认识真相:利用客观事实和主观观点可解释的谣言检测

Jiawen Li、Shiwen Ni 和 Hung-Yu Kao智能知识管理
实验室国立成功大学计算机科学与信息工程系台湾台
南 {P78073012,P78083033}@gs.ncku.edu.tw,hykao@mail.ncku.edu。 tw

抽象的

目前的对抗训练大多是为了最小化输入扰动的最大风险,这已被证明是一种提高模型泛化能力的正则化方法。然而,只有输入攻击过于单一,我们将攻击扩展到神经网络的权重参数。在这项工作中,我们提出了一种新的对抗训练方法 DropAttack,它受到 dropout 思想的启发,让模型的某个权重参数以一定的概率受到攻击,通过最小化来提高神经网络的泛化能力重量攻击的对抗风险。为了验证所提方法的有效性,我们使用自然语言处理和计算机视觉领域的五个公共数据集进行实验测试。实验结果表明,与不使用 DropAttack 的神经网络相比,DropAttack 提高了所有数据集的泛化性能。此外,我们将所提出的方法与其他对抗训练方法和正则化方法进行了比较,我们的方法在所有数据集上都达到了最先进的水平。

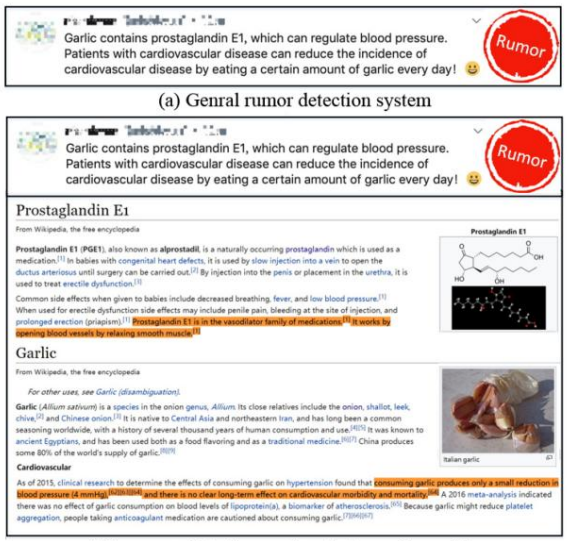

图1:两种不同系统的谣言检测结果对比图。图 1 (b) 中的橙色高亮部分是从维基百科中检索到的证据。从这些证据句中,读者可以很容易地判断给定的说法是否是半真半假,并清楚地理解为什么该说法是谣言。

1 简介

随着社交媒体平台的盛行,谣言已成为严重的社会问题。值得注意的是,现有的谣言检测方法粗略地将此任务表述为自然语言分类任务。该任务的目标是简单地将给定的文本声明标记为谣言或非谣言。然而,仅仅对可疑言论作出定论,不足以让人们理解和推理为什么一个说法是谣言。例如,图1是现有谣言检测方法与提供证据的谣言检测方法的对比图。图 1 中的说法半真半假,极具欺骗性。对于这样的谣言,仅仅提供一个标签是难以令人信服的。因此,我们相信一个好的谣言检测系统应该具备2个基本功能,即谣言识别功能和证据提供功能。

提供证据的谣言检测具有以下好处:(1)提高检测性能。 (2) 提升用户体验。 (3)为人工审核提供依据。 (4)提高早期谣言检测的准确性。 (五)拦截类似谣言的传播。

尽管有许多优点,但提供证据的谣言检测非常困难。
如果谣言检测训练数据集中没有包含标记的证据信息,则深度学习网络不太可能自行生成这些文本证据内容。不幸的是,目前用于谣言检测的数据集不能作为证据。

表1:主客观信息对比特征表。

主观性信息	客观信息需要抓取
容易访问广泛	
	稀有的
片面性	综合一致性高纯度
冲突有噪音	

为了找出可以用作证据的信息类型,本部分讨论了两种不同类型的信息,即主观信息和客观信息（Merigo 等人，2016 年； Zorio-Grima 和 Merello， 2020 年）。在谣言检测领域,主观信息是指来源推文、评论等,而客观信息是指维基百科或百度百科等信息。通过我们的综合分析,我们发现主观信息和客观信息表现出明显的区别- 不同的特征,总结在表 1 中。

客观信息一致性好,纯度高,可以作为证据,主观信息也包含一定的破绽线索

传闻。

为了同时利用主观信息和客观信息,本文提出了一种新模型 LOSIRD。这是众所周知的挑战,困难在于：(1)模型应该具有强大的检索能力。 (2) 模型应具有自然语言推理（NLI)能力。

(3) 模型需要能够处理拓扑信息。

图 2 显示了其架构的高级视图。
该模型分为两个模块,即ERM （证据检索模块)和RDM（谣言检测模块）。受迁移学习概念的启发,我们的 LOSIRD 模型采用了两阶段训练方法。在第一个培训阶段,广泛使用的事实核查数据库被用于培训企业风险管理模块。在第二个训练阶段,使用两个谣言检测数据集来训练和评估模型。

本文的主要贡献有四个方面：

1.这项研究首次提出了可以提供证据的谣言检测模型。

2.我们率先提出了两个新颖的图形对象来模拟传播布局的推文和嵌入证据的关系和索赔谣言检测任务。

3.我们的 LOSIRD在谣言检测任务中实现了最高的检测精度,并且优于最先进的模型。

4.我们的 LOSIRD在谣言的早期检测中更具泛化性和鲁棒性。

2 相关工作

2.1 证据检索

证据检索任务与谣言检测任务高度相关。 FEVER1 是使用最广泛的证据检索数据集之一。

大多数研究人员按照 FEVER 组织者的管道方法处理发烧分享任务,分三个步骤检索和验证证据（Hanselowski 等人，2018 年； Malon， 2018 年）。

周等。 (2019a)将声明验证制定为图推理任务,并提供两种注意力。刘等人。 (2020)提出了结合边缘内核和节点内核的 KGAT ,以更好地嵌入和过滤证据。钟等。 (2020)构建了两个语义级拓扑来增强验证性能。米田等。 (2018)为发烧共享任务采用了四阶段模型。

2.2 谣言检测

现有的谣言检测深度学习方法可以分为三类,特征驱动法、内容驱动法和混合驱动法。

特征驱动方法,如机器学习方法,依靠各种特征来识别谣言。拉斯等人。 (2017)提出了一种新的可信度概念,用于自动识别传播谣言的用户。

内容驱动方法是一种基于自然语言处理的方法。许多研究人员采用深度学习模型来处理这项任务（Rath 等人， 2017 年； Ma 等人， 2016 年； Yu 等人， 2017 年； Chen 等人， 2018 年； Ma 等人， 2018 年）。蒙蒂等。 (2019)提出了GCN基于传播的假新闻检测。 Nguyen (2019)使用多模式社交图检测谣言。苏贾娜等人。 (2020)提出了一种用于假新闻检测的多损失分层 BiL STM 模型。

[1] http://fever.ai/task.html

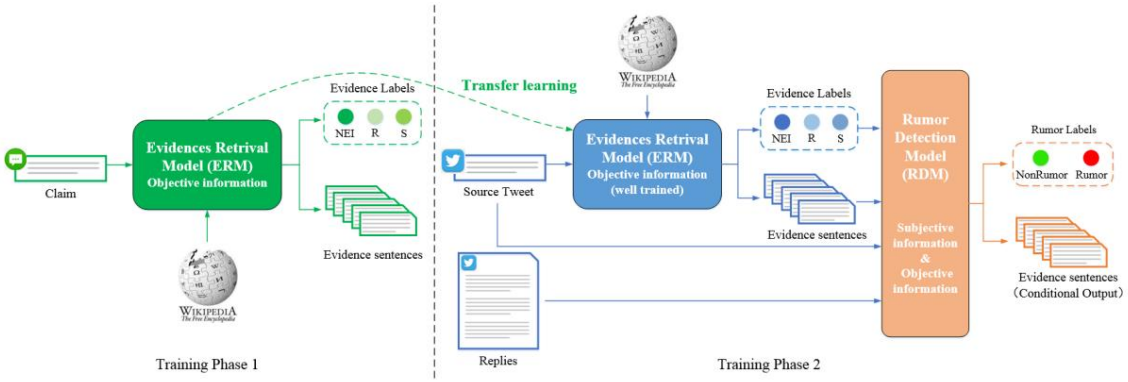

图 2:我们的 LOSIRD 模型的架构。声明和源推文在本文中本质上是同一件事，“声明”表示模型正在使用 Fever 数据集进行训练，而“源推文”表示使用 PHEME 数据集。

混合驱动方法结合了特征工程和文本信息表示来检测谣言（Liu 和 Wu， 2018； Yang 等人， 2018）。 Ruchansky 等人。 (2017)提出了一种称为 CSI 的谣言检测模型，该模型使用文章并提取用户特征来揭穿谣言。 Lu 和 Li (2020)通过从用户的个人资料和社交互动中提取用户特征来对谣言进行分类。李等。 (2020b)使用 Graph SEGA 对对话结构进行编码。李等。 (2020a)抓取用户-关注者信息，并基于关注者关系构建友好网络。卡斯蒂略等人。 (2011)使用推文和重新发布信息来检测谣言。

Kochkina 等人。 (2018);李等。 (2019)提出了一种多任务学习方法来联合训练主任务和辅助任务，提高模型的谣言检测性能。刘等人。 (2015)汇总了 Twitter 用户的常识和调查性新闻，用于谣言检测。嘛

等。 (2017)用于谣言检测的编码帖子的传播结构。通过内核学习使用传播结构检测微博帖子中的谣言。

2.3 比较

我们模型的亮点包括提供证据，覆盖两个异构结构图信息，结合证据线索和回复信息检测谣言。我们的模型表现出更强的模拟能力、更好的可扩展性和更好的说服能力。

3 LOSIRD 模型

3.1 问题陈述

我们将这个谣言检测任务制定为结合证据检索子任务的混合任务

和谣言预测子任务。

证据检索子任务定义为：

给定一个声明，此子任务的目标是匹配来自维基百科的文本证据，并将这些潜在证据句子与给定声明之间的关系推理为“支持”、“反驳”或“信息不足（NEI）”。

我们将维基百科定义为客观信息语料库： W iki = {D1, D2, ..., D|w|}， Di作为来自维基百科的文档。一份文件包含描述维基百科中一个实体的几个句子。该子任务的目标是检索证据，对证据与给定声明C之间的关系进行分类，即fERM : C → {(ye, E); E ∈ W iki}， ye是声明的预测证据标签， E是声明的检索证据集，包含多个句子级别的证据。

谣言预测子任务定义为：

给定一个声明、该声明的回复、该声明检索到的证据集和证据标签，该模型检测该声明是否为谣言

或非谣言并提供证据。我们将此子任务中的谣言数据集定义为Ψ = {T1, T2, ..., T|Ψ|}，其中Ti是数据集中的推文。Ti = {Ci , , yei}，其中Ci是谣言数据集中第 i 个来源Pi文， Pi是来源推文的回复帖子， Ei和yei是对应的Ci检索到的证据集和证据标签。鉴于

一条推文T这个任务的函数被定义为

fRDM : T → {(yr, E); E ∈ W iki}， yr是预测的谣言标签。

3.2 企业风险管理

主要遵循(Liu et al., 2020)和(Hanselowski et al., 2018)，我们采用了一个三步管道模块来检索证据,称为 ERM。ERM 的体系结构如图3 所示。它包含三个主要步骤,即文档检索:采用关键字匹配算法来爬取维基百科中的相关文件。

证据检索:提取句子级证据
从检索到的文章。 Claim verification:基于句子级别的证据,预测claim与证据的关系为“支持”、“反驳”或“NEI”。具体来说， ERM 首先利用语义 NLP 工具包从给定的声明中提取潜在实体。

通过解析的实体， MediaWiki API 过滤了前 k 个排名最高的维基百科文章。然后,从这些检索到的文档中,ERM以相关的句子形式提取客观事实作为预测证据

为理赔。最后,ERM 的验证组件对给定的陈述和检索到的证据进行预测,并验证声明和证据之间的关系是支持、反驳还是 NEI。

3.3 资源管理

图 4 显示了 RDM 的结构。由于源推文在回复和证据之间形成不同的拓扑结构,因此在RDM中构建了两个异构图对象,即对话树形图和证据星形图。

3.3.1 两个异构图

对话树形结构是社交媒体自然形成的一种奇特的回复关系拓扑结构,承载着对话的重要线索。

mor 检测（Belkaroui 等人， 2014 年； Pace 等人，2016 年）。值得注意的是,对话结构是树形的。其中树的根是源推文,每个节点代表一条评论,每个节点由其回复关系连接。

证据星形结构表明每条证据都是对源推文的补充描述,因此与源推文直接相关的每个证据句子形成星形拓扑。在这个星形结构中,节点

源推文的位置在中心,所有证据节点都围绕着源推文,代表星形结构中的一个角度。

3.3.2 谣言检测模块

谣言检测模块包含四个组件：(1)词向量编码组件。 (2) 句子嵌入组件。 (3)图形处理组件。 (4) 分类器组件。

在 RDM 中,深度 BiLSTM 被用来提取单词之间的信息并生成句子表示。将获得的句子向量传递到图形处理组件，GraphSAGE （Hamilton 等人， 2017)模型用作其主干。GraphSAGE 有效地处理了变量图。由于前一个组件的输出是一组句向量,不包含结构信息。那里

因此,在将此信息传递给图形处理组件之前,分别构建了两个图形对象,即会话树形对象和证据星形对象。

会话图对象的创建：

$$\begin{aligned} Gp &= (Vp, Ep) \\ Vp &= [c, p1, p2, .., pj\,] \\ Ep &= \{(c, p1), ...(pn, pm), ...\} \end{aligned} \tag{1}$$

其中Gp是第 i 个事件的对话图对象,它的顶点集是Vp ，边集是Ep。顶点集包括事件中的所有帖子,边集Ep表示每个帖子之间的回复关系。 c和pj是BiLSTM 组件的推文嵌入结果,我们选择BiLSTM 的最后一个隐藏状态作为句子嵌入结果。

证据图对象的创建：

$$\begin{aligned} Ge &= (Ve, Ee) \\ Ve &= [c, e1, e2, .., ek] \\ Ee &= \{(c, e1),(c, e2), ...,(c, ek)\} \end{aligned} \tag{2}$$

其中Ge是由顶点集Ve和边集Ee组成的证据图对象。顶点集Ve包括源帖子c和证据句子,而边集Ee表示证据与源帖子之间的关系， ek表示来自 BiLSTM 组件的证据句子嵌入结果。

在前向传播步骤开始时,将每个节点的特征分配给节点

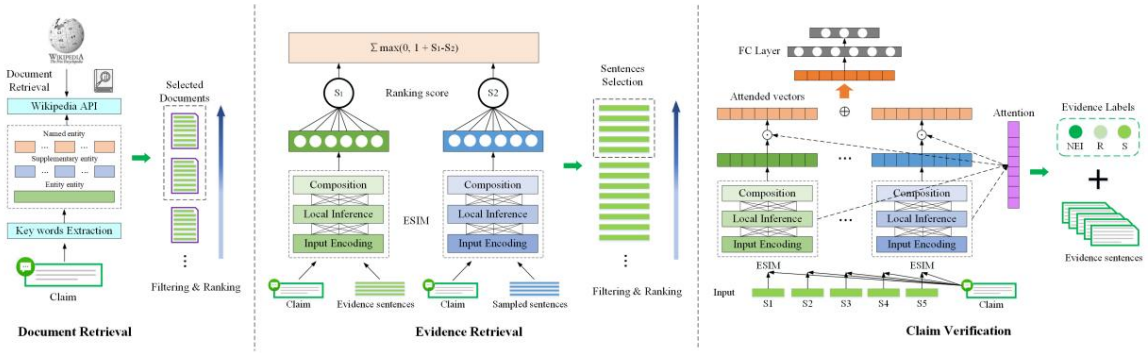

图 3:企业风险管理的架构。

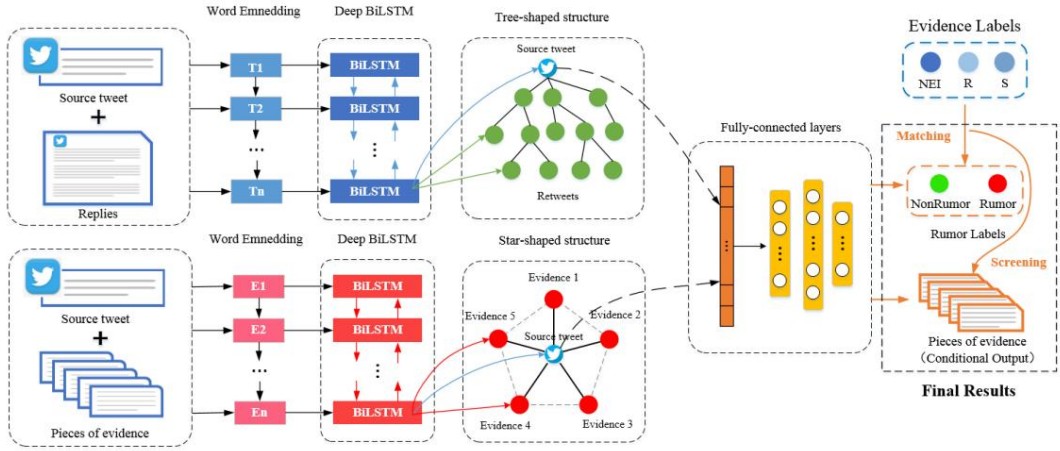

图 4:RDM 的架构。

在隐藏状态下如下：

$$[H^0_{p0}, h0_{p1}, ..., h0_{pj}] \leftarrow [c, p1, p2, .., pj]$$
$$[H^0_{e0}, h0_{e1}, ..., h0_{恩}] \leftarrow [c, e1, e2, .., ek] \quad (3)$$

其中h^0_{pj} , $h0_{恩}$ 是的初始隐藏状态

GraphSAGE 中对话图对象和证据图对象的节点。

节点在 GraphSAGE 中的隐藏状态通过不断聚合其直接邻居的隐藏状态,将它们与自己的状态结合并生成新的隐藏状态来更新。这个过程使节点获得越来越丰富的信息（Hamilton 等人，2017）：

$$h^k_{pN(v)} \leftarrow 聚合池(\{h^{k-1}_{浦}, \forall u \in N(Vp)\})$$
$$h^k_{pv} \leftarrow \sigma(Wk_p \cdot CON(h^{k-1}_{光伏}, h^{k-1}_{香港 pN(v)})) \quad (4)$$

$$h^k_{eN(v)} \leftarrow 聚合池(\{h^{k-1}_{欧盟}, \forall u \in N(Ve)\}) \quad (5)$$
$$h^k_{EV} \leftarrow \sigma(Wk_{e} \cdot CON(h^{k-1}_{,}, h^{k-1}_{香港 电子 (v)}))$$

其中$h^k_{pN(v)}$, $h^{k-1}_{香港 电子 (v)}$ 是他们聚合后的邻域向量， k是信息传输更新的深度（次数图信息被更新）， N是邻域函数， N(v)是节点的直接邻域的集合， AGGpool是聚合函数， CON是连接函数。

GraphSAGE 中提供了三个聚合器,在本文中我们选择了 Max Pooling 聚合器。这是公式：

$$AGGpool^k = \max(\{\sigma(Wpool \cdot h^k_{N(v)} + graph_{pool}), \forall u \in N(v)\}) \quad (6)$$

其中max是逐元素最大算子， σ 是非线性激活函数。

基于会话结构和星型结构进行k次信息传递迭代后,得到会话嵌入结果和证据嵌入结果的最终表示：

$$k\,p \leftarrow h_{pv}, \forall v \in Vp$$
$$k\,e \leftarrow h_{ev}, \forall v \in Ve \quad (7)$$

p,e是第 i 个事件的回复和证据。

表 2:FEVER 数据集的统计数据。

拆分	SUPPORTED	REFUTED	NEI	
火车	80,035	29,775	35,659	6,666 6,666 6,666 6,666
开发	6,666	6,666		
测试				

表 3:谣言数据集的统计数据。

统计	PHEME2017	PHEME2018
用户	49,345	50,593
帖子	103,212	105,354
事件	5,802	6,425
平均帖子/事件	17.8	16.3
谣言	1,972	2,402
非谣言	3,830	4,023

最大聚合器用于将信息聚合成固定大小。

此后,将这两个部分的信息连接在一起,然后传递到多层感知器中进行最终预测。公式如下:

$$yr = Sof\ tmax(V \cdot (p \oplus e) + by) \quad (8)$$

其中 V 和by是输出层中的参数。

4 实验与结果

4.1 数据集

发烧数据集用于训练证据重新
三级模块。 FEVER数据集的统计
如表 2 所示。两个广泛使用的谣言数据集 PHEME 2017 和 PHEME 20181用于训练和评估整个提议的模型,如表32所示。 ,是

4.2 实验设置

为了评估我们模型的谣言检测性能,我们将我们提出的模型与其他流行的谣言检测模型进行了比较,包括一些当前最先进的模型。在文本处理阶段,我们通过去除无用的表达和符号、统一大小写等来清理文本信息。我们使用 Twitter 27B 预训练的GloVe 数据,具有
200 维的词嵌入

ding 并将最大词汇量设置为 80,000。
对于谣言检测模块隐藏尺寸
BiLSTM 为 128,层数为

[1] https://figshare.com/articles/dataset/ 2本
研究符合访问和使用条件
这些谣言数据集。

2. graphSAGE的batch size为64,我们使用学习率为0.0015的Adam来优化模型,dropout rate设置为0.5。对于证据检索,我们将 ESIM 中的学习率设置为 0.002,丢弃率为 0,批量大小为 64,激活函数为 relu。对于索赔验证,我们设置 ESIM 中的学习率为 0.002,丢弃率为 0.1,批大小为 128,激活函数为 relu。我们拆分数据集,保留 10% 的事件作为验证集,并将其余按 3:1 的比例进行训练和测试分区。

- CNN:用于谣言检测的卷积神经模型（Chen 等人， 2017 年）。
- BiLSTM:用于揭穿谣言的双向 LSTM 模型（Augenstein 等人， 2016 年）。
- BERT:一种用于检测谣言的微调 BERT （Devlin 等人， 2019 年）。
- CSI:一种最先进的模型,通过根据用户的行为对用户进行评分来检测谣言（Ruchansky 等人， 2017 年）。
- DEFEND:最先进的模型学习源文章的句子和用户配置文件之间的相关性（Shu 等人， 2019 年）。
- RDM:一种结合GRU 和强化学习以在早期阶段检测谣言的最先进模型（Zhou 等人， 2019b）。
- CSRD:一种通过建模对话结构来检测谣言的最先进模型（Li 等人， 2020b）。
- LOSIDR:我们的模型利用客观事实和主观观点来检测可解释的谣言。

4.3 实验结果

主要实验结果如表 4 所示。LOSIRD 在 PHEME 17 和 PHEME 18 上的表现优于其他最佳竞争方法。其准确率在 PHEME 2017 中为 91.4%,在 PHEME 2018 中为 92.5%。此外,准确率、召回率和F1在两个数据集中都高于 90% 。这些有希望的结果证实了证据信息和拓扑消息处理方法在谣言检测中的有效性。

对于 CNN、BiLSTM、DEFEND 和 RDM 模型,它们通常根据发布时间将帖子连接成一行,同时忽略

表 4:主要实验结果。最好的模型和最好的竞争对手用粗体和下划线突出显示。

方法	2017年菲美				菲美2018			
	加速器	前	记录	F1	加速器	前	记录	F1
美国有线电视新闻网	0.787	0.737	0.702	0.710	0.795	0.731	0.673	0.686
双LSTM	0.795	0.763	0.691	0.725	0.794	0.727	0.677	0.701
伯特	0.865	0.859	0.851	0.855	0.844	0.834	0.835	0.835
CSI	0.857	0.843	0.859	0.851	0.851	0.836		
0.855	0.845	防御	0.868	0.867	0.859	0. 863	0.863	0.857
0.859	0.858	0.873	0.817	0.823	0.820	0.858	0.847	0.859
0.852								
RDM								
CSRD	0.900	0.893	0.869	0.881	0.919	0.892	0.923	0.907
洛西德	0.914	0.915	0.900	0.906	0.925	0.922	0.924	0.923

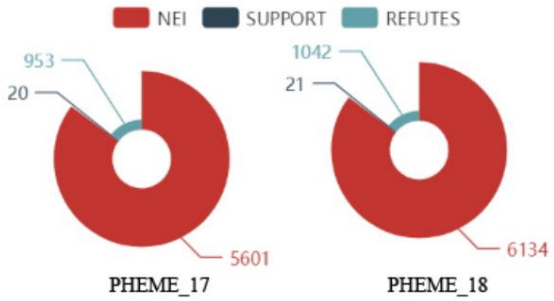

图 5:检索到的证据的分布。

会话结构信息。尽管如此,该结构对于将帖子编码为全面和精确的表示至关重要。 CSI 和 CRNN 处理拓扑信息,但这些模型仅采用主观信息,导致信息提取不足。

4.4 证据影响研究

在本节中,我们讨论了证据是否有助于谣言检测并确定了前者

证据在揭穿谣言中的影响。值得注意的是,评估的数据集是PHEME 2017 和 PHEME 2018。

4.4.1 检索证据的分布

为了准确评价检索到的证据,分析了检索到的证据基于证据标签的分布。构建了两个饼图来反映它们的分布情况。

如图 5 所示,大部分检索到的证据与给定的主张无关,约 14.8%的检索到的证据句子具有足够的信息来支持或反驳给定的主张。

尽管支持和反对的比例都不大,但这个结果还是值得称道的,好于我们的预期。

表 5:检索证据概率分析结果。

数据集原始反驳增量		
PHEME 17 33.30% 75.80% 42.50%		
PHEME 18 36.60% 70.92% 34.30%		

4.4.2 取证概率分析

我们通过统计计算原始数据中的谣言与被标记反驳的数据中的谣言之间的概率差距,进一步评估了证据的影响。结果如表5所示。

两个数据集中原始数据中谣言的概率约为35%,而标记为反驳的数据中谣言的概率约为73%,远高于原始数据。

具体来说,在 PHEME 17 上标记反驳的数据中的谣言增加到 42.5%,在

PHEME 18. 这有力地证实了检索到的证据是谣言检测的重要线索

化。

4.4.3 证据对深度学习模型的影响分析

为了进一步说明证据对谣言检测的影响以及分析证据对深度学习模型的影响,本小节特意选择了 CNN、BILSTM 和 BERT三种 NLP 模型作为检测模型。我们将可疑声明及其证据语句连接起来,分别将它们输入到三个模型中。实验结果如图 6 所示,横轴代表不同数量的证据句,0 表示只有源推文,而 1 到 5 表示源推文加上 1 到 5 个证据句。此外,本文分析了证据过滤前后的性能,用两个图表表示

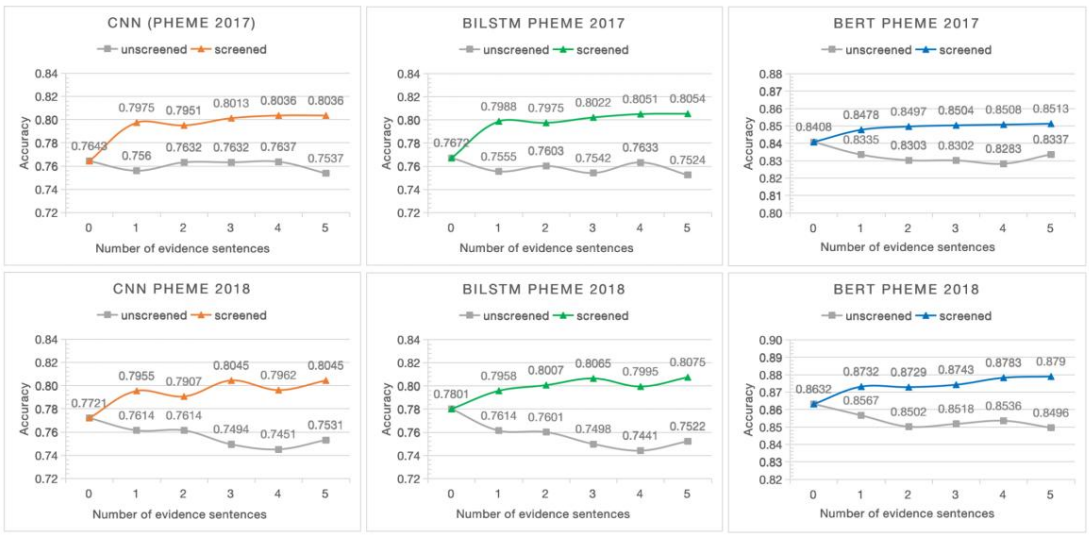

图 6:证据对深度学习模型性能的影响分析。

行,即,一条用于未筛选的证据（过滤 NEI 证据）,另一行用于筛选的证据。所有未屏蔽的断线

图表显示下降趋势。这表明
NEI 证据通过增加检测过程的难度来包含一定数量的无用信息。此外,在删除NEI 证据后,所有模型的准确性平均提高了 5%。

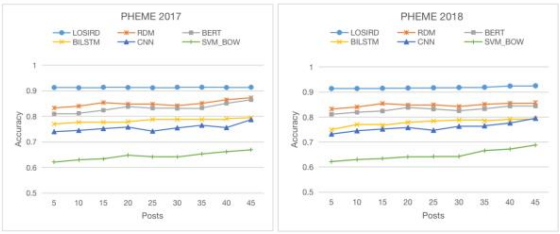

图 7:早期谣言检测性能。

这表明过滤后的证据可以显着帮助深度学习模型揭穿事实
谣言。

4.5 早期检测性能

为了评估我们模型的早期谣言检测性能,创建了 9 个反映谣言在 Twitter 上传播的真实场景的测试集。每个测试集包含不同数量的

回复,介于 5 条回复和 45 条回复之间。
测试子集是根据发布时间戳进行采样的。如图 7 所示,即使帖子数量只有 5 个,我们的 LOSIRD 模型在 PHEME 2017 数据集和 PHEME 2018 数据集上的准确率都超过 91%。此外,折线图显示我们模型的曲线非常稳定,表明在早期谣言检测中具有令人满意的鲁棒性和高性能。此外,我们的模型有效地利用了来自维基百科的客观信息,因此它不依赖于用户回复中的主观信息,从而在谣言传播的早期阶段取得了令人满意的性能。

5 结论

在本文中,我们提出了 LOSIRD,这是一种用于谣言检测的新型解释模型。值得注意的是， LOSIRD 辟谣机制取决于客观事实和主观观点。从 5,416,537 篇维基百科文章中检索到的客观事实句子被充分利用来帮助 LOSIRD分析可疑声明的真实性。

同时,基于会话结构模拟主观观点的传播,提取主观观点中的信息。

两个公共 Twitter 数据集的结果表明,与最先进的基线相比,我们的模型在一定程度上提高了谣言检测性能。此外,我们分析了客观事实对谣言检测的影响,并分析了对话结构的有效性。实验表明,客观事实和主观看法都是辟谣的重要线索。此外,我们相信我们的模型将用于谣言检测和其他文本分类

社交媒体上的任务。

致谢

这项工作部分由 Qualcomm通过台湾大学研究合作项目资助,部分由台湾科技部资助,资助 MOST 109-2221-E-006-173和 NCKU B109-K027D。

参考

Isabelle Augenstein、Tim Rocktaschel、Andreas Vlachos 和 Kalina Bontcheva。 2016.双向条件编码的姿态检测。在 2016 年自然语言处理经验方法会议论文集中,第 876-885 页,德克萨斯州奥斯汀。计算林学协会。

Rami Belkaroui、Rim Faiz 和 Aymen Elkhlifi。 2014. 社交网站的对话分析。 2014 年第十届信号图像技术和基于互联网的系统国际会议,第 172-178 页。 IEEE。

Carlos Castillo、Marcelo Mendoza 和 Barbara Poblete。 2011.推特上的信息可信度。 在第 20 届万维网国际会议记录中,WWW 2011,印度海得拉巴,2011 年 3 月 28 日至 4 月 1 日,第 675-684 页。美国计算机协会。

Tong Chen, Xue Li, Hongzhi Yin, and Jun Zhang. 2018. 引起对谣言的关注:用于早期谣言检测的基于深度注意力的递归神经网络。在关于知识发现和数据挖掘的亚太会议上,第 40-52 页。施普林格。

Yi-Chin Chen、Zhao-Yang Liu 和 Hung-Yu Kao。 2017. IKM 在 SemEval-2017 任务 8:用于姿态检测和谣言验证的卷积神经网络。在第 11 届国际语义评估研讨会 (SemEval-2017) 的会议记录中,第 465-469 页,加拿大温哥华。计算语言学协会。

Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。 2019. BERT:用于语言理解的深度双向转换器的预训练。在计算语言学协会北美分会 2019 年会议记录中:人类语言技术,第 1 卷（长文和短文）,第 4171-4186 页,明尼苏达州明尼阿波利斯。计算语言学协会。

William L. Hamilton、Zhitao Ying 和 Jure Leskovec。 2017.大图的归纳表示学习。神经信息处理系统进展 30:2017 年神经信息处理系统年会,2017 年 12 月 4-9 日,美国加利福尼亚州长滩,第 1024-1034 页。

Andreas Hanselowski、Hao Zhang、Zile Li、Daniil Sorokin、Benjamin Schiller、Claudia Schulz 和 Iryna Gurevych。 2018. UKP-athene:用于索赔验证的多句文本蕴含。在第一届事实提取和验证研讨会 (FEVER) 的会议记录中,第 103-108 页,布鲁塞尔,比利时。计算林学协会。

Elena Kochkina、Maria Liakata 和 Arkaitz Zubi aga。 2018. 多合一:用于 ru mour 验证的多任务学习。在第 27 届国际计算语言学会议论文集中,第 3402–3413 页,美国新墨西哥州圣达菲。作为计算语言学协会。

Jiawen Li、Shiwen Ni 和 Hung-Yu Kao。 2020a。鸡毛蒜皮的谣言在一起?探索社交媒体谣言的同质性和对话结构

检测。 IEEE 访问,8:212865–212875。

Jiawen Li、Yudianto Sujana 和 Hung-Yu Kao。 2020b。 利用微博对话结构来检测谣言。在第 28 届国际计算语言学会议论文集中,第 5420-5429 页,西班牙巴塞罗那（在线）。国际计算语言学委员会。

李全志、张琼、罗思。 2019.利用用户可信度信息、注意力和多任务学习进行谣言检测。在第 57 届计算语言学协会年会会议记录中,第 1173-1179 页,意大利佛罗伦萨。计算语言学协会。

Xiaomo Liu、Armineh Nourbakhsh、Quanzhi Li、Rui Fang 和 Sameena Shah。 2015. Twitter 上的实时谣言揭穿。在第 24 届 ACM 信息和知识管理国际会议论文集中,CIKM 2015,澳大利亚维多利亚州墨尔本,2015 年 10 月 19 日至 23 日,第 1867-1870 页。ACM。

杨柳和吴怡芳布鲁克。 2018.通过循环网络和卷积网络的传播路径分类,早期检测社交媒体上的假新闻。在第 32 届 AAAI 人工智能会议 (AAAI-18)、第 30 届人工智能创新应用 (IAAI-18) 和第 8 届 AAAI 人工智能教育进展研讨会 (EAAI-18) 的会议记录中, 美国路易斯安那州新奥尔良,2018 年 2 月 2 日至 7 日,第 354-361 页。 AAAI 出版社。

Zhenghao Liu、Chenyan Xiong、Maoson Sun 和 Zhiyuan Liu。 2020.核图注意力网络的细粒度事实验证。在计算语言学协会第 58 届年会会议记录中,第 7342-7351 页,在线。计算语言学协会。

Yi-Ju Lu 和 Cheng-Te Li。 2020. GCAN:用于社交媒体上可解释的假新闻检测的图形感知共同关注网络。在第 58 届会议记录中

计算语言学协会年会,第 505-514 页,在线。计算语言学协会。

Jing Ma、Wei Gao、Prasenjit Mitra、Sejeong Kwon、Bernard J. Jansen、Kam-Fai Wong 和 Meeyoung Cha。 2016.用递归神经网络检测微博谣言。第 25 届国际人工智能联合会议论文集,IJCAI 2016,美国纽约州纽约市,2016 年 7 月 9 日至 15 日,第 3818-3824 页。 IJCAI/AAAI 出版社。

Jing Ma、Wei Gao 和 Kam-Fai Wong。 2017.通过内核学习使用传播结构检测微博帖子中的谣言。在计算语言学协会第 55 届年会论文集（第 1 卷:长文）,第 708-717 页,加拿大温哥华。计算语言学协会。

Jing Ma、Wei Gao 和 Kam-Fai Wong。 2018.使用树结构递归神经网络在 Twitter 上检测 Ru mor。在第 56 届计算语言学协会年会会议记录（第 1 卷:长篇论文）,第 1980-1989 页,澳大利亚墨尔本。计算语言学协会。

克里斯托弗·马隆。 2018. Team papelo： FEVER 的跨前网络。在第一届事实提取和验证研讨会 (FEVER) 的会议记录中,第 109–113 页,比利时布鲁塞尔。计算语言学协会。

Jose M Merigo、Daniel Palacios-Marques 和 Shouzhen Zeng。 2016. 语言多标准群体决策中的主观和客观形成。欧洲运筹学杂志,248(2):522–531。

Federico Monti、Fabrizio Frasca、Davide Eynard、Da mon Mannion 和 Michael M Bronstein。 2019.
使用几何深度学习在社交媒体上检测假新闻。

清潭阮。 2019. 基于图的社交媒体谣言检测。技术报告。

Stefano Pace、Stefano Buzzanca 和 Luciano Fratoc chi。 2016. 社交网络对话的结构:对话和辩证线索之间。

国际信息管理杂志,36(6):1144–1151。

Bhavtosh Rath、Wei Gao、Jing Ma 和 Jaideep Srivas tava。 2017.从转发到可信度:利用信任来识别 Twitter 上的谣言传播者。在 2017 年 IEEE/ACM 国际会议论文集中,2017 年社交网络分析和挖掘进展,澳大利亚悉尼,2017 年 7 月 31 日至 8 月 3 日,第 179-186 页。美国计算机协会。

Natali Ruchansky、Sungyong Seo 和 Yan Liu。 2017.
CSI:一种用于假新闻检测的混合深度模型。
在 2017 ACM 会议记录中

关于信息和知识管理,CIKM 2017,新加坡,2017 年 11 月 6 日至 10 日,第 797-806 页。美国计算机协会。

Kai Shu、Limeng Cui、Suhang Wang、Dongwon Lee 和 Huan Liu。 2019.捍卫:可解释的假新闻检测。在第 25 届 ACM SIGKDD 知识发现与数据挖掘国际会议记录中,KDD 2019,美国阿肯色州安克雷奇,2019 年 8 月 4 日至 8 日,第 395-405 页。美国计算机协会。

Yudianto Sujana、Jiawen Li 和 Hung-Yu Kao。 2020.
使用具有衰减因子的多重损失分层 BiLSTM 检测 Twitter 上的谣言。在计算语言学协会亚太分会第一次会议和第 10 届国际自然语言处理联合会议论文集,第 18-26 页,中国苏州。计算林学协会。

杨洋、郑雷、张家伟、崔庆才、李周军和 Philip S Yu。 2018. Ti-cnn：用于假新闻检测的卷积神经网络。 arXiv 预印本 arXiv:1806.00749。

Takuma Yoneda、Jeff Mitchell、Johannes Welbl、Pon tus Stenetorp 和 Sebastian Riedel。 2018. UCL 机器阅读小组:事实发现的四因素框架 (HexaF)。在关于事实提取和验证 (FEVER) 的第一个工作坊的记录中,第 97-102 页,比利时布鲁塞尔。计算语言学协会。

冯宇、刘强、舒武、梁王、谭铁牛。 2017.错误信息识别的卷积方法。在第 26 届国际人工智能联合会议论文集中,IJCAI 2017,澳大利亚墨尔本,2017 年 8 月 19 日至 25 日,第 3901-3907 页。ijcai.org.

Wanjun Zhong、Jingjing Xu、Duyu Tang、Zenan Xu、Nan Duan、Ming Zhou、Jiahai Wang 和 Jian Yin。
2020.用于事实检查的语义级图推理。在计算语言学协会第 58 届年会会议记录中,第 6170-6180 页,在线。计算语言学协会。

Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2019a.
GEAR:基于图形的证据聚合和推理以验证事实。在第 57 届计算语言学协会年会会议记录中,第 892-901 页,意大利佛罗伦萨。

计算语言学协会。

Kaimin Zhou、Chang Shu、Binyang Li 和 Jey Han Lau。 2019b。早期谣言检测。在计算语言学协会北美分会 2019 年会议记录中:人类语言技术,第 1 卷（长文和短文）,第 1614-1623 页,明尼苏达州明尼亚波利斯。计算语言学协会。

Ana Zorio-Grima 和 Paloma Merello。 2020. 消费者信心:因果关系与主观和客观信息来源。技术预测与社会变革,150:119760。