 月 15—18 日

会议网站:https://conf.ccf.org.cn/ncca2024

论文投稿网站:https://conf.ccf.org.cn/ncca2024/paper

CCF 计算机应用专业委员会

中科国鼎数据科学研究院