

融合外部知识和用户交互特征的虚假新闻检测

刘 帅¹ 傅丽芳²

¹(东北农业大学工程学院 哈尔滨 150038)

²(东北农业大学文理学院 哈尔滨 150038)

摘要:【目的】针对虚假新闻在社交媒体中肆意传播这一现象,通过融入外部知识特征和用户交互特征,构建多维度数据分类模型以提高虚假新闻检测的效率和准确性。【方法】提取虚假新闻文本的背景知识,通过维基知识图谱引入外部知识检测新闻内容与既有知识体系的内在一致性,同时根据心理学中相似效应理论分析传播链上的用户交互,通过改进图卷积网络的连接边权更真实地体现用户间相互影响,构建了一个融合外部知识、新闻内容、传播链特征与用户交互关系的多维度数据虚假新闻检测模型。【结果】在两个公开数据集 Twitter15、Twitter16 上验证模型的性能,与 5 个类似模型进行对比分析,该模型的准确率分别达到 0.901 和 0.927。【局限】未考虑新闻附加内容中隐藏的知识信息和语言表达等其他特征,模型的可解释性也需要进一步提高。【结论】外部知识和传播链用户交互特征等多维度数据信息融合的检测模型能够有效提高虚假新闻的识别准确率。

关键词: 虚假新闻检测 特征工程 网络社交媒体 知识图谱

分类号: G250 TP393

DOI: 10.11925/infotech.2096-3467.2022.1144

引用本文: 刘帅, 傅丽芳. 融合外部知识和用户交互特征的虚假新闻检测[J]. 数据分析与知识发现, 2023, 7(11): 79-87. (Liu Shuai, Fu Lifang. Identifying Fake News with External Knowledge and User Interaction Features[J]. Data Analysis and Knowledge Discovery, 2023, 7(11): 79-87.)

1 引 言

随着互联网的发展和普及,网络社交媒体逐渐成为发布、传播新闻的主流平台。社交媒体的蓬勃发展使得网络上新闻发布数量呈指数型增长,随之而生的是虚假新闻泛滥问题。若不加以遏制,不仅会损害媒体的公信力,还会给国家和社会带来严重危害。为遏制虚假新闻所带来的一系列问题,首要任务是虚假新闻的检测。

针对如何科学有效地检测虚假新闻,许多学者做了大量研究。这些研究大体可以分为两类:基于新闻文本的虚假新闻检测和基于社交媒体的虚假新闻检测。前者主要是从新闻内容入手,通过检测文本的语言风格^[1]、图文语义一致性^[2]等来检测虚假新

闻。后者是从新闻的社交语境中提取有效信息辅助模型检测新闻真实性,如传播路径^[3]、用户信息^[4]等。研究新闻的文本信息和社交语境是至关重要的。然而已有研究忽略了转发虚假新闻的社交用户之间存在的内在联系,也忽视了现有的外部知识体系对于检测虚假新闻所能带来的帮助。

针对以上问题,本文提出一种融合外部知识特征和用户交互特征的多维度数据虚假新闻检测模型(Multidimensional-data Fake News Detection, MFND)。一方面,该模型挖掘虚假新闻文本中隐含的背景知识,并基于维基知识图谱引入外部知识作为辅助信息,分析新闻背景知识与既有知识体系的内在一致性;另一方面,根据心理学中相似效应理论分析传播链中用户的相互影响作用,构建加权图卷

通讯作者(Corresponding author): 傅丽芳(Fu Lifang), ORCID: 0000-0003-2298-2378, E-mail: lifangfu@neau.edu.cn。

积神经网络(Graph Convolutional Network, GCN)模型,更真实地模拟和刻画虚假新闻传播用户之间的交互关系。

2 相关工作

早期的虚假新闻检测主要依靠人工检测,但是这种方法检测的新闻数量十分有限。人工智能技术的发展给虚假新闻检测带来了新的机遇,其入手点一般分为两类:基于文本的检测和基于社交媒体的检测^[5]。

2.1 基于文本的虚假新闻检测

基于文本的虚假新闻检测方法主要依靠分析新闻的内容来检测新闻报道的真实性^[6]。张国标等^[2]利用虚假新闻的文字和图像之间存在的差异构建跨模态语义偏差检测模型,利用语义距离测度,实现新闻内容特征的语义一致性检测。Przybyla^[7]创建了基于风格的分类器,对新闻文本进行主题提取和风格分析,并与基于卷积神经网络(Convolutional Neural Network, CNN)的分类器进行比较,证明了基于风格的分类器对于新闻文本中情感元素的捕捉效果。Karimi等^[8]提出基于分层语篇级结构的模型用于虚假新闻检测,该模型以一种自动化、数据驱动的方式学习和构建真假新闻的语篇级结构。刘华玲等^[9]使用结合位置编码的自注意力机制和预训练的CNN模型分别提取文本和图像的语义特征,并引入混合注意力机制融合文本、图像特征用于虚假新闻检测。元鹏等^[10]提出一种语义增强的虚假新闻检测方法,利用预训练语言模型理解新闻图像的深层语义,在此基础上提取不同语义层次的视觉特征。刘赏等^[11]研究新闻标题和正文文本在语义和情感上的差异性特征,基于异构图注意网络构建了新闻差异性检测模型。

2.2 基于社交媒体的虚假新闻检测

基于社交媒体的虚假新闻检测方法主要是利用虚假新闻在社交媒体上传播所形成的社会语境进行分析检测。为了研究虚假新闻的传播(Propagation)结构和散布(Dispersion)结构,Bian等^[12]使用自上而下和自下而上的图卷积神经网络构建了双向卷积检测模型。Chen等^[13]将检测的重点放在对虚假新闻来源网站的检测上,利用推特用户共享结构挖掘政

治类网站,并使用topic-agnostic分类器对网站进行打分、排序并分类,构建了虚假新闻域主动检测模型。Liang等^[14]注重于研究用户特征对虚假新闻检测模型的影响,通过提取5个新的用户特征,并将其与已经过验证的用户有效特征进行结合,构建全新的用户特征集,并将其应用于微博真实性检测。

综上,在新闻文本方面,大多数现有模型对虚假新闻内容的挖掘和提炼不够充分,通常只关注对文本语义特征的挖掘,忽略了隐藏在新闻内容中的背景知识及其与外部知识体系的内在关系;在传播链条方面,现有模型往往只强调来自上一传播节点的单一链条,忽略了现实中用户之间错综复杂的交互关系和相互影响。因此,本文基于维基知识图谱引入外部知识特征,并根据“相似效应”建立加权图卷积网络分析用户交互信息,构建了一个融合多维度数据信息的模型MFND用以检测虚假新闻。

3 MFND模型建立

MFND模型通过知识挖掘和实体链接技术,基于维基知识图谱引入外部知识体系作为辅助信息。采用GCN对处于传播链上的用户之间的关系网络进行分析。为了使模型更贴近新闻传播时社交用户间的真实交互关系,本文引入心理学中“相似效应”的概念改进拉普拉斯矩阵,通过计算用户间的相似度构建加权图卷积网络。同时,分别使用BERT(Bidirectional Encoder Representations from Transformers)和GRU(Gate Recurrent Unit)对新闻文本内容和传播序列进行特征提取。针对上述多维度信息,本文采用co-attention机制进行融合,并输入分类器中进行概率预测。模型框架如图1所示。

3.1 文本预处理与用户数据收集

预处理新闻文本,删除冗余信息和噪声,并使用NLTK进行分词处理。经过预处理后,新闻文本数据被表示为 $S = \{s_1, s_2, s_3, \dots, s_n\}$,其中 $s_i(i = 1, 2, 3, \dots, n)$ 表示第 i 个预处理过的新闻文本。

用户信息是实现虚假新闻智能化检测的重要信息之一。对于新闻文本集合 S ,收集转发用户组成用户集合 $U = \{u_1, u_2, u_3, \dots, u_m\}$,收集每个用户的个人信息组成用户特征矩阵 $X_u = \{X_{u1}, X_{u2}, X_{u3}, \dots, X_{um}\}$,其中 u_j 表示第 j 个用户、 X_{uj}

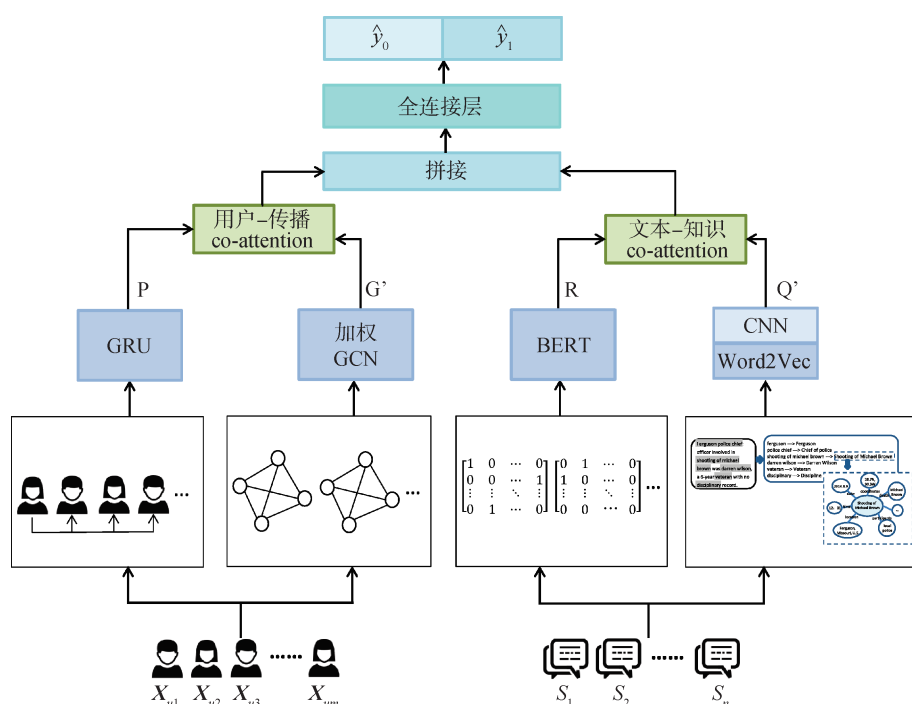


图1 MFND模型的结构框架

Fig.1 The Framework of MFND Model

表示用户 u_j 对应的特征向量。

用户特征向量 \mathbf{X}_{u_j} 包含以下信息:① x_1 :用户昵称的字数;② x_2 :个人简介的字数;③ x_3 :用户的粉丝数量;④ x_4 :用户的关注账号的数量;⑤ x_5 :用户发布的推文数量;⑥ x_6 :是否允许分享地理定位;⑦ x_7 :是否经过实名认证;⑧ x_8 :账号使用时长(单位:年);⑨ x_9 :从新闻发布到用户转发的时间差;⑩ x_{10} :账号是否受到保护。最终形成 $\mathbf{X}_{u_j} \in \mathbb{R}^v$, 其中 v 代表特征数量, 本文中 $v = 10$ 。

3.2 源信息文本知识挖掘及特征提取

基于经过预处理的新闻文本,引入外部知识作为辅助信息。由于虚假新闻经常包含一些与现有知识体系相矛盾的错误信息,因此引入外部知识体系检测虚假新闻具有积极意义。本文基于维基知识图谱引入外部知识,维基知识图谱是自由开放的,其内容都是结构化信息,能够为本文模型提供有力支撑。外部知识的引入通常分为三个步骤:

(1)文本知识抽取:知识抽取的目的在于从新闻文本中挖掘出可能成为知识概念的词汇。针对经过预处理的新闻文本数据,使用 Python 第三方库

TagMe抽取文本中的知识概念。以一条具体的推文为例,其最终抽取出来的知识概念如图2所示。

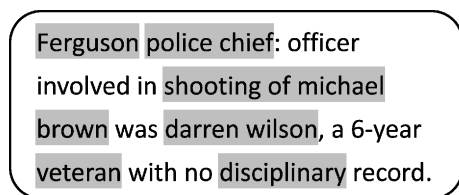


图2 知识抽取结果

Fig.2 Knowledge Extraction Results

(2)知识实体链接:知识实体链接是指将文本中抽取出来的知识概念(Mention)与知识图谱中的知识实体(Entity)相互链接的过程。实体链接工作需要克服两个难点:一个知识概念对应不同的知识实体;一个知识实体对应不同的知识概念。对此本文采用 Shen 等^[15]提出的 Candidate Entity Generation (CEG)和 Entity Disambiguation (ED)方法来解决,获得每篇新闻的知识实体序列 $\mathbf{E} = \{e_1, e_2, e_3, \dots, e_n\}$, 其中 n 表示每篇新闻包含的实体数量。同时,为了获得更加丰富的外部知识,对于每个知识实体 e_i ,

本文选择出10个与之联系紧密的相邻实体组成实体上下文,如公式(1)所示。

$$en(e_i) = \{en | (en \sim e_i) \in KG\} \quad (1)$$

其中, $en \sim e_i$ 表示实体 en 与实体 e_i 之间存在一种相互关系, KG 表示知识图谱, $en(e_i)$ 表示知识实体 e_i 的实体上下文。对于知识实体序列 $E = \{e_1, e_2, e_3, \dots, e_n\}$, 获取每个实体的实体上下文后, 可以组成实体上下文序列 $EN = \{en(e_1), en(e_2), en(e_3), \dots, en(e_n)\}$ 。以图2所展示的新闻为例, 本文基于维基知识图谱对其进行知识实体链接, 链接结果如图3所示。可以看出, “police chief”与实体“Chief of police”相关联, “shooting of michael brown”与实体“Shooting of Michael Brown”相关联。以实体“Shooting of Michael Brown”为例, 本文绘制知识网络图, 用以展示实体上下文挖掘结果。实体上下文是指目标实体在知识图谱中的邻近节点。本文选择在维基知识图谱中与目标实体只有一跳(One Hop)距离的10个相邻实体组成实体上下文。由于篇幅限制, 仅在网络图中绘制出6个相邻实体。维基知识图谱的基本组成单位是“实体-关系-实体”三元组, 该网络图也是根据这一原则进行绘制的。如三元组“Shooting of Michael Brown-death-Michael Brown”表达了这一客观事实: Michael Brown枪击案中死者为 Michael Brown。

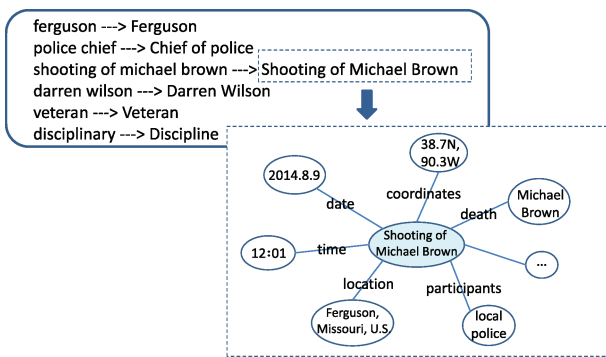


图3 实体链接结果展示
Fig.3 Entity Link Result

(3)知识特征提取: 针对知识实体序列 E 和实体上下文序列 EN , 使用 Word2Vec 得到实体嵌入 $E' = \{e'_1, e'_2, e'_3, \dots, e'_n\}$ 和实体上下文嵌入 $EN' =$

$\{en'(e_1), en'(e_2), en'(e_3), \dots, en'(e_n)\}$, 并且 $E', EN' \in R^{n \times d}$, 其中 n 为实体个数, d 为嵌入维数。由于实体上下文 $en(e_i)$ 是由多个实体 en 组成的, 因此计算 $en'(e_i)$ 是个求均值的过程, 具体如公式(2)所示。

$$en'(e_i) = \frac{1}{num(en(e_i))} \sum_{j=0}^{num(en(e_i))} en'_j \quad (2)$$

其中, $en'(e_i)$ 表示实体 e_i 的实体上下文嵌入向量, $num(en(e_i))$ 表示 $en(e_i)$ 所包含的相邻实体数量, en'_j 表示第 j 个相邻实体对应的嵌入向量。随后, 获得同时包含知识实体和实体上下文的知识嵌入 EE' 。为了简化计算, 通过计算 E' 和 EN' 平均值的方法获得知识嵌入序列 $EE' = \{ee'_1, ee'_2, ee'_3, \dots, ee'_n\}$, 计算方法如公式(3)所示。最后, 将 EE' 输入 CNN, 输出同时包含知识实体和实体上下文信息知识特征矩阵 Q' 。

$$ee'_i = \frac{1}{2} (e'_i + en'(e_i)) \quad (3)$$

3.3 基于BERT的文本特征和基于GRU的传播特征提取

对经过预处理的新闻文本数据 s_i 进行 One-Hot 编码, 获得高维 0-1 矩阵。由于文本长度不同, 需要设定最大文本长度 L 来统一维度, 小于最大文本长度的部分用 0 填充, 得到 $OH = \{oh_1, oh_2, \dots, oh_L\}$, 其中 oh_i 表示第 i 个单词的 One-Hot 编码。将 OH 代入公式(4)中以计算词嵌入。

$$V = \tanh(W_{oh}OH + b_{oh}) \quad (4)$$

其中, V 表示词嵌入, $V = \{v_1, v_2, \dots, v_L\}$; W_{oh} 表示权重矩阵; b_{oh} 表示偏置项。最后, 将 V 输入 BERT 获得新闻的文本特征 $R = \{r_1, r_2, \dots, r_L\}$, $R \in R^{d \times L}$, 如公式(5)所示。其中, d 表示维数。

$$r_i = BERT(v_i) \quad (5)$$

针对用户信息 X_u , 使用 GRU 模型可以获得新闻的传播特征 $P = \{p_1, p_2, \dots, p_\eta\}$, $p_i = GRU(X_{ui})$ 。需要注意的是: 对于每条新闻, 用户数量应该相同, 数量设置为 η 。 X_{ui} 的顺序应与转发的时间顺序一致。

3.4 基于加权GCN的用户交互特征提取

为了获得所有转发同一篇新闻 s_i 的用户间的潜在交互信息, 本文使用 GCN 提取用户交互特征。以往的文献大多研究用户与传播链上前一位用户(即

被转发用户)之间的交互关系,但在实际情况中,用户决定转发新闻的行为并非只受到前一位用户的影响,而是综合地受到在其之前所有转发过该新闻的用户的影响。因此,本文对 GCN 加权。首先,创建用户社交网络图 $G=(V, \varepsilon)$,其中 V 是转发过新闻 s_i 的用户的集合, ε 是用户两两之间存在的交互关系。在实际的传播过程中,很难获得用户间的交互关系,为此本文引入社会心理学中“相似效应”概念,即人类更倾向于接受那些与自己相似的人的观点^[16]。本文创建的用户社交网络图 G 是完全连通图,每条边 ε_{ij} 都与一个权重 w_{ij} 相关联。通过计算用户 i 和用户 j 的特征向量 $\mathbf{X}_{ui}, \mathbf{X}_{uj}$ 的余弦相似度获得权重 w_{ij} 的具体数值,计算方法如公式(6)所示。最终获得权重矩阵 $\mathbf{W}_\varepsilon=[w_{ij}] \in \mathbf{R}^{\eta \times \eta}$,其中 η 表示用户的数量。

$$w_{ij} = \frac{\mathbf{X}_{ui} \cdot \mathbf{X}_{uj}}{\|\mathbf{X}_{ui}\| \|\mathbf{X}_{uj}\|} \quad (6)$$

随后计算出对应的依存矩阵 $\check{\mathbf{A}}$ 和度矩阵 $\check{\mathbf{D}}$,并使用 GCN 进行分析。在 GCN 网络中,层与层之间的传播如公式(7)所示。其中, $\check{\mathbf{D}}^{-\frac{1}{2}} \check{\mathbf{A}} \check{\mathbf{D}}^{-\frac{1}{2}}$ 是拉普拉斯矩阵, $\mathbf{H}^{(l)}$ 是第 l 层的特征向量, $\mathbf{H}^{(0)}$ 被设置为用户特征矩阵 \mathbf{X}_u , $\mathbf{W}^{(l)}$ 是每一层的权重矩阵, σ 是非线性激活函数。通过上述操作,可以获得用户交互特征 $\mathbf{G}' = \{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_\eta\}$ 。

$$\mathbf{H}^{(l+1)} = \sigma(\check{\mathbf{D}}^{-\frac{1}{2}} \check{\mathbf{A}} \check{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)}) \quad (7)$$

3.5 特征融合与新闻分类

对于上述操作提取的知识特征(3.2节)、文本特征(3.3节)、传播特征(3.3节)和用户交互特征(3.4节),本文使用 Lu 等^[6]的方法构建 co-attention 机制分别融合文本、知识信息以及传播、用户信息,得到 \mathbf{PG}' 和 \mathbf{RQ}' ,如公式(8)和(9)所示。

$$\mathbf{PG}' = \text{co-attention}(\mathbf{P}, \mathbf{G}') \quad (8)$$

$$\mathbf{RQ}' = \text{co-attention}(\mathbf{R}, \mathbf{Q}') \quad (9)$$

其中, \mathbf{P} 表示传播特征; \mathbf{G}' 表示用户交互特征; \mathbf{PG}' 表示融合传播特征和用户交互特征后得到的传播-交互信息。同理, \mathbf{R} 表示文本特征; \mathbf{Q}' 表示知识特征; \mathbf{RQ}' 表示融合文本特征和知识特征后得到的文本-知识信息。

拼接 \mathbf{PG}' 和 \mathbf{RQ}' ,并输入一个由全连接神经网络构成的分类器中,得到最终的分类概率分布 $y =$

$[\hat{y}_0, \hat{y}_1]$,其中 \hat{y}_0 和 \hat{y}_1 分别表示目标新闻被分类为真新闻 y_0 和假新闻 y_1 的概率。损失函数如公式(10)所示。

$$\Gamma = -y \log(\hat{y}_1) - (1 - y) \log(1 - \hat{y}_0) \quad (10)$$

4 实验结果与分析

4.1 数据集

本文使用的数据集来自于公共数据集 Twitter15 和 Twitter16^[17]。Twitter15 和 Twitter16 数据集常被用作验证虚假新闻检测模型有效性的支撑数据^[6,9,14,18]。这两个数据集分别收集了超过 10 万用户发布的新闻信息,具体内容包括:新闻 ID、新闻文本信息、每条新闻的传播链(传播链的基本组成结构为[用户 ID,新闻 ID,用户转发新闻的时间])、新闻标签(分为 4 类: true, false, non-rumor, unverified)。根据实验目的,选择新闻标签为“true”和“false”两种情形的数据作为本文的实验数据,具体数据结构如表 1 所示。出于对用户个人信息的保护,数据集中没有公开用户的个人信息,需要实验者根据数据集中的用户 ID 使用推特 API 自行爬取。爬取的用户信息同 3.1 节说明。数据集按照 7:3 的比例分为训练集和测试集。

表 1 数据集的结构

Table 1 Structure of the Dataset

对比项目	Twitter15	Twitter16
源推文数量	742	412
标签为真	372	205
标签为假	370	207
用户数量	190 868	115 036

4.2 参数设置

本次实验中模型的参数设置如表 2 所示。由于每个推特的文本长度和转发用户数量不同,因此本文设置最大文本长度为 30,每条文本的用户数设置为 40。对于转发用户数大于 40 的文本,只保留前 40 个用户;对于转发用户数小于 40 的文本,需要对转发该新闻的现有用户进行随机抽样来填补空缺。Kipf 等^[19]指出 GCN 的层数应控制在 1~3 层,因此本文的层数设置为 2,并采用 ReLU 函数作为 GCN 的激活函数。选择 Softmax 函数作为全连接神经网络

分类器的激活函数。

表2 模型的参数设置
Table 2 Parameter Setting

阶段	参数	值
预处理阶段	每个推文的用户数	40
	最大文本长度	30
	知识嵌入的维度	300
特征获取阶段	CNN输出维度	32
	过滤器尺寸	3
	BERT输出维度	32
	GRU输出维度	32
	GCN输出维度	32
	GCN层数	2
训练阶段	Optimizer	Adam
	Epoch	100
	学习率	0.001

4.3 不同模型的比较

将MFN模型与目前最先进的部分模型以及具有代表性的基线方法进行比较,以展示本文模型在虚假新闻检测方面的性能。对比模型如下:

(1)DTC^[20]:基于决策树(Decision Tree,DT)的检测模型,通过标记新闻内容的主题并提取相关的主题词特征达到自动评估新闻可信度的目的。

表3 实验结果

Table 3 Experimental Results

模型	Twitter15				Twitter16			
	F1-score	Recall	Precision	Accuracy	F1-score	Recall	Precision	Accuracy
DTC	0.495	0.481	0.496	0.495	0.562	0.537	0.575	0.561
SVM-TS	0.519	0.519	0.520	0.520	0.692	0.691	0.693	0.693
mGRU	0.510	0.515	0.515	0.555	0.556	0.562	0.560	0.661
CSI	0.717	0.687	0.699	0.699	0.630	0.631	0.632	0.661
GCAN	0.825	0.830	0.826	0.877	0.759	0.763	0.759	0.908
MFND	0.890	0.886	0.894	0.901	0.879	0.878	0.880	0.927

(1)使用深度学习方法的模型其各项指标普遍优于那些使用人工提取特征的模型(如DTC、SVM-TS)。这是因为基于深度学习的模型能够提取更高层次的新闻文本和用户特征。

(2)CSI模型综合了文本特征、用户特征以及转发序列特征,其各项指标明显优于分析单一模态特征或两种模态特征的模型,例如DTC、SVM-TS、mGRU。这也证明了虚假新闻检测的重点在于对多

(2)SVM-TS^[21]:基于支持向量机(Support Vector Machine,SVM)构建的虚假信息检测模型,该模型捕捉时间序列特征并与源推文相结合,共同分析新闻的真实性。

(3)mGRU^[22]:基于改进的递归神经网络(Recursive Neural Network,RNN)构建的检测模型,用来捕捉推文的背景信息随时间的变化,学习时间特征和文本特征。

(4)CSI^[23]:多模态虚假新闻检测模型,将源文本信息、用户信息与参与转发新闻的用户群体信息这三种信息结合分析,共同作用于虚假新闻检测任务,是一种较为先进的模型。

(5)GCAN^[6]:基于图感知网络的模型。通过分析源文本和对应的转发用户序列预测目标新闻的真实性,是一种较为先进的虚假新闻检测模型。

本文选择F1分数(F1-score)、召回率(Recall)、精确率(Precision)、准确率(Accuracy)作为统一的评价指标。对比实验的具体结果如表3所示。可见,MFND模型在所有指标上均优于其他模型。在Twitter15和Twitter16数据集上,其各项指标均值与GCAN相比分别平均提高5.3和9.4个百分点,证明了MFND模型检测虚假新闻的有效性。

模态特征的选择和分析。构建更为丰富的多模态信息特征是提高模型检测性能的有效途径。

(3)CGAN与CSI一样对文本、用户和转发序列特征进行分析。然而,相比CSI把用户关系简化为欧氏数据,CGAN构建了用户社交网络图,并将图神经网络引入模型中,创建了图感知型虚假新闻检测模型,其性能与CSI相比也有很大的提升。同时这也证明了在检测新闻真实性的过程中,对复杂数据

的简化会在一定程度上影响模型的检测效果,改进特征提取手段能够为模型检测虚假新闻带来帮助。在所有的模型中,MFND 具有最高的分类准确率,这也证明了引入外部知识特征作为辅助信息以及改进 GCN 挖掘更真实的用户交互特征能够为虚假新闻检测带来帮助,这也为虚假新闻的检测与分类提供了新的研究思路和模型。

4.4 消融实验

为了检验 MFND 模型中各个模块对于检测新闻真实度的有效性,基于原有模型进行消融实验。在 MFND 模型的基础上创建出 MFND/G、MFND/K 和 MFND/ALL,具体解释如下:

(1)MFND/G:删除原有模型的图感知模块。将加权 GCN 改为 CNN,并且不再创建用户社交关系图,而是直接使用 CNN 对用户数据进行卷积和分析。模型其他部分保持不变。

(2)MFND/K:删除原有模型的知识感知模块。随着外部知识信息的删除,也需要对应删除文本-知识 co-attention 机制。

(3)MFND/ALL:删除包括外部知识、用户交互以及传播链在内的全部辅助信息,模型仅对新闻的文本内容进行分析。

其中,MFND/G 模型中,CNN 的过滤器尺寸设置为 3,输出维度设置为 32。消融实验的结果如图 4 所示。

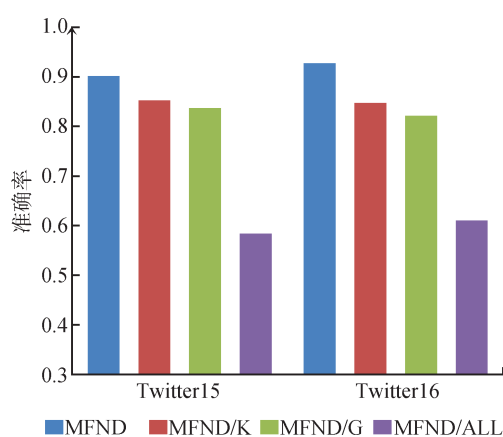


图 4 消融实验的结果展示

Fig.4 Results of Ablation Experiments

通过观察可知,模型准确率从高到底依次是:MFND>MFND/K>MFND/G>MFND/ALL。这说明

了原有模型的图感知模块、知识感知模块都对虚假新闻的检测提供了很大的帮助。引入外部知识体系以及提取更加贴近真实传播过程的用户交互信息作为辅助信息是可行的方法。且删除所有的辅助信息后,模型的分类准确率有明显下降,这也证明了辅助信息对模型检测虚假新闻带来的积极影响是不容忽视的。

5 结 语

研究虚假新闻的智能化检测有助于维护网络健康以及社会稳定,具有重要的现实意义。多维度信息融合的虚假新闻检测是当前的研究重点。然而,现有模型侧重于新闻文本语义信息和语言表达特征的提取,没有充分利用隐藏在新闻内容中的背景知识及其与外部知识体系之间的内在联系。同时,对于虚假新闻传播链的研究往往只关注前后用户节点间的单一链条,忽略了现实中多名用户之间复杂的相互影响和交互特征。针对上述问题,本文提出一个融合外部知识和用户交互特征的多维度数据虚假新闻检测模型 MFND。通过挖掘文本中隐藏的知识概念,并基于维基知识图谱引入外部知识实体作为辅助信息检测虚假新闻。同时,本文受到心理学中“相似效应”概念的启发,通过计算用户特征相似度构建加权图卷积网络提取传播链上用户的真实交互特征。该模型在公开数据集 Twitter15 和 Twitter16 上进行实验,准确率分别达到 0.901 和 0.927。在其他评价指标上与 5 个对比模型也具有较为显著的优势。本文研究为虚假新闻的检测提供了新的研究思路。

本文提出的虚假新闻检测模型还存在一定的局限性,主要包含以下两点:

(1)仅对新闻的文本内容进行知识挖掘,没有充分利用新闻的附加内容中所隐含的知识信息,例如隐藏在图片、视频或音频中的背景知识。

(2)对模型的可解释性研究不够充分,这是未来虚假新闻检测研究的热点问题之一。

在未来工作中,笔者将尝试从新闻的附加内容入手挖掘知识信息,获取更多的知识实体,进一步丰富模型的知识体系。同时致力于实现虚假新闻的可解释性检测,从新闻内容、传播新闻的用户特征、外

部知识等方面分析目标新闻被判断为假(或真)的原因,把复杂运算转化为人能够理解的信息,增强人们对分类结果的信任。

参考文献:

- [1] Popat K. Assessing the Credibility of Claims on the Web[C]// Proceedings of the 26th International Conference on World Wide Web Companion. 2017: 735-739.
- [2] 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.)
- [3] 段大高,白宸宇,韩忠明,等.基于多传递影响力的社交媒体谣言检测方法[J]. 计算机工程, 2022, 48(10): 138-145, 157. (Duan Dagao, Bai Chenyu, Han Zhongming, et al. Social Media Rumor Detection Method Based on Multi-Transmit Influence [J]. Computer Engineering, 2022, 48(10): 138-145, 157.)
- [4] Shu K, Wang S, Liu H. Understanding User Profiles on Social Media for Fake News Detection[C]// Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2018: 430-435.
- [5] 刘鹏飞. 基于多模态特征及语义增强的虚假新闻检测算法的研究与应用[D]. 青岛: 山东科技大学, 2020. (Liu Pengfei. Research and Application of Fake News Detection Algorithm Based on Multi-modal Feature and Semantic Enhancement [D]. Qingdao: Shandong University of Science and Technology, 2020.)
- [6] Lu Y J, Li C T. GCAN: Graph-aware Co-attention Networks for Explainable Fake News Detection on Social Media[OL]. arXiv Preprint, arXiv: 2004.11648.
- [7] Przybyla P. Capturing the Style of Fake News[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(1): 490-497.
- [8] Karimi H, Tang J. Learning Hierarchical Discourse-level Structure for Fake News Detection[OL]. arXiv Preprint, arXiv: 1903.07389.
- [9] 刘华玲,陈尚辉,乔梁,等.多模态混合注意力机制的虚假新闻检测研究[J]. 计算机工程与应用, 2023, 59(9): 95-103. (Liu Hualing, Chen Shanghui, Qiao Liang, et al. Multimodal False News Detection Based on Fusion Attention Mechanism[J]. Computer Engineering and Applications, 2023, 59(9): 95-103.)
- [10] 元鹏,曹娟,盛强.语义增强的多模态虚假新闻检测[J]. 计算机研究与发展, 2021, 58(7): 1456-1465. (Qi Peng, Cao Juan, Sheng Qiang. Semantics-Enhanced Multi-modal Fake News Detection [J]. Journal of Computer Research and Development, 2021, 58(7): 1456-1465.)
- [11] 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.)
- [12] Bian T, Xiao X, Xu T, et al. Rumor Detection on Social Media with Bi-directional Graph Convolutional Networks[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(1): 549-556.
- [13] Chen Z, Freire J. Proactive Discovery of Fake News Domains from Real-time Social Media Feeds[C]//Companion Proceedings of the Web Conference 2020. 2020: 584-592.
- [14] 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.
- [15] Shen W, Wang J, Han J. Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(2): 443-460.
- [16] 朱彤. 人际交往中的心理学[M]. 北京:金城出版社出版, 2009.
- [17] Ma J, Gao W, Wong K F. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017.
- [18] Lai Y, Zhang L, Han D, et al. Fine-Grained Emotion Classification of Chinese Microblogs Based on Graph Convolution Networks[J]. World Wide Web: Internet and Web Information Systems, 2019, 23(5): 2771-2787.
- [19] Kipf T N, Welling M. Semi-supervised Classification with Graph Convolutional Networks[C]//Proceedings of the International Conference on Learning Representations (ICLR 2017). 2016.
- [20] Castillo C, Mendoza M, Poblete B. Information Credibility on Twitter[C]//Proceedings of the 20th International Conference on World Wide Web. 2011: 675-684.
- [21] Ma J, Gao W, Wei Z, et al. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites[C]// Proceedings of the 24th ACM International Conference on Information and Knowledge Management. 2015: 1751-1754.
- [22] 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. 2016: 3818-3824.
- [23] Ruchansky N, Seo S, Liu Y. CSI: A Hybrid Deep Model for Fake News Detection[C]//Proceedings of the 26th ACM International Conference on Information and Knowledge Management. 2017: 797-806.

作者贡献声明:

刘帅:进行实验,采集、清洗和分析数据,起草论文;
傅丽芳,刘帅:提出研究思路,设计研究方案,论文最终版本修订。

支撑数据:

[1] Ma J. Twitter15、Twitter16 数据集 .<https://aclanthology.org/P17-1066>.

利益冲突声明:

所有作者声明不存在利益冲突关系。

收稿日期:2022-11-01

收修改稿日期:2023-03-04

Identifying Fake News with External Knowledge and User Interaction Features

Liu Shuai¹ Fu Lifang²

¹(College of Engineering, Northeast Agricultural University, Harbin 150038, China)

²(College of Letters and Science, Northeast Agricultural University, Harbin 150038, China)

Abstract: [Objective] This paper proposes a multidimensional-data classification model to improve the efficiency of fake news detection. The new model incorporates external knowledge features and user interaction features to reduce fake news spreading in social media. [Methods] First, we extracted the background knowledge of fake news. Then, we introduced external knowledge through the Wikipedia knowledge graph to detect the consistency between the news content and the existing knowledge system. Third, we analyzed the user interaction on the communication chain according to the psychological “similarity effect”. Finally, we improved the connection edge weight of the graph convolutional network to reflect the interaction between users. [Results] We examined the new model’s performance with two public datasets, Twitter15 and Twitter16. Compared with the other five similar models, our model’s accuracy reached 0.901 and 0.927. [Limitations] We did not consider features like knowledge information and language expression hidden in the additional news content. The model’s interpretability needs to be further improved. [Conclusions] By integrating news content, external knowledge, and user interaction characteristics of the communication chain, the proposed model can effectively detect fake news.

Keywords: Fake News Detection Feature Engineering Online Social Media Knowledge Graph