

基于新闻标题-正文差异性的虚假新闻检测方法^{*}

刘 赏 沈逸凡

(天津财经大学理工学院 天津 300222)

摘要:【目的】为解决在虚假新闻检测中新闻评论难以收集、新闻文本简短难以提取有效信息的问题,本文给出一种基于新闻标题和正文差异性的虚假新闻检测方法。【方法】首先,设计 Cos-Gap 差异性计算方法以获取新闻标题-正文在文本和情感上的差异性特征;然后,根据获得的差异性特征,以异构图注意网络为基础,构建新闻差异性异构图网络 NDHN。该网络既包含基于差异性特征构造的边,也包含基于语义特征和情感特征构建的标题、正文和情感三种类型节点。【结果】在 GossipCop 开放数据集上的实验结果显示,本文提出的检测方法在分类准确率上提升约 2.7 个百分点,F1 指标提升约 3.2 个百分点。【局限】本文方法适用于带有标题的新闻,对于微博、Twitter 等无标题文本存在局限。【结论】融合新闻差异性特征可以有效提升虚假新闻检测准确率,为社交媒体快速检测出虚假新闻提供有力支持。

关键词: 虚假新闻检测 异构图网络 差异性特征 舆情分析

分类号: TP391

DOI: 10.11925/infotech.2096-3467.2022.0293

引用本文: 刘赏, 沈逸凡. 基于新闻标题-正文差异性的虚假新闻检测方法[J]. 数据分析与知识发现, 2023, 7(2): 97-107.(Liu Shang, Shen Yifan. Detecting Fake News Based on Title-Content Difference[J]. Data Analysis and Knowledge Discovery, 2023, 7(2): 97-107.)

1 引 言

社交媒体中存在大量的虚假新闻,这些虚假新闻传播范围更广、速度更快、程度更深^[1],在某些情况下其扩散程度与影响远远超过了真实新闻;另外,虚假新闻的传播湮没了可靠的信息来源,愈发导致公众困惑,最终可能以更大的扩散效应造成严重的社会问题^[2]。因此,在社交媒体中快速有效地检测出虚假新闻具有重要意义。

现有虚假新闻检测研究大多基于新闻正文和评论展开,但读者越来越倾向于仅阅读新闻标题而非正文,读者评论愈发难以收集。Kwak 等^[3]指出,社交媒体上的大多数读者倾向于简单地分享信息,而不留下任何评论,为每一篇新闻收集评论并通过评

论进行虚假新闻检测是难以实现的。同时,伴随着读者阅读习惯的改变,在信息过载的数字环境中,标题对于新闻十分重要,多数情况下,读者只阅读新闻标题而不阅读正文^[4],即使阅读了正文后,标题也会给读者留下深刻持久的印象^[5]。因此,虚假新闻往往会以断章取义、夸大其词等方式制造标题,以达到吸引读者眼球、增加阅读量与曝光率、制造轰动效应的目的。现有研究指出:标题和正文的差异是识别虚假新闻的重要特征^[6],虚假新闻和真实新闻的标题与正文内容之间的显著区别,是有效识别虚假新闻的重要特征。另外,很多虚假新闻仅基于标题进行转载,而忽略了正文内容。可见,新闻标题分析在虚假新闻检测中起到至关重要的作用,从新闻标题

通讯作者(Corresponding author): 刘赏(Liu Shang), ORCID: 0000-0002-3797-7339, E-mail: liushangw@tjufe.edu.cn。

*本文系教育部人文社会科学研究规划基金项目(项目编号: 19YJA630046)、天津市自然科学基金项目(项目编号: 20JCQNJC00970)和天津市艺术科学规划项目(项目编号: C22030)的研究成果之一。

The work is supported by Humanities and Social Sciences Research Planning Project of the Ministry of Education (Grant No. 19YJA630046), Natural Science Foundation of Tianjin (Grant No. 20JCQNJC00970), Tianjin Art Science Planning Project (Grant No. C22030).

入手能够有效进行新闻的真实性鉴别。

为解决在虚假新闻检测中新闻评论难以收集、新闻文本简短难以提取有效信息的问题,本文利用新闻标题和正文在文本和情感上的特征,结合新闻标题和正文在文本和情感两方面的差异性特征,通过改进的异构图注意网络^[7],提出利用新闻差异性异构图网络(News Differential Heterogeneous Graph Network, NDHN)进行虚假新闻检测的方法。首先,使用BERT及Transformer网络提取新闻标题-正文的文本和情感的特征;然后,提出基于Cos-Gap(Cosine Similarly-Emotion Gap)方法的文字-情感差异性计算模块以计算差异性特征,并以此为依据构建异构图中边的关联,解决传统异构图不能体现边权重特征的问题;最后,对提取的多种差异性特征进行融合,并作为新闻差异性异构图网络的输入,训练虚假新闻检测分类模型。

2 相关研究

近年来,虚假新闻检测引起了人们极大的关注。2011年,Castillo等^[8]最先基于手工特征方法辨别真假新闻,结合传统机器学习方法对虚假新闻进行检测。Shu等^[9]应用随机森林和SVM等监督学习方法进行真假新闻二分类。由于利用手工特征的方法需先验知识指导,缺乏全面性和灵活性,需要大量人工参与特征提取,提取海量的数据时存在较大的局限性。因而,使用自动学习数据特征的深度学习方法日益成为虚假新闻检测的热点。Ma等^[10]首次将循环神经网络用于虚假新闻检测。

虚假新闻检测主要基于文本词法、句法等特征以及用户及其传播特征进行。例如,Potthast等^[11]依据写作风格及其一致性进行虚假新闻检测。Shu等^[12]联合构建新闻内容和用户评论间的关系,进行可解释性的虚假新闻检测。Ruchansky等^[13]将真实新闻与虚假新闻的语言特征进行比较,通过找出虚假新闻的语言特征对文本的真实性进行鉴定。Liu等^[14]通过提取新闻在社交媒体传播路径上的全局和局部变化特征以检测虚假新闻。以上研究收集每条新闻报道的用户评论进行虚假新闻检测。然而,社交媒体上的大多数用户倾向于简单地重新分享源信息,而不留下任何评论^[3],因此本文基于新闻的标题

和正文进行虚假新闻的检测分析,能够在一定程度上克服上述问题。

现有研究表明,标题和正文差异是识别虚假新闻的重要特征。Horne等^[6]通过分析虚假新闻的标题和正文发现,虚假新闻的传播受众焦点是那些不阅读正文、只关心标题的读者。而且,虚假新闻和真实新闻的标题与正文内容在文本词法和句法中存在显著区别。Shrestha等^[15]使用统计方法证实了虚假新闻和真实新闻的标题与正文存在的特征差异,并发现虚假新闻的标题和正文部分比真实新闻表达了更多的负面情感,而真实新闻比虚假新闻更具描述性。

另外,情感特征会对虚假新闻检测产生一定作用,虚假新闻通常使用能够引起读者强烈情感共鸣的方式迷惑读者,试图使其相信新闻的真实性,有利于该类新闻在社交网络中传播。Ajao等^[16]指出新闻真实性与发布文本的情感之间存在着关系,基于情感特征(负面和正面单词数量的比率)能够对文本形式的虚假新闻进行检测。基于此,Giachanou等^[17]进一步从新闻内容中提取情感词汇,以丰富情感特征,并用于虚假新闻检测。Wu等^[18]提出一种从内容和评论中提取情感信息并嵌入自适应融合网络的虚假新闻检测方法。Zhang等^[19]发现真实新闻和虚假新闻中内容与评论的情感是有所区别的,通过提取新闻和评论中的双重情感有效提升了虚假新闻检测的性能。Ghanem等^[20]发现情感因素在影响读者辨别新闻真假中起着关键作用,并且已有实证表明,该因素可以成功地应用于虚假信息检测。Vosoughi等^[1]调查Twitter上的新闻,发现虚假新闻会引发恐惧、厌恶和惊讶的情感,而真实新闻则会引发喜悦、悲伤、信任和期待的情感。可见,情感是适用于虚假新闻检测的重要特征之一。

基于以上研究,本文发现:提取新闻标题-正文的差异性特征对有效识别虚假新闻提供了帮助。虚假新闻检测领域研究现已取得了一定成果,但大多基于新闻标题-正文差异性的研究停留在理论和统计分析方法上,少有研究使用新闻标题-正文差异性特征进行虚假新闻检测。因此,本文融合新闻标题-正文差异性特征,在异构图注意网络^[7]的基础上,提出融合新闻差异性特征的新闻差异性异构图网络NDHN进行虚假新闻检测。不同于异构图注意网

络,本文将差异性特征融合到异构图邻接矩阵的构造中,此处的差异性包含文本差异性和情感差异性。另外,本方法中的新闻标题-正文差异性解决了短文本因其稀疏性造成的难以提取有效信息的问题,也进一步丰富了新闻的特征表示。

3 研究框架与方法

3.1 问题定义

“虚假新闻”为故意发布且可以被证实为虚假的新闻^[9]。虚假新闻检测任务可以定义为一个二元分类问题,旨在将社交媒体上的新闻分类为虚假新闻和真实新闻。其形式化定义如下:给定一组来自社

交媒体的新闻 $D = \{p_1, p_2, \dots, p_n\}$, 其中, p_i 是一个由一组新闻标题和正文信息组成的新闻, n 表示新闻的数量。通过学习模型 $f: D \rightarrow Y$, 将每条新闻 p_i 分类为预定义类别 $Y = \{0, 1\}$, 其中, 0 表示真实新闻, 1 表示虚假新闻。

3.2 新闻差异性异构图网络 NDHN

新闻标题和正文的差异性是识别虚假新闻的重要特征^[6], 因此本文融合差异性特征在异构图神经网络^[7]的基础上提出虚假新闻检测的方法: 新闻差异性异构图网络 NDHN。该方法基于新闻标题和正文的文本差异性与情感差异性两方面进行分析以识别虚假新闻, 如图 1 所示。

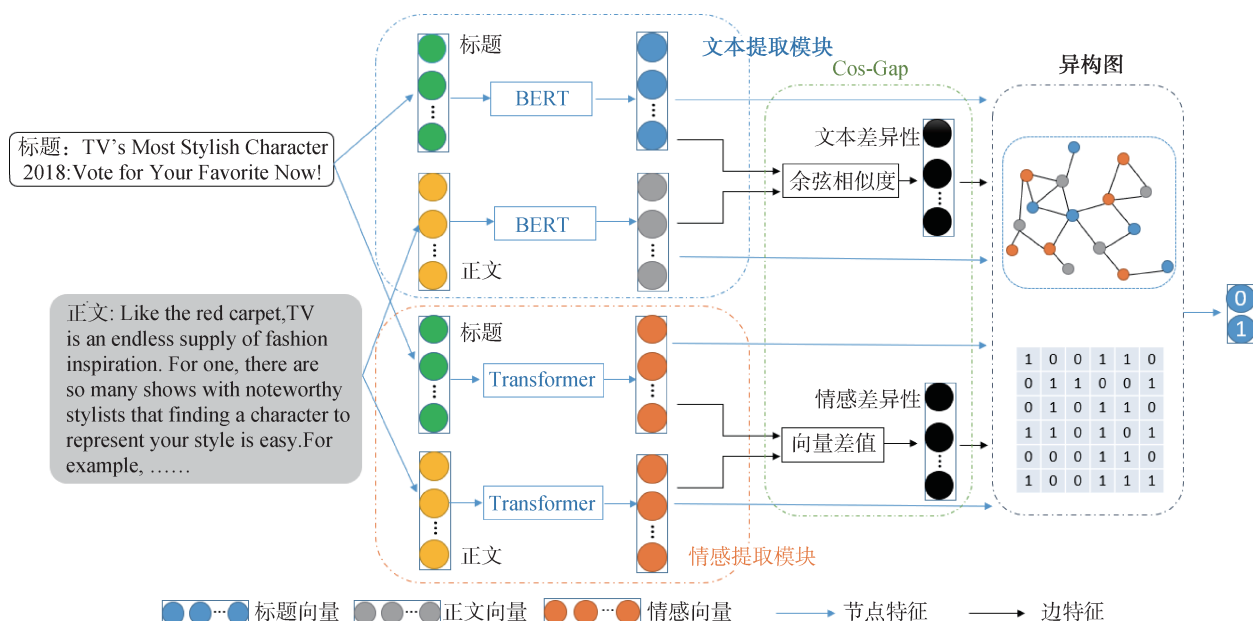


图 1 NDHN 结构

Fig.1 Structure of NDHN

NDHN 使用标题、正文和情感三种特征进行虚假新闻检测,其对应于异构图网络中的三种不同节点。使用异构图卷积^[7]将不同类型节点的特征空间联合到一起构造一个新的更大的特征空间;考虑到不同类型特征的差异性,利用各自的变换矩阵将其投影到一个隐式空间中。异构图网络充分利用了特征信息,使其沿图传播,提取节点间丰富的关系信息,解决了短文本稀疏性的问题。该网络可以灵活添加任何额外的信息,并可以获取短文本与附加信

息间丰富的关系,使得短文本的稀疏性得到了平衡。

(1) 新闻标题-正文文本和情感特征提取模块

①新闻标题-正文的文本特征提取

NDHN 模型采用 BERT 模型学习新闻标题和正文的文本特征,将新闻中的句子转换为表示向量。BERT 预训练语言模型^[21]使用一种双向 Transformer 编码结构,使得模型能够充分获取输入文本中的潜在语义信息,更好地进行特征表示。

对于每一篇新闻,本文分别提取出标题句和正

文句,使用BERT预训练模型将每一个句子表示为一个句向量作为新闻文本特征。新闻标题文本特征表示为 $T = \{t_1, t_2, \dots, t_{d_m}\}$, $T \in R^{d_m}$;新闻正文文本特征表示为 $C = \{c_1, c_2, \dots, c_{d_n}\}$, $C \in R^{d_n}$,其中, d_m 表示BERT句向量维度。

②新闻标题-正文的情感特征提取

本文使用NVIDIA^[22]发布的开源情感分类模型中的Transformer模型对标题和正文分别进行情感分类,将其分为愤怒、期待、厌恶、恐惧、喜悦、悲伤、惊讶和信任8个类别。Transformer模型在学习语义信息的同时,更专注于词的不同位置,理解句子中词之间的顺序和距离信息,同时使用自注意力机制关注输入句子的其他相关部分,使模型更加充分地利用句子的各种特征信息,从而有效识别出句子的情感极性。

对于每一篇新闻,将新闻标题和正文分别作为一个句子输入Transformer模型中,模型先通过编码器获取输入文本的向量表示,然后将向量输入解码器进行分类,即可得到基于8种类别的新闻标题情感特征 $E_t = \{E_t^1, E_t^2, \dots, E_t^8\}$ 和新闻正文情感特征 $E_c = \{E_c^1, E_c^2, \dots, E_c^8\}$,其中, $E_t \in R^8$, $E_c \in R^8$, E_t^1 和 E_c^1 表示愤怒情绪的特征值, E_t^2 和 E_c^2 表示期待情绪的特征值, E_t^3 和 E_c^3 表示厌恶情绪的特征值, E_t^4 和 E_c^4 表示恐惧情绪的特征值, E_t^5 和 E_c^5 表示喜悦情绪的特征值, E_t^6 和 E_c^6 表示悲伤情绪的特征值, E_t^7 和 E_c^7 表示惊讶情绪的特征值, E_t^8 和 E_c^8 表示信任情绪的特征值。

(2) 基于Cos-Gap方法的文本-情感差异性计算模块

本文提出Cos-Gap(Cosine Similarity-Emotion Gap)方法用于虚假新闻检测任务,以获取新闻标题和正文之间的差异性信息。余弦相似度是衡量文本间相似度的度量标准,传统的余弦相似度多用于求取单一特征的相似度,本文根据新闻中标题-正文差异性的特点,融合文本差异性和情感差异性双重特征,提出Cos-Gap文本-情感差异性提取方式,具体如下:

①新闻标题-正文的文本差异性计算

由标题-正文的文本和情感特征提取模块获得新闻标题的文本特征: $T = \{t_1, t_2, \dots, t_{d_m}\}$, $T \in R^{d_m}$;新闻正文的文本特征 $C = \{c_1, c_2, \dots, c_{d_n}\}$, $C \in R^{d_n}$ 。采

用余弦相似度获取新闻标题句向量 T 及新闻正文句向量 C 文本上的差异特征,如公式(1)所示。

$$s = \frac{\sum_{i=1}^{d_m} (c_i \times t_i)}{\sqrt{\sum_{i=1}^{d_m} (c_i)^2} \times \sqrt{\sum_{i=1}^{d_n} (t_i)^2}} \quad (1)$$

其中, $s \in R$ 。两个向量的余弦相似度值越大,则表明这两个向量所指的新闻标题-正文差异性越小,相似度越高,即该新闻为虚假新闻的可能性更小。

②新闻标题-正文的情感差异性计算

为丰富特征表示,获取新闻标题-正文情感的差异性,计算标题情感特征 $E_t = \{E_t^1, E_t^2, \dots, E_t^8\}$ 与正文情感特征 $E_c = \{E_c^1, E_c^2, \dots, E_c^8\}$ 之间的向量差值,并取绝对值作为情感差异 $E_{gap} = \{E_{gap}^1, E_{gap}^2, \dots, E_{gap}^8\}$,如公式(2)所示。

$$E_{gap} = |E_t - E_c| \quad (2)$$

其中, $E_{gap} \in R^8$ 。通过这种方法,可以得到情感之间的差异。对于情感相同的新闻而言,情感差异向量的绝对值很小(接近零);反之,若情感差异向量的绝对值大,则说明其标题-正文情感差异性大,则该新闻为虚假新闻的可能性更大。

(3) 基于异构图网络的虚假新闻识别模块

本文在异构图注意网络^[7]的基础上设计了融合新闻差异性的异构图 $G = (V, E, \tau)$,其中, V 表示节点集, E 表示边集,即节点之间的关系, τ 表示节点类型, $\tau \in \{T, C, EM\}$ 。具体来说,共建立三种类型的异构节点,新闻标题类型节点 T ,新闻正文类型节点 C 和情感类型节点 EM ,即 $V = T \cup C \cup EM$ 。建立异构图的边 E ,在这些节点中,不同类型节点之间会产生下述情况的紧密联系:一是在新闻标题与新闻正文类型节点间依据文本差异性特征建立边;二是在新闻标题与其相应的情感之间依据情感特征建立边;三是在新闻正文与其相应的情感之间依据情感特征建立边。通过Cos-Gap方法将新闻标题-正文差异性特征作为异构图边 E 的权值。

①节点初始化

以 n 篇新闻文章为例,为丰富新闻特征表示,采用三种类型的节点对新闻进行描述,其节点初始化方法如下。

1)对于新闻标题节点,每一个新闻标题作为一

个标题节点,同时使用新闻标题句向量作为节点的主要表示状态。由此,可以得到 n 个标题节点表示 $T = \{T^1, T^2, \dots, T^n\}$, $T \in R^{n \times d_m}$, 其中 d_m 表示嵌入句向量的维度。

2) 对于新闻正文节点,每一个新闻正文作为一个正文节点,同时使用新闻正文句向量作为节点的主要表示状态。由此,可以得到 n 个正文节点表示 $C = \{C^1, C^2, \dots, C^n\}$, $C \in R^{n \times d_m}$, 其中 d_m 表示嵌入句向量的维度。

3) 对于情感节点,由于新闻标题及正文分别进行了情感 8 分类,构造愤怒 (Anger)、期待 (Anticipation)、厌恶 (Disgust)、恐惧 (Fear)、喜悦 (Joy)、悲伤 (Sad)、惊讶 (Surprise)、信任 (Trust) 8 个情感节点,使用 One-Hot 编码对其进行初始化,每个节点的向量表示中只有一位处于状态 1,其余位都是 0。则情感节点表示为 $EM = \{EM_{anger}, EM_{anticipation}, EM_{disgust}, EM_{fear}, EM_{joy}, EM_{sadness}, EM_{surprise}, EM_{trust}\}$, $EM \in R^{8 \times 8}$ 。

通过结合标题、正文和情感三类节点,融合了三类节点间的关系特征,丰富了短文本的语义,从而更有利于新闻的分类任务。

②边的初始化

图卷积使用图结构中节点之间的邻接关系描述特征之间的依赖关系,邻接矩阵 A 即为特征依赖图。不同于普通邻接矩阵(节点与节点间存在边即为 1,不存在边即为 0),本文设计了融合 Cos-Gap 方法计算出的新闻标题-正文差异性特征的邻接矩阵,将差异性特征融合到邻接矩阵边的权值中。依据不同类型的节点,共定义三种邻接矩阵:标题-正文邻接矩阵、情感-标题邻接矩阵和情感-正文邻接矩阵。将三种邻接矩阵拼接到一起,构建出 NDHN 模型的邻接矩阵。通过构建相应的邻接矩阵以初始化图模型的边,具体过程如下。

1) 基于标题-正文文本差异性的邻接矩阵构建

利用前文得到的标题-正文的文本差异性特征 $s, s \in R$, 即可表示出 n 篇新闻的文本差异性特征 $S = \{s_1, s_2, \dots, s_n\}$, $S \in R^n$ 。将其按文章顺序分布在标题-正文的邻接矩阵 $A_{title-content} \in R^{n \times n}$ 的主对角线中。对应的依赖图如图 2 所示, $A_{title-content}$ 的主对角线上的值均为 s_i 的值,即 $A_{ii} = s_i$ 。由于新闻标题只与其对

应的正文存在依赖关系,因此,除了主对角线有值以外,其他均为 0。

		title			
		T^1	T^2	...	T^n
content	C^1	s_1			
	C^2				
	...		s_2		
	C^n			...	s_n

图 2 标题-正文邻接矩阵 $A_{title-content}$

Fig.2 Title-Content Adjacency Matrix $A_{title-content}$

2) 基于标题-正文情感差异性的邻接矩阵构建

本文网络建立了标题、正文和情感三类节点,标题与正文均与情感节点相连接,因此情感差异性邻接矩阵分为情感-标题邻接矩阵 $A_{emotion-title}$ 以及情感-正文邻接矩阵 $A_{emotion-content}$ 两种。

利用前文得到的标题-正文情感差异 $E_{gap}, E_{gap} \in R^8$, 本文提出情感-标题邻接矩阵 $A_{emotion-title}$ 的计算方法如公式(3)所示。具体来说,通过将标题情感特征 E_t 与情感差异 E_{gap} 相加,做转置运算得到情感-标题邻接矩阵 $A_{emotion-title}$ 中的每一列向量 A_j 。

$$A_j = (E_t + E_{gap})^T \quad (3)$$

其中, $A_{emotion-title} \in R^{8 \times n}$, n 为新闻数量, T 表示转置运算。对应的依赖图如图 3 所示, $A_{emotion-title}$ 中的每一列向量 A_j 对应第 j 个新闻标题的 8 类情感特征的关联表示。

		title			
		T^1	T^2	...	T^n
emotion	EM_{anger}	e^1_{t1}	e^1_{t2}	...	e^1_{tn}
	$EM_{anticipation}$	e^2_{t1}	...		
	$EM_{disgust}$	e^3_{t1}			
	EM_{fear}	e^4_{t1}		...	
	EM_{joy}	e^5_{t1}			
	$EM_{sadness}$	e^6_{t1}			...
	$EM_{surprise}$	e^7_{t1}			
	EM_{trust}	e^8_{t1}			e^8_{tn}

图 3 情感-标题邻接矩阵 $A_{emotion-title}$

Fig.3 Emotion-Title Adjacency Matrix $A_{emotion-title}$

利用前文得到的标题-正文情感差异 E_{gap} , $E_{gap} \in R^8$, 本文提出情感-正文邻接矩阵 $A_{emotion-content}$ 的计算方法如公式(4)所示。具体来说, 通过将正文情感特征 E_c 与情感差异 E_{gap} 相加, 做转置运算得到情感-正文邻接矩阵 $A_{emotion-content}$ 中的每一列向量 A_j 。

$$A_j = (E_c + E_{gap})^T \quad (4)$$

其中, $A_{emotion-content} \in R^{8 \times n}$, n 为新闻数量, T 表示转置运算。对应的依赖图如图4所示, $A_{emotion-content}$ 中的每一列向量 A_j 对应第 j 个新闻正文的8类情感特征的关联表示。

		content			
		C^1	C^2	...	C^n
emotion	EM_{anger}	e_{c1}^1	e_{c2}^1	...	e_{cn}^1
	$EM_{anticipation}$	e_{c1}^2	...		
	$EM_{disgust}$	e_{c1}^3			
	EM_{fear}	e_{c1}^4		...	
	EM_{joy}	e_{c1}^5			
	$EM_{sadness}$	e_{c1}^6			...
	$EM_{surprise}$	e_{c1}^7			
	EM_{trust}	e_{c1}^8			e_{cn}^8

图4 情感-正文邻接矩阵 $A_{emotion-content}$

Fig.4 Emotion-Content Adjacency Matrix $A_{emotion-content}$

3) 邻接矩阵融合

依据新闻标题-正文在文本和情感两方面的差异性特征, 得到三种邻接矩阵: $A_{title-content} \in R^{n \times n}$, $A_{emotion-title} \in R^{8 \times n}$, $A_{emotion-content} \in R^{8 \times n}$, 其中 n 为新闻数量。为融合差异性特征并满足邻接矩阵的方阵特性, 将节点组合在一起, 形成包含以上三种邻接矩阵的融合邻接矩阵, 用于各个节点间关联关系的表示。

本文按照正文类型节点、标题类型节点和情感类型节点的顺序进行融合邻接矩阵的构建。由于正文类型节点、标题类型节点、情感类型节点的个数分别是: $n, n, 8$, 因此, 异构图共有 $(2n+8)$ 个节点, 即该矩阵具有 $(2n+8)$ 行和 $(2n+8)$ 列。对三个矩阵按照节点类型顺序依次将 $A_{title-content}$ 存入对应标题类型行、正文类型列的位置, $A_{emotion-content}$ 存入情感类型行、正文类型列的位置, $A_{emotion-title}$ 存入情感类型行、标题类型

列的位置, 形成下三角矩阵 A_{fusion} 。

根据 Kipf 等^[23]提出的无向图对称思想和自循环思想, 邻接矩阵应为对称矩阵, 每个词都应与自身连接, 且矩阵对角线元素的值为1, 因此在 A_{fusion} 中设置 $A_{ii} = 1$ 且 $A_{ij} = A_{ji}$, 由此融合三种邻接矩阵得到最终的邻接矩阵 A , 如图5所示。 $A \in R^{(2n+8) \times (2n+8)}$, n 为新闻数量, $(2n+8)$ 为异构图的节点数。

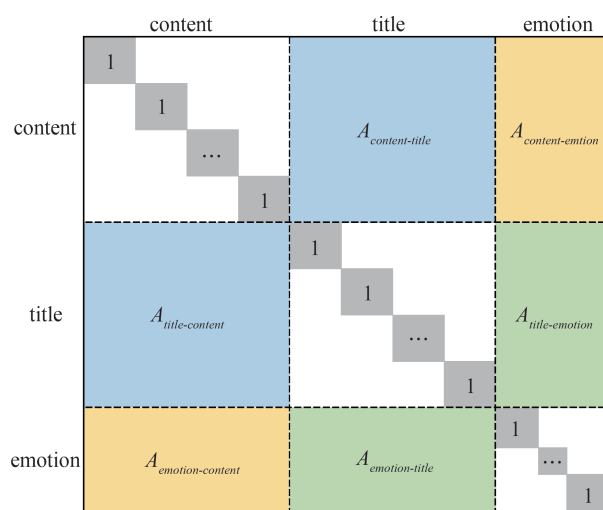


图5 NDHN网络的邻接矩阵

Fig.5 Adjacency Matrix of the NDHN Network

③异构图卷积

为解决多特征异构问题, 本文引入异构图神经网络的编码部分^[7]对所构建的异构图从三种类型的节点和关系中提取特征。它考虑了不同类型信息的差异, 并利用各自的变换矩阵将它们投影到一个隐式公共空间中, 如公式(5)所示。

$$H^{(t+1)} = \sigma \left(\sum_{\tau \in T} A_{\tau} \cdot H_{\tau}^{(t)} \cdot W_{\tau}^{(t)} \right) \quad (5)$$

其中, $A_{\tau} \in R^{|\mathcal{V}| \times |\mathcal{V}|}$ 是 A 的子矩阵, 行表示所有节点, 列表示类型为 τ 的相邻节点。节点 $H^{(t+1)}$ 是通过使用不同的变换矩阵 $W_{\tau}^{(t)} \in R^{q^{(t)} \times q^{(t+1)}}$ 从不同类型 τ 的相邻节点 $H_{\tau}^{(t)}$ 特征中聚合信息来表示的。变换矩阵 $W_{\tau}^{(t)}$ 定义不同特征空间的差异, 并将它们投影到一个隐式的公共空间 $R^{q^{(t+1)}}$ 中。最初, $H_{\tau}^{(0)} = X_{\tau}$ 。

④双重注意力机制

一般情况下, 对于一个特定的节点, 不同类型的相邻节点可能会对其产生不同的影响。例如, 相同类型的相邻节点可能产生更多有用的信息。另外,

同一类型的不同相邻节点也可能有不同的重要性。针对节点层次和类型层次的不同重要性,充分利用新闻中标题-正文的差异性特征,引入双重注意力机制^[7],其能够捕捉不同邻居节点的重要性,降低噪声信息,还能够提取出不同类型节点的重要性,具体过程如下。

1) 类型级注意力机制

给定一个特定的节点 v , 类型级注意力学习邻近节点不同类型的权重。将 τ 类型进行嵌入表示, 如公式(6)所示。

$$\mathbf{h}_\tau = \sum_{v'} \mathbf{A}_{vv'} \mathbf{h}_{v'} \quad (6)$$

其中, \mathbf{h}_τ 为相邻节点特征 $\mathbf{h}_{v'}$ 的和, 其中节点 $v' \in N_v$ 且具有 τ 类型, $\mathbf{A}_{vv'}$ 为节点 v 和节点 v' 的邻接矩阵。

基于当前节点嵌入 \mathbf{h}_v 和类型嵌入 \mathbf{h}_τ 计算类型水平注意力评分, 如公式(7)所示。

$$a_\tau = \sigma(\boldsymbol{\mu}_\tau^\top [\mathbf{h}_v \parallel \mathbf{h}_\tau]) \quad (7)$$

其中, $\boldsymbol{\mu}_\tau$ 为 τ 型的注意力向量, \parallel 表示连接, $\sigma(\cdot)$ 表示激活函数, 如 Leaky ReLU。

利用 Softmax 函数对所有类型的注意力得分进行归一化, 得到类型级注意力权重, 如公式(8)所示。

$$\alpha_\tau = \frac{\exp(a_\tau)}{\sum_{\tau' \in T} \exp(a_{\tau'})} \quad (8)$$

2) 节点级注意力机制

应用节点级注意力机制^[7]获取不同相邻节点的重要性, 并减少噪声节点的权重。给定一个类型 τ 的特定节点 v 和类型 τ' 的相邻节点 $v' \in N_v$, 基于节点嵌入 \mathbf{h}_v 和 $\mathbf{h}_{v'}$, 节点 v' 的类型级注意力权重为 $\alpha_{\tau'}$, 则节点级注意力分数计算方法如公式(9)所示。

$$b_{vv'} = \sigma(v^\top \alpha_{\tau'} [\mathbf{h}_v \parallel \mathbf{h}_{v'}]) \quad (9)$$

其中, v 是注意力向量。

利用 Softmax 函数对节点级注意力分数进行归一化, 如公式(10)所示。

$$\beta_{vv'} = \frac{\exp(b_{vv'})}{\sum_{v' \in N_v} \exp(b_{vv'})} \quad (10)$$

替换公式(5), 将包括类型级和节点级注意力在内的双层注意力机制整合到异构图卷积中, 每层之间的传播规则如公式(11)所示。

$$\mathbf{H}^{(l+1)} = \sigma\left(\sum_{\tau \in T} \mathbf{B}_\tau \cdot \mathbf{H}_\tau^{(l)} \mathbf{W}_\tau^{(l)}\right) \quad (11)$$

其中, \mathbf{B}_τ 表示注意力矩阵, 其第 v 行第 v' 列中的元素为公式(10)中的 $\beta_{vv'}$ 。

⑤ 虚假新闻检测模型训练

在经过一个 l 层 NDHN 之后, 可以获得节点(包括短文本)的嵌入。采用文献[7]中模型训练方法将短文本嵌入 $\mathbf{H}^{(l)}$ 提供给 Softmax 函数进行激活, 输出新闻分类概率, 如公式(12)所示。

$$\mathbf{Z} = \text{softmax}(\mathbf{H}^{(l)}) \quad (12)$$

模型训练过程中, 为防止过拟合, 在交叉熵公式后加入 l_2 正则项, 交叉熵损失函数表示如公式(13)所示。

$$\mathcal{L} = -\sum_{i \in D_{train}} \sum_{j=1}^C Y_{ij} \cdot \log Z_{ij} + \eta \|\Theta\|_2 \quad (13)$$

其中, C 为类数, D_{train} 为用于训练的短文本索引集, Y 为对应的标签指标矩阵, Θ 为模型参数, η 为正则化因子。在进行模型优化时, 采用梯度下降算法。

4 实验与结果分析

4.1 实验数据集

采用 GossipCop 开放语料作为实验数据集, 该数据集属于虚假新闻检测标准数据集 FakeNewsNet^[24]。FakeNewsNet 数据集是从事实证网站收集到的新闻网页数据; GossipCop 数据集包含从娱乐新闻媒体网站 GossipCop3 收集的虚假新闻, 事实核查的评估结果来自 GossipCop 网站上的评分。作为虚假新闻检测的常用数据集, 其采用分类准确率(Accuracy)和 F1 值作为分类评估指标。在清理了缺失新闻正文或标题的数据后, 总共获得 7 612 篇新闻文章。

4.2 实验环境

本实验在 Ubuntu 18.04 操作系统下进行, 模型训练在 NVIDIA RTX 3090 显卡上使用 PyTorch 1.8 框架完成。采用 BERT 预训练语言模型, 以一个句子为例, 模型随机遮蔽输入的 15% 单词, 然后预测被遮蔽的单词。同时, 采用 NVIDIA 预先训练好的 Transformer 模型提取情感标签, 为保证模型的鲁棒性, 随机采样 100 个实例, 由三个注释者手动独立标记, NVIDIA 模型准确率为 86%。因此, 在提取情感进行虚假新闻检测方面, 该分类器被认为是可靠的。

NDHN 的隐含维度 d 设置为 512, 预先训练的词

嵌入维度设置为768,层数 l 设置为2。对于模型训练,将学习率设定为0.001,Dropout设定为0.8,正则化因子设定为 $5e-6$ 。实验中使用自适应矩估计优化器对损失函数进行优化求解。以分类准确率和F1值作为实验结果的评价指标。

4.3 对比实验

为综合评价本文模型,选取目前最具代表性的三种基于机器学习的方法和三种基于深度学习的方法进行实验结果对比。

(1)SVM^[25]:利用支持向量机分类器,基于从新闻中提取的特征进行虚假新闻检测。

(2)RFC^[26]:随机森林分类器,利用新闻的识别特征进行虚假新闻检测。

(3)DTC^[10]:基于各种手工特征,使用决策树分

类器进行虚假新闻检测和识别。

(4)GRU-2^[10]:基于GRU的GRU-2模型,通过添加第二个GRU层,捕获不同时间步长之间更高层次的特征交互进行虚假新闻检测。

(5)B-TransE^[27]:B-TransE模型结合正、负两种单一模型,基于新闻内容和知识图检测虚假新闻。

(6)KAN^[28]:KAN是一个先进的模型,利用Transformer将知识图中的外部知识用于虚假新闻检测。识别新闻内容中提到的实体,并将其与知识图中的实体对齐,将实体及其上下文用作外部知识来提供补充信息。

在Gossip Cop数据集上利用SVM、RFC、DTC、GRU-2、B-TransE、KAN等6种方法和NDHN进行虚假新闻检测的结果比较如表1所示。

表1 不同模型实验对比结果
Table 1 Results of Different Model Experiments

数据集	指标	SVM	RFC	DTC	GRU-2	B-TransE	KAN	NDHN
GossipCop	准确率	0.664 3	0.691 8	0.695 9	0.718 0	0.739 4	0.776 6	0.803 9
	F1	0.595 5	0.669 1	0.691 9	0.707 9	0.734 0	0.771 3	0.803 7

NDHN方法的分类准确率为0.803 9,高于对比方法中的最优值0.776 6(KAN);NDHN方法虚假新闻识别结果的F1值为0.803 7,高于对比方法中的最优值0.771 3(KAN)。由此可见,本文方法在分类准确率上提升约2.7个百分点,F1指标提升约3.2个百分点,显著优于其他检测方法,验证了本文方法在虚假新闻检测分类中的有效性。

4.4 消融实验

为验证本文模型中各模块的有效性,在数据集上对NDHN进行消融实验,主要包括三部分。

(1)-Word Similarity:移除标题-正文文本差异性模块,标题-正文文本仅依靠文章关系进行关联,标题与对应文本存在关系则矩阵对应值置为1,无关则置为0。用于验证文本差异性的有效性。

(2)-Emotion Gap:移除标题-正文情感差异性模块,标题和正文只与其相对应的情感标签为1的情感节点相连,而非与所有节点相连,且邻接矩阵对应值为其对应情感标签的概率值。用于验证情感差异性的有效性。

(3)-Attention:移除双重注意力模块,即不使用

注意力机制。用于验证本文引入的双重注意力与异构图卷积网络融合的有效性。

消融实验得到的分类准确率和F1值如表2所示,所有变量中的超参数都由验证集确定。

表2 NDHN的消融实验结果
Table 2 Experimental Results of Ablation of NDHN

方法	准确率	F1
-Word Similarity	0.789 5	0.789 4
-Attention	0.800 0	0.799 1
-Emotion Gap	0.801 3	0.800 3
NDHN	0.803 9	0.803 7

当消除标题-正文文本差异性时,模型准确率和F1值下降1.43个百分点。该结果表明,标题-正文文本差异性信息有助于提取新闻特征,便于理解新闻整体的结构,提升新闻检测准确性。当忽略标题-正文情感差异性时,模型性能比NDHN下降0.3个百分点左右,可见情感差异性特征在新闻检测中发挥着重要作用。将新闻文本及情感差异性纳入对虚假新闻检测的指标具有重要意义。当忽略双重注意力时,模型性能下降0.4个百分点,证明了本文提出

的新闻差异性异构图网络的有效性,它考虑新闻标题、正文及情感三种信息类型的异质性,并通过类型级和节点级双重注意力机制捕获不同信息的重要性,减少了噪声信息的权重对新闻检测的影响,进一步提升了NDHN的性能。

4.5 数据分析

(1) 8种不同情感标签下新闻标题-正文差异性的可视化

所有新闻标题-正文在8种不同情感标签下产生的差异数目如图6所示。整体来看,虚假新闻产生的差异总和远多于真实新闻,除了期望和信任情感标签外,虚假新闻的差异性均大于真实新闻,由此可知,新闻标题-正文情感差异性特征有利于检测其新闻的真实与虚假。情感上展现出厌恶、快乐、惊喜类型的新闻具有一定的敏感性,相关机构应重视此类新闻,进行重点监测,新闻的情感很大程度上会带动网民的情感波动,从而带动虚假新闻的传播,因而一旦发现虚假信息,应在第一时间对其进行干预,避免其传播。

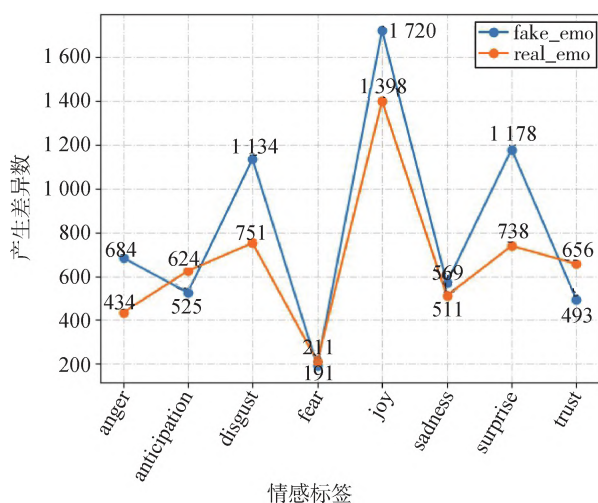


图6 不同情感标签下新闻标题-正文差异数目

Fig.6 The Number of News Title-Content Differences Under Different Emotion Labels

(2) Cos-Gap方法差异频数直方图

使用Cos-Gap方法依据数据集信息绘制文本差异频数直方图和情感差异频数直方图,如图7和图8所示。

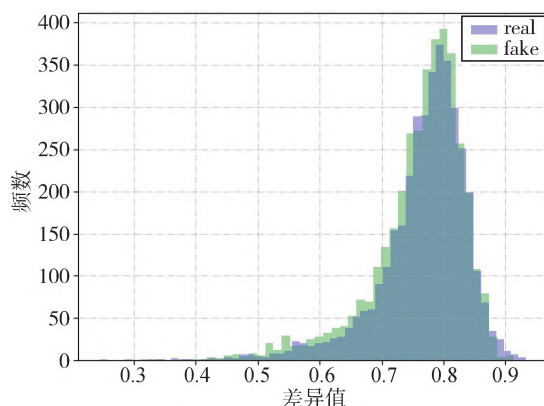


图7 新闻标题-正文文本差异频数直方图

Fig.7 Histogram of News Title-Content Text Difference Frequency

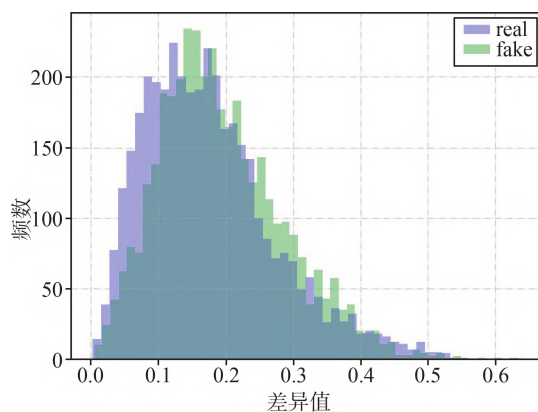


图8 新闻标题-正文情感差异频数直方图

Fig.8 Histogram of News Title-Content Emotion Difference Frequency

虚假新闻在文本差异和情感差异上的整体频数都超出了真实新闻,由此可证明,差异性亦能作为虚假新闻的判断指征之一,进一步证明了本文提取文本与情感两种差异性特征进行虚假新闻检测的合理性。

5 结 语

虚假新闻的传播不仅会污染网络社区,还会对媒体和当事人心理健康及名誉造成损害。针对新闻文本简短、评论难以获取的特点,本文基于图异构注意力网络提出新闻差异性异构图网络NDHN模型,仅利用新闻标题和正文即可进行虚假新闻检测。模型利用新闻标题和正文在文本与情感上的特征,结合新闻标题和正文在文本与情感两方面的差异性特征

丰富特征表示,通过异构图建立标题、正文、情感三种类型节点,同时提出 Cos-Gap 文本-情感差异性计算模块,计算三类节点的边之间的关联关系,充分考虑文本和情感差异性特征交互,提升了虚假新闻检测的效率。在 GossipCop 数据集上与其他代表性模型进行对比实验和消融实验,验证了 NDHN 模型能够有效地进行虚假新闻检测。本文方法适用于带有标题的新闻,对于微博、Twitter 等无标题文本则存在局限。未来将融合图像等信息丰富特征表示,同时构建更大的数据集进行虚假新闻实验验证,提高模型对于多场景中虚假新闻识别的泛化能力,并将细粒精准分析“标题党”对于虚假新闻识别的影响作为后续研究的主要目标。

参考文献:

- [1] Vosoughi S, Roy D, Aral S. The Spread of True and False News Online[J]. Science, 2018, 359(6380): 1146-1151.
- [2] Mian A, Khan S. Coronavirus: The Spread of Misinformation[J]. BMC Medicine, 2020, 18(1): Article No.89.
- [3] Kwak H, Lee C, Park H, et al. What is Twitter, a Social Network or a News Media? [C]//Proceedings of the 19th International Conference on World Wide Web. 2010: 591-600.
- [4] Gabielkov M, Ramachandran A, Chaintreau A, et al. Social Clicks: What and Who Gets Read on Twitter?[C]//Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science. 2016: 179-192.
- [5] Ecker U K H, Lewandowsky S, Chang E P, et al. The Effects of Subtle Misinformation in News Headlines[J]. Journal of Experimental Psychology: Applied, 2014, 20(4): 323-335.
- [6] Horne B, Adali S. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire Than Real News[C]//Proceedings of the 2nd International Workshop on News and Public Opinion at ICWSM. 2017.
- [7] Hu L M, Yang T C, Shi C, et al. Heterogeneous Graph Attention Networks for Semi-Supervised Short Text Classification[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. 2019: 4823-4832.
- [8] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter[C]//Proceedings of the 20th International Conference on World Wide Web. ACM, 2011: 675-684.
- [9] Shu K, Sliva A, Wang S, et al. Fake News Detection on Social Media: A Data Mining Perspective[C]//Proceedings of the 2017 ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2017: 22-36.
- [10] 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. ACM, 2016: 3818-3824.
- [11] Potthast M, Kiesel J, Reinartz K, et al. A Stylometric Inquiry into Hyperpartisan and Fake News[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018: 231-240.
- [12] Shu K, Cui L M, Wang S H, et al. dEFEND: Explainable Fake News Detection[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 395-405.
- [13] 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. 2017: 2931-2937.
- [14] 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 2018 AAAI Conference on Artificial Intelligence. 2018: 254-261.
- [15] Shrestha A, Spezzano F. Textual Characteristics of News Title and Body to Detect Fake News: A Reproducibility Study[C]//Proceedings of the 2021 European Conference on Information Retrieval. 2021: 120-133.
- [16] Ajao O, Bhowmik D, Zargari S. Sentiment Aware Fake News Detection on Online Social Networks[C]//Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing. 2019: 2507-2511.
- [17] Giachanou A, Rosso P, Crestani F. Leveraging Emotional Signals for Credibility Detection[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2019: 877-880.
- [18] Wu L W, Rao Y. Adaptive Interaction Fusion Networks for Fake News Detection[C]//Proceedings of the 24th European Conference on Artificial Intelligence. 2020: 2220-2227.
- [19] Zhang X Y, Cao J, Li X R, et al. Mining Dual Emotion for Fake News Detection[C]//Proceedings of the 2021 International Conference on World Wide Web. ACM, 2021: 3465-3476.
- [20] Ghanem B, Rosso P, Rangel F. An Emotional Analysis of False Information in Social Media and News Articles[J]. ACM Transactions on Internet Technology, 2020, 20(2): Article No.19.
- [21] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019: 4171-4186.
- [22] Kant N, Puri R, Yakovenko N, et al. Practical Text Classification with Large Pre-Trained Language Models[OL]. arXiv Preprint,

arXiv: 1812.01207.

- [23] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[OL]. arXiv Preprint, arXiv: 1609.02907.
- [24] Shu K, Mahudeswaran D, Wang S H, et al. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media[J]. Big Data, 2020, 8(3): 171-188.
- [25] Yang F, Liu Y, Yu X H, et al. Automatic Detection of Rumor on Sina Weibo[C]//Proceedings of the 2012 ACM SIGKDD Workshop on Mining Data Semantics. 2012: Article No.13.
- [26] Kwon S, Cha M, Jung K, et al. Prominent Features of Rumor Propagation in Online Social Media[C]//Proceedings of the IEEE 13th International Conference on Data Mining. 2013: 1103-1108.
- [27] Pan J Z, Pavlova S, Li C, et al. Content Based Fake News Detection Using Knowledge Graphs[C]//Proceedings of the 17th International Semantic Web Conference. 2018: 669-683.
- [28] Dun Y, Tu K, Chen C, et al. KAN: Knowledge-Aware Attention

Network for Fake News Detection[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. 2021: 81-89.

作者贡献声明:

刘赏:确定研究框架,修订论文;
沈逸凡:设计研究方案,进行实验,撰写论文。

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:

[1] 刘赏. GossipCop数据集. DOI:10.57760/sciencedb.j00133.00085.

收稿日期:2022-04-02

收修改稿日期:2022-08-08

Detecting Fake News Based on Title-Content Difference

Liu Shang Shen Yifan

(School of Science & Technology, Tianjin University of Finance and Economics, Tianjin 300222, China)

Abstract: [Objective] This paper proposes a fake news detection method based on the difference between news titles and contents, aiming to address the issues of extracting features from short news texts or retrieving comments. [Methods] Firstly, we designed the Cos-Gap calculation method to obtain the difference between news titles and contents' textual and emotional features. Then, we constructed a News Differential Heterogeneous Graph Network (NDHN) based on the obtained differential features and the Heterogeneous Graph Attention Networks. The NDHN contains edges constructed based on differential features and nodes constructed based on semantic and emotional features of title, content, and emotion. [Results] We examined the proposed model on the GossipCop dataset and found that the NDHN can improve the classification accuracy by 2.7% and the F1 by 3.2%. [Limitations] This method is suitable for analyzing the news with title and has limitations for untitled texts from Sina Weibo or Twitter. [Conclusions] The new model could effectively detect fake news from social media.

Keywords: Fake News Detection Heterogeneous Graph Network Differential Features
Public Opinion Analysis