计算机科学与探索
Journal of Frontiers of Computer Science and Technology
1673-9418/2022/16(06)-1334-09
doi: 10.3778/j.issn.1673-9418.2101030

传播用户代表性特征学习的谣言检测方法

谢欣彤[1,2],胡悦阳[2,5],刘譞哲[1,2],赵耀帅[3,4],姜海鸥[5,6+]
1. 北京大学 信息科学技术学院,北京 100871
2. 高可信软件技术教育部重点实验室(北京大学),北京 100871
3. 中国民航信息网络股份有限公司,北京 101318
4. 中国民用航空局 民航旅客服务智能化应用技术重点实验室,北京 101318
5. 北京大学 软件与微电子学院,北京 102600
6. 北京大学(天津滨海)新一代信息技术研究院,天津 300452
+ 通信作者 E-mail: seagullwill@foxmail.com

摘 要:谣言的及时发现和有效管控,是互联网+政务服务中公共舆情治理的重要组成部分。互联网和移动互联网的发展,提高了民众沟通交流的便利度,同时也加速了谣言的传播速度和广度,极大地提高了谣言的影响力和危害力,给民众的生产生活带来干扰,也严重影响社会秩序。现有的网络平台辟谣工作大多依赖于人工举报筛查,往往耗费大量的时间和精力。而利用数据挖掘、机器学习技术实现的谣言检测算法大多基于文本信息,常用于追溯性谣言检测,不适用于谣言扩散早期数据量不足的情况。首先收集最新的网络平台数据进行标注构造数据集Weibo2020,对其中用户特征分布进行统计分析并选择具有代表性的用户特征,进而提出了基于传播用户代表性特征学习的早期谣言检测方法(RPPC)。经实验验证,RPPC与同样基于传播路径的算法在同等条件下,在输入数据规模减少50%的同时,将准确率提高了2.57个百分点。此外,该方法能对5 min内发布的消息进行检测,快速发现互联网内容中的疑似谣言且准确率达到近80%。因此可以认为提出的方法在现有的数据集中有较好的表现,能够在一定程度上辅助政府部门的舆情治理工作,从而提高政务服务的时效及质量。

关键词:谣言检测;机器学习;特征分析;传播路径;互联网+政务服务;舆情治理
文献标志码:A **中图分类号:**TP181

Rumor Detection Based on Representative User Characteristics Learning Through Propagation

XIE Xintong[1,2], HU Yueyang[2,5], LIU Xuanzhe[1,2], ZHAO Yaoshuai[3,4], JIANG Hai'ou[5,6+]
1. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China
2. Key Laboratory of High Confidence Software Technologies of Ministry of Education (Peking University), Beijing 100871, China
3. TravelSky Technology Limited, Beijing 101318, China
4. Key Laboratory of Intelligent Passenger Service of Civil Aviation, Civil Aviation Administration of China, Beijing 101318, China
5. School of Software and Microelectronics, Peking University, Beijing 102600, China
6. Peking University Information Technology Institute (Tianjin Binhai), Tianjin 300452, China

基金项目:国家重点研发计划(2018YFB1004400);北京高等学校卓越青年科学家项目(BJJWZYJH01201910001004)。
This work was supported by the National Key Research and Development Program of China (2018YFB1004400) and the Beijing Outstanding Young Scientist Program (BJJWZYJH01201910001004).
收稿日期:2021-01-07 **修回日期:**2021-04-25

Abstract: Effective rumor detection and management has become an essential part of Internet plus government services initiative. The Internet era brings great convenience to people's communication as well as speeds up the propagation of rumors, which not only interferes people's normal living but also does harm to the social confidence system. Existing work of rumor debunking on the Internet is mostly based on manual work of public tip-offs and screening, which is time consuming and demanding. Meanwhile, work on algorithm of rumor detection based on data mining and machine learning depends heavily on text content, which is deficient during the early stage of rumor propagation. This paper constructs latest dataset Weibo2020, composed of both rumors and normal information, and extracts representative user characteristics from the perspective of statistics, then proposes an algorithm of early-stage rumor detection based on brief propagation path, named RPPC (representative propagation path classification). The experimental results indicate that the proposed method can improve the prediction accuracy by 2.57 percentage points while reducing the input data scale by 50%. Meanwhile, the proposed method can predict the authenticity of news released in 5 minutes and achieve an accuracy of about 80%. Therefore, the proposed method achieves good results in a limited size of dataset and can to some degree help with network public opinion governance and improve the efficiency and quality of government service.

Key words: rumor detection; machine learning; characteristic analysis; propagation path; Internet plus government services initiative; public opinion management

近年来，互联网技术改变了千家万户的生活习惯，成为了人们获取信息、互动交流的重要渠道。在中国互联网信息中心2020年4月发布的第45次中国互联网发展统计报告（http://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm）中称，截至2020年3月，我国网民数量已超9.04亿，互联网普及率达到64.5%。

然而，互联网在带来便利的同时，也为谣言的传播提供了环境。谣言是在社会中出现并流传的未经官方公开证实或已经被官方辟谣的信息[1]，其特点是所根据的事实较少，主观的补充与改造较多。尤其在疫情期间，大量制造恐慌、捕风捉影、伪科学消息在网络上涌现。中国互联网联合辟谣平台数据统计显示，2020年4月“粮食短缺，赶紧囤米抢油”相关信息达437 186条，“新冠抗体可使人免受‘二次感染’”相关信息达205 187条，这样广泛散布的谣言消息无疑将在一定程度上影响社会秩序。

互联网已经成为了思想文化信息的集散地和社会舆论的放大器，网络空间中传播的信息有着日益强大的社会影响力。如何有效地对网络空间进行公共舆情管理，是对现代化政府治理能力提出的考验。中共中央、国务院印发的《新时代公民道德建设实施纲要》中也提到，为适应新时代新要求，抓好网络空间道德建设十分关键。

信息技术是一把双刃剑，其发展同样推动了互联网与政府公共服务体系，特别是政务服务的深度融合，也加快了互联网+政务服务模式创新进程。网络空间中的公共舆情治理，是互联网+政务服务中重要的一环，而及时有效地开展网络平台辟谣工作，更是公共舆情治理尤为关键的一步。

现有的网络平台辟谣工作大多依赖于人工举报筛查机制。新浪公司成立了“微博辟谣”账号及社区管理中心，开放用户对存疑消息的举报渠道，跟进有关部门的查证工作并进行结果发布。而为了提高平台内容可靠性，过滤编造、假新闻等低质内容，今日头条公司在2018年已有4 000名内容审核编辑，人员规模仍在进一步扩大，未来预期达到10 000名。但是仅仅依靠人工进行举报、筛查，不仅耗费大量时间和精力，辟谣的时效性也有很高的局限性，因为往往在谣言的传播具有一定规模时，对社会公共秩序产生较大影响时才能引起有关部门工作人员的注意。

基于这样的背景，为了帮助推进互联网+政务服务公共舆情治理工作，本文提出以高时效性谣言自动检测过滤代替传统的人工举报筛查机制，辅助辟谣工作人员捕捉网络平台上发布的海量消息中疑似的谣言，进而推动互联网治理进一步精准化和精细化。本文的主要工作是收集最新的数据集并进行真实性标注，对其中用户特征分布进行统计分析进行特征选取并提出了基于传播用户代表性特征的早期谣言检测方法RPPC，再通过实验验证该方法的有效性。实验结果表明，RPPC能够在消息传播初期过滤疑似谣言，在一定程度上辅助政府部门的舆情治理工作，从而提高政务服务的时效及质量。

1 相关工作

谣言检测算法方面的研究大多围绕着提取谣言的消息内容及传播中的趋势特点来展开。可以根据处理方式分为基于分类的机器学习方法和基于对比的方法[2]。

基于对比的检测方法将待检测的消息与真实性可察的消息对象进行比照[2]。此类方法虽能有效地提高检测时效性,但准确率普遍较低,因此本章主要介绍基于分类的检测方法及相关工作。

基于分类的方法,大多借助各类机器学习算法,利用带标签的数据训练分类器,从而得到检测模型。然而,输入特征在很大程度上影响着分类器的准确度。谣言检测领域的开创性研究团队Castillo等人[3]提出包括消息、用户、话题和传播等方面的一系列特征。在此基础上,后续工作大多通过对特征的取舍及创新来提高分类器的表现。下面对基于常见类型特征的相关工作进行介绍。

文本特征主要分为显性特征和隐性特征。其中,显性特征分析从语法角度出发,主要包括词语、符号和简单情感特征[4]等。谣言检测相关的早期研究大多借助于对显性特征进行机器学习分类。文献[3]提取的文本特征包括内容长度、字母数量、符号数;Takahashi等人[5]提出将真实消息和谣言信息中的词频分布作为检测谣言的文本特征;Ratkiewicz等人[6]提取文本中的标签、链接和提问作为特征。但研究发现独特的显性文本特征常局限于特定的话题,分类模型不具有普适性。基于语义的隐性特征包括潜在语义[7-8]、情感(词向量[9]、分类器[10]等)和消息间关联特征(语义相似性计算[11])等。这类方法在预测的准确率方面优于基于语法的显性特征提取类方法,但总体而言,基于文本特征的方法常借助于大量对于消息评论文本、转发文本的挖掘,因此由于谣言扩散早期文本信息不足,常用于追溯性谣言检测,即时性检测表现不佳。

多媒体信息特征包括图片、音视频等内容,具有较强的吸引力和误导性(Sun等人[12]的研究结果表明80%的谣言都含有图片信息)。文献[13]提出了从基于图片本身的视觉特征(像素、清晰度、相关性、区分度)和基于事件的统计特征(图片数、含图片消息比率、图像与消息数量比例)两个角度识别图片类虚假消息,且在各类分类器上实验表明,图片类特征的检测效果优于常见的其他特征。然而,当前基于多媒体特征检测谣言大多需要在模型中引入文本特征[14-15]及其他外部知识来印证内容,模型输入及结构较复杂,也未考虑到多媒体信息中包含的元数据(文件名、创建时间及地点等),同时很少运用基于相关的多媒体处理技术识别深层的语义特征[2]。

基于用户行为特征的方法主要对信息的发布者、传递者和接受者及其交互行为进行分析。此类方法大多通过搜集发布用户的动态数、转发数、关注数、粉丝数及异常行为模式等特征作为判别依据。Wu等人[16]对消息的传播模式进行分析,指出谣言的传播模式与其他消息存在明显差异。文献[17]使用了聚类的方法对用户的转发及评论行为进行分析。文献[18]创新性地引入五个特征(日均关注数、日均动态数、发布相似内容的用户数、质疑性质评论比、纠正性质评论比),实验结果表明选取的新特征效果显著。Li等人[19]引入了用户的可靠性特征,同时也结合了大量文本信息数据作为输入。Liu等人[20]将消息传播中的转发用户特征作为输入,在中文及英文的社交媒体平台数据集的早期谣言检测中均取得了较好的检测效果。

受其启发,本文试图探究基于用户行为特征的谣言检测方法的可移植性。例如一些综合资讯类应用,虽然没有集成度高的转发功能,评论区信息却很丰富;与此同时,此类应用的用户信息完善度不及传统社交媒体。因此,本文考虑从更改采集的数据源、精简输入特征两方面入手,初步探究基于用户行为特征的检测方法是否具有移植可能。

2 基于传播用户代表性特征的谣言检测方法

本文设计了一种基于传播用户代表性特征的谣言检测方法(representative propagation path classification,RPPC),通过提取发布及评论用户具有代表性的特征向量作为输入,对消息的真实性进行分类。

2.1 问题定义

现有消息集合 $\mathcal{A}=\{a_1,a_2,\cdots,a_m\}$,$U=\{u_1,u_2,\cdots,u_{|U|}\}$ 为参与消息传播的发布及评论用户,其中每个用户 u_i 都可以由用户特征组成的信息向量 $\boldsymbol{x}_j\in\mathbf{R}^d$ 来表示,u_1 为发布消息的用户。定义对于指定消息 a_i,其传播路径为一可变长度的时间序列 $P(a_i)=\left\langle\cdots,(\boldsymbol{x}_j,t),\cdots\right\rangle$,其中每个元组 $(\boldsymbol{x}_j,t)$ 表示用户 u_j 在 t 时刻发布/传播了消息 a_i。

而每个消息 a_i 都对应着标签 $L(a_i)\in\{0,1\}^r$,用于表示该消息的真实性,目标是得到模型 f,当给定消息 a_i 的传播路径 $P(a_i)$ 时,能预测得到消息的真实性,即 $L(a_i)=f(P(a_i))$。本文目标是检测消息为谣言与否,当 $r=1$ 时,$L(a_i)=0$ 表示消息属实,而 $L(a_i)=1$ 表示其为谣言。当 $r>1$ 时,标签可以表示多级别的真实性,

如真实、虚假、不明等。

为了保证实用时效性，应尽可能快地在消息开始散布时获得检测结果，即模型应该能够利用消息传播早期阶段的传播路径得出结果。这里定义指定消息 a_i 的传播路径片段为 $P(a_i,T)=\left\langle (\boldsymbol{x}_j,t<T)\right\rangle$，其中 T 为检测截止时间。模型具体要解决的问题是，基于给定消息 a_i 的传播路径片段，给出其真实性的预测值，即 $L(a_i)=f(P(a_i,T))$。

2.2 数据集构造

本文所构造的数据集 Weibo2020 如表 1 所示，由两部分组成：谣言消息及真实消息。其中谣言消息来自微博社区管理中心 2016 年 8 月 2 日至 2020 年 3 月 23 日所判定的不实信息，以及中国互联网联合辟谣平台、腾讯新闻较真平台中公布的谣言反向搜索得到的谣言微博。真实消息采集自 3 月 20 日微博热门内容中的社会、国际、科技、健康等板块爬取实时发布的微博。筛去已删除的微博及互动数为 0 的条目，共收集谣言消息 3 688 条，真实信息 3 460 条。

表 1 数据集 Weibo2020 统计情况

Table 1 Statistics of dataset Weibo2020

统计信息	真实消息	谣言消息
消息数量	3 460	3 688
发布用户	2 871	3 317
评论用户	206 518	74 192

本文方法主要关注的是参与消息传播的用户特征，数据集包含的主要用户字段如表 2 所示。

表 2 数据集 Weibo2020 包含的用户特征

Table 2 User characteristics in dataset Weibo2020

字段名称	含义
bi_followers_count	互相关注用户数
user_description	个人简介
user_avatar	头像图片链接
friends_count	关注数
followers_count	粉丝数
verified	是否为认证用户
statuses_count	动态数
user_created_at	注册时间
favourites_count	收藏数量
screen_name	用户名
gender	性别
user_geo_enabled	地理标识开启
urank	微博等级

该数据集的标签为“真”或“假”，微博社区管理中心等判定的谣言信息标定为“假”，采集的实时微博为“真”。

2.3 消息传播用户特征分析

在问题定义中，本文用参与传播的发布及评论用户的特征作为传播路径的向量表示，关注用户特征的选取。本文将消息的发布及评论行为作为传播路径，对 Weibo2020 进行统计分析，结果显示在消息的发布用户和评论群体中，用户的注册时间、认证情况、粉丝数、动态数四个特征分布有明显差异。

图 1、图 2 为用户注册时间分布情况，其中横坐标为用户注册时间戳，纵坐标为用户比例。可以看出，发布用户中，普通用户的注册高峰出现较早，谣言用户群体的注册时间则比较平均。而在评论用户中，普通评论用户的注册时间则普遍偏早于评论谣言用户。

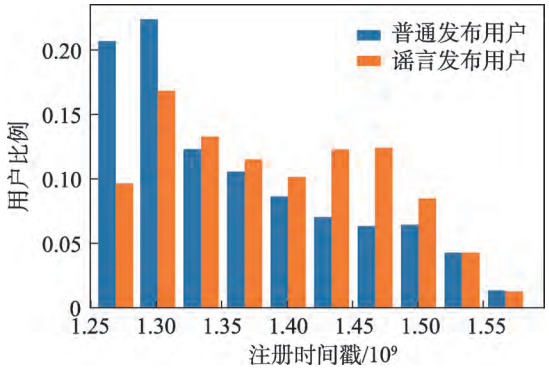

图 1 发布用户注册时间戳

Fig.1 Publishers' registration timestamp

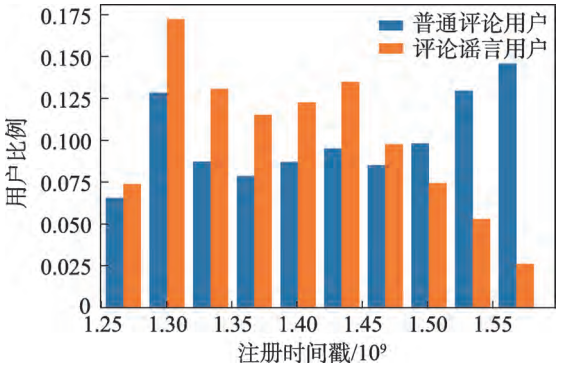

图 2 评论用户注册时间戳

Fig.2 Commentators' registration timestamp

图 3 为用户群体认证情况统计。在评论用户群体中，用户的认证情况分布较为相近。但在发布用户群体的认证情况分布上，两个群体比例存在显著差异，一个可能的原因是认证用户所发布的内容更容易出现在热门板块，但是发布用户的认证与否仍然极可能有助于谣言的检测。

图 4 为发布用户及评论用户粉丝数分布箱线图，

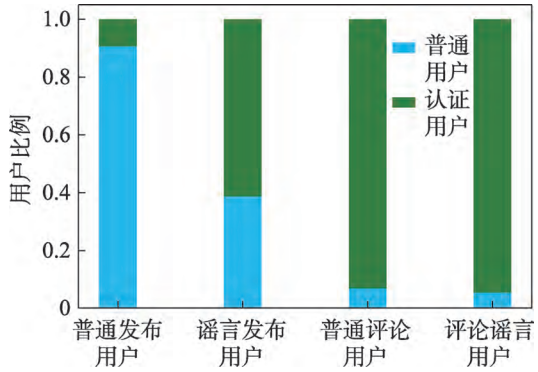

图3 用户认证情况

Fig.3 Verification of users

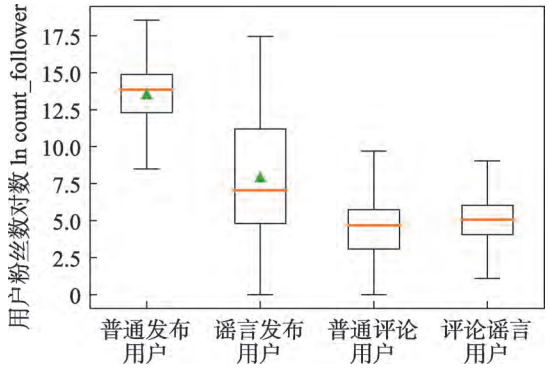

图4 用户粉丝数

Fig.4 User follower count

可以看出普通发布用户的粉丝数明显高于谣言发布用户。

图5为用户动态数分布情况。在发布群体中，普通发布用户相较于发布谣言用户有更多的发表动态

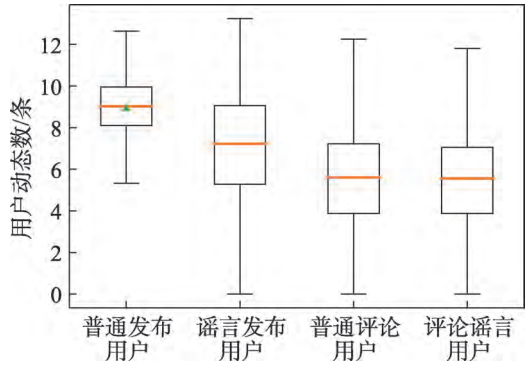

图5 用户动态数

Fig.5 User status count

表现，因此传播路径中用户的动态发布数也很可能成为判断消息真实性的重要特征。

2.4 基于传播用户代表性特征学习的谣言检测算法

本文算法RPPC模型结构如图6所示，主要由四部分构成：传播路径构造与转换模块、基于门控循环单元的特征提取模块、基于卷积神经网络的特征提取模块和传播路径向量分类模块。

其中传播路径构造与转换模块将消息的传播过程处理为固定的输入模式，基于门控循环单元、卷积神经网络的模块对其进行学习，拼接后得到传播路径向量，最终交由传播路径向量分类模块给出消息真实性预测结果。

2.4.1 传播路径构造与转换模块

给定一个在社交媒体上传播的消息，为构造其传播路径，首先要选定参与其传播的用户特征向量

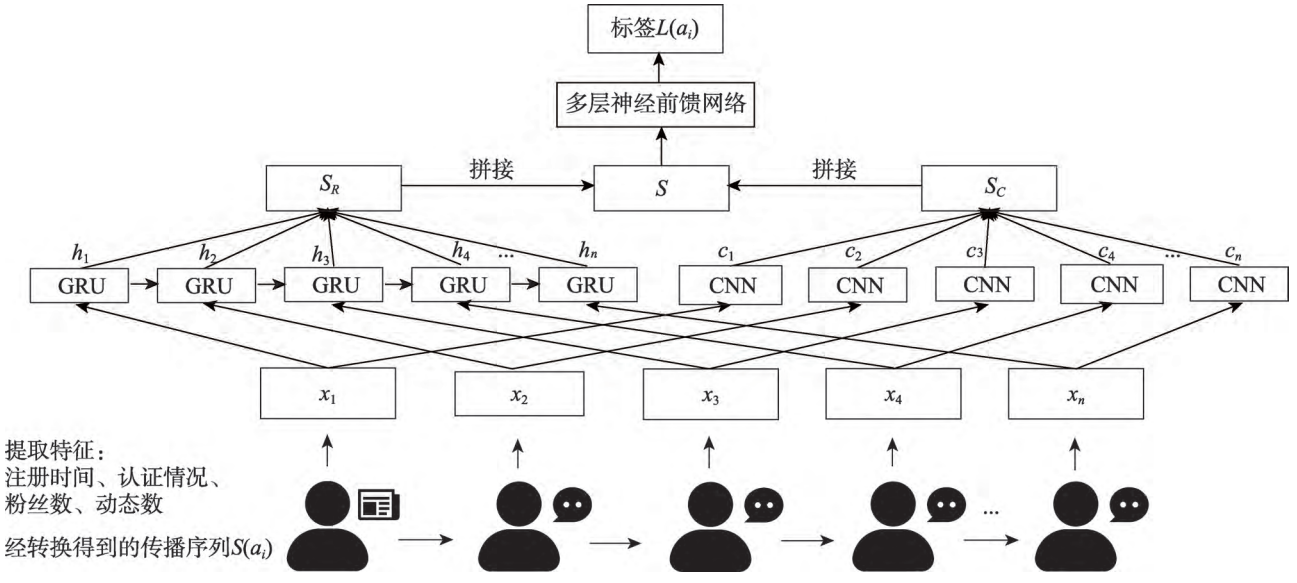

图6 算法RPPC框架示意图

Fig.6 Workflow for RPPC

$\boldsymbol{x}_i$，包括注册时间、认证情况、粉丝数及动态数。从发布及进行传播的用户资料中提取了用户特征后，可以根据各用户参与传播的时间 t 排序，得到传播路径 $P(a_i)=\langle\cdots,(\boldsymbol{x}_j,t),\cdots\rangle$。由于每条消息的传播路径 $P(a_i)$ 是不定长的，此后需要将其转换成定长的序列 $S(a_i)$，令长度为 n，则 $S(a_i)=\langle\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots\rangle$，若原传播路径长度大于 n，则将其截断取前 n 个元组；若长度小于 n，则随机从其中采样扩充以保证最终得到的序列长度为 n。

2.4.2 基于门控循环单元的特征提取模块

2014年，Chung等人提出了门控循环单元（gated recurrent unit，GRU）[21]。GRU将LSTM中的遗忘门和输入门合成为更新门，混合了神经元状态和隐藏状态，每个单元由控制更新的门控 z_t、控制重置的门控 r_t、候选隐藏层 $\hat{\boldsymbol{h}}_t$、最终的隐藏层 $\boldsymbol{h}_t$ 组成，比标准的LSTM模型更简单，但在模型准确率方面同样有突出的表现。

$$z_t=\sigma(\boldsymbol{W}_z\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]) \tag{1}$$

$$r_t=\sigma(\boldsymbol{W}_r\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]) \tag{2}$$

$$\hat{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}_h\cdot[r_t\cdot\boldsymbol{h}_{t-1},\boldsymbol{x}_t]) \tag{3}$$

$$\boldsymbol{h}_t=(1-z_t)\cdot\boldsymbol{h}_{t-1}+z_t\cdot\hat{\boldsymbol{h}}_t \tag{4}$$

依次将第 t 个用户向量 $\boldsymbol{x}_t$ 作为输入，获得GRU单元输出的序列 $\langle\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_n\rangle$，最后对其做平均值池化得到向量 $\boldsymbol{s}_R\in\mathbb{R}^m$，可以作为传播序列 $S(a_i)$ 全局视角下的向量组成部分。

2.4.3 基于卷积神经网络的特征提取模块

模型选用卷积神经网络（convolutional neural network，CNN）对传播序列 $S(a_i)$ 的向量表示进行学习。首先用滤波器 $\boldsymbol{W}_f\in\mathbb{R}^{h\times4}$ 对 h 个连续的用户向量 $\langle\boldsymbol{x}_t,\boldsymbol{x}_{t+1},\cdots,\boldsymbol{x}_{t+h-1}\rangle$ 进行一维卷积，产生对这段传播序列的定量表示 $\boldsymbol{c}_t$：

$$\boldsymbol{c}_t=\mathrm{ReLU}(\boldsymbol{W}_f*\boldsymbol{X}_{t:t+h-1}+b_f) \tag{5}$$

其中，$\boldsymbol{X}_{t:t+h-1}$ 的行是用户向量，b_f 为偏差。ReLU为线型整流函数。共使用 k 个滤波器获得特征向量 $\boldsymbol{c}_t\in\mathbb{R}^k$，对所有连续 h 个向量进行卷积可以获得 $\langle\boldsymbol{c}_1,\boldsymbol{c}_2\cdots,\boldsymbol{c}_{n-h+1}\rangle$。最后，取平均值得到基于卷积神经网络模型的传播序列 $S(a_i)$ 向量表示 $\boldsymbol{s}_C$。

$$\boldsymbol{s}_C=\frac{1}{n}\sum_{t=1}^{n-h+1}\boldsymbol{c}_t \tag{6}$$

2.4.4 传播路径向量分类模块

通过门控循环单元及卷积神经网络模块获得 $\boldsymbol{s}_R$、$\boldsymbol{s}_C$ 后，将其拼接起来成为一个向量 $\boldsymbol{s}\in\mathbb{R}^{m+k}$：

$$\boldsymbol{s}=\mathrm{Concatenate}(\boldsymbol{s}_R,\boldsymbol{s}_C) \tag{7}$$

再将其输入多层前馈神经网络获得对于消息的预测。

$$\boldsymbol{l}_j=\mathrm{ReLU}(\boldsymbol{W}_j\boldsymbol{l}_{j-1}+\boldsymbol{b}_j),\forall j\in[q] \tag{8}$$

RPPC使用Softmax函数作为神经网络的最后一层，并选取概率最大的作为预测目标值。

$$z=\mathrm{Softmax}(\boldsymbol{l}_q) \tag{9}$$

其中，q 为隐藏层的数量，$\boldsymbol{l}_j$ 为第 j 个隐藏层的输出，$\boldsymbol{W}_j$、$\boldsymbol{b}_j$ 为第 j 层的权重矩阵及偏差，z 为最终的输出，代表对于该消息传播路径的可信度预测值。

3 实验及结果分析

本章对RPPC算法进行实验验证。将RPPC算法和现有工作中在早期谣言检测表现突出的谣言检测算法PPC（propagation path classification）[20]进行比较，并对特征及传播路径长度选取对算法表现的影响进行实验。

3.1 实验参数选取

在模型结构设计部分，与PPC[20]一致，选取了GRU输出维度及CNN滤波器数量均为32，因此经过循环神经网络及卷积循环网络处理后得到的向量表示长度均为32，其中CNN滤波器长度为3。传播路径分类部分的多层前馈神经网络中每层神经元数为20，进行实验后设定层数为4。

本文选择的批量（batchsize）大小为32，优化算法为Adam，学习率为1E－4，momentum为0，多层前馈神经网络激活函数为ReLU。

为了更好地评估模型表现，本文进行了五折交叉验证。

3.2 实验结果与分析

将传播路径定义为在同条微博下的评论用户特征向量序列。Weibo2020中，单条微博下的评论数量分布如图7所示。仅有不到25%的微博评论不足10

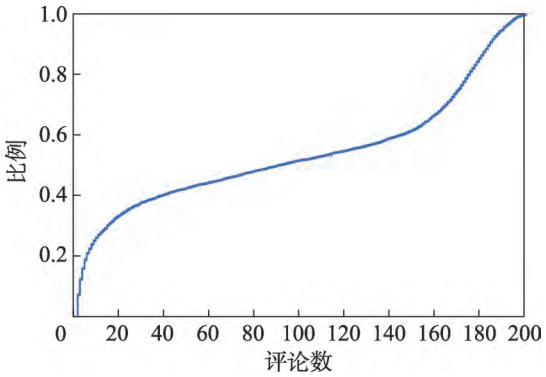

图7 数据集评论数分布

Fig.7 Distribution of dataset comment count

条，即超过75%的微博的评论数超过10。为了保证实验结果对绝大多数微博有效，对传播路径长度为10的情况进行实验。

本实验与PPC一致，将PPC_RNN+CNN模型作为基线，本文提出的将注册时间、认证情况、粉丝数、动态数四个特征作为输入的模型记为“RPPC_RNN+CNN”。本文同时也实现了模型的两个轻量级版本，只使用单一的循环神经网络或者卷积神经网络，分别记为“RPPC_RNN”及“RPPC_CNN”。为了验证模型特征选取是否合理，也在原有四个特征基础上依次添加了个人简介长度、用户名长度、关注用户数的模型进行实现，记为“RPPC_RNN+CNN_5”“RPPC_RNN+CNN_6”及“RPPC_RNN+CNN_7”，实验结果如表3。

表3 实验结果对比

Table 3 Comparison of experimental results %

模型	类别	准确率	精确率	召回率	*F*1值
PPC_RNN+CNN	谣言	76.86	75.62	79.79	77.25
	真实消息		79.50	74.25	76.24
RPPC_RNN	谣言	60.11	55.95	99.58	71.65
	真实消息		97.88	19.63	32.71
RPPC_CNN	谣言	73.13	69.78	82.33	75.54
	真实消息		78.03	63.77	70.19
RPPC_RNN+CNN	谣言	**79.43**	**77.51**	**82.30**	**79.74**
	真实消息		**81.81**	**76.73**	**79.09**
RPPC_RNN+CNN_5	谣言	78.98	78.32	79.65	78.86
	真实消息		80.00	78.18	78.94
RPPC_RNN+CNN_6	谣言	79.69	78.79	80.96	79.82
	真实消息		80.76	78.35	79.48
RPPC_RNN+CNN_7	谣言	79.36	78.79	79.65	79.17
	真实消息		80.04	79.01	79.48

结果显示，本文提出的模型“RPPC_RNN+CNN”在准确率等指标上超过了基于转发路径并使用了8个用户特征的基线模型“PPC_RNN+CNN”，即在提高了迁移至其他应用平台可能性的同时兼顾了检测效果。同时，模型的表现也明显优于基于单一神经网络的“RPPC_CNN”及“RPPC_RNN”，说明将两类神经网络集成于模型中在当前问题中是具有意义的。此外，与“RPPC_RNN+RNN_X”系列模型的对比结果显示，增加模型使用的特征对模型表现几乎没有影响。因此本文认为提出的模型“RPPC_RNN+CNN”结构设计合理、特征选取得当，在检测效果上具有很好的表现。

3.3 传播路径长度对模型的影响

RPPC对消息的分类基于传播路径，而路径长度越长，输入数据所包含的信息量越大，模型的表现则可能会得到提升。因此本文也对选取不同长度的传播路径对模型表现的影响进行探究，并对实际应用中的模型选取进行讨论。

基于图8对于Weibo2020中微博评论数量随时间增长的情况统计，发现在检测时间1 h内，平均一条微博会收到60条评论，因此本文对传播路径长度在10~60之间的模型表现进行实验。

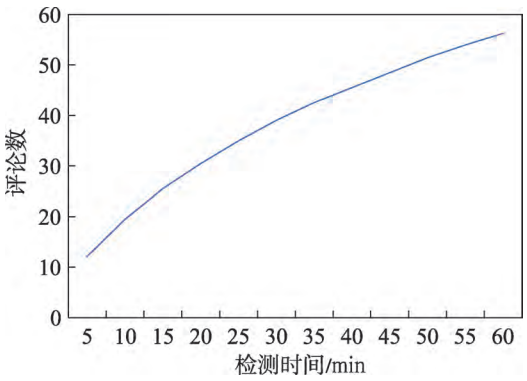

图8 微博评论数随时间增长情况

Fig.8 Weibo comment increasement with time

选用不同长度传播路径的模型运行结果如图9所示。

实验结果显示，总体而言传播路径长度对RPPC表现的影响并不大，因此本文认为选用输入传播路径长度为10的模型，便可以对5 min内发布消息的真实性进行预测，具有很好的时效性，符合本文场景的需要。

4 总结与展望

本文针对目前辟谣工作中大量依靠人工举报筛查、工作量大而时效性不高的情况，提出以高时效性谣言自动检测分析代替传统的人工举报筛查机制，推进互联网+政务服务，帮助提升政府的公共舆情治理能力。具体工作如下：

收集最新的数据集Weibo2020并进行真实性标注，通过对其中用户群体的特征分布进行特征选取，在此基础上设计并实现了基于传播用户代表性特征的谣言检测算法RPPC，其具有迁移至社交媒体类之外应用平台可能性，并通过实验测试该方法的有效性。实验结果表明，RPPC与同规模的基于传播路径的算法，在输入数据规模减少了50%的同时，提高准确率2.57个百分点，能对5 min内发布的消息进行真实性预测，且准确率达到约80%。

同时，也必须指出本文工作使用数据集的局限性。首先，由于采集的数据集规模有限，受当前较为

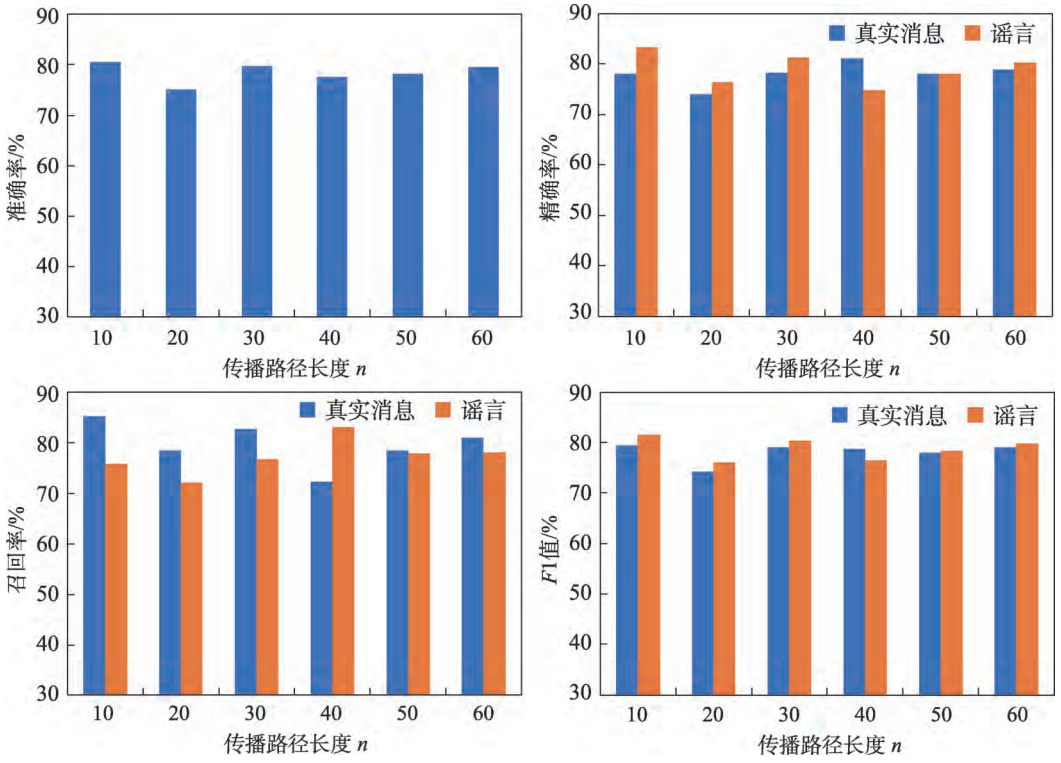

图9 传播路径长度对模型表现的影响

Fig.9 Influence of propagation length on model performance

特殊的时间环境背景影响较大，在与Liu等人工作[20]的比较中很可能存在偏差，算法的性能表现还需要在未来工作中构造规模更大、覆盖面更全的数据集，进而进行更全面的测试、调整。此外，由于在实际运用场景中，谣言与真实消息的存在比例远小于数据集中所选取的1∶1，在进行实时过滤时可能会出现将较多普通消息判断为谣言的情况，目前本文模型RPPC的检测结果仅作为对消息真实性的初步判断。

在未来的工作中，为了能够帮助提供更好的服务质量，可以考虑从扩大数据集规模、调整数据集构造比例等方面进一步对算法性能进行测试；同时，为了提高服务覆盖面及服务质量，应构造综合资讯类应用平台数据集，实地验证该方法的可迁移性，并考虑使用多种检测方法相结合的方式，对处于各个传播阶段、包含信息量不同的消息提供更有针对性、准确率更高的检测。

参考文献：

[1] SCHRAVENDIJK J P V. Rumeurs, le plus vieux média du monde[M]. KAPFERER J N. Paris: Editions du Seuil, 1987.

[2] 陈燕方, 李志宇, 梁循, 等. 在线社会网络谣言检测综述[J]. 计算机学报, 2018, 41(7): 1648-1677.
CHEN Y F, LI Z Y, LIANG X, et al. Review on rumor detection of online social networks[J]. Chinese Journal of Computers, 2018, 41(7): 1648-1677.

[3] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter[C]//Proceedings of the 20th International Conference on World Wide Web, Hyderabad, Mar 28-Apr 1, 2011. New York: ACM, 2011: 675-684.

[4] ANDREEVSKAIA A, ANDREEVSKAIA A, BERGLER S, et al. Mining WordNet for fuzzy sentiment: sentiment tag extraction from WordNet glosses[C]//Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, 2006. Stroudsburg: ACL, 2006: 209-216.

[5] TAKAHASHI T, IGATA N. Rumor detection on Twitter[C]//Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems, and the 13th International Symposium on Advanced Intelligent Systems, Kobe, Nov 20-24, 2012. Piscataway: IEEE, 2012: 452-457.

[6] RATKIEWICZ J, CONOVER M, MEISS M, et al. Detecting and tracking the spread of astroturf memes in microblog streams[J]. arXiv:1011.3768, 2010.

[7] QAZVINIAN V, ROSENGREN E, RADEV D R, et al. Rumor has it: identifying misinformation in microblogs[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Jul 27-31, 2011. Stroudsburg: ACL, 2011: 1589-1599.

[8] 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, New York, Jul 9-15, 2016. Menlo Park: AAAI, 2016: 3818-3824.

[9] 毛二松, 陈刚, 刘欣, 等. 基于深层特征和集成分类器的微博谣言检测研究[J]. 计算机应用研究, 2016, 33(11): 3369-3373.
MAO E S, CHEN G, LIU X, et al. Research on detecting micro-blog rumors based on deep features and ensemble classifier[J]. Application Research of Computers, 2016, 33(11): 3369-3373.

[10] 郭凯. 基于评论情感的微博谣言检测研究[D]. 大连: 大连理工大学, 2014.
GUO K. The research of Microblog rumors detection based on comments sentiment[D]. Dalian: Dalian University of Technology, 2014.

[11] ZHANG Q, ZHANG S Y, DONG J, et al. Automatic detection of rumor on social network[C]//LNCS 9362: Proceedings of the 4th Natural Language Processing and Chinese Computing, Nanchang, Oct 9-13, 2015. Cham: Springer, 2015: 113-122.

[12] SUN S Y, LIU H Y, HE J, et al. Detecting event rumors on Sina Weibo automatically[C]//LNCS 7808: Proceedings of the 15th Asia-Pacific Web Conference on Web Technologies and Applications, Sydney, Apr 4-6, 2013. Berlin, Heidelberg: Springer, 2013: 120-131.

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

[14] WANG Y Q, MA F L, JIN Z W, et al. EANN: event adversarial neural networks for multi-modal fake news detection [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, 2018. New York: ACM, 2018: 849-857.

[15] KHATTAR D, GOUD J S, GUPTA M, et al. MVAE: multimodal variational autoencoder for fake news detection[C]// Proceedings of the 2019 World Wide Web Conference, San Francisco, May 13-17, 2019. New York: ACM, 2019: 2915-2921.

[16] WU K, YANG S, ZHU K Q. False rumors detection on Sina weibo by propagation structures[C]//Proceedings of the 31st 2015 IEEE International Conference on Data Engineering, Seoul, Apr 13-17, 2015. Piscataway: IEEE, 2015: 651-662.

[17] CAI G Y, WU H, LV R. Rumors detection in Chinese via crowd responses[C]//Proceedings of the 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Beijing, Aug 17-20, 2014. Washington: IEEE Computer Society, 2014: 912-917.

[18] LIANG G, HE W, 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.

[19] LI Q Z, ZHANG Q, SI L. Rumor detection by exploiting user credibility information, attention and multi-task learning[C]// Proceedings of the 57th Conference of the Association for Computational Linguistics, Florence, Jul 28-Aug 2, 2019. Stroudsburg: ACL, 2019: 1173-1179.

[20] 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 the 32nd AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, Feb 2-7, 2018. Menlo Park: AAAI, 2018: 354-361.

[21] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv:1412.3555, 2014.

谢欣彤(1998—),女,广东大埔人,硕士研究生,主要研究方向为数据分析。
XIE Xintong, born in 1998, M.S. candidate. Her research interest is data analysis.

胡悦阳(1997—),男,安徽蚌埠人,硕士研究生,主要研究方向为大数据、区块链。
HU Yueyang, born in 1997, M.S. candidate. His research interests include big data and block chain.

刘譞哲(1980—),男,甘肃兰州人,博士,副教授,主要研究方向为服务计算、系统软件。
LIU Xuanzhe, born in 1980, Ph.D., associate professor. His research interests include services computing and system software.

赵耀帅(1977—),男,山东济宁人,硕士,高级软件工程师,主要研究方向为大数据、人工智能。
ZHAO Yaoshuai, born in 1977, M.S., senior software engineer. His research interests include big data and artificial intelligence.

姜海鸥(1987—),女,辽宁丹东人,博士,助理研究员,主要研究方向为云计算、大数据、机器学习。
JIANG Hai'ou, born in 1987, Ph.D., assistant researcher. Her research interests include cloud computing, big data and machine learning.