

基于边推断增强对比学习的社交媒体谣言检测模型

刘楠, 张凤荔, 尹嘉奇, 陈学勤, 王瑞锦

引用本文

刘楠, 张凤荔, 尹嘉奇, 陈学勤, 王瑞锦. 基于边推断增强对比学习的社交媒体谣言检测模型[J]. 计算机科学, 2023, 50(11): 49-54.

LIU Nan, ZHANG Fengli, YIN Jiaqi, CHEN Xueqin, WANG Ruijin. Rumor Detection Model on Social Media Based on Contrastive Learning with Edge-inference Augmentation [J]. Computer Science, 2023, 50(11): 49-54.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多粒度对比学习的聊天对话摘要模型](#)

Chat Dialogue Summary Model Based on Multi-granularity Contrastive Learning
计算机科学, 2023, 50(11): 192-200. <https://doi.org/10.11896/jsjx.230300241>

[骨架数据增强和双重最近邻检索自监督动作识别](#)

Self-supervised Action Recognition Based on Skeleton Data Augmentation and Double Nearest Neighbor Retrieval
计算机科学, 2023, 50(11): 97-106. <https://doi.org/10.11896/jsjx.230500158>

[基于对比学习的多关系属性图聚类方法](#)

Clustering Method Based on Contrastive Learning for Multi-relation Attribute Graph
计算机科学, 2023, 50(11): 62-70. <https://doi.org/10.11896/jsjx.220900166>

[基于图对比学习的多模态交通流量协同预测方法](#)

Co-Forecasting for Multi-modal Traffic Flow Based on Graph Contrastive Learning
计算机科学, 2023, 50(10): 135-145. <https://doi.org/10.11896/jsjx.230700127>

[基于ConvNeXt热图定位和对比学习的细粒度图像分类研究](#)

Study on Fine-grained Image Classification Based on ConvNeXt Heatmap Localization and Contrastive Learning
计算机科学, 2023, 50(10): 119-125. <https://doi.org/10.11896/jsjx.220900196>

基于边推断增强对比学习的社交媒体谣言检测模型

刘楠 张凤荔 尹嘉奇 陈学勤 王瑞锦

电子科技大学信息与软件工程学院 成都 610054

(ln@std.uestc.edu.cn)

摘要 近年来,为了应对谣言广泛传播所带来的一系列社会问题,研究者开发了许多基于深度学习的谣言检测方法。虽然这些方法通过从传播结构中学习谣言的高级表征实现了较优的检测性能,但它们都忽略了在构造传播网络时边的不确定性,导致模型的可靠性降低,出现累积误差。针对该问题,提出了边推断增强对比学习的社交媒体谣言检测模型(Edge-Inference Contrastive Learning, EICL)。首先, EICL 基于消息转发(评论)时间戳为给定消息构建传播图;然后,利用新设计的边权重调整策略进行事件传播图数据增强以捕获传播结构边的不确定性;最后,利用对比学习方法解决原数据集本身存在的稀疏性问题,提高模型泛化能力。实验结果表明,与其他基准模型相比,模型 EICL 在公开数据集 Twitter15 和 Twitter16 上的准确率分别提高了 2.0% 和 3.0%,证明其可显著提升社交媒体谣言检测效果。

关键词: 谣言检测;对比学习;数据增强;因果推断

中图分类号 TP181

Rumor Detection Model on Social Media Based on Contrastive Learning with Edge-inference Augmentation

LIU Nan, ZHANG Fengli, YIN Jiaqi, CHEN Xueqin and WANG Ruijin

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract In recent years, in order to deal with various social problems which are caused by the wide spreading of rumors, researchers have developed many deep learning-based rumor detection methods. Although these methods improve detection performance by learning the high-level representation of rumor from its propagation structure, they still suffer the problem of lower reliability and cumulative errors effect, due to the ignoring of edges' uncertainty when constructing the propagation network. To address such a problem, this paper proposes the edge-inference contrastive learning (EIC) model. EICL first constructs a propagation graph based on timestamps of retweets (comments) for a given message. Then, it augments the event propagation graph to capture the edge uncertainty of the propagation structure by a newly designed edge-weight adjustment strategy. Finally, it employs the contrastive learning technique to solve the sparsity problem of the original dataset and improve the model generalization. Experimental results show that the accuracy of EICL is improved by 2.0% and 3.0% on Twitter15 and Twitter16, respectively, compared with other state-of-the-art baselines, which demonstrate that it can significantly improve the performance of rumor detection on social media.

Keywords Rumor detection, Contrastive learning, Data augmentation, Causal inference

1 引言

在过去的十年里,出现了大量的在线社交媒体平台,如 Twitter、微博、Facebook、Instagram、抖音等,它们逐渐成为人们日常生活中获取信息的主要来源,并从根本上改变了信息的传统共享方式,打破了大众之间的交流屏障。然而,这些

平台有时却成为伤害他人的直接利器,如新冠疫情期间,谣言不断,“某地要封城”“某地要进行空中消杀”“某某隔离人员逃跑了”等不实言论出现,大量的键盘侠借助这些平台渠道的快捷性,加速了那些未经证实的谣言的传播,使得在扰乱抗疫秩序的同时,也对当事人造成了伤害。因此,对社交媒体上的网络信息进行谣言检测是非常必要的。

到稿日期:2022-10-09 返修日期:2022-12-30

基金项目:国家自然科学基金(62271128);四川省区域创新合作项目(2020YFQ0018);四川省科技计划重点研发项目(2022ZDZX0004, 2023YFG0029, 2023YFG0150, 2023ZHCG004, 2022YFG0212, 2021YFS0391, 2021YFG0027)

This work was supported by the National Natural Science Foundation of China(62271128), Sichuan Regional Innovation Cooperation Project(2020YFQ0018) and Sichuan Science and Technology Program(2022ZDZX0004, 2023YFG0029, 2023YFG0150, 2023ZHCG004, 2022YFG0212, 2021YFS0391, 2021YFG0027).

通信作者:尹嘉奇(yinjiaqi@Gmail.com)

传统的谣言检测方法主要是基于监督学习,通过大量的标签数据训练得到分类器,如决策树^[1]、随机森林^[2]和支持向量机(SVM),它们的检测准确性依赖于手工标注的特征。随着图神经网络的研究兴起,人们发现谣言在传播的过程中,其文本内容和传播模式的变化存在一定的规律。其中,Vo-soughi等^[3]已经证实谣言比真相传播得更远、更快、更深、更广,为从传播结构进行谣言检测分析提供了研究的现实意义。

此外,大量的学者基于深度学习对谣言检测模型进行深入研究,但部分研究^[4-5]仅关注传播序列中的时间特征。其中,部分研究从谣言传播结构中捕获序列特征^[6-7],以学习谣言传播和扩散过程中存在的高级表示;部分研究^[8-9]将传播结构建模为树结构,以捕捉结构特征。随着图表示学习的进一步发展,基于图表示学习的谣言检测已是研究热潮。Bian等^[7]、Wei等^[8]基于回复或转发关系构造图,并通过边聚合邻居的特征。然而,大部分基于深度学习的谣言检测方法,都是基于可靠的传播拓扑结构和大样本的数据集进行分析挖掘,这不仅忽略了对传播结构不稳定性的建模,同时在数据预处理和模型训练部分也造成了大量的成本浪费。

社交网络上的谣言传播结构具有很强的信息表达能力,现有基于深度学习的谣言检测模型在对传播结构进行分析时,忽略了其在真实社交网络中的不确定性,导致谣言进行多级传播后易出现严重误差累积。此外,现有的模型大多数都是针对大型公开数据集进行训练,由于谣言的传播结构呈长尾分布,实际有效数据并不多,这在计算上是昂贵和低效的^[12-13]。

为了刻画和表达真实网络中谣言传播者不可靠关系导致的检测不确定性问题,以缩短训练时间和提高训练效率,我们提出了一种基于边推断增强对比学习的社交媒体谣言检测模型,通过对传播结构不确定性的捕获,设计基于先验概率的自适应边权值调整的数据增强策略,以保留事件的关键性特征。结合对比学习相似样本数据之间的内在强相关性的特性,利用原数据与增强后的数据进行对比预训练,获得谣言事件级更深层次的特征表示。最后通过少量的标签数据对模型进行微调,完成模型的优化。

本文的主要贡献如下:

(1)提出了一种基于对比学习框架的边推理谣言检测模型 EICL,用于学习事件级谣言的深度刻画,从而实现小样本学习,并提高谣言检测结果的精度。

(2)基于谣言传播结构中存在的 uncertainty,创新性地设计了边权值自适应作为事件图的增强策略。

(3)在两个公开数据集上的实验表明,本文提出的谣言检测模型 EICL 的鲁棒性表现最好,且比以往的方法具有更高的预测准确率。

2 相关工作

2.1 谣言检测

常见谣言检测方法可以概括为两类:传统方法和人工智能方法。传统方法是通过对谣言相关数据进行统计分析以定义关键特征。例如,Castillo等^[1]通过分析用户属性、谣言传播路径和文本,提出了大量谣言检测特征。一些研究人员从

不同的角度介绍了各种特征。随着人工智能的发展,一些学者开始尝试利用深度学习来识别谣言。Ma等^[14]采用了循环神经网络(RNN)来依次处理谣言传播序列中的每个时间步骤。之后,许多研究人员通过注意力机制^[15]和卷积神经网络^[16]获得更为牢靠的长期依赖。随着图神经网络的兴起,部分学者开始针对谣言传播拓扑结构的潜在规律进行研究。其中,Ma等^[14]提出了同时包含基于自下而上和自上而下的传播树结构的递归神经网络方法(RvNN)。受图卷积网络(GCN)的启发,Ran等^[17]和Nguyen等^[18]将传播结构表述为图,且Bian等^[7]首先应用了两个基于传播图和分散图的GCNs。

然而,这些对传播结构的分析都默认了公开数据集的可靠性,认为其所包含的拓扑链接为可靠链接,忽略了实际社交网络中非个人主观意志行为造成的不可靠关系,导致模型的鲁棒性不足,且在一定程度上影响了检测精度。

2.2 对比学习

对比学习(Contrastive Learning)作为自监督学习中一个非常重要的部分,被广泛运用在计算机视觉、自然语言处理等领域。其目标是通过嵌入空间使一个样本和经过增强的新样本尽可能靠近,同时让不同样本之间尽可能远离。这与学习映射到某些标签的判别模型和重建输入样本的一般模型有着本质上的区别。

对比学习旨在最大化相同类型实例之间的一致性,并区分不同类型的其他实例^[19]。近年来,对比学习在无监督视觉表征学习方面取得了巨大成功^[20-21]。除计算机视觉外,最近的研究表明,其在语义文本相似性^[22]、姿态检测^[23]、短文本聚类、未知意图检测^[21]等方面仍具有研究前景。然而,上述对比学习框架是专为增强非结构化文本数据(如句子和文档)而提出的,这些数据不适用于复杂社交网络中的谣言检测任务。受它们的启发,本文将从传播结构的不确定性出发,结合对比学习框架进行探索,设计鲁棒性更好的谣言检测模型。

3 问题描述

本章将对部分概念、问题进行定义与解释。在本章中,我们提供了形式化的问题定义,并对所研究的问题进行了简要的说明。表1列出了本文使用到的符号。

表1 符号
Table 1 Notations

符号	描述
C	事件主题集合
c^i, r^i, x_j^i	特定事件的相关帖子集合、源帖文本和相关帖子文本集
F, A	事件有向传播图的特征矩阵和邻接矩阵
$A^{(l)}$	最大 l 阶邻域的图边权重更新以后的邻接矩阵
d	文本特征维度
$W_s^{(l)}, \psi, W_c, b_c$	均是可学习参数
G	整个数据集构成的图集

假设本文收集了一些特定事件主题集,其中 c^i 表示第 i 个事件主题。每一个事件主题 $c^i = \{r^i, \{x_j^i, j \in [1, n-1]\}\}$,涵盖了事件的源帖 r^i ,以及对 r^i 进行评论或者转发操作的一系列帖子 x_j^i 。

定义1(事件传播图) 以事件主题的源帖为源节点,然后以评论或者转发主被动性作为帖子之间的传播方向,所有

参与的帖子可以看作一个节点,整个传播过程形成的图网络结构即为事件传播图 $G_i = \langle V_i, E_i \rangle$ 。其中, $V_i = \{r^i, x_1^i, x_2^i, \dots, x_{n_i-1}^i\}$ 即事件的相关帖子均为图中是节点, $E_i = \{e_{st}^i \mid s, t = 0, \dots, n_i - 1\}$ 表示帖子传播的有向边集,即假设 x_2^i 是 x_1^i 的回复帖,那么节点 x_2^i 与 x_1^i 之间将构造一条有向边 e_{12}^i 。

本文谣言检测任务针对的是事件级的,通过对同一个事件主题下相关帖子的语义分析,以及对整个事件在社交网络中发酵情况的分析,来捕获区分事件真实性的核心特征,以达到对事件类谣言的早期检测。本文将谣言检测看

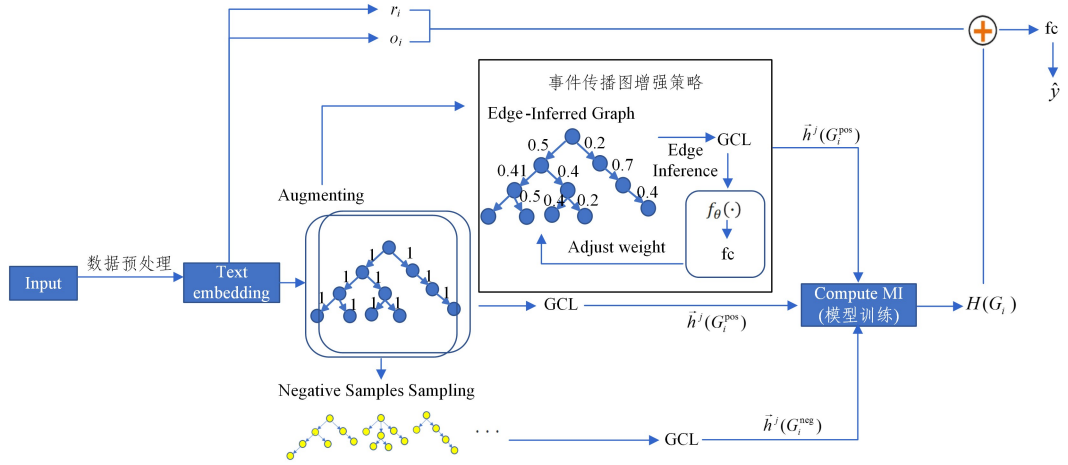


图1 模型总体结构图

Fig. 1 Overall structure of the proposed model

4.1 数据预处理

本文进行的是事件主题谣言检测,因此本节需要对原始数据集进行图构建和文本嵌入操作,从而为后续实验实现数据的预处理。

首先,基于原始数据集中已划分好的事件类,根据特定事件相关帖子之间的转发和回复操作,以事件的源帖为根节点构建自顶向下的初始传播图,具体的构造类似于之前的工作^[10]。构建完毕以后,每个事件的传播过程都有对应的事件有向图,其初始化邻接矩阵为 $\mathbf{A} \in \{0,1\}^{|\mathcal{V}| \times |\mathcal{V}|}$ 。

然后,本文利用当前较为流行的 Word2Vec 算法对所有帖子的文本进行预训练,将得到的初始值作为图中节点的初始特征,整个图的初始特征矩阵表示为 $\mathbf{F} \in \mathbb{R}^{|\mathcal{V}| \times d}$, d 是文本特征的维度。

4.2 事件传播图增强

真实社交网络环境中,总是存在部分用户故意传播虚假谣言,这通常是由金钱或非金钱利益(如权力和流量)所驱动,同时还有部分用户思想和情绪上易受外界影响,他们无意中助推了虚假谣言的传播。因此,等价看待每一位谣言事件的参与者是不明智的。本文通过修改图的边权重值,以获取谣言传播结构的不确定性。

本节将通过对比图邻接矩阵值的自适应调整,实现对图的原边关系权重值的调整,从而弥补以往谣言传播结构不确定性造成的损失。为提高模型的鲁棒性,本文设计边权重推断调整策略进行事件传播图增强。

针对已构建的事件初始图,基于前一层节点表示,结合

作是一个多分类问题,目标是训练一个模型以预测特定信息的标签 \hat{y}_i ,本文采用细粒度的 NR,FR,TR,UR(即非谣言、虚假谣言、真实谣言和未经验证的谣言)作为分类标签。

4 模型设计

本章将详细介绍基于边推断增强对比学习的社交媒体谣言检测模型,其模型如图1所示。该模型主要包括4部分:数据预处理、事件传播图增强、模型训练和优化目标函数。模型的最终目标是得到分类结果。

变换函数 $f_c(\cdot; \theta_l)$ 计算推断出相邻节点之间的新权重值,则邻接矩阵 $\mathbf{A}^{(l)}$ 可以通过式(1)和式(2)计算更新。

$$g_l^{(l)} = f_c(\|\mathbf{h}_i^{(l-1)} - \mathbf{h}_j^{(l-1)}\|; \theta_l) \quad (1)$$

$$\mathbf{A}^{(l)} = \sum_{i=1}^T \sigma(\mathbf{W}_i^{(l)} g_l^{(l)} + \mathbf{b}_i^{(l)}) \cdot \mathbf{A}^{(l-1)} \quad (2)$$

其中, $f_c(\cdot; \theta_l)$ 由卷积层和激活函数组成, $\mathbf{h}_i^{(l-1)}$ 指第 $(l-1)$ 层传播节点的特征, T 指谣言传播过程中潜在关系类型的数量, $\sigma(\cdot)$ 是 sigmoid 函数, $\mathbf{W}_i^{(l)}$ 是可学习的参数。

4.3 基于对比学习的模型训练

4.2节中采用的事件图增强策略在某种程度上增加了噪声数据,可以提高模型的鲁棒性,然而,社交网络的信息传播数据集稀疏且满足长尾分布,在有限的数据集下,训练的模型泛化能力较差。然而,特定事件的谣言数据集采集工作的时间和人力成本是非常昂贵的。

因此,本节将从节点级和整图级两个角度,利用对比学习将训练数据集中输入事件图表示和增强事件传播图表示之间的互信息(Mutual Information)最大化,以提升模型的泛化能力。

具体而言,对于谣言传播图数据集,图 G 中的节点 j 的自监督学习过程如式(3)和式(4)所示:

$$\mathbf{h}_j^k = \text{GIN}(\mathbf{h}_j^{k-1}) \quad (3)$$

$$\mathbf{h}^j = \text{concat}(\{\mathbf{h}_j^k\}_{k=1}^K) \quad (4)$$

其中, \mathbf{h}_j^k 是第 k 层节点 j 的特征向量。GIN^[24] 是本文所选用的图卷积编码器,其将所有深层次的特征向量汇总为节点级特征向量 \mathbf{h}^j 。 \mathbf{h}^j 包含了每个节点从不同阶域捕获到的特征集合。

然后,利用式(5)计算获得整个事件图的整图特征表示。

$$H(G) = \sigma \left(\frac{1}{|V|} \sum_{j=1}^{|V|} \mathbf{h}^j \right) \quad (5)$$

其中, $H(G)$ 指特定事件图的整图级特征表示, σ 是一个 sigmoid 函数。谣言传播图数据集上的最大化互信息(MI)可由式(6)计算得到。

$$I_{\psi}(\mathbf{h}^i(G); H(G)) = \mathbb{E}[-s\mathcal{P}(-T_{\psi}(\tilde{\mathbf{h}}^i(G_i^{\text{pos}}), H(G_i)))] - \mathbb{E}[s\mathcal{P}(T_{\psi}(\tilde{\mathbf{h}}^i(G_i^{\text{neg}}), H(G_i)))] \quad (6)$$

$$s\mathcal{P}(z) = \log(1 + e^z) \quad (7)$$

其中, T_{ψ} 是一个用于预估互信息 I_{ψ} 的判别器; G_i 是输入的特定事件图样本; G_i^{pos} 是 G_i 的正样本, 即 G_i 通过 4.2 节增强操作得到的事件图样本; G_i^{neg} 是 G_i 的负样本, 即同 batch 中的其他事件图输入样本; ψ 表示神经网络中的一组参数, 可通过学习得到。

4.4 优化目标函数

谣言检测任务可以被视为一个多分类问题, 因此, 本文采用分类预测值与标签真实值的交叉熵作为目标损失函数。

为提高谣言检测的准确性, 本节将 4.1 节嵌入方式得到的事件相关文本特征表示和基于互信息最大化得到的事件图特征表示进行组合, 作为事件的最终表示, 其计算过程如式(8)和式(9)所示:

$$\mathbf{o}_i = \frac{1}{|V_i|} \left(\sum_{j=1}^{|V_i|-1} \mathbf{x}_j^i + \mathbf{r}_i \right) \quad (8)$$

$$S_i = H(G_i) \oplus \mathbf{o}_i \oplus \mathbf{r}_i \quad (9)$$

其中, \mathbf{o}_i 是特定事件 i 的相关文本特征表示, \mathbf{x}_j^i 是关于事件 i 的帖子 j 的文本特征表示, \mathbf{r}_i 是关于事件 i 的源帖文本特征表示。

基于事件的最终表示, 其对应的类标签概率可以通过全连接层和 softmax 函数计算得到, 计算流程如式(10)和式(11)所示:

$$\hat{y}_i = \text{softmax}(\mathbf{W}_c S_i + \mathbf{b}_c) \quad (10)$$

$$\zeta(y, \hat{y}) = \sum_{i=1}^{|C|} y_i \log \hat{y}_i + \lambda \|\Theta\|_2^2 \quad (11)$$

其中, \mathbf{W}_c 和 \mathbf{b}_c 是可学习参数, \hat{y}_i 是预测值, $\|\cdot\|_2^2$ 是 L_2 的正则化, Θ 是模型所有参数, λ 是折衷系数。

EICL 模型算法的伪代码如算法 1 所示。

算法 1 EICL 模型算法

输入: $G = (V, E, T)$

输出: 每个事件源帖的类别标签概率

1. /* 对所有事件图进行预处理, 再进行文本嵌入, 得到图的初始特征矩阵 \mathbf{F} 和初始邻接矩阵 \mathbf{A} , 每一个图对应一个特定的事件, 源帖文本特征表示为 \mathbf{r}_i , 按照式(8)计算特定事件 i 的相关文本特征表示 \mathbf{o}_i */
2. for $i=1$ to $|C|$ do
3. $\mathbf{o}_i = \frac{1}{|V_i|} \left(\sum_{j=1}^{|V_i|-1} \mathbf{x}_j^i + \mathbf{r}_i \right)$
4. /* 计算加入不确定性的新邻接矩阵 $\mathbf{A}^{(1)}$ */
5. $\mathbf{g}_i^{(1)} = f_c(\|\mathbf{h}_i^{(1-1)} - \mathbf{h}_j^{(1-1)}\|; \theta_i)$
6. $\mathbf{A}^{(1)} = \sum_{i=1}^T \sigma(\mathbf{W}_t^{(1)} \mathbf{g}_i^{(1)} + \mathbf{b}_t^{(1)}) \cdot \mathbf{A}^{(1-1)}$
7. /* 通过对比学习中互信息最大化方式, 得到事件图特征表示 $H(G)$ */

$$8. H(G) = \sigma \left(\frac{1}{|V|} \sum_{j=1}^{|V|} \mathbf{h}^j \right)$$

$$9. I_{\psi}(\mathbf{h}^i(G); H(G)) = \mathbb{E}[-s\mathcal{P}(-T_{\psi}(\tilde{\mathbf{h}}^i(G_i^{\text{pos}}), H(G_i)))] - \mathbb{E}[s\mathcal{P}(T_{\psi}(\tilde{\mathbf{h}}^i(G_i^{\text{neg}}), H(G_i)))]$$

$$10. s\mathcal{P}(z) = \log(1 + e^z)$$

$$11. /* 计算 i 事件的最终表示, 并预测其对应的类标签概率 \hat{y}_i */$$

$$12. S_i = H(G_i) \oplus \mathbf{o}_i \oplus \mathbf{r}_i$$

$$13. \hat{y}_i = \text{softmax}(\mathbf{W}_c S_i + \mathbf{b}_c)$$

14. end for

5 实验分析

本章将本文提出的 EICL 方法与部分传统方法进行实验比较, 从数据集、基线模型、超参和评价指标设定、实验结果分析 4 个部分进行介绍。

5.1 数据集

本文实验采用的是主流社交媒体平台的两类数据集: Twitter15^[25] 和 Twitter16^[25]。相关的数据集统计如表 2 所列。Twitter15 和 Twitter16 分别包含 1490 个事件和 818 个事件, 每条源帖都注释了对应的类别标签。标签一共分为 4 类, 即非谣言(NR)、假谣言(FR)、真谣言(TR)或未经证实的谣言(UN)。

表 2 数据集统计

Table 2 Statistics of datasets

统计	Twitter15	Twitter16
事件主题	1490	818
非谣言(NR)	374	205
虚假谣言(FR)	370	205
未经验证的谣言(UN)	374	203
真实谣言(TR)	372	205
帖子总数量	331612	204820

所有数据集随机分为 5 部分, 并进行 5 次交叉验证, 以获得稳健的结果。

5.2 超参和评价指标设定

在本文实验中, 通过对 NLP 相关研究进行调研以及考虑到本研究适用的应用范围, 本文将词嵌入向量的维度值 d 设置为 5000。模型的参数更新采用随机梯度下降方法, 并使用 Adam 算法优化模型。GIN 网络提取的隐藏特征向量维数为 64。通过测试, 发现潜在关系类型的数量 T 设置为 2, 3 或者 4 时, 模型的性能最好。自我监督预训练过程迭代 25 个 epochs, 而监督微调过程迭代 100 个 epochs, 当验证精度增加 10 个 epochs 时, 训练提前停止。

为了评估本文模型的有效性, 基于 Twitter15 和 Twitter16 这两个公开数据集, 实验采用与对比实验相同的评价指标(准确率 Acc. 和 F1 值)。

5.3 对比模型

将本文提出的模型与所选取的基准模型在相同的数据集上开展实验, 本文选取的 6 个基准模型具体描述如下:

(1)GRU-RNN: 利用循环神经网络的多种变体组合, 充分捕获帖子在传播过程中的时序结构变化特征。

(2)PPC: 由递归神经网络和卷积神经网络组成的传播路径分类器建模用户特征序列。

(3)RvNN:基于从叶子节点到根节点的传播树遍历方向的递归神经网络,捕获扩散线索和内容语义。

(4)StA-PLAN:利用 transformer 网络更好地捕获帖子在网络传播结构较远邻域之间的影响。

(5)BiGCN:该模型分别使用 GCN 从信息传播结构和信息扩散结构进行特征学习,然后通过融合的方式将其作为整个传播图的结构特征,完成分类。

(6)EBGCN:该模型利用神经网络优化图结构中关系的不确定性,再结合 GCN 学习整个图结构的特征表示,最后进行分类。

5.4 实验结果分析

为了验证本文模型的有效性,我们将所提模型与其他的6个基准模型进行比较。基于数据集 Twitter15 和 Twitter16 的实验对比结果分别如表 3 和表 4 所列。为了保证比较的公平性,实验引用了文献[26]和文献[6]的实验结果。

表 3 Twitter15 数据集上的实验结果

Table 3 Results on Twitter15 dataset

methods	Acc.	F1 值			
		NR	FR	UR	TR
GRU-RNN	0.641	0.684	0.634	0.571	0.688
PPC	0.697	0.689	0.760	0.696	0.645
RvNN	0.723	0.682	0.758	0.654	0.821
StA-PLAN	0.808	0.930	0.758	0.808	0.812
BiGCN	0.836	0.791	0.842	0.801	0.887
EBGCN	0.855	0.831	0.857	0.903	0.816
EICL	0.875	0.825	0.885	0.871	0.918

从表 3 可以看出,EICL 模型在 Twitter15 数据集上的准确率为 0.875,且在 TR 类谣言的检测上表现较佳,F1 值达到 0.918。从表 4 可以看出,EICL 模型在 Twitter16 数据集上的准确率为 0.910,且在 FR 类谣言的检测上表现较佳,F1 值达到 0.937。

表 4 Twitter16 数据集上的实验结果

Table 4 Results on Twitter16 dataset

methods	Acc.	F1 值			
		NR	FR	UR	TR
GRU-RNN	0.636	0.617	0.715	0.527	0.577
PPC	0.702	0.608	0.711	0.664	0.816
RvNN	0.737	0.662	0.743	0.708	0.835
StA-PLAN	0.852	0.840	0.846	0.837	0.884
BiGCN	0.864	0.788	0.859	0.864	0.932
EBGCN	0.880	0.823	0.878	0.937	0.875
EICL	0.910	0.850	0.937	0.912	0.935

分析表 3 和表 4 可以得出,相比传统的时序网络模型,StA-PLAN 模型因可以捕获更深层次的时序特征,其性能明显优于 GRU-RNN 模型,在 Twitter15 和 Twitter16 的数据集上分别达到了 0.808 和 0.852 的准确率。相比深度学习模型,传统学习模型性能明显不佳,EBGCN 模型通过捕获全局结构边的不确定关系,性能得到了明显的提升。而本文模型在已有的研究基础上,融合了对比学习思想,在 Twitter15 和 Twitter16 数据集上分别达到了 0.873 和 0.912 的准确率,比 StA-PLAN 模型分别提升了 0.065 和 0.060,比 EBGCN 模型分别提升了 0.020 和 0.030,实验结果证明了本文提出的模型的有效性。

综上,本文提出的 EICL 模型在两个常用基准数据集上的扩展实验证明了在传播结构中建模不确定性的有效性,且在谣言检测和早期谣言检测任务上,所提方法的准确性显著优于现有的基线方法。虽然此模型具有一定的鲁棒性,但是只有在标记数据较少的情况才能凸显优势。

结束语 本研究提出了一个新的谣言检测模型,创新性地将两种增强策略,以提取有意义的谣言传播模式,并结合对比学习编码器,实现了小样本即可充分学习复杂谣言事件的相关内容语义信息。后续,我们可以从特征融合的方式以及损失函数的设定方面对该模型进行进一步研究和改进,并通过实施社会特征工程对该模型进行扩展,以分析和跟踪谣言。

参考文献

- [1] CASTILLO C,POBLETE B,MENDOZA M,et al. Information Credibility on Twitter[C]// Proceedings of the 20th International Conference on World Wide Web. 2011:675-84.
- [2] KWON S,CHA M,JUNG K,et al. Prominent Features of Rumor Propagation in Online Social Media[C]// 2013 IEEE 13th International Conference on Data Mining. 2013:1103-1108.
- [3] VOSOUGHI S,ROY D,ARAL S. The Spread of True and False News Online[J]. Science,2018,359(6380):1146-1151.
- [4] MA J,GAO W,PRASENJIT M,et al. Detecting Rumors from Microblogs with Recurrent Neural Networks[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. 2016:3818-3824.
- [5] KOTTETI C M M,DONG X,QIAN L. Ensemble deep learning on time-series representation of tweets for rumor detection in social media[J]. Applied Sciences,2020,10(21):7541.
- [6] KHOO L M S,CHIEU H L,QIAN Z,et al. Interpretable rumor detection in microblogs by attending to user interactions[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020,34(5):8783-8790.
- [7] BIAN T,XIAO X,XU T,et al. Rumor detection on social media with bi-directional graph convolutional networks[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020,34(1):549-556.
- [8] WEI P,XU N,MAO W. Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity[J]. arXiv:1909.08211,2019.
- [9] SHARMA K,QIAN F,JIANG H,et al. Combating fake news: A survey on identification and mitigation techniques[J]. ACM Transactions on Intelligent Systems and Technology(TIST),2019,10(3):1-42.
- [10] ZHOU X,ZAFARANI R. A survey of fake news: Fundamental theories,detection methods,and opportunities[J]. ACM Computing Surveys(CSUR),2020,53(5):1-40.
- [11] LIU X,NOURBAKHSH A,LI Q,et al. Real-time rumor debunking on twitter[C]// Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015:1867-1870.
- [12] SHELKE S,ATTAR V. Rumor Detection in Social Network Based on User,Content and Lexical Features[J]. Multimedia

- Tools and Applications, 2022, 81(12):347-368.
- [13] ZHANG H, QIAN S, FANG Q, et al. Multi-modal Meta Multi-task Learning for Social Media Rumor Detection [J]. IEEE Transactions on Multimedia, 2021, 24, 1449-1459.
- [14] MA J, GAO W, WONG K. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks [C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2018:1980-1989.
- [15] CHEN T, LI X, YIN H, et al. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection [C] // Pacific-Asia Conference on Knowledge Discovery and Data Mining. Cham: Springer, 2018:40-52.
- [16] MISHRA R, SETTY V. Sadhan: Hierarchical attention networks to learn latent aspect embeddings for fake news detection [C] // Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval. 2019:197-204.
- [17] RAN H, JIA C, ZHANG P, et al. MGAT-ESM: Multi-channel graph attention neural network with event-sharing module for rumor detection [J]. Information Sciences, 2022, 592:402-416.
- [18] NGUYEN V H, SUGIYAMA K, NAKOV P, et al. Fang: Leveraging social context for fake news detection using graph representation [C] // Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020:1165-1174.
- [19] WANG T, ISOLA P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere [C] // International Conference on Machine Learning. PMLR, 2020:9929-9939.
- [20] LIN H, MA J, CHENG M, et al. Rumor detection on twitter with claim-guided hierarchical graph attention networks [J]. arXiv:2110.04522, 2021.
- [21] LIN H, YAN Y, CHEN G. Boosting Low-resource Intent Detection with In-scope Prototypical Networks [C] // ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing, 2021:7623-7627.
- [22] GAO T, YAO X, CHEN D. Simcse: Simple contrastive learning of sentence embeddings [J]. arXiv:2104.08821, 2021.
- [23] MOHTARAMI M, GLASS J, NAKOV P. Contrastive language adaptation for cross-lingual stance detection [J]. arXiv:1910.02076, 2019.
- [24] ZHUANG C, MA Q. Dual graph convolutional networks for graph-based semi-supervised classification [C] // Proceedings of the 2018 World Wide Web Conference. 2018:499-508.
- [25] MA J, GAO W, WONG K. Detect Rumors in Microblog Posts Using Propagation Structure Via Kernel Learning [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2017:708-717.
- [26] WEI L, HU D, ZHOU W, et al. Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection [J]. arXiv:2107.11934, 2021.



LIU Nan, born in 1995, Ph.D. Her main research interests include social network analysis and deep representation learning.



YIN Jiaqi, born in 1986, bachelor, engineer. His main research interests include big data and environmental informatization.

(责任编辑:何杨)