

基于用户传播网络与消息内容融合的谣言检测模型

薛海涛¹, 王莉^{1*}, 杨延杰¹, 廉颢²

(1. 太原理工大学 大数据学院, 太原 030600; 2. 北方自动控制技术研究所, 太原 030006)

(* 通信作者电子邮箱 wangli@tyut.edu.cn)

摘要:针对社交媒体平台上消息内容普遍很短、传播结构中存在大量空转发、用户角色与内容间的失配等条件约束,提出了一种基于传播网络中的用户属性信息和消息内容的谣言检测模型 GMB_GMU。首先以用户属性为节点、传播链为边构建用户传播网络,并引入图注意力网络(GAT)得到用户属性的增强表示;同时,基于此用户传播网络,利用 node2vec 得到用户的结构表征,并使用互注意机制对其进行增强。另外,引入 BERT 建立源帖内容表征。最后,利用多模态门控单元(GMU)对用户属性表征、结构表征和源帖内容表征进行融合,从而得到消息的最终表征。实验结果表明,GMB_GMU 模型在公开的 Weibo 数据上的准确率达到 0.952,能够有效识别谣言事件,效果明显优于基于循环神经网络(RNN)和其他神经网络基准模型的传播算法。

关键词:谣言检测;用户属性;图注意力网络;多模态门控单元;传播网络

中图分类号:TP391 **文献标志码:**A

Rumor detection model based on user propagation network and message content

XUE Haitao¹, WANG Li^{1*}, YANG Yanjie¹, LIAN Biao²

(1. College of Data Science, Taiyuan University of Technology, Taiyuan Shanxi 030600, China;

2. North Automatic Control Technology Institute, Taiyuan Shanxi 030006, China)

Abstract: Under the constraints of very short message content on social media platforms, a large number of empty forwards in the transmission structure, and the mismatch between user roles and contents, a rumor detection model based on user attribute information and message content in the propagation network, namely GMB_GMU, was proposed. Firstly, user propagation network was constructed with user attributes as nodes and propagation chains as edges, and Graph Attention neTwork (GAT) was introduced to obtain an enhanced representation of user attributes; meanwhile, based on this user propagation network, the structural representation of users was obtained by using node2vec, and it was enhanced by using mutual attention mechanism. In addition, BERT (Bidirectional Encoder Representations from Transformers) was introduced to establish the source post content representation of the source post. Finally, to obtain the final message representation, Gated Multimodal Unit (GMU) was used to integrate the user attribute representation, structural representation and source post content representation. Experimental results show that the GMB_GMU model achieves an accuracy of 0.952 on publicly available Weibo data and can effectively identify rumor events, which is significantly better than the propagation algorithms based on Recurrent Neural Network (RNN) and other neural network benchmark models.

Key words: rumor detection; user attribute; Graph Attention neTwork (GAT); Gated Multimodal Unit (GMU); propagation network

0 引言

社交媒体的即时性和便利性使得用户可以便捷获取信息并交流观点,但也为谣言传播提供了便利平台。谣言传播误导公众舆论,破坏社会安定,谣言检测与监管已经成为一个迫切需要解决的社会问题。

谣言检测的早期研究中,研究者一般通过手工构造消息文本的特征,利用决策树、随机森林、支持向量机(Support Vector Machine, SVM)等机器学习的方法进行分类^[1-2],以判别真假。这类工作依赖于特征工程,无法适应社交媒体丰富多变的内容表现以及不断演化的特点,而且社交媒体上的消

息篇幅普遍较短,其包含的语义特征有限,一定程度限制了算法模型性能。后来一些研究工作利用消息传播过程^[3-4]进行谣言检测,性能有所提升,但由于消息转发中存在大量空转发,这类基于转发结构及内容的模型在实际场景中有很大的局限性。还有一些工作将用户属性引入,以用户的权威度、可信度等作为判别所发消息真假的证据,但是由于当前谣言更具迷惑性,一些权威用户有时也会被蒙蔽而转发谣言,所以,仅依靠用户权威度评价是不够的。

实际应用中我们发现,虽然消息转发中用户有时会空转发,但是传播链中用户的参与信息在一定程度上表达了对于所转发消息的认同,同时,用户的粉丝数、发表推文数及关注

收稿日期:2021-05-12;修回日期:2021-06-25;录用日期:2021-07-04。 基金项目:国家自然科学基金资助项目(61872260)。

作者简介:薛海涛(1997—),男,山西介休人,硕士研究生,主要研究方向:自然语言处理、谣言检测; 王莉(1971—),女,山西太原人,教授,博士,CCF高级会员,主要研究方向:大数据计算与分析、数据挖掘; 杨延杰(1995—),男,山西原平人,硕士研究生,主要研究方向:自然语言处理、数据挖掘; 廉颢(1987—),男,山西太原人,硕士,主要研究方向:软件开发、数据挖掘。

数等属性信息在一定程度上反映了用户的权威度,因此,传播链中的用户属性信息可以帮助判别谣言。同时,消息内容对于检测谣言也非常重要。鉴于以上观察分析,本文提出了一种融合传播链中的用户属性和消息内容的谣言检测模型 GMB_GMU。

受上述工作启发,本文使用用户属性、传播结构和消息内容进行谣言检测。首先,构建转发图,节点特征为用户属性,边表示转发关系。结合转发图,利用图注意力网络(Graph Attention neTwork, GAT)获取增强的用户属性表征。在另一部分,使用 node2vec 算法获取转发图用户结构表征,而不依赖于节点属性,借助互注意机制以得到增强的结构表征。最后一部分使用源帖文本的 BERT (Bidirectional Encoder Representations from Transformers) 向量,取得增强的源帖内容表征。将这三种角度的表征输入到多模态门控单元(Gated Multimodal Unit, GMU)进行选择与融合,得到信息的最终表征。在公共数据集上进行实验,结果验证了本文模型的判别精度和泛化能力,并检测了模型在谣言早期检测任务上的表现性能。本文主要工作如下:

1)融合传播链中的用户属性信息和消息内容进行谣言检测,避免了信息传播结构计算中空转发带来的信息缺失问题。

2)引入 GAT 生成用户属性表征,采用 node2vec 得到结构表征,采用互注意机制建立用户结构增强表征;同时基于 BERT 建立源帖内容表征,最终将用户属性表征、结构表征和源帖内容表征三者融合得到信息的最终表征,以支持谣言检测。

3)在公共数据集上进行实验,验证了模型的性能,并通过早期检测任务,验证了模型的有效性。

1 相关工作

目前的研究主要可分为基于内容语义和基于传播结构的谣言检测方法。

1.1 基于内容语义的谣言检测

随着深度学习的发展,卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)以及长短期记忆(Long Short-Term Memory, LSTM)网络等被应用到自然语言处理任务中进行文本表征学习^[5],以挖掘文本语义。段大高等^[6]提出了一种融合用户属性和帖子内容特征的 BP 神经网络模型,通过 Word2Vec 算法进行表征计算,然后输入到神经网络模型中提取文本特征。刘政等^[7]提出了基于 CNN 的谣言检测算法,使用句向量结合多层卷积网络,能有效地捕捉到特定范围的上下文语义信息。Mikolov 等^[8]将双向 RNN 应用到文本特征提取中,获取到了前后语义之间的联系。对于其他神经网络方法,虽然可以准确地把握整体信息,但对于较长的文本,仍然存在上下文信息丢失的情况。

为了最大限度地捕捉到上下文信息,Fang 等^[9]将注意力算法引入到谣言检测工作中,取得了较好的检测结果。Vaibhav 等^[10]提出了一种基于图神经网络的虚假新闻检测模型,该模型对新闻中所有句子对之间的语义关系进行建模,从而进行谣言检测。Wang 等^[11]依赖文本内容,提出了 SemSeq4FD 模型来检测虚假信息,该模型同时考虑了新闻中

句子之间的全局语义关系和局部上下文顺序特征。上述所有基于内容语义的方法最大的局限性在于这些算法更适用于长文本处理,模型表现优良的前提在于提供大量长文本进行训练,但是,社交媒体上的帖子大多都是短文本,难免造成数据稀疏问题,从而影响了该类方法的检测性能。

1.2 基于传播结构的谣言检测

由于谣言涉及的领域繁杂,普遍存在学科交叉的现象,使得研究谣言传播成为一个复杂的问题。在相关工作中,不仅涉及传播的覆盖范围,还涉及传播的速度。其中较常用的是传染病模型和其变体模型,如 SIS 模型和 SIR 模型,每个传播结构中的用户可被分为“感染者(S)”“被感染者(I)”以及“停止感染者(R)”状态,由一传到多传多,可以描述谣言的传播模式。但由于信息传播的时间等特点,传染病相比信息传播在衰退效应、信息内容和联系强度等方面都有明显的区别,不适用于谣言传播。

当前,研究者倾向于关注传播网络本身的拓扑结构,从而揭示其传播特征和规律。Ma 等^[4]提出了通过评估不同类型的传播树结构的相似性来捕获新闻事件的高阶表征。Liu 等^[12]提取了信息传播过程中传播树的宽度与深度等特征。Shu 等^[13]利用社交媒体上的新闻分层传播网络进行了假新闻检测,从宏观和微观两个层面建立了一个层次化的传播网络,并比较分析了传播网络的结构特征和时间特征。Bian 等^[14]提出了一种使用双向 GCN 进行谣言检测的方法,该方法结合了谣言的序列传播和横向扩散,捕捉到了传播结构的全局特征,获得了较好的检测结果。现有的基于传播的方法使用消息传播结构信息,但是消息传播中存在大量空转发,导致消息中存在较多噪声,影响模型性能。在另一方面,这些方法没有充分使用源帖信息,在检测性能方面受到了很大限制。

2 问题定义

给定新闻 C ,其中包含 m 个参与新闻传播的用户,如图 1 所示,该新闻的用户集合为 $U = \{u_1, u_2, u_3, \dots, u_m\}$,其中 u_1 为发布原始文本的用户, u_2, u_3, \dots, u_m 为参与转发的用户,包含一阶转发用户、二阶转发用户和更高阶的转发用户,目标为训练一个网络 $f: (C, U) \rightarrow y$,将新闻 C 分类为谣言或非谣言。其中,标签值 $y \in \{0, 1\}$,0 表示非谣言,1 表示谣言。信息传播一般表现为树型转发扩散的模式。

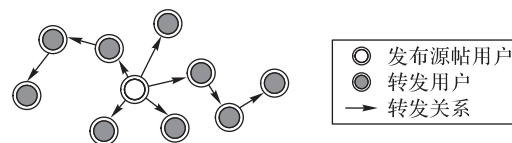


图1 信息传播模式示意图

Fig. 1 Schematic diagram of information propagation mode

3 谣言检测模型 GMB_GMU

本文提出的谣言检测模型 GMB_GMU 如图 2 所示,由四个部分组成:用户信息编码器、结构信息编码器、源文本编码器和融合分类器,图中 FC 为全连接神经网络层。

3.1 用户信息编码器

在实际场景下,用户很多时候仅仅转发原始内容,而不发表意见,转发内容不适合作为节点的属性特征。引入用户属

性作为节点特征,可以在一定程度上弥补转发结构上信息的缺失。

首先构建新闻 C 的有向转发图 $\mathcal{G}=(\mathcal{V},\mathcal{E})$, 其中 $\mathcal{V}=\{u_1, u_2, u_3, \dots, u_m\}$ 表示新闻传播中的全部用户, $\mathcal{E}=\{e_1, e_2, e_3, \dots, e_o\}$ 表示用户之间的转发关系, 定义 $H \in \mathbb{R}^{m \times t}$ 为节点集 \mathcal{V} 的特征矩阵, 用户节点 u_i 的特征向量为 $h_i \in \mathbb{R}^t$, t 为属性个数, 包含用户的粉丝数、发表推文数量和关注数等属性

信息。定义 $A \in \mathbb{R}^{m \times m}$ 为邻接矩阵, 若用户 j 转发了用户 i 的推文, 邻接矩阵 A 中元素 $a_{ij} = 1$; 若无转发关系, 则 $a_{ij} = 0$ 。

由于用户各项属性量纲和变化范围不同, 所以往往对特征矩阵进行归一化, 使得模型可以快速收敛。具体的, 通过以下公式得到 H 归一化后的特征矩阵 H^s :

$$H^s = \frac{(H - H_{\max}(\text{axis} = 0))}{(H_{\max}(\text{axis} = 0) - H_{\min}(\text{axis} = 0))} \quad (1)$$

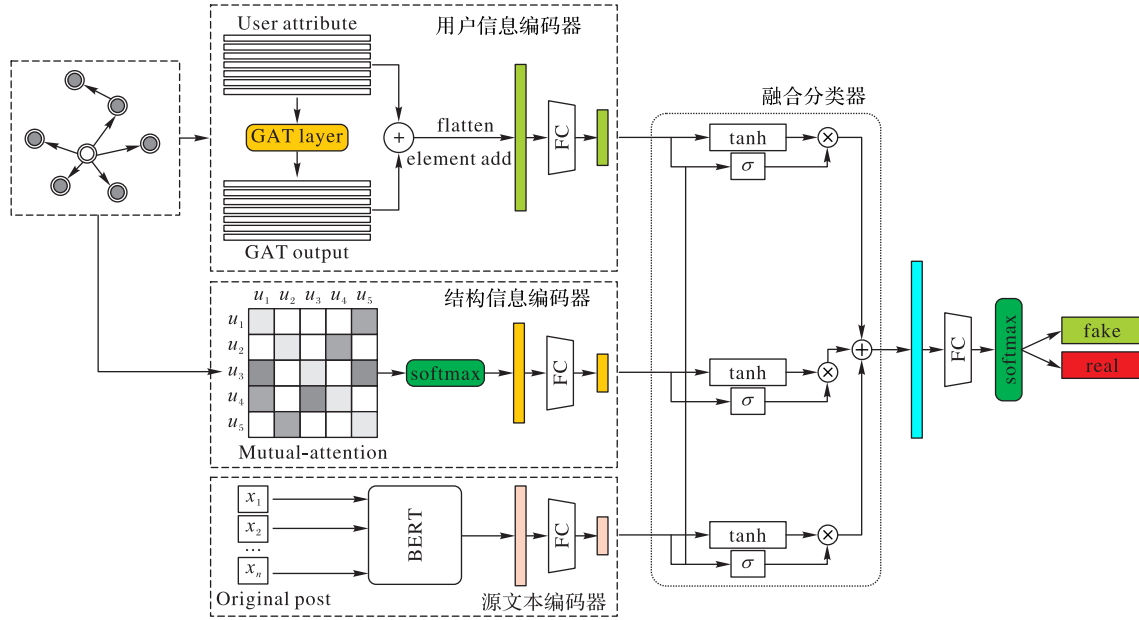


图2 GMB_GMU模型

Fig. 2 GMB_GMU model

接下来, 将 $H^s = \{h_1^s, h_2^s, \dots, h_m^s\}$ 输入到 $GAT^{[15]}$ 中, $h_i^s \in \mathbb{R}^t$, m 为节点个数, GAT 会产生一个包含结构信息的节点特征表示, 其输出为 $H^{s'} = \{h_1^{s'}, h_2^{s'}, \dots, h_m^{s'}\}$, $h_i^{s'} \in \mathbb{R}^t$ 。

为了保留一定程度的原始信息, 将 H^s 与 $H^{s'}$ 对应元素相加, 得到用户信息矩阵 H^{add} 。

$$H^{\text{add}} = H^s \oplus H^{s'} \quad (2)$$

将 H^{add} 铺平, 得到用户信息的最终表征 $V^{\text{add}} \in \mathbb{R}^{m \times t}$, 整个用户信息编码器如图3所示。

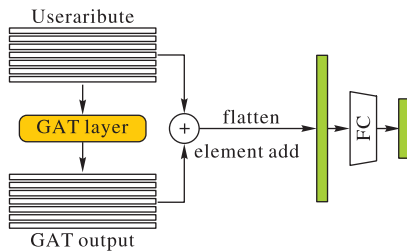


图3 用户信息编码器

Fig. 3 User information encoder

3.2 结构信息编码器

使用 node2vec^[16] 算法可在不受节点属性干扰的情况下获取转发图结构信息, 将 node2vec 得到的节点表征作为互注意机制的输入, 以增强结构表征。

首先, 使用 node2vec 算法得到转发图中各个节点的表征, 其目标函数为:

$$\max_f \sum_{u \in \mathcal{V}} \left[-\log Z_u + \sum_{n_i \in N_f(u)} f(n_i) \cdot f(u) \right] \quad (3)$$

转发图中 u_i 的 node2vec 表征为 $p_i \in \mathbb{R}^d$, 然后, 计算用户之间的相互注意分数, 类似于文献[17]中的方法, 如图4所示, 使用全连接网络得到两两用户表征之间的注意力分数 M_{ij}^A :

$$M_{ij}^A = W_{\text{cat}}([p_i || p_j]) + b_{\text{cat}} \quad (4)$$

其中: 权重矩阵 $W_{\text{cat}} \in \mathbb{R}^{2d}$; 偏置项 $b_{\text{cat}} \in \mathbb{R}$; $||$ 为拼接操作, 依次计算用户之间的注意力分数, 得到相互注意分数矩阵 $M^A \in \mathbb{R}^{m \times m}$, m 为转发网络中的节点数。

使用 softmax 函数和行最大池化对互注意分数矩阵进行运算, 得到节点的互注意分数向量 M_s^A :

$$M_s^A = \text{Softmax}(\max_{\text{row}}(M^A)) \quad (5)$$

$M_s^A \in \mathbb{R}^m$ 的每个值表示各个节点的注意力权重, 而互注意表征 $M_R^A \in \mathbb{R}^d$ 由式(6)计算:

$$M_R^A = \sum_{i=1}^m (p_i \cdot (M_s^A)_i) \quad (6)$$

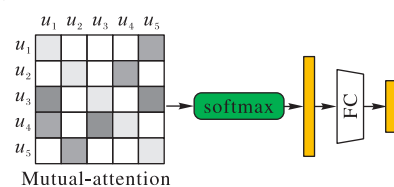


图4 结构信息编码器

Fig. 4 Structural information encoder

3.3 源文本编码器

源文本包含丰富的信息,有助于增强模型表现。为了最大限度地提取文本上下文信息,引入BERT^[18]算法获取源文本表征,该算法通过双向Transformer编码器^[19]生成的字向量,可以

完全融合字词左右上下文信息,与CNN或RNN相比,BERT可以更加充分地表征词语的多重含义,由图5所示,BERT生成的词向量由字符向量、句向量和位置向量组成。输入为新闻 C ,输出源文本表征 $X_{\text{output}} \in \mathbb{R}^r$,代表源文本的全部语义信息。

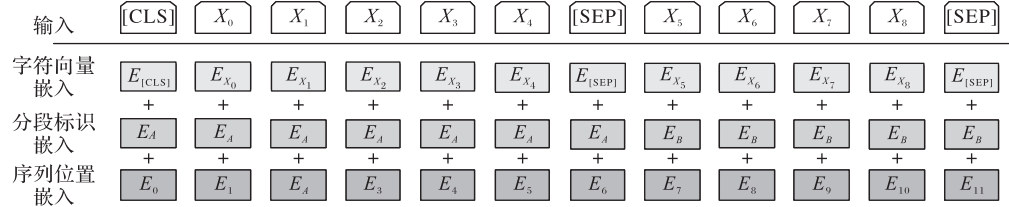


图5 BERT模型的输入表征

Fig. 5 Input representation of BERT model

3.4 融合分类器

以上三部分计算得到的三个表征分别表示传播模式中的用户属性、结构信息和源帖内容信息,要对这三个表征进行融合,使得最终表征可以最大限度地捕捉到谣言传播的多方面特征。此处通过多模态门控单元^[20]对上述三个表征进行融合,在不同来源数据组合基础上找到最优表示。

由于用户属性表征 $V^{\text{add}} \in \mathbb{R}^{m \times t}$,结构表征 $M_R^A \in \mathbb{R}^d$ 以及源帖内容表征 $X_{\text{output}} \in \mathbb{R}^r$ 的维度不同,首先定义对应权重矩阵 $W_{V^{\text{add}}}^{\Phi} \in \mathbb{R}^{g \times (m \times t)}$, $W_{M_R^A}^{\Phi} \in \mathbb{R}^{g \times d}$, $W_{X_{\text{output}}}^{\Phi} \in \mathbb{R}^{g \times r}$,将三个不同维度的表征投影到目标维度 g ,其过程为:

$$V^{\text{add}'} = W_{V^{\text{add}}}^{\Phi} \cdot V^{\text{add}} \quad (7)$$

$$M_R^{A'} = W_{M_R^A}^{\Phi} \cdot M_R^A \quad (8)$$

$$X'_{\text{output}} = W_{X_{\text{output}}}^{\Phi} \cdot X_{\text{output}} \quad (9)$$

接下来,将 $V^{\text{add}'}$ 、 $M_R^{A'}$ 和 X'_{output} 输入到GMU多模态门控单元中:

$$h_{\text{add}} = \tanh(W_{\text{add}} \cdot (V^{\text{add}'}))^T \quad (10)$$

$$h_{\text{mutual}} = \tanh(W_{\text{mutual}} \cdot (M_R^{A'})^T) \quad (11)$$

$$h_X = \tanh(W_X \cdot (X'_{\text{output}}))^T \quad (12)$$

$$z_1 = \sigma(W_{z_1} [V^{\text{add}'} \| M_R^{A'}]^T) \quad (13)$$

$$z_2 = \sigma(W_{z_2} [V^{\text{add}'} \| X'_{\text{output}}]^T) \quad (14)$$

$$z_3 = \sigma(W_{z_3} [M_R^{A'} \| X'_{\text{output}}]^T) \quad (15)$$

$$h_1 = z_1 h_{\text{add}} + (1 - z_1) h_{\text{mutual}} \quad (16)$$

$$h_2 = z_2 h_{\text{add}} + (1 - z_2) h_X \quad (17)$$

$$h_3 = z_3 h_{\text{mutual}} + (1 - z_3) h_X \quad (18)$$

h_1, h_2 和 h_3 的维度都为 g ,将三个表征拼接起来:

$$h = \text{concat}[h_1, h_2, h_3] \quad (19)$$

则 $h \in \mathbb{R}^{3g}$ 为模型的最终表征,将该向量输入到全连接网络(FC)和分类函数中,得到最终分类结果。

$$\mathcal{R} = \text{FC}(W_p h + b_p) \quad (20)$$

$$\hat{y} = \text{Softmax}(W_{cl} \mathcal{R} + b_{cl}) \quad (21)$$

3.5 优化器

在本文中,模型训练目标为最小化交叉熵损失函数。

$$L = - \sum_{C \in NS} (y_C \log(y_C) + (1 - y_C) \log(1 - \hat{y}_C)) \quad (22)$$

其中: L 为损失函数, NS 为新闻集合, y_C 为新闻 C 的真实标签, \hat{y}_C 为新闻 C 的预测标签。

4 实验与分析

为了评估本文模型在真实社交场景下的检测性能,在真实世界的数据集进行对比实验。

4.1 实验数据

Weibo数据集^[21]来源于新浪微博。在该数据集中,由帖子、转发和用户构成,有两个标签: $\{\text{谣言}, \text{非谣言}\}$,具体介绍如表1所示。

表1 实验数据集统计

Tab. 1 Experimental dataset statistics

统计信息	数量
事件数量	4 664
真实事件数量	2 351
虚假事件数量	2 313
用户数量	2 746 818
帖子数	3 805 656

4.2 实验设计

在实验中,使用5折交叉验证。其中node2vec训练的节点表征维数 $d = 256$,BERT训练的文本信息表征维数 $r = 768$,将三种信息表征投影到目标维度 $g = 256$ 。

为了减少模型的运算量,截取了前200名传播用户,以此构建转发图。若传播网络中用户数不足200,则根据实际状况构建转发图,用户信息最终表征 V^{add} 维数固定为1 800,若不足,则填充0值。在实验中,epoch设置为60,batch size设置为128,学习率为0.000 1。

在用户信息编码器部分,使用用户常见的9个特征作为转发图节点的特征, $t = 9$,如表2所示。

表2 节点特征

Tab. 2 Node features

特征	描述
reposts_count	帖子的转发数
bi_followers_count	互相关注的数量
friends_count	关注数
followers_count	粉丝数
statuses_count	发表帖子数
verified	是否验证
favourits_count	最喜欢的帖子数
comments_count	评论数
t	用户转发时间戳

在用户信息编码器部分,所提出方法具备用户排序不变性,即用户信息编码器的输出与用户排列顺序无关。图注意力网络 GAT 作为图神经网络算法,节点表征更新依赖邻接矩阵,满足平移不变性,与节点的顺序无关。将用户信息矩阵 H^{add} 铺平后得到 V^{add} ,目的是能够进入全连接层,填充 0 值的作用是补齐向量,使得进入全连接层时规格统一。即使铺平顺序发生变化,全连接层可以灵活地更新权重矩阵,用户铺平顺序对最终结果不会造成影响。综上所述,将 H^{add} 铺平不会对用户排序不变性造成影响。

4.3 对比方法

为了验证本文模型的优越性,与以下 6 个模型进行比较:

1)DTC(Decision-Tree Classifier)^[1]:该模型使用文本的手工特征和其他统计特征,利用决策树模型来得到检测结果。

2)SVM-RBF(SVM classifier with RBF kernel function)^[22]:该模型使用基于径向基核函数(Radial Basis Function, RBF)的 SVM 模型得到新闻的预测标签。

3)TD-RvNN(Tree-structured Recursive Neural Network)^[4]:该模型使用递归神经网络处理树结构的新闻传播路径得到分类结果。

4)PPC-RNN+CNN(Propagation Path Classification with RNN and CNN)^[3]:模型提出了新闻传播的多元时间序列,使用 GRU 和 CNN 的结合模型进行分类。

5)HiMap-HO+Text(Higher-order user to user Mutual-Attention Progression method)^[17]:作者使用 LSTM 计算转发链的节点表征,并结合用户的互注意表征得到新闻的检测结果。

6)BiGCN(Bi-directional GCN)^[14]:利用信息传播时的双向传播结构使用图卷积网络进行谣言检测的模型。

4.4 实验结果

实验结果如表 3 所示,显示了本文模型和比较模型在 Weibo 数据集上的表现。首先,深度学习算法的表现远超机器学习算法,表明深度学习算法学习到的表征优于人工选择的谣言特征;其次,GMB_GMU 模型在各性能指标上明显优于 TD-RvNN 方法,因为 TD-RvNN 节点特征为转发文本,但真实场景下存在很多空转发,而本文模型选择使用用户属性作为节点特征,弥补了转发结构上信息的缺失。最后,GMB_GMU 模型优于 HiMap-HO+Text,在 HiMap-HO+Text 中,使用转发链计算节点二到四阶的相互注意力,数据收集和处理有极大的难度,而本文模型仅使用转发图中的一阶用户相互注意力表征,实验表现远超 HiMap-HO+Text。

表 3 Weibo 数据集上的实验结果

Tab. 3 Experimental results on Weibo dataset

方法	准确率		精确率		召回率		F1 值	
	谣言	非谣言	谣言	非谣言	谣言	非谣言	谣言	非谣言
DTC	0.831	0.815	0.847	0.825	0.815	0.819	0.831	0.831
SVM-RBF	0.879	0.579	0.777	0.708	0.656	0.615	0.708	0.708
TD-RvNN	0.908	0.904	0.912	0.918	0.897	0.911	0.905	0.905
PPC_RNN+CNN	0.913	0.927	0.884	0.901	0.932	0.907	0.922	0.922
HiMap-HO+Text	0.892	0.892	0.892	0.878	0.910	0.896	0.888	0.888
BiGCN	0.905	0.919	0.894	0.895	0.916	0.900	0.898	0.898
GMB_GMU	0.952	0.934	0.968	0.967	0.939	0.950	0.953	0.953

4.5 早期检测实验

为了在谣言传播的早期检测到谣言,及时遏制其传播,分

别测试了模型在 5 min、10 min 等时间点的谣言检测表现,其实验结果如图 6 所示。

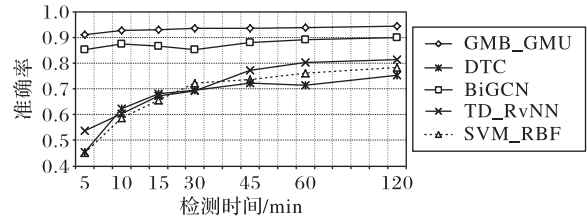


图 6 早期检测实验结果

Fig. 6 Experimental results of early detection

在源帖发布的 5 min 内,GMB_GMU 模型的准确率超过 0.9,表现优于其他比较模型,表明仅存在较少用户信息时,GMB_GMU 仍能保持较为准确的早期检测能力。当传播时间增大,转发结构逐渐复杂,信息噪声增多时,不同于 BiGCN 等比较模型,GMB_GMU 模型性能呈稳定上升趋势,体现了模型处理复杂数据的能力,具有较好的稳定性和鲁棒性。

4.6 消融实验

为了验证模型中各部分的作用,将模型可拆分为以下子模型:1)用户属性 GAT;2)源帖 BERT 网络;3)GMB 网络,即基于用户属性 GAT-互注意机制-BERT 的简单融合网络;4)互注意机制。实验结果如图 7 所示。

用户属性 GAT 可以取得 0.865 的准确率,结果远低于 GMB_GMU,表明仅使用用户属性,不考虑内容语义信息,会严重影响谣言检测效果。使用源帖 BERT 网络取得了 0.914 的准确率,结果仍低于 GMB_GMU,仅仅依靠源帖语义信息来检测谣言存在的问题。最后,基于用户属性 GAT-互注意机制-BERT 的简单融合网络将 GMU 替换为简单拼接方式,达到 0.937 的准确率,说明门控多模态单元可以对多个模态信息进行选择组合,更好地表征联合信息。

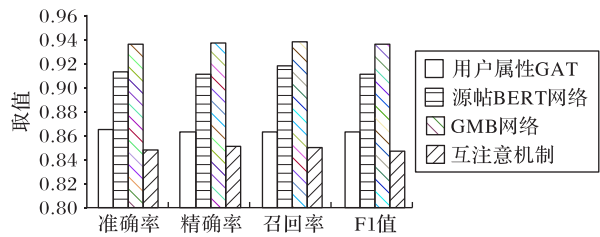


图 7 GMB_GMU 的消融实验结果

Fig. 7 Ablation experimental results of GMB_GMU

5 结语

本文提出了一种基于用户传播网络与消息内容融合的谣言检测模型 GMB_GMU,引入 GAT 增强用户属性表征。在转发图的基础上计算节点的 node2vec 表征,并通过互注意机制增强结构信息;另外,引入源帖文本的 BERT 向量,增强源帖信息。最后,使用 GMU 门控网络将三种角度的谣言表征恰当地结合起来,提取到消息的最终表征,在公共数据集上取得了优异的实验结果,并通过早期检测和消融实验说明了模型在谣言早期传播的有效性和每个部分的重要程度。在谣言的传播过程中,一些有影响力的用户会导致谣言扩散呈指数级增长,加重了谣言负面影响。在未来工作中,通过发现这些有影响力的用户,进而提升谣言检测模型的性能。

参考文献 (References)

- [1] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter[C]// Proceedings of the 20th International Conference on World Wide Web. New York: ACM, 2011:675-684.
- [2] QIAN F, GONG C Y, SHARMA K, et al. Neural user response generator: fake news detection with collective user intelligence[C]// Proceedings of the 27th International Joint Conference on Artificial Intelligence. [S.l.]: IJCAI Organization, 2018: 3834-3840.
- [3] LIU Y, WU Y F B. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks [C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018:354-361.
- [4] MA J, GAO W, WONG K F. Rumor detection on Twitter with tree-structured recursive neural networks [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Stroudsburg, PA: Association for Computational Linguistics, 2018: 1980-1989.
- [5] RASHKIN H, CHOI E, JANG J Y, et al. Truth of varying shades: analyzing language in fake news and political fact-checking [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2017:2931-2937.
- [6] 段大高,谢永恒,盖新新,等. 基于神经网络的微博虚假消息识别模型[J]. 信息安全, 2017(9): 134-137. (DUAN D G, XIE Y H, GAI X X, et al. A rumor detection modal based on neural network[J]. Netinfo Security, 2017(9):134-147.)
- [7] 刘政,卫志华,张韧弦. 基于卷积神经网络的谣言检测[J]. 计算机应用, 2017, 37(11): 3053-3056, 3100. (LIU Z, WEI Z H, ZHANG R X. Rumor detection based on convolutional neural network [J]. Journal of Computer Applications, 2017, 37(11): 3053-3056, 3100.)
- [8] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2013:3111-3119.
- [9] FANG Y, GAO J, HUANG C, et al. Self multi-head attention-based convolutional neural networks for fake news detection [J]. PLoS ONE, 2019, 14(9): No. e0222713.
- [10] VAIBHAV V, ANNASAMY R M, HOVY E. Do sentence interactions matter? leveraging sentence level representations for fake news classification [C]// Proceedings of the 13th Workshop on Graph-Based Methods for Natural Language Processing. Stroudsburg, PA: Association for Computational Linguistics, 2019: 134-139.
- [11] WANG Y H, WANG L, YANG Y J, et al. SemSeq4FD: integrating global semantic relationship and local sequential order to enhance text representation for fake news detection[J]. Expert Systems with Applications, 2020, 166: No. 114090.
- [12] LIU Y H, JIN X L, SHEN H W, et al. Do rumors diffuse differently from non-rumors? a systematically empirical analysis in Sina Weibo for rumor identification [C]// Proceedings of the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNCS 10234. Cham: Springer, 2017: 407-420.
- [13] SHU K, MAHDESWARAN D, WANG S H, et al. Hierarchical propagation networks for fake news detection: investigation and exploitation [C]// Proceedings of the 14th International AAAI Conference on Web and Social Media. Palo Alto, CA: AAAI Press, 2020: 626-637.
- [14] BIAN T, XIAO X, XU T Y, et al. Rumor detection on social media with bi-directional graph convolutional networks [C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2020: 549-556.
- [15] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks [EB/OL]. (2018-02-04) [2020-10-10]. <https://arxiv.org/pdf/1710.10903.pdf>.
- [16] GROVER A, LESKOVEC J. node2vec: scalable feature learning for networks [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855-864.
- [17] MISHRA R. Fake news detection using higher-order user to user mutual-attention progression in propagation paths [C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 2775-2783.
- [18] DAVLIN 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, Volume 1 (Long and Short Papers). Stroudsburg, PA: Association for Computational Linguistics, 2019: 4171-4186.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017:6000-6010.
- [20] AREVALO J, SOLORIO T, MONTES-Y-GOMÓZ M, et al. Gated multimodal networks [J]. Neural Computing and Applications, 2020, 32(14): 10209-10228
- [21] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks [C]// Proceedings of the 25th International Joint Conference on Artificial Intelligence. [S.l.]: IJCAI Organization, 2016: 3818-3824.
- [22] YANG F, LIU Y, YU X H, et al. Automatic detection of rumor on Sina Weibo [C]// Proceedings of the 18th ACM SIGKDD Workshop on Mining Data Semantics. New York: ACM, 2012: No. 13.

This work is partially supported by the National Natural Science Foundation of China (61872260).

XUE Haitao, born in 1997, M. S. candidate. His research interests include natural language processing, rumor detection.

WANG Li, born in 1971, Ph. D., professor. Her research interests include big data computation and analysis, data mining.

YANG Yanjie, born in 1995, M. S. candidate. His research interests include natural language processing, data mining.

LIAN Biao, born in 1987, M. S. His research interests include software development, data mining.