

融合多模态内容语义一致性的社交媒体虚假新闻检测^{*}

张国标^{1,2} 李 洁³

¹(武汉大学信息管理学院 武汉 430072)

²(武汉大学信息检索与知识挖掘研究所 武汉 430072)

³(苏州大学社会学院 苏州 215000)

摘要:【目的】实现社交媒体虚假新闻早期检测,遏制虚假信息的广泛传播。【方法】在同时利用图像与文本特征的基础上,通过将图像映射为语义标签,设计了一种图像与文本内容语义一致性计算方法,构建虚假新闻检测模型,并采用虚假新闻检测标准数据集 FakeNewsNet 验证模型的性能。【结果】融合新闻图像与文本语义一致性特征的全特征模型在 PolitiFact 数据上的检测 F1 值达到 0.775,在 GossipCop 数据上的 F1 值达到 0.879,说明该模型具有良好的检测效果。【局限】由于现有图像语义标注模型标注能力的局限性,尚无法准确描述图像内容,所计算的语义一致性存在误差。【结论】多模态特征融合能够有效提升虚假新闻检测性能,本文构建的新闻文本与图像语义一致性特征能够丰富和拓展虚假新闻检测依据。

关键词: 虚假新闻检测 社交媒体 多模态特征融合 语义一致性 深度学习

分类号: TP393

DOI: 10.11925/infotech.2096-3467.2020.0884

引用本文: 张国标, 李洁. 融合多模态内容语义一致性的社交媒体虚假新闻检测[J]. 数据分析与知识发现, 2021, 5(5): 21-29. (Zhang Guobiao, Li Jie. Detecting Social Media Fake News with Semantic Consistency Between Multi-model Contents[J]. Data Analysis and Knowledge Discovery, 2021, 5(5): 21-29.)

1 引言

随着社交媒体的飞速发展,资讯的分享与获取达到前所未有的便利,虚假新闻(Fake News)也乘势泛滥。同时,人工智能技术的不断应用,越来越多的研究人员寄希望于通过技术手段实现虚假新闻的自动检测,以扼制虚假新闻的传播。目前,研究人员分别从新闻内容^[1]、来源^[2]、传播链路^[3]等角度出发设计了大量虚假新闻判别特征,并结合机器学习分类方法及深度学习网络设计了不同的虚假新闻自动检测模型。其中,针对新闻内容的检测,因数据易获取

且能够实现虚假新闻的早期检测,受到越来越多的关注,但由于内容特征语言模式高度依赖于特定事件及其领域场景,难以适用于普适性的虚假新闻早期检测任务。为克服这一难点,本文利用文本和视觉信息之间的差异,构建更具内容特征表示普适性的虚假新闻检测模型,聚焦文本和图像特征的跨模态语义偏差问题,利用语义距离测度,实现多模态内容特征的语义一致性检测。在虚假新闻的检测过程中,如果一则新闻并非真实发生而是人为编造的,与故事内容相配的图像往往都是有误导性或篡改过

通讯作者(Corresponding author): 李洁(Li Jie), ORCID: 0000-0002-6178-4377, E-mail: allison_lijie@163.com。

*本文系苏州大学 2020 年人文社会科学优秀学术团队(项目培育)项目(项目编号: NH33711520)的研究成果之一。

The work is supported by Soochow University 2020 Humanities and Social Sciences Excellent Academic Team Project (Grant No. NH33711520)。

的^[4]。例如,图1是一则关于美国政府购买30 000台断头台的虚假新闻,由于没有真实照片,其配图选用了法国玛丽王后被砍头的绘画,新闻文字与配图是不一致的。针对这种文字内容和图像内容之间存在的语义偏差现象,本文采用语义一致性特征作为虚假新闻判别的新依据,进行跨模态语义一致性的虚假新闻早期检测。



图1 虚假新闻图文内容语义不一致示例

Fig.1 An Example of Image and Text Semantic Inconsistency of Fake News

2 相关研究

早期的虚假新闻检测研究主要是利用新闻文字内容捕捉虚假新闻与真实新闻的写作差异,构建虚假新闻检测模型。Castillo等^[5]通过统计文本中的字数、标点符号、表情、超链接等的出现频次,设计出一种简单的推特信息真实性评估模型。Rashkin等^[6]使用更为复杂的语法信息与心理语言特征工具LIWC设计了多种语言特征,并结合LSTM网络构建虚假信息识别模型。这些方法依赖手工设计特征,费时费力,且需要专业领域知识,无法满足大数据时代数据频繁处理的需求。深度学习的发展为自动特征提取提供了一种解决方案,研究人员纷纷采用深度学习方法构建虚假新闻检测模型。Ma等^[7]通过抽取相关事件推文组成新闻事件,利用词嵌入和RNN模型对虚假新闻事件进行检测,证明了RNN模型在虚假新闻检测中的有效性。Popat等^[8]利用新闻言论语句和外部证据语句结合Bi-LSTM和注意

力机制,设计了一种端到端的言论验证模型。无论是手工设计的特征还是深度学习自动抽取的特征,都只能识别文本中的虚假信息,无法识别由图像虚假造成的虚假新闻。

描述同一新闻事件的不同模态数据,其内容往往是相互联系的,不同模态数据能够起到相互补充的作用。研究人员不断尝试将新闻文章中的视觉信息融合到虚假新闻检测任务中,以实现图文特征互补。Jin等^[9]利用深度神经网络将新闻图像信息引入到虚假新闻检测中,通过预训练的VGG19模型抽取与事件相关的图像语义特征,采用注意力机制抽取文本与社交上下文中的关键信息,并与图像语义特征进行逐元素相乘,以调整视觉语义特征的权重。实验表明,该方法可以发现许多单一模态下难以判别的虚假新闻案例。然而,多模态特征表示仍然高度依赖于数据集中的特定事件,这些特征难以进行迁移并且会降低模型的泛化能力,导致无法识别有关新事件的虚假新闻。因此,Wang等^[10]提出一种基于对抗网络的端到端模型,认为应当引导模型学习更具泛化能力的事件无关特征。该模型基于VGG19抽取视觉模态语义特征,同时采用Text-CNN抽取文本模态特征,并融合两种模态特征作为虚假新闻的内容特征,可同时判断该新闻是否为虚假新闻以及该新闻来自哪一类事件。Khattar等^[11]认为文本模态特征与视觉模态特征的简单拼接难以充分表达两个模态之间的交互与关联,因此,研究使用了编码-解码的方式构建多模态的特征表达。在多模态可变自动编码模型中,文本模态与视觉模态的拼接特征被编码为一个中间表达,利用重构损失保证编码后的中间表达可以解码回原状态,并利用该中间表达的向量进行虚假新闻检测。Sing等^[12]分别从内容、组织、情感和操纵4个维度手工设计了文本和图像特征,并采用特征串联的方式对各种特征进行融合,利用多种传统的机器学习分类算法实现了虚假新闻检测。

综上,已有的虚假新闻检测研究已经开始综合利用文本与图像,从跨模态角度设计可应用于虚假新闻检测的判别特征,但对于深挖新闻文本与图像特征之间的关联关系,充分利用图文间的语义偏差现象进行虚假新闻的有效识别与检测,还未形成系

统研究。而文本与图像内容语义是否一致可以为检测虚假新闻提供重要的判别依据。因此,本文从新闻文本与图像内容的语义一致性出发,利用虚假新闻相较于真实新闻的低语义一致性假设,构建融合图文内容语义一致性的虚假新闻检测模型,实现虚假新闻的早期检测。

3 新闻图文内容语义一致性映射

作为对特定现实的陈述,文字和图像均为传播信息、完成叙事的重要媒介。同样,新闻文本与图像在陈述特定现实的过程中也存在一种相互补充、相互呼应的共同叙事关系,并遵循图像附属于文字、图像与文字互仿的基本逻辑^[13]。新闻图像通常显示的是事件发生时的核心信息,新闻文字则对该事件进行语言描述,两者从不同侧面表达同一语义。从认知心理来看,文字和图像实际上对应于不同的接受心理,而且不同的图文组合方式又对应于不同的信

息加工机制。在信息的呈现方式上,文字表征只能沿着固定的线性逻辑展开,而图像表征是在空间维度上延伸的^[14]。文字遵循的是时间逻辑,而图像则遵循空间逻辑,两者之间存在逻辑性和异质性差异。因此,不可直接度量两者语义一致性。

为实现文字与图像语义一致性的度量,研究采用基于语义分布的映射方法,其核心是充分利用图像高层语义组的语义标签信息,将图像转换为文字标签,利用预设定的语义标签将不同形态数据映射到同一语义空间,以方便对比语义一致性。具体地,运用 ImageNet 数据集^[15]提供的语义标签组以及在 ImageNet 数据集上预训练的卷积神经网络(Convolutional Neural Network, CNN)^[16]模型对新闻图像进行语义标注,从而实现新闻图像到文本的转换,即图像到文本的语义映射。在通过 CNN 进行语义映射的过程中,能够得到图像的语义标签及对应的标签概率,如图 2 所示。

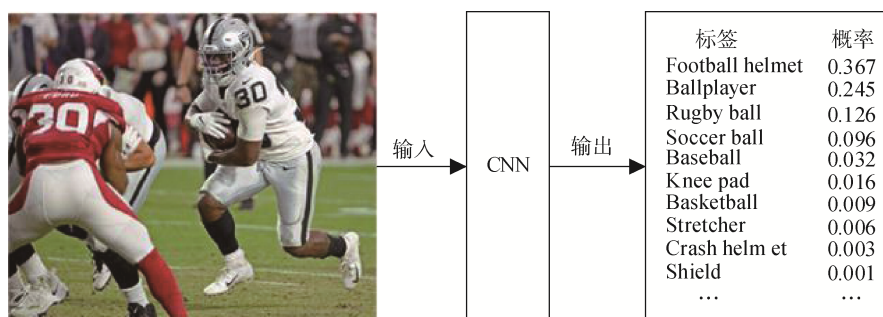


图 2 图像标签映射过程

Fig.2 Image Label Mapping Process

对于同时包含图像和文本内容的社交媒体新闻内容,其之间的语义一致性可定义为:一条新闻包含文本内容 T 和视觉内容 V ,其中文本可以表示为由 n 个词组成的集合 $T = \{w_1, w_2, \dots, w_n\}$,并采用 Word2Vec^[17] 方式将每个词转化为向量 $E_i = [e_{i1}, e_{i2}, \dots, e_{ij}]$,其中 i 表示第 i 个词, j 表示词向量的维度;对于视觉内容 V ,通过预训练的 CNN 将其映射为由 m 个标签组成的标签集合 $V = \{t_1, t_2, \dots, t_m\}$,同样通过 Word2Vec 方式转化为词向量。Word2Vec 能够将词汇嵌套于一个连续的向量空间中,语义近似的词汇可以被映射为相邻的数据点,从而可以体现不同词汇的语义相似度。为避免矩阵维数灾难和输入文本长度的变化,使用词向量相加的方法,通过计

算每个维度的均值,得到与单个词相同维数的向量作为文字内容或图像内容的语义向量 **Embedding**^[18],如公式(1)所示。

$$\text{Embedding} = [e_1, e_2, \dots, e_j] = \sum_{i=1}^n \frac{[e_{i1}, e_{i2}, \dots, e_{ij}]}{n} \quad (1)$$

文本与图像内容语义一致性可以通过计算文本语义向量 **Embedding**(T) 和图像语义向量 **Embedding**(V) 之间的余弦距离来度量,如公式(2)所示。

$$\text{Consistency}(T, V) = \frac{\text{Embedding}(T) \cdot \text{Embedding}(V)}{\|\text{Embedding}(T)\| \times \|\text{Embedding}(V)\|} \quad (2)$$

4 融合图文语义一致性的社交媒体虚假新闻检测模型

本研究的主要目标是基于社交媒体新闻推文中的文字和图像数据,使用深度学习方法构建虚假新闻检测模型,以此识别虚假新闻。由于虚假新闻本质上是发布者对新闻报道的扭曲与偏见,根据Gentzkow等^[19]对媒体偏差理论的研究,这种扭曲和

偏见通常被建模为二元分类问题。模型输入数据为包含文字和图片的推文,输出为对应的真假新闻标签,0为真实新闻,1为虚假新闻。为实现文本与图像特征的有效提取,以及文本与图像内容语义一致性的有效衡量,分别采用不同的特征抽取方法抽取文本、图像和两者语义一致性特征,最终将融合后的特征向量用于虚假新闻检测,模型总体框架如图3所示。

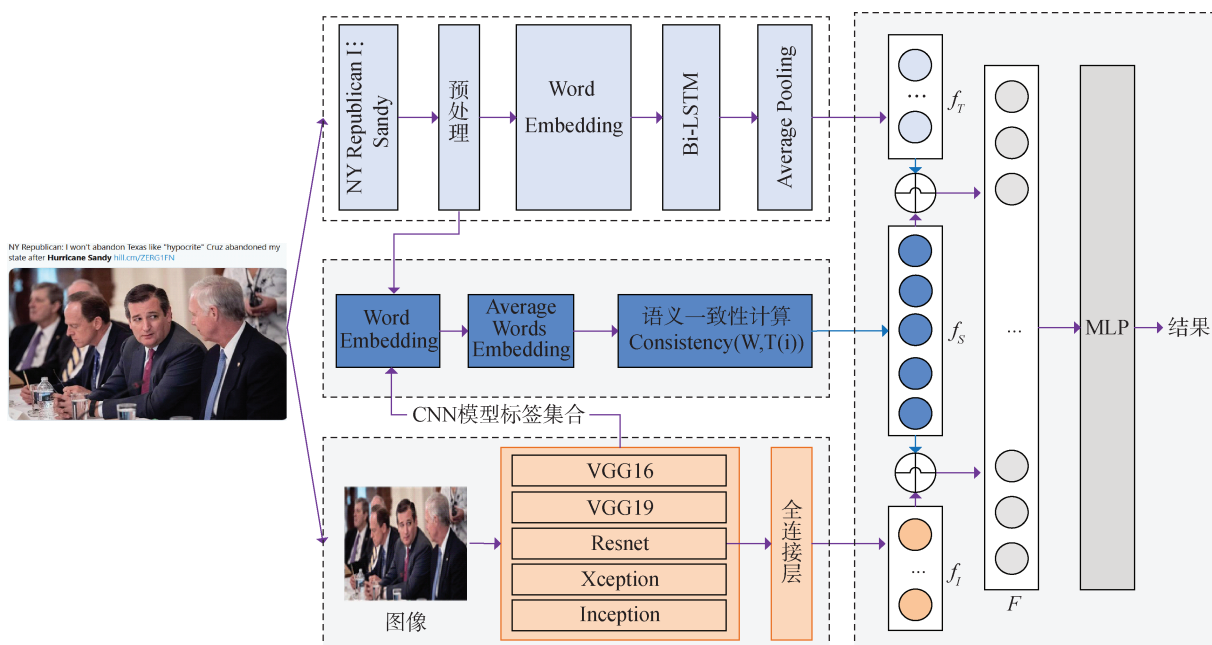


图3 多模态特征融合的社交媒体虚假新闻检测模型

Fig.3 Social Media Fake News Detection Model Based on Multi-modal Feature Fusion

4.1 新闻内容特征抽取

(1) 文本特征抽取

文本特征在文本分析中扮演着至关重要的角色,直接影响到虚假新闻检测的准确率。对于每一条推文文本,在通过分词、去停用词等预处理之后,采用双向长短期记忆(Bi-directional Long Short Term Memory, Bi-LSTM)模型^[20]来抽取文本特征,以捕捉文本中潜在的依存关系。Bi-LSTM模型不仅可以捕获前面词汇对后面词汇的影响,同时可以捕捉后面词汇对前面词汇的影响,从而提高模型提取句子语义的能力,其输入为通过词向量(Word Embedding)转化的文本词向量 W ,输出为每一步的隐状态集合 H 。最终,将均值池化(Average Pooling)

后的结果作为文本特征 f_T 。计算过程如公式(3)和公式(4)所示。

$$H = Bi - LSTM([W]) \quad (3)$$

$$f_T = AveragePooling(H) \quad (4)$$

(2) 图像特征抽取

人类大脑处理视觉信息的速度是处理文字信息的60 000倍^[21],视觉信息不仅能够缩小沟通差距帮助人类理解信息内容,而且能够增强人类记忆^[22]。因此,模型同时融合了图像特征,并采用目前图像分类效果最好的Resnet网络^[23]抽取图像特征,并将Resnet网络输出的数据输入到一个全连接层以降低图像特征维度,得到图像特征 f_I ,其过程可以表示为如公式(5)所示。

$$f_i = \sigma(w \cdot \text{Resnet} + b) \quad (5)$$

其中, w 、 b 分别表示全连接层的权重和偏差, σ 表示激活函数。

(3) 文本与图像语义一致性特征抽取

文本与图像语义一致性特征通过计算推文文字与图像内容语义相似度来抽取。根据前文设计的图像与文本语义一致性计算方法, 对文本内容采用谷歌提供的 300 维的 GoogleNews^[24] 词向量实现文本向量化表示; 对于图像内容, 分别采用 5 种预训练的 CNN 模型 (VGG16^[25], VGG19^[25], Resnet, Inception^[26], Xception^[27]) 对图像进行语义标注, 并选取各模型输出的概率大于 0.01 的标签作为图像语义标签。此 5 种 CNN 模型, 分别采用不同的卷积核数量和卷积核大小实现图像数据的抽象表示, 能够识别图像不同层面的语义信息, 增强图像语义表达的全面性。同样采用 GoogleNews 词向量对图像语义标签进行向量化表示, 将单词映射至向量空间, 使用向量余弦计算文本与图像语义一致性特征, 将文本向量分别与 5 种模型的标签向量进行相似度计算, 最终得到包含 5 个元素的语义一致性特征 f_s , 如公式 (6) 所示。

$$f_s = [\text{Consistency}_1, \text{Consistency}_2, \text{Consistency}_3, \text{Consistency}_4, \text{Consistency}_5] \quad (6)$$

4.2 特征融合与新闻分类

特征融合即输入两个模态的特征向量, 输出融合后的向量。研究采用串联拼接的方式实现特征融合。对于每一条推文, 分别用 f_t 表示文本特征, f_i 表示图像特征, f_s 表示语义一致性特征, 融合后的推文特征可以表示为 F , 如公式 (7) 所示。

$$F = [f_t \oplus f_i \oplus f_s] \quad (7)$$

最终, 采用包含三层全连接网络的多层感知器 (Multi-Layer Perceptrons, MLP) 作为新闻分类器, 通过 Softmax 函数输出分类结果。

5 实验与分析

5.1 实验数据

为评估所构建的社交媒体虚假新闻检测模型性能, 采用虚假新闻检测标准数据集 FakeNewsNet^[28] 所提供的网页新闻数据作为原始数据, 并以其推特转发数据验证模型的有效性。FakeNewsNet 数据集是从事实验证网站 (politifact.com 和 gossipcop.com)

收集到的新闻网页数据, 包含政治和娱乐明星两种新闻, 其中政治类虚假新闻、真实新闻数量分别为 432 和 624 条, 娱乐明星类虚假新闻、真实新闻数量分别为 5 323 和 16 817 条。由于 GossipCop 数据中虚假新闻和真实新闻数据严重不平衡, 随机抽取与 GossipCop 虚假新闻相同数量的 5 323 条真实新闻作为实验中的真实新闻数据。并按照 7:1:2 的比例将新闻数据划分为训练集、验证集和测试集, 采用 Python 和 Twitter API 获取新闻的推文数据, 并对没有图像数据的推文进行过滤, 得到 8 081 条 PolitiFact 新闻推文数据和 46 655 条 GossipCop 新闻推文数据, 最终实验数据为推特上的新闻转发数据, 如表 1 所示。

表 1 FakeNewsNet 实验数据

Table 1 FakeNewsNet Experimental Data

项目	PolitiFact		GossipCop	
	Fake	True	Fake	True
训练集	2 466	3 190	14 737	17 922
验证集	352	456	2 105	2 560
测试集	705	912	4 210	5 121
总计	3 523	4 558	21 052	25 603

5.2 实验设置

实验在 Ubuntu16.4 操作系统和 Python3.6 编程环境下进行, 采用 TensorFlow 深度学习框架构建虚假新闻检测模型并进行模型训练, 为防止过拟合并提高模型鲁棒性, 采用指数衰减法优化深度学习速率, 并在模型的全连接层添加了 Dropout。在训练过程中采用 Early Stop 策略, 并使用自动调参工具 Talos^[29] 优化模型参数, 具体参数如表 2 所示。

表 2 实验参数设置

Table 2 Experimental Parameter Settings

参数	参数值
Epoch	50
Dropout	0.4
Batch_size	32
激活函数	ReLU
学习率	0.0001
图像全连接层神经元个数	200
MLP 各层神经元个数	500, 200, 100

5.3 实验结果与分析

实验分别运用 PolitiFact 和 GossipCop 数据检验

不同特征模型的性能,比较不同特征的检测性能。采用分类任务常用的准确度(Accuracy)、精确率(Precision)、召回率(Recall)和F1指标评价虚假新闻检测模型的性能,使用宏平均方式计算各类别指标数值,得到最终评价指标数值。通过设置对照组分别检验不同特征模型的性能,并选取多模态虚假新闻检测任务中最具代表性的EANN模型^[10]作为基准模型,实验结果如表3所示。

表3 虚假新闻检测结果

Tabel 3 Fake News Detection Results

特征类型	PolitiFact				GossipCop			
	准确度	精确率	召回率	F1	准确度	精确率	召回率	F1
文本特征	0.761	0.768	0.773	0.753	0.836	0.810	0.821	0.815
图像特征	0.540	0.520	0.560	0.520	0.654	0.704	0.702	0.653
语义一致性特征	0.520	0.450	0.524	0.480	0.564	0.530	0.545	0.548
文本与图像特征	0.782	0.784	0.813	0.770	0.857	0.827	0.838	0.836
全部特征	0.791	0.792	0.803	0.775	0.883	0.864	0.853	0.879
EANN	0.776	0.764	0.798	0.768	0.841	0.814	0.796	0.806

从表3中可以看出,本文所提出的虚假新闻检测模型在PolitiFact和GossipCop两种数据中表现性能均优于对比模型EANN,在使用全部特征时模型准确率和F1值均取得了最大值。对比仅使用单一特征时的检测结果,可以发现文本特征检测性能明显优于图像特征和语义一致性特征。这一结果说明新闻文本是实现虚假新闻检测的主要依据,新闻图像和语义一致性特征仅能起到一定辅助作用。这是因为图像的功能在于给人以直观的印象,而文字的效用则是准确地表达信息内容和含义,说明更深层次的概念,揭示隐藏于直观表象之下的内在涵义,是新闻作者表达观点思想的主要媒介^[12]。语义一致性特征的检测效果低于其他特征,这是因为现有CNN模型所采用的语义标签集有限,尚无法完整准确地标识图像内容语义,所计算的图文语义一致性存在偏差。通过对比组合特征可以发现,语义一致性特征能够有效提升模型检测性能,且在GossipCop数据上性能提升更为明显。这是因为政治类新闻信息更为庞大、内容语义更为复杂,语义一致性特征对政治类新闻图文语义差异表示能力较弱。

5.4 语义一致性特征有效性分析

为具体分析研究所设计的反映新闻文本与图像语义一致性的特征的有效性,分别统计了两类数据中的真实新闻和虚假新闻在使用5种CNN模型标注图像语义时计算所得的语义一致性的均值,结果如图4所示。

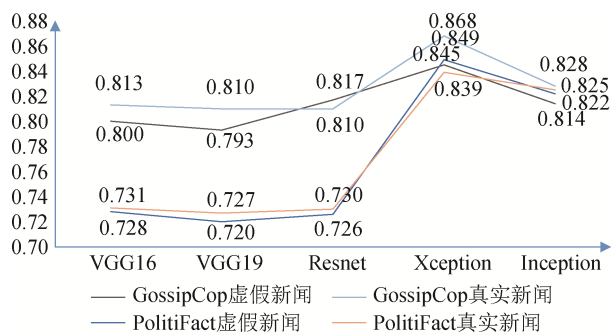


图4 各CNN模型语义一致性均值

Fig.4 Average Value of Semantic Consistency of Each CNN Model

从图4可以发现,两类数据均有4种CNN模型表现出真实新闻语义一致性均值高于虚假新闻,在娱乐明星类GossipCop数据中仅Resnet模型出现真实新闻语义一致性均值小于虚假新闻的情况;在政治类PolitiFact数据中仅Xception模型出现真实新闻语义一致性均值小于虚假新闻的情况。通过对比语义一致性均值在两类新闻数据中的差异,可以发现娱乐明星类数据中真实新闻与虚假新闻语义一致性差异较大,政治类真假新闻语义一致性差距较小,且在实验结果表3中仅使用语义一致性特征时娱乐明星类数据的检测效果要优于政治类新闻的检测效果。这一现象可以说明,政治类新闻所用配图差异性小于娱乐明星类新闻,语义一致性特征更适用于娱乐明星类新闻。

此外,研究还选取了具体新闻案例进行分析,如图5所示,分别给出了各类真假新闻案例及对应的新闻图像与文本语义一致性数值。

通过观察发现,虚假新闻的语义一致性低于真实新闻是因为虚假新闻往往无法找到真实的现场图片,虚假新闻文字内容无法针对现场图片进行描述,例如图5(a)和图5(b)所示,新闻作者分别采用多幅照片拼接而成的图像和漫画,对应文字内容分别采



图 5 新闻图文语义一致性案例

Fig.5 Examples of News Text and Image Semantic Consistency

用疑问和回答式的语句。而在真实新闻中,例如图 5(c)和图 5(d)的新闻作者直接采用了真实的现场图片,且文字内容均为描述性的陈述语句,从而使得文字内容与图像内容语义一致性较高。

总体上,本研究所提出的虚假新闻语义一致性低于真实新闻的假设在政治类和娱乐明星类新闻数据中得到了验证,且多能够从真实现场图片的缺失和篡改等问题中找到依据。

6 结 语

虚假新闻的泛滥不仅损害了新闻媒体的公信力,而且还污染了网络信息环境,对社交媒体健康发展有着巨大的危害。本文针对社交媒体虚假新闻检测这一挑战,在虚假新闻文本内容基础上,综合考虑新闻图像内容、文本与图像语义一致性等判别信息,构建了虚假新闻检测模型,有效提升了虚假新闻检测性能。通过比较不同特征的实验结果和不同类型

新闻语义一致性均值,发现文本与图像语义一致性特征能够有效区分虚假新闻和真实新闻,并且在不同类型新闻中的表现效果不同,在娱乐明星类新闻中的表现要优于政治类新闻。然而,由于预训练的 CNN 模型所提供的语义标签有限,无法完全描绘图像内容,还需探索新的图像语义描述方法。因此,如何有效揭示新闻文本与图像内容关联关系,并在更多类型的新闻数据中验证其有效性将是未来基于新闻内容的虚假新闻检测研究的重点。

参考文献:

- [1] Aldwairi M, Alwahedi A. Detecting Fake News in Social Media Networks[J]. Procedia Computer Science, 2018, 141: 215-222.
- [2] Kim A, Moravec P L, Dennis A R. Combating Fake News on Social Media with Source Ratings: The Effects of User and Expert Reputation Ratings[J]. Journal of Management Information Systems, 2019, 36(3): 931-968.
- [3] Shu K, Mahudeswaran D, Wang S, 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. 2020.
- [4] Qi P, Cao J, Yang T, et al. Exploiting Multi-domain Visual Information for Fake News Detection[C]//Proceedings of the 19th IEEE International Conference on Data Mining (ICDM), Beijing, China. USA: IEEE, 2019.
- [5] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter[C]//Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India. New York, USA: ACM, 2011.
- [6] 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, Copenhagen, Denmark. USA: ACL, 2017.
- [7] 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 (IJCAI), New York, USA. New York, USA: ACM, 2016.
- [8] Popat K, Mukherjee S, Yates A, et al. DeClarE: Debunking Fake News and False Claims Using Evidence-Aware Deep Learning [C]//Proceeding of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP), Brussels, Belgium. USA: ACL, 2018:22-32.
- [9] Jin Z, Cao J, Guo H, et al. Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs[C]// Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, USA. New York, USA: ACM, 2017: 795-816.
- [10] Wang Y, Ma F, Jin Z, 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 & Data Mining, London, UK. New York, USA: ACM, 2018.
- [11] 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. ACM, 2019.
- [12] Sing V K, Ghosh I, Sonagara D. Detecting Fake News Stories via Multimodal Analysis[J]. Journal of the Association for Information Science and Technology, 2021, 72(1):3-17.
- [13] 鲍远福. 新媒体文本表意论:从“语图关系”到“语图间性”[J]. 南京邮电大学学报(社会科学版), 2016, 18(1): 11-22. (Bao Yuanfu. Ideographic Text of New Media: From “Language-icon Relationship” to “Language-Icon Intertextuality” [J]. Journal of Nanjing University of Posts and Telecommunications (Social Science), 2016, 18(1):11-22.)
- [14] Gombrich E H. The Image and the Eye: Further Studies in the Psychology of Pictorial Representation[M]. Oxford: Phaidon Press, 1982: 150.
- [15] Deng J, Dong W, Socher R, et al. ImageNet: A Large-scale Hierarchical Image Database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision & Pattern Recognition, Miami, USA. USA: IEEE, 2009.
- [16] Krizhevsky A, Sutskever I, Hinton G E. Imagenet Classification with Deep Convolutional Neural Networks[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, USA. 2012:1097-1105.
- [17] 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. 2013:3111-3119.
- [18] Maas A L, Daly R E, Pham P T, et al. Learning Word Vectors for Sentiment Analysis[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, USA. New York, USA: ACM, 2011:142-150.
- [19] Gentzkow M, Shapiro J M, Stone D F. Media Bias in the Marketplace: Theory[R]. National Bureau of Economic Research, Inc., 2014: 623-645.
- [20] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [21] Jibril T A, Abdullah M H. Relevance of Emoticons in Computer-Mediated Communication Contexts: An Overview[J]. Asian Social Ence, 2013, 9(4):201-207.
- [22] Yoon J, Chung E. Image Use in Social Network Communication: A Case Study of Tweets on the Boston Marathon Bombing[J]. Information Research, 2016, 21(1): 106-116.
- [23] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA. USA: IEEE, 2016.
- [24] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[OL]. arXiv Preprint, arXiv: 1301.3781.
- [25] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-scale Image Recognition[OL]. arXiv Preprint, arXiv: 1409.1556.
- [26] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions [C]//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, USA. USA: IEEE, 2015.
- [27] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA. USA: IEEE, 2017.
- [28] Shu K, Mahudeswaran D, Wang S, et al. Fakenewsnet: A Data

Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media[OL]. arXiv Preprint, arXiv: 1809.01286.

[29] Autonomio Talos[EB/OL]. [2020-11-07]. <http://github.com/autonomio/talos>.

作者贡献声明:

张国标:提出研究思路,实验验证,论文撰写与修订;
李洁:数据预处理,论文修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储, E-mail: zgb0537@whu.edu.cn。

[1] 张国标. FakeNewsNet_Twitter_data.rar. 虚假新闻推文数据。

[2] 张国标. FakeNewsDetection_results_examples_data.rar. 图 1-图 5、表 1-表 3 的原始数据和生成数据。

收稿日期:2020-09-08

收修改稿日期:2020-11-08

Detecting Social Media Fake News with Semantic Consistency Between Multi-model Contents

Zhang Guobiao^{1,2} Li Jie³

¹(School of Information Management, Wuhan University, Wuhan 430072, China)

²(Institute for Information Retrieval and Knowledge Mining, Wuhan University, Wuhan 430072, China)

³(School of Sociology, Soochow University, Suzhou 215000, China)

Abstract: [Objective] This study aims to detect fake news on social media earlier and curb the dissemination of mis/dis-information. [Methods] Based on the features of news images and texts, we mapped the images to semantic tags and calculated the semantic consistency between images and texts. Then, we constructed a model to detect fake news. Finally, we examined our new model with the FakeNewsNet dataset. [Results] The F1 value of our model was up to 0.775 on PolitiFact data and 0.879 on GossipCop data. [Limitations] Due to the limits of existing annotation methods for image semantics, we could not accurately describe image contents, and calculate semantic consistency. [Conclusions] The constructed model could effectively detect fake news from social media.

Keywords: Fake News Detection Social Media Multi-modal Feature Fusion Semantic Consistency Deep Learning