



主办:中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图形象报
图形学

2021
07
VOL.26

ISSN1006-8961
CN11-3758/TB



中国多媒体大会(ChinaMM2020)专刊



第26卷第7期（总第303期）

2021年7月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院空天信息创新研究院

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 吴一戎

编辑出版 《中国图象图形学报》编辑出版委员会

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@aircas.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China

Zip code 100190

E-mail jig@aircas.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation

(P.O.Box 399, Beijing 100048, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

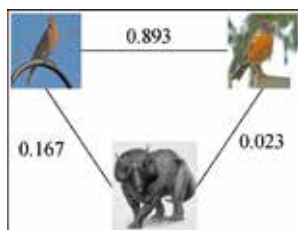
ISSN 1006-8961

CODEN ZTTXFZ

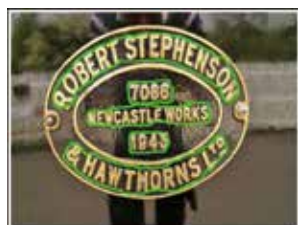
国外发行代号 M1406

国内邮发代号 82-831

国内定价 60.00元



类别语义相似性监督的小样本图像识别(第1594页)



像素聚合和特征增强的任意形状场景文本检测(第1614页)



时序特征融合的视频实例分割(第1692页)

综述

面向上行流媒体的压缩感知视频流技术前沿

刘浩, 黄荣, 袁浩东 1545

图像处理和编码

耦合保持投影哈希跨模态检索

闵康凌, 张国宾, 王磊, 李丹萍 1558

面向视觉搜索的空间局部敏感哈希方法

黄小燕, 孙彬, 杨展源, 朱映映, 田奇 1568

图像分析和识别

融合视觉关系检测的电力场景自动危险预警

高明, 左红群, 柏帆, 田清阳, 葛志峰, 董兴宁, 甘甜 1583

类别语义相似性监督的小样本图像识别

徐鹏帮, 桑基韬, 路冬媛 1594

局部双目视差回归的目标距离估计

张羽丰, 李昱希, 赵明璧, 喻晓源, 占云龙, 林巍峭 1604

像素聚合和特征增强的任意形状场景文本检测

师广琛, 巫义锐 1614

相位一致性指导的全参考全景图像质量评价

夏雨蒙, 王永芳, 王闯 1625

图像理解和计算机视觉

特征金字塔结构的时序行为识别网络

何嘉宇, 雷军, 李国辉 1637

多模态多层次事件网络的谣言检测

李莎, 张怀文, 钱胜胜, 方全, 徐常胜 1648

多模态零样本人体动作识别

吕露露, 黄毅, 高君宇, 杨小汕, 徐常胜 1658

足球视频球员感知跟踪算法

冯思佳, 宋子恺, 于俊清, 何云峰, 管涛 1668

融合时空图卷积的多人交互行为识别

成科扬, 吴金霞, 王文杉, 荣兰, 詹永照 1681

时序特征融合的视频实例分割

黄泽涛, 刘洋, 于成龙, 张加佳, 王轩, 漆舒汉 1692

医学图像处理

多尺度深度特征提取的肝脏肿瘤CT图像分类

毛静怡, 宋余庆, 刘哲 1704

3D多尺度深度卷积神经网络肺结节检测

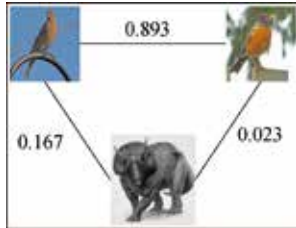
孙华聪, 彭延军, 郭燕飞, 张晓庆 1716

结合多通道注意力的糖尿病性视网膜病变分级

顾婷菲, 郝鹏翼, 白琮, 柳宁 1726

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Few shot image recognition based on class semantic similarity supervision(P1594)



Arbitrary shape scene-text detection based on pixel aggregation and feature enhancement (P1614)



Video instance segmentation based on temporal feature fusion (P1692)

Review

Survey on compressive sensing video stream for uplink streaming media

Liu Hao, Huang Rong, Yuan Haodong 1545

Image Processing and Coding

Structure-preserving hashing with coupled projections for cross-modal retrieval

Min Kangling, Zhang Guobin, Wang Lei, Li Danping 1558

Locality-sensitive hashing approach based on semantic space for visual retrieval

Huang Xiaoyan, Sun Bin, Yang Zhanyuan, Zhu Yingying, Tian Qi 1568

Image Analysis and Recognition

Visual relationship detection-based emergency early-warning description generation in electric power industry

Gao Ming, Zuo Hongqun, Bai Fan, Tian Qingyang, Ge Zhifeng, Dong Xingning, Gan Tian 1583

Few shot image recognition based on class semantic similarity supervision

Xu Pengbang, Sang Jitao, Lu Dongyuan 1594

Object distance estimation based on stereo regional disparity regression

Zhang Yufeng, Li Yuxi, Zhao Mingbi, Yu Xiaoyuan, Zhan Yunlong, Lin Weiyao 1604

Arbitrary shape scene-text detection based on pixel aggregation and feature enhancement

Shi Guangchen, Wu Yirui 1614

Phase consistency guided full-reference panoramic image quality assessment algorithm

Xia Yumeng, Wang Yongfang, Wang Chuang 1625

Image Understanding and Computer Vision

Temporal action detection based on feature pyramid hierarchies

He Jiayu, Lei Jun, Li Guohui 1637

Multi-modal multi-level event network for rumor detection

Li Sha, Zhang Huaiwen, Qian Shengsheng, Fang Quan, Xu Changsheng 1648

Multimodal-based zero-shot human action recognition

Lyu Lulu, Huang Yi, Gao Junyu, Yang Xiaoshan, Xu Changsheng 1658

Players-aware tracking algorithm in soccer video

Feng Sijia, Song Zikai, Yu Junqing, He Yunfeng, Guan Tao 1668

Multi-person interaction action recognition based on spatio-temporal graph convolution

Cheng Keyang, Wu Jinxia, Wang Wenshan, Rong Lan, Zhan Yongzhao 1681

Video instance segmentation based on temporal feature fusion

Huang Zetao, Liu Yang, Yu Chenglong, Zhang Jiajia, Wang Xuan, Qi Shuhan 1692

Medical Image Processing

CT image classification of liver tumors based on multi-scale and deep feature extraction

Mao Jingyi, Song Yuqing, Liu Zhe 1704

3D multi-scale deep convolutional neural networks in pulmonary nodule detection

Sun Huacong, Peng Yanjun, Guo Yanfei, Zhang Xiaoqing 1716

Diabetic retinopathy grading based on multi-channel attention

Gu Tingfei, Hao Pengyi, Bai Cong, Liu Ning 1726

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)07-1648-10

论文引用格式: Li S, Zhang H W, Qian S S, Fang Q and Xu C S. 2021. Multi-modal multi-level event network for rumor detection. Journal of Image and Graphics, 26(07): 1648-1657 [李莎, 张怀文, 钱胜胜, 方全, 徐常胜. 2021. 多模态多层次事件网络的谣言检测. 中国图象图形学报, 26(07): 1648-1657] [DOI:10.11834/jig.200499]

多模态多层次事件网络的谣言检测

李莎¹, 张怀文², 钱胜胜², 方全², 徐常胜^{2*}

1. 郑州大学, 郑州 450000; 2. 中国科学院自动化研究所模式识别国家重点实验室, 北京 100190

摘要: **目的** 自动检测谣言至关重要, 目前已有多种谣言检测方法, 但存在以下两点局限: 1) 只考虑文本内容, 忽略了可用于判断谣言的辅助多模态信息; 2) 只关注时间序列模型捕捉谣言事件的时间特征, 没有很好地研究事件的局部信息和全局信息。为了克服这些局限性, 有效利用多模态帖子信息并联合多种编码策略构建每个新闻事件的表示, 本文提出一种新颖的基于多模态多层次事件网络的社交媒体谣言检测方法。**方法** 通过一个多模态的帖子嵌入层, 同时利用文本内容和视觉内容; 将多模态的帖子嵌入向量送入多层次事件编码网络, 联合使用多种编码策略, 以由粗到细的方式描述事件特征。**结果** 在 Twitter 和 PHEME 数据集上的大量实验表明, 本文提出的多模态多层次事件网络模型比现有的 SVM-TS (support vector machine—time structure)、CNN (convolutional neural network)、GRU (gated recurrent unit)、CallAtRumors 和 MKEMN (multimodal knowledge-aware event memory network) 等方法在准确率上提升了 4 % 以上。**结论** 本文提出的谣言检测模型, 对每个事件的全局、时间和局部信息进行建模, 提升了谣言检测的性能。

关键词: 多模态; 谣言检测; 社交媒体; 多层次编码策略; 事件网络

Multi-modal multi-level event network for rumor detection

Li Sha¹, Zhang Huaiwen², Qian Shengsheng², Fang Quan², Xu Changsheng^{2*}1. Zhengzhou University, Zhengzhou 450000, China; 2. National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: **Objective** The proliferation of social media has revolutionized the way people acquire information. A growing number of people choose to share information, and express and exchange opinions through social media. Unfortunately, because a large number of users do not carefully verify the released content when posting information and sharing their opinions, various rumors have been fostered on social media platforms. The extensive spread of these rumors is expected to bring new threats to the political, economic, and cultural fields and affect people's lives. To strengthen the detection of rumors and prevent their spread, many approaches to rumor detection have been proposed. An early rumor detection platform (e.g., snopes.com) mainly reported through users, and then invited experts or institutions in related fields to confirm. Although these methods can achieve the purpose of rumor detection, the timeliness of detection has obvious limitations. Thus, how to detect rumors automatically has become a key research direction in recent years. To date, many automatic detection approaches have been proposed to improve the efficiency of rumor detection, including feature construction-based and neural network-based methods. The feature construction-based methods rely on hand-craft features to train rumor

收稿日期: 2020-08-21; 修回日期: 2021-02-09; 预印本日期: 2021-02-16

* 通信作者: 徐常胜 csxu@nlpr.ia.ac.cn

基金项目: 国家自然科学基金项目 (61802405, 61832002)

Supported by: National Natural Science Foundation of China (61802405, 61832002)

classifiers and neural network-based methods using neural networks to automatically extract deep features. Compared with traditional methods, models based on deep neural networks can automatically learn the underlying deep representation of rumors and extract more effective semantic features. However, these methods may suffer from the following limitations. 1) At post level, many existing methods only consider the text content. In fact, posts often contain various types of information (e.g., text and images), and the visual information are often used as an auxiliary information to judge the credibility of posts in reality. Therefore, the key to detecting rumors is obtaining the multi-modal information of the posts and systematically integrating the textual and visual information. 2) At the event level, existing approaches typically only use the temporal sequence model to capture temporal features of events. Local and global information has not been well investigated yet. In practice, local and global features are important because the former helps distinguish between posts of subtle differences, and the latter helps capture features that repeatedly present in the event. Therefore, based on encoding the temporal information of the event, local and global information should be exploited to obtain a fine-grained feature of the event for event encoding collaboratively. **Method** To overcome these limitations, this paper presents a novel multi-modal multi-level event network (MMEN) for rumor detection, which can effectively use multi-modal post information and combine multi-level encoding strategies to construct a representation of each news event. MMEN employs an encoding network that jointly exploits multiple encoding strategies such as mean pooling, recurrent neural networks, and convolutional networks to model the global, temporal, and local information of each event. Then, these various types of information are combined into a unified deep model. Specifically, our model consists of the following three components: 1) The multi-modal post embedding layer employs bidirectional encode representations from transformers (BERT) to generate the text content embedding vector and use Visual Geometry Group-19 (VGG-19) to obtain the visual content. 2) The multi-level event encoding network utilizes three-level encodings to capture global, temporal, and local information. The first level is a global encoder through the mean pooling, which represents the elements that are repeatedly present in the posts. The second is a temporal encoder that exploits a bidirectional recurrent neural network to use past and future information of a given post sequence. The third level is a local encoder by utilizing more subtle local representation of events. Then, the encoding results are combined to describe the events in a coarse-to-fine fashion. 3) The rumor detector layer aims to classify each event as either fake or authentic. The detector exploits a fully connected layer with corresponding activation function to generate predicted probability to determine whether the event is a rumor or not. **Result** In this study, the public datasets PHEME and Twitter are used to evaluate the effectiveness of the MMEN. The quantitative evaluation metrics included accuracy, precision, recall, and F1 score. We also perform five-fold cross-validation throughout all experiments. The experiments demonstrate that our proposed MMEN has improved accuracy by more than 4% over current best practices. MMEN has an accuracy of 82.2% on the PHEME dataset and 87.0% on the Twitter dataset. We compare our model MMEN with five state-of-the-art baseline models. Compared with all the baselines, the MMEN achieves the best performance and outperforms other rumor detection methods in most cases. To examine the usefulness of each component in the MMEN and demonstrate its effectiveness, we compare variants of MMEN. The experiment results show that the multi-modal features learned by the multimodal post embedding layer can improve the accuracy of rumor detection by nearly 0.2% on the two datasets. The experimental results also show that the temporal encoder has a stronger effect on detection accuracy. **Conclusion** In this study, we design a novel MMEN for rumor detection. Experiments and comparisons demonstrate that our model is more robust and effective than state-of-the-art baselines based on two public datasets for rumor detection. We attribute the superiority of MMEN to its two properties. The MMEN takes advantage of the multiple modalities of posts, and the proposed multi-level encoder jointly exploits multiple encoding strategies to generate powerful and complementary features progressively.

Key words: multi-modal; rumor detection; social media; multi-level encoding strategy; event network

0 引言

社交媒体平台的激增,彻底改变了人们获取信

息的方式。越来越多的人选择通过社交媒体分享信息、表达和交流意见。然而,由于大量用户在发布信息 and 分享意见时没有仔细核实发布的内容,社交媒体网站上滋生了各种谣言。这些谣言的广泛传播,

扰乱了政治、经济和文化等领域的正常秩序,影响了人们的正常生活。因此,社交媒体上的自动谣言检测器对于缓解谣言的不良影响,防止谣言传播,构建良好的网络环境具有重要意义。为了加强对谣言的检测,防止谣言传播,相继提出了许多谣言检测方法。早期的辟谣平台(如 snopes.com)主要通过用户举报,然后邀请相关领域专家或机构进行确认,虽然可以达到谣言检测的目的,但在时效性方面存在局限性。因此,如何实现谣言的自动检测成为近年来来的一个重点研究方向。

现有的谣言自动检测方法包括基于手工特征构建的方法(Castillo 等,2011;Kwon 等,2013)和基于深度神经网络的方法(Jin 等,2017;Ma 等,2016;Yu 等,2017)。基于特征构建的方法是依靠手工特征来训练谣言分类器。例如,Yang 等人(2012)首先提取每个微博中发布的内容、使用的微博客户端程序、回复和转发次数等 19 个特征,然后利用这些特征训练具有 RBF(radial basis function)核函数的支持向量机分类器。基于深度神经网络的方法利用神经网络自动提取深层特征。例如,Ma 等人(2016)提出一种基于循环神经网络(recurrent neural network, RNN)的谣言检测模型,用以捕获随时间变化的帖子的上下文信息。Yu 等人(2017)引入卷积操作从帖子文本内容中学习高层次表示用以检测谣言。

与传统方法相比,基于深度神经网络的模型可以自动学习谣言的深层表示,提取更有效的语义特征,但大多只关注文本内容,忽略了社交媒体平台中的帖子具有的文本和图像等多模态信息。同时,目前主流谣言检测算法仅使用时间序列模型进行谣言检测(Ma 等,2016;Chen 等,2018)。但多媒体内容和事件的全局和局部特征在谣言检测中具有重要作用,不应予以忽视。因此,建立有效的谣言检测框架,需要应对以下挑战:

1)在帖子层面,现有许多方法只考虑文本内容。事实上,帖子往往包含文本和图像等各种信息,如图 1 所示,其中 June、July 和 August 是对图中橘子的命名。而视觉信息在实际应用中往往作为判断帖子可信度的辅助信息。因此,如何获取帖子的多模态信息,并系统地整合文本信息和视觉信息是谣言检测的关键。

2)在事件层面,现有方法通常只使用时间序列



图 1 多模态帖子的示例

Fig. 1 Examples of multi-modal posts

模型来捕捉事件的时间特征(如 RNN),还没有很好地研究事件的全局与局部信息。在实际应用中,全局与局部特征至关重要,前者有助于捕捉事件中重复出现的特征,后者有助于区分细微差别的帖子特征。因此,在对事件的时间信息进行编码的基础上,还需要利用事件的全局和局部信息来获取事件的细粒度特征,以便协同进行事件编码。

针对上述局限性,本文提出一种基于多模态多层次事件网络的谣言检测方法,由两个关键部分组成:1)在帖子层面,为了利用不同模态的信息,设计了一个多模态帖子嵌入层来提取文本内容和视觉内容;2)在事件层面,联合利用均值池化、递归神经网络和卷积神经网络这 3 种编码策略的编码网络,对每个事件的全局、时间和局部信息进行建模,并将这些信息组合到统一的深层模型。在得到事件的统一表示之后,利用具有相应激活函数的全连接层产生的预测概率来确定事件是否为谣言。总体而言,本文工作有如下贡献:1)提出一个多模态帖子嵌入方法,提取谣言帖子的文本语义、图像信息并融合为统一的帖子嵌入向量。2)提出一种用于社交媒体谣言检测的多模态多层次事件编码方法。与以往只考虑事件单一特征的编码方法不同,该方法采用多层次编码器同时捕获帖子的全局信息、时间信息和局部信息。3)在两个公共基准数据集实验结果表明,与当前最好的方法相比,本文模型在检测准确率上提升了 4% 以上。

1 相关工作

1.1 谣言检测

现有的谣言检测研究将谣言检测任务看做是一种二分类任务。早期方法(Castillo 等,2011;Yang

等,2012)大多使用文本内容特征训练谣言分类器,并设计了大量手工制作的特征检测谣言。Yang 等人(2012)使用从微博客户端程序和活动地点中制作的特征。虽然这些人工选择的特征提高了谣言检测的性能,但这些方法通常需要大量的预处理和特征的手工制作/选择工作。随着深度学习技术在自然语言处理、计算机视觉等领域的优越表现,研究者希望利用深度神经网络自动学习谣言潜在的深层表征,提取更有效的语义特征。Ma 等人(2016)提出一种循环神经网络从相关帖子的文本内容中学习隐含的表示。在 RNN 的基础上,Chen 等人(2018)将 RNN 与注意力机制结合来区分不同文本特征的重要性。Yu 等人(2017)提出一个用于错误信息识别的卷积方法,在事件的嵌入向量上应用卷积运算提取深层特征,进行谣言的早期检测。Ma 等人(2018b)基于自下而上和自上而下树状结构提出两种递归神经模型,用于谣言表示学习和分类。

虽然已有方法取得了较好效果,但现有方法主要侧重于捕捉事件的时间特征,而忽略了事件全局和局部特征。本文引入一种新的多模态多层次事件网络,通过联合使用均值池化、递归神经网络和卷积神经网络这3种编码策略来提高谣言检测任务的性能。

1.2 多模态学习

随着深度学习在学习图像和文本表示方面取得的巨大成功,研究人员意识到视觉特征在识别谣言中发挥着重要作用。近年来,出现了多种基于多模态内容进行谣言检测的方法(Khattar 等,2019; Singhal 等,2019; Zhang 等,2019; Jin 等,2017),并取得优越的性能。Jin 等人(2017)提出一种基于多模态的谣言检测模型,提取包括视觉、文本和社交上下文特征在内的多模态信息,并通过注意力机制进行融合。Khattar 等人(2019)提出一种多模态变分自编码器,学习文本和图像两者的共享表示。Singhal 等人(2019)介绍一种多模态的假新闻检测框架,考虑来自文本和图像两个不同模式的特征,然后串联在一起,以便在不考虑任何其他子任务的情况下对事件进行分类。Zhang 等人(2019)提出一个多模态知识感知事件记忆网络,在文本特征和图片特征之外加入外部知识来补充帖子的语义表示。为了充分利用帖子的多模态信息,本文提出一个有效的多模态帖子嵌入层来融合文本与视觉特征。

2 方法

2.1 基本符号

社交媒体谣言检测主要分为基于帖子的检测和基于事件的检测两种类型。前者识别单个帖子是否为谣言,后者对由一组帖子组成的事件进行分类。本文考虑事件级别的谣言检测。

对一组给定的 n 个新闻事件,每个事件由一系列相关帖子组成,每个帖子对应一个时间戳,则

$$\mathbf{E} = \{\mathbf{E}^1, \mathbf{E}^2, \dots, \mathbf{E}^n\} \quad (1)$$

$$\mathbf{E}^i = \{\mathbf{p}_1^i, \mathbf{p}_2^i, \dots, \mathbf{p}_T^i\} \quad (2)$$

式中, T 是每个事件集合的长度,事件 i 中的帖子 $\mathbf{p}_t^i$ 由文本信息 s_t^i 与视觉信息 $\mathbf{v}_t^i$ 组成。根据先前的工作(Ma 等,2018a; Wu 等,2019),谣言检测任务可以定义为一个二分类问题,每个事件 $\mathbf{E}^i$ 对应一个真实标签 $Y \in \{0, 1\}$, 0 表示非谣言,1 表示谣言。因此,谣言检测的目标是学习一个映射

$$F_d(\mathbf{E}^i) \rightarrow Y \quad (3)$$

2.2 整体框架

本文提出的多模态多层次事件网络结构如图2所示,由多模态帖子嵌入层、多层次事件编码层和谣言检测层3部分组成。通过利用3个层次的编码策略,可以在每个事件的多模态表示中捕获全局、时间和局部信息。

2.3 多模态帖子嵌入层

给定一系列帖子,多模态帖子嵌入层旨在将它们转换为一系列嵌入向量,每个向量代表其对应的帖子。如图3所示,多模态帖子嵌入层包括文本特征提取器与视觉特征提取器两部分,用以处理不同类型的输入。

2.3.1 文本特征提取器

BERT(bidirectional encoder representations from transformers)在问答、翻译、阅读理解和文本分类(Devlin 等,2019; Sanh 等,2019; Sun 等,2019)等许多领域已证明是有效的。为了精确建模单词的语义和上下文信息,使用 BERT(Devlin 等,2019)作为文本特征提取器的核心模块。即引入改进的预训练 BERT 模型——DistilBERT(distilled version of BERT)(Sanh 等,2019)以提取帖子的文本表示。以第 t 个帖子 $\mathbf{p}_t$ 为例,给定该帖子中的一系列文本单

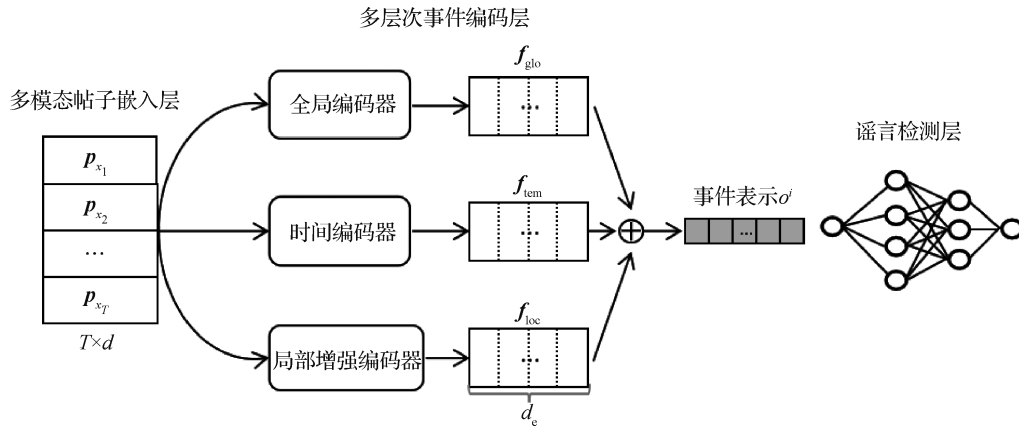


图2 多模态多层次事件网络结构

Fig. 2 The structure of multi-modal multi-level event network

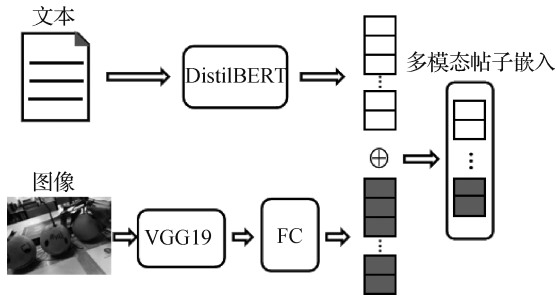


图3 多模态帖子嵌入层结构

Fig. 3 The structure of multi-modal post embedding layer

词 $s_t = \{w_{t1}, w_{t2}, \dots, w_{tm}\}$, 帖子 p_t 的文本表示可通过预训练的 DistilBERT 计算, 具体为

$$p_{s_t} = \varphi_{\text{dis}}(s_t) \quad (4)$$

式中, $\varphi_{\text{dis}}(\cdot)$ 表示 DistilBERT 运算, $p_{s_t} \in \mathbf{R}^{d_s}$ 是 BERT 中分类标记的最后一层, d_s 是文本嵌入的维度。

2.3.2 视觉特征提取器

本文使用预训练的 VGG19 (visual geometry group-19) (Simonyan 和 Zisserman, 2015) 提取视觉特征, 并且为了调整最终视觉特征表示的维度, 在 VGG19 的最后一层增加一个全连接层。对给定帖子 p_t 附带的图像 v_t , 视觉特征提取器的最后一层操作可以表示为

$$p_{v_t} = \sigma(W_v \cdot \varphi_{\text{VGG}}(v_t)) \quad (5)$$

式中, $\varphi_{\text{VGG}}(\cdot)$ 表示 VGG 卷积操作, $p_{v_t} \in \mathbf{R}^{d_v}$, W_v 是视觉特征提取器中的全连接层的权重矩阵, d_v 是视觉特征的维度。

将文本特征表示 p_{s_t} 与视觉特征表示 p_{v_t} 串联起

来, 即形成帖子的多模态特征表示, 即

$$p_{x_t} = p_{s_t} + p_{v_t} \quad (6)$$

式中, $p_{x_t} \in \mathbf{R}^d$, $d = d_s + d_v$, 同时, 预训练的 DistilBERT 和 VGG19 的参数保持静态, 以避免在训练期间过拟合。

2.4 多层次事件编码层

本文采用完全不同的多层次编码结构产生统一的事件表示。多层次事件编码层主要由全局编码器、时间编码器和局部增强编码器 3 部分组成, 如图 4 所示。

2.4.1 全局编码器

对于每个事件, 全局编码器模块旨在捕获帖子中重复出现的内容, 这些内容具有全局特性。本文采用提取全局特征最简单的方法——均值池化, 通过简单平均每个帖子的特征来表示事件。在全局编码器中, 第 i 个事件的全局特征表示为

$$f_{\text{glo}}^i = \frac{1}{T} \sum_{t=1}^T p_{x_t}^i \quad (7)$$

2.4.2 时间编码器

每个事件包括转发和评论等一系列相关的帖子, 并且每个帖子都与一个时间相关联。为了提取事件的时间特征, 使用双向递归神经网络获取给定帖子序列的过去和未来信息。首先, 经过多模态帖子嵌入层, 将第 i 个事件 E^i 映射为一个多模态嵌入特征 $E^i = \{p_{x_1}^i, p_{x_2}^i, \dots, p_{x_T}^i\}$, $p_{x_t}^i$ 是第 i 个事件中第 t 个帖子的多模态向量。然后, 将这些嵌入向量送入双向门控递归单元 (bi-directional gated recurrent unit, Bi-GRU), 以捕捉事件的时间信息。前向 GRU (gate recurrent unit) 用于编码正向序列特征, 后向

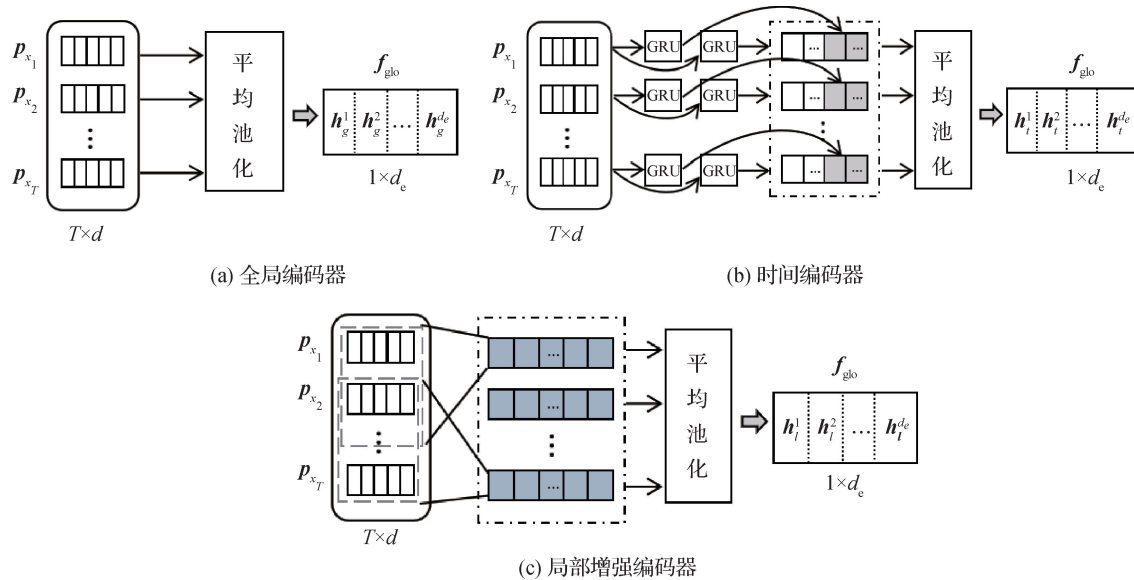


图4 多层次事件编码层结构

Fig.4 The structure of multi-level event encoding layer ((a) global encoder; (b) temporal encoder; (c) local enhanced encoder)

GRU 用于编码反向序列特征。具体为

$$\vec{h}_t = \text{GRU}(\vec{p}_{x_t}, \vec{h}_{t-1}) \quad (8)$$

$$\overleftarrow{h}_t = \overleftarrow{\text{GRU}}(\overleftarrow{p}_{x_t}, \overleftarrow{h}_{t-1}) \quad (9)$$

式中, $\vec{h}_t \in \mathbf{R}^{d_h}$, $\overleftarrow{h}_t \in \mathbf{R}^{d_h}$ 。进一步将每一个时间 t 处的前向隐藏状态 $\vec{h}_t$ 与后向隐藏状态 $\overleftarrow{h}_t$ 相串联, 获得整个 Bi-GRU 单元的隐藏状态 h_t , 即

$$h_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (10)$$

式中, $h_t \in \mathbf{R}^{d_t}$, $d_t = 2d_h$ 。

最后, 在第 i 个事件的所有帖子隐藏状态上应用均值池化来获取事件 i 的时间特征, 表示为

$$f_{\text{tem}}^i = \frac{1}{T} \sum_{t=1}^T h_t^i \quad (11)$$

2.4.3 局部增强编码器

前两部分的编码器获得了事件的全局和时间特征, 但忽略了更细粒度的特征, 这些特征有助于区分帖子间的细微差异。为了获得更细微事件的局部表示, 提出一种局部增强编码器来捕获每个事件中帖子之间的局部时间动态行为。局部增强编码器包含一个卷积核宽度为 k 的 1 维卷积和一个非线性 ReLU (rectified linear unit) 激活函数。对于事件 E^i 中的每一个帖子, 1 维卷积捕获输入帖子的 k 个相邻帖子之间的交互。具体来说, $k=3$ 的卷积核允许事件 i 中的 3 个相邻帖子彼此交互, 而具有较大 k 值的卷积核可以同时利用更多的相邻帖子。即

$$l_k = \text{CONV1d}_k(E^i) = E^i * \Gamma_k \quad (12)$$

式中, $*$ 表示卷积操作, Γ_k 表示大小为 k 的卷积核, $\text{CONV1}()$ 表示 1 维卷积操作。本文分别使用 3 个不同的卷积核 ($k=3, 4, 5$) 来获得 3 种尺度的相邻帖子交互, 并将它们串联起来送入最大池化层来获得事件 i 的局部特征。具体为

$$f_{\text{loc}}^i = \varphi_{\max}(\sigma([l_3, l_4, l_5])) \quad (13)$$

式中, $\sigma(\cdot)$ 代表 ReLU, $\varphi_{\max}$ 代表 max pooling。

通过特定的编码策略, 可以从全局特征 f_{glo}^i 、时间特征 f_{tem}^i 以及局部特征 f_{loc}^i 等 3 个层次对事件进行由粗到细的表示, 将这 3 个层次的编码结果连接起来, 可得到事件的多层次编码特征, 即

$$o^i = [f_{\text{glo}}^i, f_{\text{tem}}^i, f_{\text{loc}}^i] \quad (14)$$

2.5 谣言检测层

谣言检测器的目标是识别特定事件是否为谣言。在得到事件的多层次表示后, 利用具有相应激活函数的全连接层来预测事件的真假。具体为

$$r^i = \sigma(W_r \cdot o^i + b) \quad (15)$$

式中, σ 为 softmax 激活函数, r^i 表示第 i 个事件的预测概率, W_r 是全连接层的权重矩阵。使用 r^i 作为第 i 个事件的真实标签, 并且利用交叉熵损失作为检测损失, 则

$$L(\Theta) =$$

$$- \left[\sum_{i=1}^n -y^i \log(r^i) + (1 - y^i) \log(1 - r^i) \right] \quad (16)$$

式中, n 是事件的数量。

2.6 优化

为了构建一个端到端的网络,首先将帖子的图像和文本特征编码成多模态向量。然后,将多模态向量送入多层次编码器中,根据式(7)(11)(13)分别得到事件的全局、时间以及局部特征。最后,通过式(15)得到事件多层次编码特征并经过谣言检测层得到每个事件的概率,通过式(16)计算目标函数、反向传播梯度并更新网络参数。

3 实验

3.1 数据准备

使用 Twitter (Ma 等, 2016) 和 PHEME (Zubiaga 等, 2016) 两个基准数据集来验证本文模型的有效性。两个数据集都是事件级数据, 聚集了与同一事件相关的帖子, 每个事件标记为 1 (谣言) 和 0 (非谣言)。Twitter 数据集是在 snopes.com 上收集的, 每个事件都包含一系列推文。PHEME 数据集是基于 5 条突发新闻收集的, 每条新闻都包含一组事件。实验时, 首先过滤掉少于 10 条帖子的事件, 以平衡谣言数据与非谣言数据的数量。然后每个数据集保留 10% 的实例用于模型调整, 其余数据在所有实验中进行 5 折交叉验证。

3.2 比较方法

选取 SVM-TS (support vector machine-time structure) (Ma 等, 2015)、CNN (convolutional neural network) (Chen 等, 2017)、GRU (gated recurrent unit) (Ma 等, 2016)、CallAtRumors (Chen 等, 2018) 和 MKEMN (multimodal knowledge-aware event memory network)

(Zhang 等, 2019) 等 5 种先进的模型进行比较。

SVM-TS 利用帖子上下文信息的时间序列结构训练线性 SVM 分类器。CNN 使用卷积神经网络, 通过将相关帖子转化为固定长度的序列来学习谣言表示。GRU 使用递归神经网络从用户评论中自动学习时间—语言模式, 以实现有效的谣言检测。CallAtRumors 将注意力机制与 LSTM (long short-term memory) 相结合, 有选择性地学习帖子序列的隐含表示。MKEMN 将多模态内容与外部知识结合起来进行谣言检测, 并使用记忆网络来度量不同事件之间的差异性, 为了与本文模型进行比较, 实验时移除了外部知识。

3.3 评估方法与参数设置

使用准确率 (accuracy) 作为评估指标, 同时, 为了避免数据不平衡时准确率不可靠的情况, 增加了精确度 (precision)、召回率 (recall) 以及 F1 得分 (F1-score) 作为补充评估指标。

参数设置时, 将帖子嵌入维度 d 设置为 1 024, 包括文本嵌入维度 768 和视觉嵌入维度 256。时间编码器隐藏层的维度是 384。模型训练 1 000 个 epoch, 学习率为 0.01, 小批量设置为 32。

3.4 实验结果分析

3.4.1 定量分析

实验结果如表 1 和表 2 所示。可以看出, 1) SVM-TS 在所有模型中表现最差, 表明手工构造的特征较弱, 不足以识别谣言; 2) 大多数基于深度学习的模型 (如 GRU 和 CNN) 都优于基于特征工程的方法 (如 SVM-TS), 表明深度学习方法能够更好地学习事件的隐藏特征; 3) CallAtRumors 比 GRU

表 1 不同方法在 PHEME 数据集上的检测结果对比

Table 1 Comparison of detection results on PHEME dataset among different methods

方法	准确率	谣言			非谣言		
		精确度	召回率	F1 得分	精确度	召回率	F1 得分
SVM-TS	0.634	0.667	0.611	0.638	0.633	0.638	0.635
CNN	0.651	0.655	0.631	0.643	0.663	0.644	0.653
GRU	0.736	0.729	0.739	0.734	0.745	0.714	0.729
CallAtRumors	0.773	0.776	0.771	0.773	0.751	0.776	0.763
MKEMN	0.767	0.766	0.769	0.767	0.752	0.797	0.774
本文	0.822	0.835	0.896	0.864	0.792	0.691	0.738

注: 加粗字体表示各列最优结果。

表2 不同方法在 Twitter 数据集上的检测结果

Table 2 Comparison of detection results on Twitter dataset among different methods

方法	准确率	谣言			非谣言		
		精确度	召回率	F1 得分	精确度	召回率	F1 得分
SVM-TS	0.731	0.735	0.730	0.733	0.744	0.720	0.732
CNN	0.810	0.807	0.820	0.813	0.806	0.813	0.809
GRU	0.815	0.826	0.812	0.819	0.814	0.826	0.820
CallAtRumors	0.824	0.815	0.862	0.838	0.823	0.863	0.841
MKEMN	0.829	0.850	0.822	0.835	0.813	0.834	0.823
本文	0.870	0.831	0.914	0.870	0.910	0.840	0.872

注:加粗字体表示各列最优结果。

和 CNN 具有更好的性能,因为 CallAtRumors 采用了注意力机制,可以更好地提取帖子中特定的局部特征;4) MKEMN 在 Twitter 数据集上比之前的方法具有更好的性能,在 PHEME 数据集上略低于 CallAtRumors,可能是因为 PHEME 中的事件均来自于 5 条突发新闻,事件特征具有高度的相似性;5) 与所有基线相比,本文模型在大多数情况下取得了最好的性能,并且优于其他谣言检测方法。主要因为本文模型利用了帖子的多种模态,以及提出的多层次事件

编码器联合使用多种编码策略,能够逐步生成功能强大、互为补充的特征。

3.4.2 消融测试

为了检验本文提出的多模态多层次事件网络的有效性,从多模态帖子嵌入组件和不同组合事件编码策略两方面对模型的多层次事件编码层不加入多模态信息、不加入全局编码器、不加入时间编码器、不加入局部增强编码器等不同变式在两个数据集上进行比较,实验结果如表 3 和表 4 所示。

表3 模型的不同变式在 PHEME 数据集上的检测性能比较

Table 3 Comparison of performances on PHEME dataset among different variants of the models

模型	准确率	谣言			非谣言		
		精确度	召回率	F1 得分	精确度	召回率	F1 得分
不加入多模态信息	0.820	0.832	0.897	0.863	0.792	0.684	0.734
不加入全局编码器	0.818	0.836	0.889	0.862	0.782	0.696	0.736
不加入时间编码器	0.809	0.817	0.902	0.790	0.790	0.648	0.712
不加入局部增强编码器	0.812	0.822	0.898	0.858	0.789	0.660	0.719
本文	0.822	0.835	0.896	0.864	0.792	0.691	0.738

注:加粗字体表示各列最优结果。

表4 模型的不同变式在 Twitter 数据集上的检测性能比较

Table 4 Comparison of performances on Twitter dataset among different variants of the models

模型	准确率	谣言			非谣言		
		精确度	召回率	F1 得分	精确度	召回率	F1 得分
不加入多模态信息	0.866	0.829	0.900	0.865	0.905	0.831	0.867
不加入全局编码器	0.863	0.827	0.898	0.862	0.900	0.831	0.865
不加入时间编码器	0.849	0.811	0.888	0.848	0.890	0.813	0.850
不加入局部增强编码器	0.856	0.819	0.894	0.855	0.895	0.822	0.857
本文	0.870	0.831	0.914	0.870	0.910	0.840	0.872

注:加粗字体表示各列最优结果。

从表3和表4可以看出,1)在多模态帖子嵌入组件的效果方面,在两个数据集中,本文模型优于不加入多模态信息的变体,表明多模态帖子嵌入层学习的多模态特征可以提高谣言检测的性能。2)在不同组合事件编码策略的效果方面,本文模型在不加入全局特征、时间特征以及局部特征时,在两个数据集中的性能都有所下降,并且不加入时间特征的变式下降得更加明显。在包含全部编码策略时,可以在两个数据集上实现最佳性能,说明本文提出的多层次编码策略可以在编码事件特征中发挥作用。

4 结 论

现有的谣言检测方法有两个局限性,其一是主要关注事件的时间特征,忽略了与时间特征互补的全局与局部特征,其二是忽略了帖子中的多模态信息。对此,本文提出一个端到端的用于谣言检测的多模态多层次事件网络,具有两点创新:1)融合帖子的文本与视觉信息;2)提出一种多层次的事件编码策略,以更加精细的方式同时捕获事件的全局、时间以及局部信息。在两个基准数据集上的实验表明,本文提出的多模态多层次事件网络取得了较好的检测效果,比现有方法在准确率上提升了4%以上。但本文工作还存在一些不足,如:未考虑不同层次编码特征的贡献。下一步将完善本文提出的方法,采用注意力机制对3个层次的编码特征进行融合。

参考文献 (References)

- Castillo C, Mendoza M and Poblete B. 2011. Information credibility on twitter//Proceedings of the 20th International Conference on World Wide Web. Hyderabad, India; ACM: 675-684 [DOI: 10.1145/1963405.1963500]
- Chen T, Li X, Yin H Z and Zhang J. 2018. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection//Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining. Melbourne, Australia; Springer: 40-52 [DOI: 10.1007/978-3-030-04503-6_4]
- Chen Y C, Liu Z Y and Kao H Y. 2017. IKM at SemEval-2017 task 8: convolutional neural networks for stance detection and rumor verification//Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada; Association for Computational Linguistics: 465-469 [DOI: 10.18653/v1/S17-2081]
- Devlin J, Chang M W, Lee K and Toutanova K. 2019. BERT: pre-training of deep bidirectional transformers for language understanding//Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, USA; Association for Computational Linguistics: 4171-4186 [DOI: 10.18653/v1/n19-1423]
- Jin Z W, Cao J, Guo H, Zhang Y D and Luo J B. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs//Proceedings of the 25th ACM international conference on Multimedia. Mountain View, USA; ACM: 795-816 [DOI: 10.1145/3123266.3123454]
- Khattar D, Goud J S, Gupta M and Varma V. 2019. MVAE: multimodal variational autoencoder for fake news detection//Proceedings of World Wide Web Conference. San Francisco, USA; ACM: 2915-2921 [DOI: 10.1145/3308558.3313552]
- Kwon S, Cha M, Jung K, Chen W and Wang Y J. 2013. Prominent features of rumor propagation in online social media//Proceedings of the 13th IEEE International Conference on Data Mining. Dallas, USA; IEEE: 1103-1108 [DOI: 10.1109/ICDM.2013.61]
- Ma J, Gao W and Wei Z. 2015. Detect rumors using time series of social context information on microblogging websites//Proceedings of the 24th ACM International Conference on Information and Knowledge Management. Melbourne, Australia: 1751-1754 [DOI: 10.1142/9789813223615_0006]
- Ma J, Gao W, Mitra P, Kwon S, Jansen B J, Wong K F and Cha M. 2016. Detecting rumors from microblogs with recurrent neural networks//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA; IJCAI/AAAI Press: 3818-3824
- Ma J, Gao W and Wong K F. 2018a. Detect rumor and stance jointly by neural multi-task learning//Proceedings of Web Conference 2018. Lyon, France; ACM: 585-593 [DOI: 10.1145/3184558.3188729]
- Ma J, Gao W and Wong K F. 2018b. Rumor detection on Twitter with tree-structured recursive neural networks//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia; ACL: 1980-1989 [DOI: 10.18653/v1/P18-1184]
- Sanh V, Debut L, Chaumond J and Wolf T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter [EB/OL]. [2020-08-06]. <https://arxiv.org/pdf/1910.01108.pdf>
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA; ICLR
- Singhal S, Shah R R, Chakraborty T, Kumaraguru P and Satoh S. 2019. SpotFake: a multi-modal framework for fake news detection//Proceedings of the 5th IEEE International Conference on Multimedia Big Data. Singapore, Singapore; IEEE: 39-47 [DOI: 10.1109/

BigMM. 2019. 00-44]

Sun C, Qiu X P, Xu Y G and Huang X J. 2019. How to fine-tune BERT for text classification?//Proceedings of the 18th China National Conference on Chinese Computational Linguistics. Kunming, China; Springer: 194-206 [DOI: 10.1007/978-3-030-32381-3_16]

Wu L W, Rao Y, Jin H L, Nazir A and Sun L. 2019. Different absorption from the same sharing: sifted multi-task learning for fake news detection//Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China; Association for Computational Linguistics: 4644-4653 [DOI: 10.18653/v1/D19-1471]

Yang F, Liu Y, Yu X H and Yang M. 2012. Automatic detection of rumor on Sina Weibo//Proceedings of ACM SIGKDD Workshop on Mining Data Semantics. Beijing, China; ACM: 1-7 [DOI: 10.1145/2350190.2350203]

Yu F, Liu Q, Wu S, Wang L and Tan T N. 2017. A convolutional approach for misinformation identification//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia; Ijcai. org: 3901-3907 [DOI: 10.24963/ijcai.2017/545]

Zhang H W, Fang Q, Qian S S and Xu C S. 2019. Multi-modal knowledge-aware event memory network for social media rumor detection//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France; ACM: 1942-1951 [DOI: 10.1145/3343031.3350850]

Zubiaga A, Liakata M, Procter R, Hoi G W S and Tolmie P. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PLoS One, 11 (3): #e0150989 [DOI: 10.1371/journal.pone.0150989]

作者简介



李莎, 1996 年生, 女, 硕士研究生, 主要研究方向为多媒体计算分析。

E-mail: 18737909726@163.com



徐常胜, 通信作者, 男, 研究员, 主要研究方向为多媒体内容分析索引检索、模式识别与计算机视觉。

E-mail: csxu@nlpr.ia.ac.cn

张怀文, 男, 博士研究生, 主要研究方向为社交多媒体数据挖掘。E-mail: huaiwen.zhang@nlpr.ia.ac.cn

钱胜胜, 男, 副研究员, 主要研究方向为社交多媒体数据挖掘、社交事件内容分析。E-mail: shengsheng.qian@nlpr.ia.ac.cn

方全, 男, 副研究员, 主要研究方向为多媒体知识计算。

E-mail: qfang@nlpr.ia.ac.cn