·网络空间安全· 文章编号：1000-3428(2022)10-0138-08 文献标志码：A 中图分类号：TP309

基于多传递影响力的社交媒体谣言检测方法

段大高[1,2]，白宸宇[1,2]，韩忠明[1,2]，熊海涛[1]

（1.北京工商大学 国际经管学院，北京 100048；2.北京工商大学 食品安全大数据技术北京重点实验室，北京 100048）

摘 要：社交媒体谣言检测是当前研究的热点问题，现有方法多数通过获取大量用户属性学习用户特征，但不适用于谣言的早期检测，忽略了用户之间的潜在关系对信息传播的影响。提出一种基于多传递影响力的谣言检测方法，根据源微博及其对应转发(评论)之间的关系构建文本信息传播图，并通过图卷积神经网络来捕获、学习文本信息的传播特征。利用文本信息和用户传播过程中的影响力，丰富可用于谣言检测早期的检测信息。将存在转发关系的用户构成用户影响力传播图，构建一种用户节点影响力学习方法，获取用户节点影响力，以增强用户特征信息。在此基础上，将文本特征与用户特征融合以进行谣言检测，从而提升检测效果。在3个真实社交媒体数据集上的实验结果表明，该方法在谣言自动检测以及早期检测的效果都有显著提升，与目前最好的基准方法相比，在微博、Twitter15、Twitter16数据集上的正确率分别提高了2.8%、6.9%和3.4%。

关键词：谣言检测；传递影响力；图卷积神经网络；信息传播；社交媒体

开放科学(资源服务)标志码(OSID)：

中文引用格式：段大高，白宸宇，韩忠明，等.基于多传递影响力的社交媒体谣言检测方法[J].计算机工程，2022，48(10)：138-145，157.

英文引用格式：DUAN D G，BAI C Y，HAN Z M，et al.Social media trumor detection method based on multi-transmit influence[J].Computer Engineering，2022，48(10)：138-145，157.

Social Media Rumor Detection Method Based on Multi-Transmit Influence

DUAN Dagao[1,2]，BAI Chenyu[1,2]，HAN Zhongming[1,2]，XIONG Haitao[1]

（1.School of International Economics and Management，Beijing Technology and Business University，Beijing 100048，China；
2.Beijing Key Lab of Big Data Technology for Food Safety，Beijing Technology and Business University，Beijing 100048，China）

【Abstract】Social media rumor detection is a hot topic in current research.Existing methods learn user characteristics by acquiring many user attributes. However，they ignore the influence of potential relationships between users on information propagation，making them inapt for early detection of rumors. This paper proposes a Multi-Transmit Influence(MTI) model for social media rumor detection.The forwarding relationship between the source microblog and its corresponding forwards(comments) is used to construct a text information propagation graph.A graph convolution neural network is used to capture and learn the propagation characteristics of text information.In addition，the influence of users in the communication process is integrated to enrich information detection for early rumor detection.First，the users with forwarding relationships are formed into a user influence propagation graph.A user-node influence learning method is then constructed to capture the user-node influence thatenhances the user characteristic information.Finally，text and user features are fused to detect rumors more accurately. Experiments on three real social media data sets revealthat the proposed method significantly improved automatic and early rumor detection. Compared with the conventional benchmark methods，the accuracy improved by 2.8%，6.9%，and 3.4% when tested using Weibo，Twitter15，and Twitter16 data sets，respectively.

【Key words】rumor detection；transmit influence；graph convolution neural network；information propagation；social media

DOI：10. 19678/j. issn. 1000-3428. 0061592

0 概述

在现实世界中，社交网络已经与人们日常工作和生活密不可分，人们不仅通过网络获取各种信息，同时也参与到信息内容的创作中。社交网络中的信息传播具有速度快、范围广、即时性强等特点。然

基金项目：国家重点研发计划(2019YFC0507800)；北京市自然科学基金(4172016)；北京市教委科研计划一般项目(KM201710011006)。

作者简介：段大高(1976—)，男，副教授、博士，主研方向为数据挖掘、自然语言处理；白宸宇，硕士研究生；韩忠明，教授、博士；熊海涛，副教授、博士。

收稿日期：2021-05-10 **修回日期：**2021-09-30 **E-mail：**duandg@th.btbu.edu.cn

而，由于在发布信息时缺乏有效监管手段，导致社交网络平台成为谣言传播的温床[1]。网络谣言不仅会影响人们的日常生活，而且会带来严重的社会问题。例如，2016年美国大选期间，有益的谣言信息倾向于支持唐纳德·特朗普而非希拉里·克林顿，直接影响选举结果[2]。因此，研究自动高效的谣言检测方法意义重大，尤其是在信息传播早期阶段。

传统检测方法主要利用文本内容、用户特征通过手工提取特征，然后再利用分类器分类，如决策树[3]、随机森林[4]、支持向量机[5]。随着近年来深度学习的发展，越来越多的研究采用深度学习方法。除上述内容特征外，谣言的传播还存在结构特性，传播图中的节点会因为邻居及更远邻居而影响自己，关系越亲近的邻居影响更大，因此转发关系的谣言之间存在结构影响力，这将有助于对谣言的分类。虽然现有研究已经取得了部分成就，但是鉴于社交媒体下谣言检测任务的复杂性，其还存在以下问题：谣言文本包含了语义信息和传播结构信息，以往方法利用树结构学习结构影响力并不完善，谣言传播结构应是一个错综复杂的图结构；用户属性可以丰富谣言检测特征，但在传播早期很难获取大量用户信息，因此无法利用用户的关注关系描绘传播网络，但可以通过早期谣言的转发关系构建用户传播图；消息在传播过程中会受到不同用户的影响，以往方法忽略了未直接转发或评论用户存在的间接影响，而这些潜在关系可以丰富谣言检测特征。

本文提出一种基于多传递影响力（Multi-Transmit Influence，MTI）的谣言检测方法。使用转发关系对用户节点构图，根据图神经网络学习文本的结构影响力，以避免使用大规模用户信息，在此基础上通过构造基于用户传递影响力的节点表示，学习用户之间在传播过程中不同的影响力，以增强用户特征信息。

1　研究现状

目前，研究人员将谣言检测任务看作是一种分类，即判断某个消息是“虚假信息”还是“非虚假信息”，亦或是其他类别。其中一类方法为基于传统机器学习的方法，例如，文献[6]通过提取单词或短语的频率特征，选出对谣言或者是非谣言比较有代表性的词进行谣言检测。文献[7]首先按照主题分类提取用户特征，然后利用机器学习的方法进行分类。文献[8]使用了多种不同类型的特征，并通过梯度提升决策树来进行谣言检测。文献[9]提出一种基于动态时间序列的谣言检测模型，利用时间序列为谣言的社会情境特征变化进行建模，在传统机器学习中取得了较好的效果。

随着深度学习技术的快速发展，许多研究人员尝试利用深度学习来解决文本分类问题[10]。文献[11]应用递归神经网络，通过学习传播序列中的信息进行谣言检测。研究人员通过引入注意力机制的模型[12-14]和利用对抗生成网络的模型[15]都取得了一定的效果。文献[16]通过递归神经网络对谣言信息以树结构的形式，捕获自上而下和自下而上的结构信息，但是树结构学习结构影响力并不完善，谣言传播结构应是一个错综复杂的图网络，因此丢失了一些结构信息。文献[17]则分别使用了循环神经网络和卷积神经网络学习传播路径上的不同信息，但未考虑到用户间的影响力。

2　基于多传递影响力的谣言检测方法

2.1　模型整体架构

基于多传递影响力的谣言检测方法整体架构如图1所示。

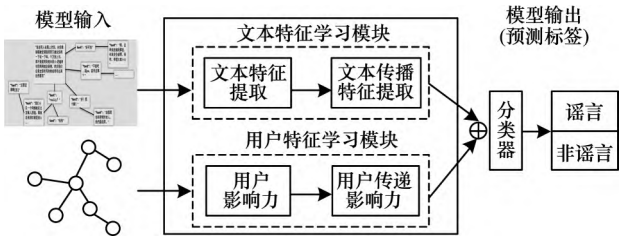

图1　谣言检测模型整体架构

Fig.1　Overall architecture of the rumor detection model

模型由文本特征学习模块和用户特征学习模块两部分构成。其中文本特征学习模块包括：1）文本特征提取，首先利用Word2Vec[18]获取词向量，将微博句子表示为一个微博词特征矩阵，再利用多头注意力机制和卷积神经网络作用于微博词特征矩阵得到微博句子特征；2）文本传播特征提取，首先构建微博文本之间的转发或评论关系图，再利用图卷积神经网络获取传播特征。用户特征学习模块包括：1）用户影响力计算，利用转发关系将微博对应的用户进行构图，获取用户向量表示，再引入注意力机制获取用户影响力；2）用户传递影响力计算，通过构造基于用户传递影响力的节点表示方法，学习用户之间在传播过程中的不同影响力。将更新后的文本特征和用户特征融合，并由分类器进行分类输出，来预测微博信息的类别。

2.2　文本特征学习模块

源谣言集合用$X=\{X_1,X_2,\cdots,X_n\}$表示，每条源谣言相关的信息用$X_i=\left\{r_i,v_1^i,v_2^i,\cdots,v_{n-1}^i\right\}$表示，其中$r$为源微博，$v$对应不同的转发。$X_i$中每条信息$v_i$包含若干词，用$v_i=\{\mathrm{Word}_1,\mathrm{Word}_2,\cdots,\mathrm{Word}_L\}$表示，其中$\mathrm{Word}_L$表示微博分词后的词组，$L$表示微博分词的长度。利用Word2Vec获取词嵌入表示，再用微博词特征矩阵$v_i\in\mathbb{R}^{L\times d}$表示每个微博句子，微博词特征矩阵如图2所示。

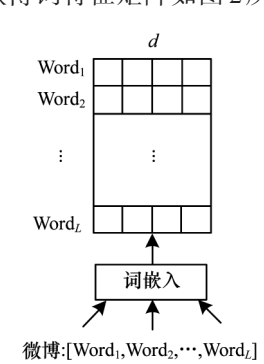

图2　微博词特征矩阵

Fig.2　Weibo word feature matrix

2.2.1 文本特征提取

在谣言检测问题中，文本信息十分重要，本文模型中文本特征提取过程如图3所示。

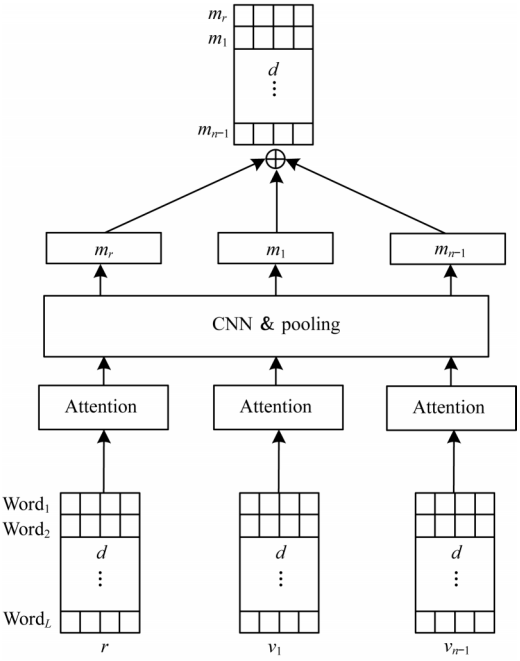

图3 文本特征提取过程

Fig.3 Text feature extraction process

本文将微博词特征矩阵作为输入，通过多头自注意力机制更新该矩阵，把更新后的微博词特征矩阵输入到卷积神经网络层和池化层提取特征，得到每条微博的句子特征，最后将不同句子特征拼接，得到源微博及相关微博的特征矩阵。

1）多头自注意力机制。在多头自注意力机制计算过程中，使句子中所有词相互影响，提取内部相关特性，获取词间依赖关系。

多头自主意力机制过程如图4所示，输入$\boldsymbol{Q}$=$\boldsymbol{K}$=$\boldsymbol{V}$，即微博句子的词特征矩阵。

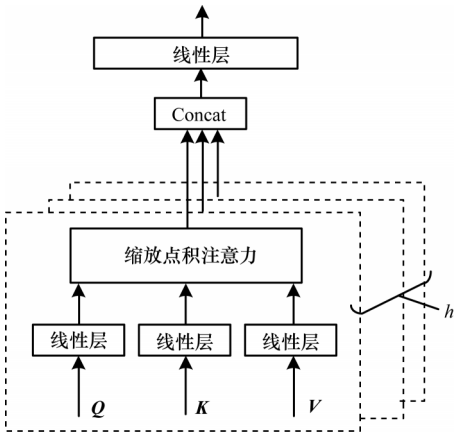

图4 多头自注意力机制过程

Fig.4 Multi-head self-attention mechanism process

线性层将$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$映射为h个不同部分，各部分进行缩放点击注意力，计算公式如式（1）所示，得到输出如式（2）所示：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V} \tag{1}$$

$$Z_i=\text{Attention}(W_i^{\boldsymbol{Q}},W_i^{\boldsymbol{K}},W_i^{\boldsymbol{V}}) \tag{2}$$

其中：$i\in[1,h]$；d表示词嵌入维度。将不同部分的结果进行拼接，再通过一个线性层得到输出，如式（3）所示：

$$\boldsymbol{Z}=\boldsymbol{W}_0\left(\text{Concat}\left(Z_1,Z_2,\cdots,Z_h\right)\right) \tag{3}$$

其中：$\boldsymbol{W}_0\in\mathbb{R}^{d\times d}$为权重矩阵，输出内容$\boldsymbol{Z}$的维度与$\boldsymbol{Q}$相同。

2）卷积池化层。通过卷积层和最大池化层捕获微博句子级的特征。将$\boldsymbol{Z}\in\mathbb{R}^{L\times d}$作为输入，文本特征卷积核为$\boldsymbol{W}\in\mathbb{R}^{h\times d}$，其中$h$表示卷积核感受野的大小，作用公式为：

$$e_i=\sigma\left(\boldsymbol{W}\times\text{Word}_{i:i+h-1}^{(\cdot)}+b\right) \tag{4}$$

其中：σ为非线性激活函数；$\boldsymbol{x}_i^{(\cdot)}$为单词的词向量表示；$b$为偏置量。经过卷积层提取得到卷积层特征$\boldsymbol{e}$，$\boldsymbol{e}=\left[e_1,e_2,\cdots,e_{L-h+1}\right]\in\mathbb{R}^{L-h+1}$。将卷积层特征输入最大池化层，对$\boldsymbol{e}\in\mathbb{R}^{(L-h+1)\times d}$进行最大池化，如式（5）所示：

$$\hat{\boldsymbol{e}}=\max\left(\boldsymbol{e}\right) \tag{5}$$

在卷积层中设置不同大小的卷积核，每种卷积核的数量为d/3。将不同卷积核对应的输出连接起来得到$m_i\in\mathbb{R}^d$，表示源微博或其转发微博的句子级特征，进而获取源微博及相关微博特征矩阵$\boldsymbol{M}=[m_r,m_1,m_2,\cdots,m_{n-1}]\in\mathbb{R}^{n\times d}$。

2.2.2 文本传播特征提取

对于谣言事件相关的信息X_i，用$G_i=\left\{E_i,V_i\right\}$表示其传播图结构。如图5所示，图中节点集合为$V_i=\left\{r_i,v_1^i,v_2^i,\cdots,v_{n_i-1}^i\right\}$，$r_i$表示源微博，边集合$E_i=\{e_{st}^i|\,s,t=0,1,0,\cdots,n_i-1\}$，其中每一条$e_{st}^i$就表示两条微博间存在着一个传播行为，用邻接矩阵$\boldsymbol{A}\in\mathbb{R}^{n\times n}$表示，邻接矩阵中对应位置元素为$a_{ij}$，微博之间存在转发或评论关系则为1，否则为0，对应关系如式（6）所示：

$$a_{ij}=\begin{cases}1,a_{ij}\in E\\0,\text{其他}\end{cases} \tag{6}$$

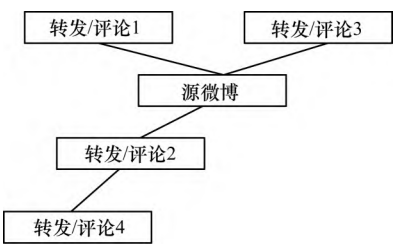

图5 微博文本传播图

Fig.5 Weibo text spread graph

利用图卷积神经网络[19]学习传播特征，将微博特征矩阵$\boldsymbol{M}$和文本传播图邻接矩阵$\boldsymbol{A}$作为输入，计算公式如式(7)所示：

$$\mathrm{Conv}_x=\sigma\left(\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{M}\boldsymbol{\theta}\right) \tag{7}$$

其中：σ为非线性激活函数；$\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{A}}\hat{\boldsymbol{D}}^{-\frac{1}{2}}$表示归一化的拉普拉斯矩阵；$\hat{\boldsymbol{A}}$为添加了自环的邻接矩阵，$\hat{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}$，$\boldsymbol{A}\in\mathbb{R}^{n\times n}$表示传播图邻接矩阵，$\boldsymbol{I}$表示单位矩阵，$\hat{\boldsymbol{D}}$为$\hat{\boldsymbol{A}}$的度矩阵，其中$\hat{D}_{ii}=\sum_j\hat{A}_{ij}$；$\boldsymbol{\theta}$表示可学习的参数矩阵。

2.3　用户特征学习模块

源谣言相关的信息用$X_i=\left\{r_i,v_1^i,v_2^i,\cdots,v_{n-1}^i\right\}$表示，$r$表示源微博，$v$对应不同的转发，每条微博对应用户使用集合$U_i=\{u_r^i,u_1^i,u_2^i,\cdots,u_{n-1}^i\}$表示。在用户特征学习模块中构造一种基于用户传递影响力的节点表示方式，传播过程的相互影响构成了用户结构影响力。使用注意力机制模拟用户间存在的影响关系，这种影响关系主要分为用户影响力和用户传递影响力两部分。

如图6所示，图中节点u_0~u_6表示转发图中的用户，实线连接的节点表示直接转发，存在直接影响力，虚线连接则表示未直接转发，存在间接影响力。例如，图中u_0节点与u_5节点、u_5节点与u_6节点都存在直接转发关系，u_5节点自身存在如粉丝数量等特征，会对u_0节点造成影响力，用S_{05}表示影响关系，同理u_6节点也会影响u_5节点，用S_{56}表示。传递影响力是为了学习未直接转发或评论的用户所造成的间接影响，即用户传递影响力。在图6中，节点u_5作为中间节点，通过S_{05}与S_{56}计算得到u_6对u_0的传递影响力S_{06}。

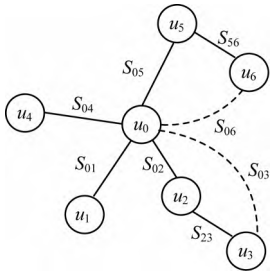

图6　用户传播图

Fig.6　User communication graph

2.3.1　用户影响力

在转发序列构成的用户传播图中，从该网络中学习得到序列中所有用户的嵌入表示：$\vec{\boldsymbol{u}}_i,\vec{\boldsymbol{u}}_i\in\mathbb{R}^d$。在得到用户嵌入后，首先学习相邻用户节点间的潜在注意力系数，通过一个全连接层将两个节点的用户嵌入信息转化为一个标量s_{ij}，如式(8)所示：

$$s_{ij}=\mathrm{LeakyReLU}(\vec{\boldsymbol{a}}^{\mathrm{T}}\left[\boldsymbol{W}\vec{\boldsymbol{u}}_i\middle\|\boldsymbol{W}\vec{\boldsymbol{u}}_j\right]) \tag{8}$$

其中：$\vec{\boldsymbol{u}}_i$与$\vec{\boldsymbol{u}}_j$分别是用户节点u_i与u_j对应的向量表示；$\boldsymbol{W}$是一个可学习的参数矩阵；$\vec{\boldsymbol{a}}^{\mathrm{T}}\in\mathbb{R}^{2d\times1}$表示注意力向量。通过$s_{ij}$的取值可以表示两个用户节点$u_i$与$u_j$之间潜在的相关性系数，构造一个相关性矩阵$\boldsymbol{M}\in\mathbb{R}^{n\times n}$，该矩阵中$n$为传播图中用户节点个数，对应位置元素为两用户间的相关性系数s_{ij}。将直接转发节点的相关性系数归一化并聚合节点的信息，如式(9)、式(10)所示：

$$e_{ij}=\mathrm{softmax}\left(s_{ij}\right)=\frac{\exp\left(s_{ij}\right)}{\sum_{k\in N_i}\exp\left(s_{ik}\right)} \tag{9}$$

$$\vec{\boldsymbol{u}}_i=\sigma\left(\sum_{j\in N_i}e_{ij}\boldsymbol{W}\vec{\boldsymbol{u}}_j\right) \tag{10}$$

其中：N_i表示转发关系中与i相连的节点；$\boldsymbol{W}$为可学习的参数矩阵；σ为激活函数，归一化相关性系数能有效反映不同节点对目标节点的影响力度。

2.3.2　用户传递影响力

上文计算只考虑了在转发关系路径中直接转发用户的影响关系，而社交网络十分复杂，在一个真实的社交网络转发序列中，未直接转发的用户(其他用户作为中间节点，间接转发)之间存在一种传递影响力，即存在一种多跳的潜在关系，这种影响力是用户信息中很重要的一部分。

通过相关性矩阵$\boldsymbol{M}$，构造传递影响力矩阵$\boldsymbol{M}'\in\mathbb{R}^{n\times n}=\boldsymbol{M}\times\boldsymbol{M}$，$n$为用户传播图中节点个数，矩阵中元素计算公式如式(11)所示：

$$s'_{ij}=\sum_C s_{ic}\times s_{cj} \tag{11}$$

其中：c表示节点i和j的中间节点；s_{ic}表示节点i和c间的用户影响力；s_{cj}表示节点c和j之间的用户影响力。$\boldsymbol{M}'$矩阵中对应位置元素s'_{ij}表示间接转发影响力系数。将间接转发节点的系数归一化并聚合节点的信息，如式(12)和式(13)所示：

$$e'_{ij}=\mathrm{softmax}\left(s'_{ij}\right) \tag{12}$$

$$\vec{\boldsymbol{u}}'_i=\sigma\left(\sum_{j\in N_i}e'_{ij}\boldsymbol{W}\vec{\boldsymbol{u}}_j\right) \tag{13}$$

其中：$\boldsymbol{W}$为可学习的参数矩阵；σ为激活函数。将学习到用户影响力和用户传递影响力的用户信息进行合并，得到最终用户特征，如式(14)所示：

$$\boldsymbol{u}''_i=\mathrm{concat}\left(\vec{\boldsymbol{u}}'_i,\vec{\boldsymbol{u}}_i\right) \tag{14}$$

2.4　谣言检测模型

基于多传递影响力的谣言检测方法如图7所示。模型由2个部分组成：1)在文本特征学习模块，首先学习谣言文本词嵌入，通过多头注意力机制和卷积神经

网络提取句子级别特征，通过图卷积神经网络学习文本结构信息最终得到节点的文本特征 m_i；2）在用户特征学习模块，首先利用谣言转发关系构建用户传播图结构，通过学习用户间直接影响力和间接影响力丰富用户信息，得到节点用户特征 $\vec{\boldsymbol{u}}_i''$。

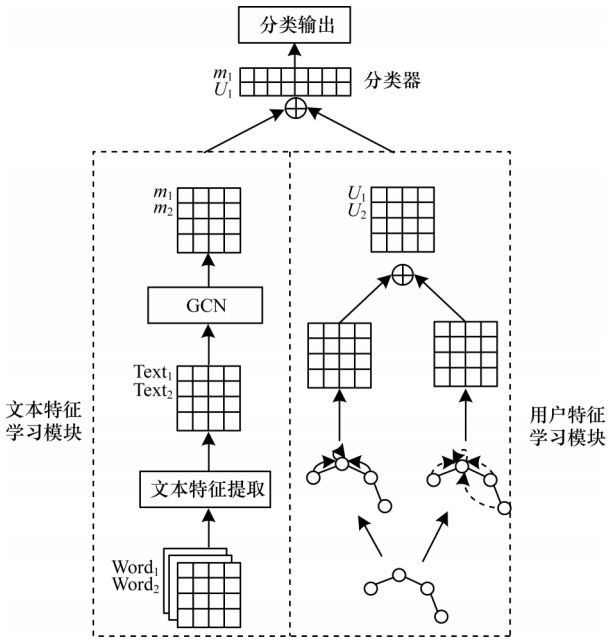

图7 基于多传递影响力的谣言检测

Fig.7 Rumor detection based on multi-transmit influence model

最后将两部分特征进行拼接得到最终的分类特征 $p_i=[m_i;\ \vec{\boldsymbol{u}}_i'']$，通过分类器模块中的全连接层和softmax层将最终表示 p_i 投影到概率空间进行分类：

$$\hat{y}_i\left(\text{class}|\vec{\boldsymbol{u}}_i^{l},m_i;\boldsymbol{\theta}\right)=\text{softmax}\left(\boldsymbol{W}^{\mathrm{T}}[m_i;\vec{\boldsymbol{u}}_i'']+b\right) \quad (15)$$

其中：$\boldsymbol{W}\in\mathbb{R}^{(d+d)\times|\text{class}|}$ 为权重矩阵；b 为偏置量；$\hat{y}$ 来表示预测 p_i 的概率分布，并利用交叉熵损失作为优化目标，如式（16）所示：

$$L\left(\text{class}^{(i)}|\boldsymbol{\theta}\right)=-\sum_i y_i\log_a\hat{y}_i(\text{class}|\vec{\boldsymbol{u}}_i'',m_i;\ \boldsymbol{\theta}) \quad (16)$$

3 实验结果与分析

3.1 数据集

本文实验采用3个真实社交媒体公共数据集，分别是Weibo谣言数据集[11]、Twitter15谣言数据集[17]和Twitter16谣言数据集[17]。Weibo数据集包含两类标签：谣言（Falserumor）和非谣言（Non-rumors），分别是2 351条和2 313条。Twitter15和Twitter16数据集包含4类标签，即谣言（FalseRumor，FR）、非谣言（Non-Rumors，NR）、未经核实的谣言（Un-verified Rumors，UR）和辟谣的谣言（True Rumors，TR），数据集内容如表1所示。

表1 实验中使用的数据集

Table 1 Datasets used in the experiment

数据集	Weibo	Twitter15	Twitter16
Source Tweets	4 664	1 490	818
Non-rumors	2 351	374	205
False rumors	2 313	370	205
Unverified rumors	0	374	203
True rumors	0	372	205
Users	2 746 818	276 663	173 487
Posts	3 805 656	331 612	204 820

3.2 对比模型和评价指标

为了验证本文提出的基于多传递影响力的谣言检测方法（MTI）的有效性，选用近年来在谣言检测任务中表现优越的模型作为对比模型与本文模型进行实验对比。

1）DTC[3]：采用决策树分类算法，利用虚假消息特征进行建模，判定数据集信息的真实性。

2）SVM-RBF[5]：采用带有RBF核的支持向量机模型算法，利用虚假消息特征进行建模，判定数据集信息的真实性。

3）SVM-TS[9]：一种线性的基于支持向量机（SVM）的分类模型，采用时间序列为虚假信息的社会情境特征变化进行建模。

4）DTR[20]：一种基于决策树（Decision Tree，DT）的算法，通过搜索判别一些有争议性的言论来识别虚假信息。

5）GRU-RNN[11]：一种基于循环神经网络的方法，通过学习随时间变化的评论特征来进行虚假信息检测。

6）PTK[21]：采用基于传播树核的方式，通过将消息传播构建为树型结构，利用支持向量机分类算法来进行虚假信息检测。

7）RvNN[16]：一种基于递归神经网络的虚假消息识别模型，通过捕获自下而上和自上而下的树结构信息实现虚假消息检测。

8）RFC[4]：一种利用随机森林（Random Forest，RF）算法构建的分类模型，采用用户、语言和结构等特征实现虚假消息检测。

9）PPC_RNN+CNN[17]：基于传播路径的虚假信息检测模型，使用了RNN和CNN来联合捕获用户特征的全局和局部信息。

10）MTI（ours）：本文提出的基于多传递影响力的谣言检测方法。

本文选用准确率（Accuracy）和F1评测值作为检测模型性能的评价指标，具体公式如式（17）和式（18）所示：

$$A_{\text{Accuracy}}=\frac{T_{\text{TP}}+T_{\text{TN}}}{T_{\text{TP}}+T_{\text{TN}}+F_{\text{FP}}+F_{\text{FN}}} \quad (17)$$

$$F1=\frac{2\cdot T_{TP}}{2\cdot T_{TP}+F_{FP}+F_{FN}} \tag{18}$$

其中：T_{TP}表示正例预测为正例；F_{FN}表示正例错分为负例；T_{TN}表示负例预测为负例；F_{FP}表示负例错分为正例。

3.3　实验设置

在本文实验中，使用的编程语言为Python，运用深度学习框架Pytorch实现提出的模型架构，版本为1.2.0。采用Adam[22-23]算法进行参数更新，参数设置β_1和β_2分别为0.9和0.999，学习率初始化为1e−3。使用Word2Vec中的Skip-Gram网络训练得到词嵌入向量，维度为300维。在文本特征学习模块中，多头自注意力机制设置K为8，即在8个部分各自进行自注意力学习。卷积层设置一维卷积核的大小为[3，4，5]，每种卷积核的个数为100个。在节点用户信息表示中，将节点的用户信息初始化为300维，将训练的批量大小设置为64，dropout为0.5。

3.4　结果分析

3.4.1　对比实验

在Twitter15和Twitter16两个数据集上的实验结果如表2和表3所示。Twitter15和Twitter16都包含了4个类别，对于每个类别列出了各模型F1指标。

表2　Twitter15数据集实验结果

Table 2　Twitter15 dataset experimental results

模型	准确率	F1			
		NR	FR	TR	UR
SVM-RBF	0.318	0.455	0.037	0.218	0.225
DTR	0.409	0.501	0.311	0.364	0.473
DTC	0.454	0.733	0.355	0.317	0.415
SVM-TS	0.544	0.796	0.472	0.404	0.483
GRU-RNN	0.646	0.792	0.574	0.608	0.592
PTK	0.750	0.804	0.698	0.765	0.733
RvNN	0.723	0.682	0.758	0.821	0.654
PPC_RNN+CNN	0.842	0.811	0.875	0.818	0.790
MIT	0.911	0.942	0.917	0.913	0.868

表3　Twitter16数据集实验结果

Table 3　Twitter16 dataset experimental results

模型	准确率	F1			
		NR	FR	TR	UR
SVM-RBF	0.321	0.423	0.085	0.419	0.403
DTR	0.414	0.394	0.273	0.630	0.344
DTC	0.465	0.643	0.393	0.419	0.403
SVM-TS	0.574	0.755	0.420	0.571	0.526
GRU-RNN	0.633	0.772	0.489	0.686	0.593
PTK	0.732	0.740	0.709	0.836	0.686
RvNN	0.737	0.662	0.743	0.835	0.708
PPC_RNN+CNN	0.863	0.820	0.898	0.843	0.837
MIT	0.897	0.893	0.909	0.929	0.847

在Weibo数据集上的对比实验结果如表4所示，分别给出了正负两类样本的准确率、召回率和F1值。

表4　Weibo数据集实验结果

Table 4　Weibo dataset experimental results

模型	类	准确率	精确率	召回率	F1
DTR	FR	0.732	0.738	0.715	0.726
	NR		0.726	0.749	0.737
SVM-RBF	FR	0.818	0.822	0.812	0.817
	NR		0.815	0.824	0.819
SVM-TS	FR	0.857	0.839	0.885	0.861
	NR		0.878	0.830	0.857
DTC	FR	0.831	0.847	0.815	0.831
	NR		0.815	0.847	0.830
RFC	FR	0.849	0.786	0.959	0.864
	NR		0.947	0.739	0.930
RvNN	FR	0.908	0.912	0.897	0.905
	NR		0.904	0.908	0.911
GRU-RNN	FR	0.910	0.876	0.956	0.914
	NR		0.952	0.864	0.906
PPC_RNN+CNN	FR	0.916	0.884	0.957	0.919
	NR		0.955	0.876	0.913
MTI	FR	0.944	0.926	0.961	0.941
	NR		0.955	0.924	0.940

实验结果分析如下：

1）对比表2~表4所有的模型可以看出，包括RvNN、GRU-RNN、PPC_RNN+CNN等在内的深度学习方法在各项评测指标上都优于基于人工构造特征的传统机器学习方法。在传统的人工特征方法中，决策树DTRANK（DTR）的效果很不理想，这是因为DTR通过将谣言的信号特征的正则表达式匹配来进行谣言检测任务，这些正则表达式与上述数据集中所能匹配的数据非常少。SVM-TS模型的结果在基于人工特征的方法中效果较好，一方面是因为SVM模型本身具有比较良好的泛化能力，可以适用于复杂的数据集，另一方面在SVM-TS模型中加入了微博事件在时间变换下的不同特征，因此使得检验性能提高。深度学习的方法表现出的优势很大，这是十分正常的现象，因为基于人工构造的特征，其局限较大，依赖于经验以及受人的主观性影响较大，对数据感知并不敏感，相比之下，RvNN、PPC_RNN+CNN以及本文模型等深度学习算法能够自动地学习到针对虚假信息检测任务的文本的高级语义表示，从而能够提取到更有效的特征。

2）相比于现有的各种方法，本文方法MTI在各项指标上均有明显提升。具体来讲，在Twitter15数据集上，相比于表现最好的PPC_RNN+CNN，本文模型在精准率上提高了6.9%，4个不同类别的F1值也都有较大的提升，分别为：NR类别提升了13%，FR类别提升了4.2%，TR类别提升了9.5%，UR类别提高了7.8%；在Twitter16数据集上，模型在准确率上提高了3.4%，各类别的F1值也都有提升；在Weibo数据集上，模型相比PPC_RNN+CNN在精准率上高出2.8%。

3）PTK和RvNN两种方法都依赖于从传播树结构中提取特征，效果优于其他线性结构方法，相比于RvNN，本文模型在Weibo数据集的准确率值指标上有3.6%的提升，在两个不同类别的F1值上也有明显提升，这表明了利用图神经网络捕获文本结构影响力的有效性。RvNN使用树型传播结构对虚假信息的传播过程进行建模，但是这种方法忽略了消息传播是一个广泛而分散的图结构而非树型结构，因此丢失了许多结构信息。

本文方法在得到微博句子级别向量后,将每条微博看作节点,利用转发关系进行构图,利用图神经网络学习传播过程中的结构影响力,得到更精细的特征,从而获得更好的谣言检测效果。

4)在3个不同的数据集上,本文方法在准确率、召回率、F1值在内的各项指标上都优于PCC_RNN+CNN模型,在Twitter15数据集上,本文方法准确率高达6.9%,优于PPC_RNN+CNN,在Twitter16和Weibo数据集上也分别有3.4%和2.8%的提升。因为PPC_RNN+CNN是利用一个时间序列上的节点向量来表示传播消息的用户特征,然而消息在传播过程中不仅相邻用户存在影响力,未直接转发或评论的用户也存在间接的影响力,PPC_RNN+CNN不能捕获这些影响力特征。本文方法通过构造能够学习不同维度的用户结构影响力,从而丰富了用户信息,使得检测精度有所提高。

3.4.2 参数分析

鉴于谣言数据中文本信息至关重要,本节将在文本特征提取过程中的卷积层部分进行一些不同的超参数设置,分别采用不同大小的卷积核验证是否影响模型的性能,实验结果如图8所示。

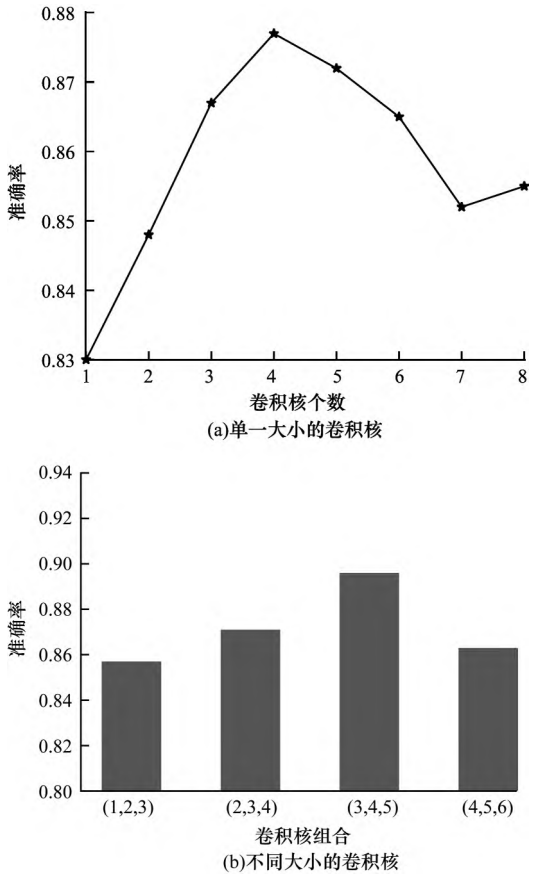

图8 不同卷积核对精度的影响

Fig.8 The impact of different convolution kernels on accuracy

从图8(a)可以看出,当卷积核设置为单核时,主要捕获单字特征,这将遗失很多信息,随着设置的卷积核增大,性能也逐渐变好,峰值为3、4、5左右,之后又继续下降。图8(b)采取的是将不同大小的卷积核进行组合的实验,实验结果表明不同大小卷积核进行组合相较于单一的卷积核性能更加优越;对比几组不同的卷积核组合,使用卷积核组合为(3,4,5)时模型性能最优,比单卷积核最佳性能要高出2个百分点,这表明不同大小卷积核的组合能捕获不同长度词语更加独特的语义信息,丰富了微博句子级别的信息表示。

3.4.3 早期检测分析

早期的谣言检测任务至关重要,因为可以更及时有效地进行预警。早期检测区别于直接检测问题,需要更快地预测谣言的真实性。本文设置一系列的检测时间点,通过只使用在检测时间点之前的相关微博来评估所提出方法的有效性。实验结果如图9所示,利用检测精度随着时间变化的曲线进行评估,横坐标表示源谣言信息出现之后的时间,设置的时间点为源消息发布后的0、4、8、12个小时,有效性则是通过准确度衡量。

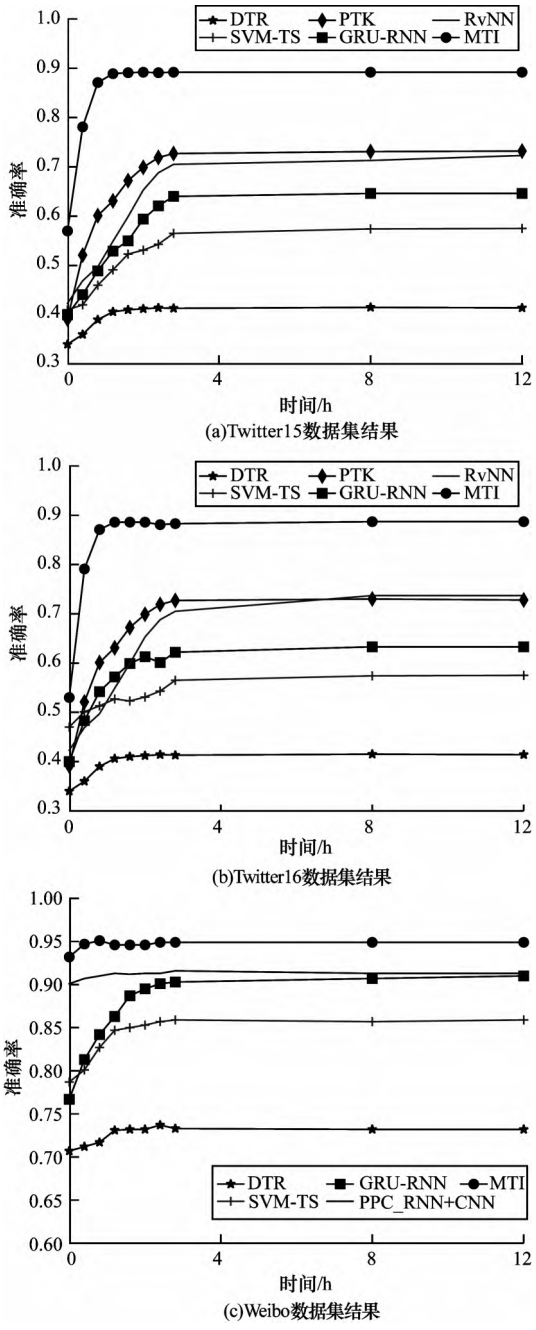

图9 早期检测结果

Fig.9 Early detection results

从图9可以看出,本文模型相比于其他基线模型,在Weibo和Twitter15、Twitter16数据集上不同截止时间点都有优异的表现。

DTR、GRU都是通过从用户评论中获取语义信息,但是DTR性能很差,这是因为在早期数据量较小的情况下,DTR可以构造的特征不够丰富,GRU能够自动获取数据中更深层的语义信息,结果优于DTR,PTK则是通过传播树结构捕获语义信息和传播结构信息,因此效果更好。所以,在传播的早期阶段,如果利用包括用户信息在内的各类信息,捕获到越多的信息会更利于检测的准确度,但是早期阶段很难获取大规模用户信息,因此本文模型通过转发结构对其进行学习,在最早的检测时间点,能够很快达到优于其他方法的性能,在Twitter15数据集上的准确率达到56%,在Twitter16数据集上的准确率达到54%,在Weibo数据集上达到93%。在之后的时间点内,本文模型准确率提升最快,能够最早达到最佳性能,这验证了本文模型以传播结构学习用户信息的有效性。同PPC_RNN+CNN方法相比,伴随着时间的增长,结构信息也会更复杂,本文模型学习到的用户传递影响力会变得更丰富,更有利于性能的提升,结果比PPC_RNN+CNN更优秀。上述实验证明了本文模型在面对复杂的语义信息时具有相对较好的稳定性和鲁棒性。因此,基于多传递影响力的谣言检测方法不仅在谣言的长期检测任务中有效,在早期检测中同样有效。

4 结束语

为提升社交媒体谣言检测精准度,本文提出一种基于多传递影响力的谣言检测方法。利用源微博和对应转发(评论)之间的传播结构关系,构建文本信息传播图和用户影响力传播图,通过图卷积神经网络捕获文本传播特征和用户节点传递影响力特征,最后将不同维度的节点信息融合,有效学习文本内容特征和用户特征,同时利用转发关系对用户节点进行构图,避免使用大规模的用户信息,对于早期检测更为有利。在3个真实数据集上的实验结果表明,本文方法具有比其他基线方法更高的谣言检测性能,并且在谣言的早期传播阶段具有良好的检测效果。本文探索了用户传递影响力在谣言检测中的作用,后续将研究更高阶用户节点信息对于检测模型性能的影响,进一步提升谣言检测精度。

参考文献

[1] 高玉君,梁刚,蒋方婷,等. 社会网络谣言检测综述[J]. 电子学报,2020,48(7):1421-1435.
GAO Y J, LIANG G, JIANG F T, et al. Social network rumor detection:a survey[J]. Acta Electronica Sinica, 2020, 48(7):1421-1435. (in Chinese)

[2] ALLCOTT H, GENTZKOW M. Social media and fake news in the 2016 election [J]. Journal of Economic Perspectives, 2017, 31(2):211-236.

[3] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on twitter [C]//Proceedings of the 20th International Conference on World Wide Web. New York, USA:ACM Press, 2011:675-684.

[4] KWON S, CHA M, JUNG K, et al. Prominent features of rumor propagation in online social media[C]//Proceedings of the 13th IEEE International Conference on Data Mining. Washington D. C. , USA:IEEE Press, 2013:1103-1108.

[5] YANG F, LIU Y, YU X H, et al. Automatic detection of rumor on sinaweibo[C]//Proceedings of ACM SIGKDD Workshop on Mining Data Semantics. New York, USA: ACM Press, 2012:1-7.

[6] JIN Z W, CAO J, ZHANG Y D, et al. Novel visual and statistical image features for microblogs news verification [J]. IEEE Transactions on Multimedia, 2017, 19(3): 598-608.

[7] 马鸣,刘云,刘地军,等. 基于主题和预防模型的微博谣言检测[J]. 北京理工大学学报,2020,40(3):310-315.
MA M, LIU Y, LIU D J, et al. Rumor detection in microblogs based on topic and prevention model [J]. Transactions of Beijing Institute of Technology, 2020, 40(3):310-315. (in Chinese)

[8] 段大高,盖新新,韩忠明,等. 基于梯度提升决策树的微博虚假消息检测[J]. 计算机应用,2018,38(2):410-414,420.
DUAN D G, GAI X X, HAN Z M, et al. Micro-blog misinformation detection based on gradient boost decision tree[J]. Journal of Computer Applications, 2018, 38(2): 410-414, 420. (in Chinese)

[9] MA J, GAO W, WEI Z Y, et al. Detect rumors using time series of social context information on microblogging websites[C]//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. New York, USA:ACM Press, 2015:1751-1754.

[10] 何力,郑灶贤,项凤涛,等. 基于深度学习的文本分类技术研究进展[J]. 计算机工程,2021,47(2):1-11.
HE L, ZHENG Z X, XIANG F T, et al. Research progress of text classification technology based on deep learning[J]. Computer Engineering, 2021, 47(2):1-11. (in Chinese)

[11] MA J, GAO W, MITRA P, et al. Detecting rumors from microblogs with recurrent neural networks[J]. Artificial Intelligence 2016, 16(8):3818-3824.

[12] 陶霄,朱焱,李春平. 基于注意力和多模态混合融合的谣言检测方法[J]. 计算机工程,2021,47(12):71-77.
TAO X, ZHU Y, LI C P. Rumor detection method based on attention and multi-modal hybrid fusion [J]. Computer Engineering, 2021, 47(12):71-77. (in Chinese)

[13] CHEN T, LI X, YIN H Z, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection[C]//Proceedings of Workshop on Trends and Applications in Knowledge Discovery and Data Mining. Berlin, Germany:Springer, 2018:40-52.

[14] 潘德宇,宋玉蓉,宋波. 一种新的考虑注意力机制的微博谣言检测模型[J]. 小型微型计算机系统,2021,42(2):348-353.
PAN D Y, SONG Y R, SONG B. New microblog rumor detection model based on attention mechanism[J]. Journal of Chinese Computer Systems, 2021, 42(2): 348-353. (in Chinese)

(下转第157页)

[21] 牛淑芬,牛灵,王彩芬,等. 一种可证安全的异构聚合签密方案[J]. 电子与信息学报,2017,39(5):1213-1218.
NIU S F,NIU L,WANG C F,et al. A provable aggregate signcryption for heterogeneous systems[J]. Journal of Electronics & Information Technology,2017,39(5):1213-1218. (in Chinese)

[22] HAN Y L, CHEN F. The multilinear maps based certificateless aggregate signcryption scheme [C]// Proceedings of International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. Washington D. C. ,USA:IEEE Press,2015:92-99.

[23] 刘建华,赵长啸,毛可飞. 高效的无证书聚合签密方案[J]. 计算机工程与应用,2016,52(12):131-135.
LIU J H,ZHAO C X,MAO K F. Efficient certificateless aggregate signcryption scheme based on XOR [J]. Computer Engineering and Applications,2016,52(12):131-135. (in Chinese)

[24] 张永洁,张玉磊,王彩芬. 具有内部安全性的常数对无证书聚合签密方案[J]. 电子与信息学报,2018,40(2):500-508.
ZHANG Y J,ZHANG Y L,WANG C F. Certificateless aggregate signcryption scheme with internal security and const pairings[J]. Journal of Electronics & Information Technology,2018,40(2):500-508. (in Chinese)

[25] 牛淑芬,牛灵,王彩芬,等. 可实现隐私保护的多接收者异构聚合签密方案[J]. 计算机工程与科学,2018,40(5):805-812.
NIU S F,NIU L,WANG C F,et al. Privacy-preserving multi-recipient aggregate signcryption for heterogeneous cryptography systems [J]. Computer Engineering & Science,2018,40(5):805-812. (in Chinese)

[26] 刘祥震,张玉磊,郎晓丽,等. 可证安全的隐私保护多接收者异构聚合签密方案[J]. 计算机工程与科学,2020,42(3):441-448.
LIU X Z,ZHANG Y L,LANG X L,et al. A provably secure privacy-preserving multi-recipient heterogeneous aggregate signcryption scheme[J]. Computer Engineering & Science,2020,42(3):441-448. (in Chinese)

[27] 王子钰,刘建伟,张宗洋,等. 基于聚合签名与加密交易的全匿名区块链[J]. 计算机研究与发展,2018,55(10):2185-2198.
WANG Z Y,LIU J W,ZHANG Z Y,et al. Full anonymous blockchain based on aggregate signature and confidential transaction [J]. Journal of Computer Research and Development,2018,55(10):2185-2198. (in Chinese)

[28] WANG Y J,DING Y,WU Q H,et al. Privacy-preserving cloud-based road condition monitoring with source authentication in VANETs [J]. IEEE Transactions on Information Forensics and Security,2019,14(7):1779-1790.

[29] 牛淑芬,李振彬,王彩芬. 适用于车载网的匿名异构聚合签密方案[J]. 计算机工程与科学,2019,41(1):80-87.
NIU S F, LI Z B, WANG C F. An anonymous heterogeneous aggregate signcryption scheme for vehicular networks[J]. Computer Engineering & Science, 2019, 41(1):80-87. (in Chinese)

[30] KIM T H, KUMAR G, SAHA R, et al. CASCF: certificateless aggregated signcryption framework for Internet-of-things infrastructure[J]. IEEE Access, 2020, 8: 94748-94756.

[31] 赖成喆,张敏,郑东. 一种安全高效的无人驾驶车辆地图更新方案[J]. 计算机研究与发展,2019,56(10):2277-2286.
LAI C Z,ZHANG M,ZHENG D. A safe and efficient map updating scheme for driverless vehicles[J]. Journal of Computer Research and Development,2019,56(10):2277-2286. (in Chinese)

编辑 薛晋栋

(上接第145页)

[15] 李奥,但志平,董方敏,等. 基于改进生成对抗网络的谣言检测方法[J]. 中文信息学报,2020,34(9):78-88.
LI A,DAN Z P,DONG F M,et al. An improved generative adversarial network for rumor detection[J]. Journal of Chinese Information Processing,2020,34(9):78-88. (in Chinese)

[16] MA J,GAO W,WONG K F. Detect rumors on twitter by promoting information campaigns with generative adversarial learning[C]//Proceedings of World Wide Web Conference. Washington D. C. USA:IEEE Press,2019:3049-3055.

[17] 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. Washington D. C. ,USA:IEEE Press,2018:2074-2085.

[18] LIU Y,WU Y F. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks[C]//Proceedings of AAAI Conference on Artificial Intelligence. New York, USA:AAAI Press,2018:354-361.

[19] MIKOLOV T,CHEN K,CORRADO G,et al. Efficient estimation of word representations in vector space [EB/OL]. [2021-04-01]. https://arxiv. org/abs/1301. 3781.

[20] KIPF T N,WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. [2021-04-01]. https://arxiv. org/abs/1609. 02907.

[21] ZHAO Z,RESNICK P,MEI Q Z. Enquiring minds:early detection of rumors in social media from enquiry posts[C]// Proceedings of the 24th International Conference on World Wide Web. Washington D. C. USA:IEEE Press,2015:1395-1405.

[22] MA J,GAO W,WONG K F. Detect rumors in microblog posts using propagation structure via kernel learning[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Washington D. C. USA:IEEE Press,2017:161-171.

[23] KINGMA D, BA J. Adam: a method for stochastic optimization[EB/OL]. [2021-04-01]. https://arxiv. org/pdf/1412. 6980. pdf.

编辑 索书志