文章编号:1000-5862(2021)04-0398-05

基于知识图谱和图像描述的虚假新闻检测研究

陈开阳,徐 凡*,王明文
(江西师范大学计算机信息工程学院,江西 南昌 330022)

摘要:针对传统虚假新闻检测方法主要采用图像统计学和图像分布式表示特征导致没有深层次挖掘图像所表达的文字含义的问题,设计了在融合知识图谱和图像描述的深度学习下的多模态虚假新闻检测模型. 该模型一方面抽取出在新闻文本中的 3 元组形式知识图谱,另一方面生成图像对应的描述文本,同时采用 Bert 框架将原文本、3 元组、图像描述文本加以集成. 在基准汉语虚假新闻语料库上的实验结果表明:该模型显著优于传统的代表性方法.

关键词:虚假新闻;知识图谱;图像描述;Bert
中图分类号:TP 311 **文献标志码**:A **DOI**:10.16357/j.cnki.issn1000-5862.2021.04.12

0 引言

随着互联网的广泛普及,相比较而言,人们从网上获取新闻比其他传统媒体更便利. 但是,在缺少有效监督的情况下,开放的互联网助长了大量虚假新闻的快速传播. 虚假新闻是错误且确实虚假的新闻内容,误导读者,给社会带来负面影响[1]. 虚假新闻具有成本低、获取方便、传播迅速等特点,容易误导社会舆论,扰乱社会秩序,损害社交媒体的公信力,侵害当事人利益,从而引发信任危机[2],如 2016 年美国总统选举导致了不良的政治影响[3].

多媒体技术推动了自媒体新闻从基于最初文本的帖子形式发展为带有图像或视频的多媒体形式,引起了消费者的更多关注,也提供了更可信的故事讲述方式. 一方面,图像、视频等视觉内容作为一种生动的描述形式比纯文本更具吸引力,从而促进了新闻传播. 例如,有图片的 tweets 比没有图片的 tweets 获得额外 18% 的点击率、89% 的喜欢率和 150% 的转发率. 另一方面,基于人们常识,视觉内容经常被用作故事的证据,这可以提高新闻的可信度. 不幸的是,这一优势也被虚假新闻发布者所利用. 为了达到快速传播目的,虚假新闻制造者通常设计包含虚假报道甚至篡改的图像或视频,以吸引和误导消费者. 因此,视觉内容已经成为虚假新闻检测不可忽视的重要组成部分,使得多媒体虚假新闻的检测成为新的挑战. 然而,现有传统虚假新闻检测方法主要采用图像统计学和图像分布式表示特征[5-22],这些方法并没有深层次挖掘出图像所表达的文字含义.

除了图像信息在虚假新闻检测中起到至关重要作用外,新闻文本内部所隐含的知识表达也同样重要. 于是,如何利用知识图嵌入技术进行虚假新闻检测,并合理利用与虚假新闻主题相关的知识图谱技术,这些都成为关键问题. 基于此,本文设计了融合知识图谱和图像描述的深度学习的多模态虚假新闻检测模型,该模型一方面抽取出在新闻文本中的 3 元组形式知识图谱,另一方面生成图像对应的描述文本,同时采用 Bert 框架将知识图谱和图像描述加以集成. 在基准汉语虚假新闻语料库上的实验结果表明该模型显著优于传统的代表性方法. 一方面,本文模型挖掘出虚假新闻文本隐含的 3 元组形式的知识图谱,并采用知识图嵌入技术得到其分布式表示形式;模型挖掘出图像对应的深层次描述文本,并将该生成的文本进行分布式向量表示;另一方面,采用 Bert 框架将上述 2 种分布式表示加以集成,并在基准汉语虚假新闻语料库上验证了该模型的有效性.

1 相关工作

目前具有代表性虚假新闻检测模型可以分为 2

收稿日期:2021-01-25
基金项目:国家自然科学基金(61772246,61876074,62162031)和江西省自然科学基金(20192ACBL21030)资助项目.
通信作者:徐 凡(1979—),男,江西万年人,副教授,博士,主要从事自然语言处理和语音信号处理的研究. E-mail:xufan@jxnu.edu.cn

大类: 单模态模型和多模态模型.

1.1 单模态模型

该类模型主要关注新闻的文本类型信息,主要方法有基于特征抽取的方法、基于核函数的方法、基于图模型的方法.

(i) 基于特征抽取的方法. 现有方法所涉及的特征可以分为基于用户的特征[5]、基于语言学的特征[6]、基于情感的特征[7]、基于位置-时间的特征[8]等. 在这些特征类型中,基于用户的特征最为常见,因为用户特征是一个较好的谣言检测指标. 同时,基于语言学的特征(如语言查询与字数统计(Linguistic Inquiry and Word Count,简称 LIWC)、文本可读性)也在虚假新闻检测中得到了广泛的应用. 此外,基于情感的特征也是一种良好的辟谣检测方法,因为用户的情感会影响对辟谣的判断. 最后,基于位置和时间的特征也能较好地反映出事件发生的地点和时间,同样是一个较好的辟谣检测指标.

(ii) 基于核函数的方法. Wu Ke 等[7-8]提出了一种基于核函数的谣言检测模型,考虑了网络谣言的传播特点. Wu Ke 等[7]发现,虚假谣言和正常信息的传播方式是不同的. 基于这一观察,他们设计了一种传播树用于谣言检测. 树的节点代表信息的发布者和评论传播信息的普通用户. 树的每条边表示任意 2 个节点之间的反应操作,每条边的权重表示为一个 3 元组,包括 2 个节点之间的认可分数、双倍分数和总体情绪分数. 然后,他们将随机游走图核(random walk graph kernel)和特征向量核集成为一个混合核,并进行谣言检测.

(iii) 基于图模型的方法. 基于图模型的方法核心思想是创建一些潜在变量,并采用超参数将先验知识融入图模型中. 一般来说,先验知识可以用真值和源权重的分布来体现. 如 Yang Shuo 等[9]将新闻真实性和用户可信度视为 2 个潜在的随机变量,并根据新的真实性来识别用户的意见. 为了解决这个问题,他们提出了一种基于 Gibbs 抽样的方法来推断新闻的真实性和用户的可信度.

1.2　多模态模型

通常而言,在社交媒体中文本内容和视觉内容会同时存在,它们都为检测虚假新闻提供了独特的线索. 因此,近年来对这一问题的研究主要集中在多模态信息的利用和有效融合上. 这些研究大多是简单地使用普通的递归神经网络(RNN)和预先训练的 CNN 来获得文本和视觉的语义特征. Jin Zhiwei 等[10]通过深度神经网络融合多模态内容,提出了一种具有注意机制的创新 RNN(attRNN),用于融合文本、视觉和社会语境特征. 对于给定的 tweets,其文本和社会背景首先与 LSTM 融合,以实现联合表示. 然后,将该表示与从预先训练的 deep CNN 中提取的视觉特征进行融合. Wang Yaqing 等[11]提出了一种端到端框架 EANN,用于基于多模态虚假新闻检测. 受对抗网络思想的启发,该模型采用事件识别器来预测训练阶段的事件辅助标签,相应的损失可用于估计不同事件之间特征表示的差异. 损失越大,差异越小. EANN 模型包括 3 个主要组件: 多模式特征提取器,假新闻检测器和事件鉴别器; 其中,多模式特征提取器与假新闻检测器配合执行识别假新闻的主要任务. 同时,多模式特征提取器试图欺骗事件判别器以学习事件不变表示. 对于多模式特征提取器,采用卷积神经网络(CNN)从帖子的文本和视觉内容中自动提取特征.

然而,现有虚假新闻检测方法主要采用图像统计学或图像分布式表示特征,并没有深层次挖掘图像所表达的文字含义. 基于此,本文设计了在融合知识图谱和图像描述的深度学习下多模态虚假新闻检测模型.

2　多模态虚假新闻检测模型

本文提出的融合知识图谱和图像描述的虚假新闻检测模型框架如图 1 所示,该模型旨在有效利用文本信息和图片信息的联合学习能力. 其中,文本语义向量的生成主要基于现有的数据集中的新闻文本信息,同时拼接上由知识关系抽取出的 3 元组信息向量. 图片信息采用图像描述的方式,生成图片对应的描述文本,将得到的图片描述向量与上述 2 种向量进行拼接后输入到 Bert 模型中进行分类预测. 接下来,分别介绍文本语义向量、知识图谱和图像描述生成、Bert 自注意力和模型训练共 4 个模块.

2.1　文本语义向量

从虚假新闻训练语料中生成一个词典 $C=\{c_1,c_2,\cdots,c_{n_1}\}$,对于中文来说,词典中的词是单个的字、单个的标点符号. 将词典中的词 c_i 都嵌入到一个行向量 d_i 中,d_i 的尺寸为 $1\times n_2$,尺寸值为 1×512. 并将所有行向量 d_i 按照顺序排列,组成矩阵 $\boldsymbol{D}=(d_1,d_2,\cdots,d_{n_1})$,其尺寸为 $n_1\times n_2$,其中尺寸值为 30 522 × 512. 给定训练样本 $X=(x_1,x_2,\cdots,x_{n_3})$. 对于 $i=1,2,\cdots,n_3$,取出 x_i 在矩阵 $\boldsymbol{D}$ 中对应行向量,记为 s_i. 将 s_i 按照从小到大的顺序排列,得到矩阵 $\boldsymbol{S}=(s_1,s_2,\cdots,s_{n_3})$,$\boldsymbol{S}$ 的尺寸为 $n_3\times n_2$,尺寸值为 128×512.

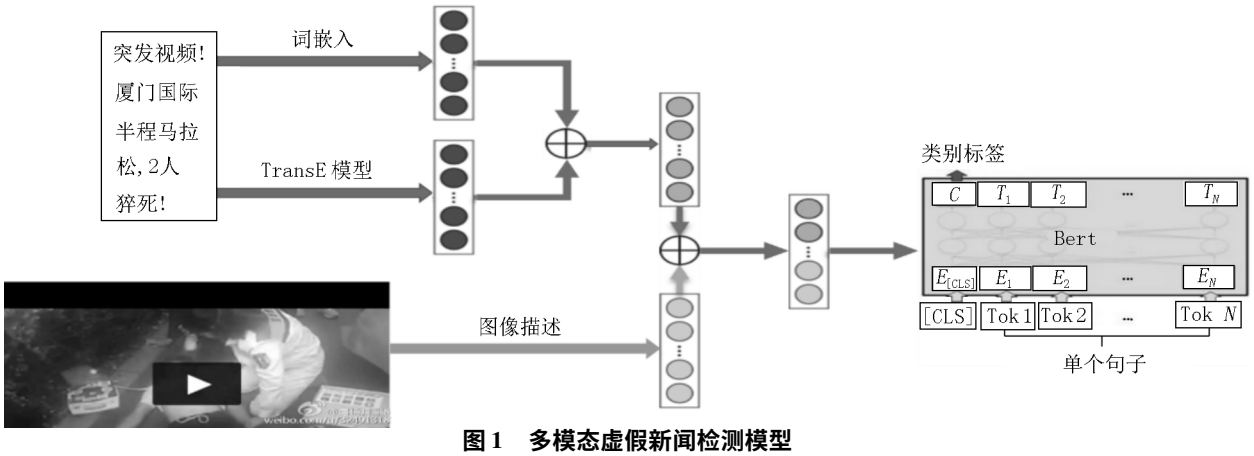

图 1 多模态虚假新闻检测模型

2.2 知识图谱和图像描述生成

本文利用 TransE 模型[12]获取虚假新闻文本的 3 元组信息. TransE 是基于实体和关系的分布式向量表示,将在每个 3 元组实例(head,relation,tail)中的关系 relation 看作从实体 head 到实体 tail 的翻译,通过不断调整 h、r 和 t(head、relation 和 tail 的向量),使($h+r$)尽可能与 t 相等,即 $h+r=t$,最终获得虚假新闻文本的 3 元组用矩阵 P 表示嵌入后的向量拼接而成的向量,矩阵 F 为使用图像描述[13]获取的图片的描述信息嵌入后的向量拼接而成的向量. 图像描述是自动从一张图片生成描述性语句,不仅能指出图片中包含的物体,而且能够表达图片中物体的相互关系、它们的属性以及它们共同参与的活动. 这有点类似于“看图说话”,但是对于机器来说却是一项很有挑战性的任务. 因为机器不仅要能检查出图像中的物体,而且要理解物体之间的相互关系,最后还要用合理的语言表达出来. 在图像描述任务中,输入的是图像,输出的是单词序列. 基于编码-解码利用图像中使用的 CNN 作为编码,提取图像的视觉特征,使用性能更好的 RNN 作为解码.

令 $Z^0=S+P+F$,其尺寸为 $n_3\times n_2$,尺寸值为 128×512,其中 Z^0 是第 1 个编码器的输入矩阵,本文中自注意力子层和全连接子层均指第 1 个编码器.

2.3 Bert 自注意力

在 Bert 的多头自注意力层中,第 i 个头的权重矩阵记为 W^{1i1},尺寸为 $n_2\times n_6$,尺寸值为 512×64,第 i 个头的偏置向量记为 b^{1i1},尺寸为 $1\times n_6$. 则 Q 矩阵为 $Q^{1i1}=Z^0W^{1i1}+b^{1i1}$,其尺寸为 $n_3\times n_6$,尺寸值为 128×64.

第 i 头的归一化分值为 $U^{i1}=\mathrm{softmax}(Q^{1i1}(K^{1i1})^{\mathrm{T}}/\sqrt{d})V^{1i1}$,其尺寸为 $n_3\times n_6$,尺寸值为 128×64.

将所有头的归一化分值连接起来,可以得到第 1 个编码器自注意力的分值为 $U^1=(U^{12},U^{12},\cdots,U^{1n_6})$,其自注意力子层的输出为 $Y^{111}=\mathrm{lnor}(U^1W^{1\cdot4}+b^{1\cdot4}+Z^0)$,其中 $W^{1\cdot4}$ 为权重矩阵,尺寸值为 512×512. $b^{1\cdot4}$ 为第 1 个编码器偏置向量,尺寸值为 1×512. Y^{111} 尺寸值为 128×512.

2.4 模型训练

对于 Bert 模型,本文采用编码器的堆叠,第 1 个编码器输入为 Z^0,输出为 Z^1,每一个编码器内部的计算过程都一致,第 2 个编码器输入为 Z^1,输出为 Z^2,以此类推,第 n 个编码器的输出为 Z^n. 对于 sequence-level 的分类任务,Bert 直接取第 1 个[CLS] token 的最终隐藏层输出 Z^n,加 1 层权重向量 $\hat{W}$ 后 softmax 预测标签的概率为 $P=\mathrm{softmax}(Z^n\hat{W})$,模型采用交叉熵作为损失函数.

3 实验

3.1 数据集

本文所采用的虚假新闻检测任务数据集来自北京市经济和信息化局、CCF 大数据专家委员会、中科院计算技术研究所共同发布,包含了文本和图片 2 种模态的信息,总共包含 61 432 幅图像,同时数据集中包含有部分新闻的评论,字段为空代表该新闻没有评论,本次实验仅使用了新闻的源文本信息进行,具体的统计数据如表 1 所示.

表 1 语料库统计

项目名称	数量
标签为 fake 的文本	14 930
标签为 true 的文本	12 719
新闻文本长度最大值	1 994
新闻文本长度最小值	1
新闻文本长度平均值	136
有图片的新闻文本数量	24 590
无图片的新闻文本数量	20 285
文本对应的图片数量最大值	9
文本对应的图片数量最小值	1

数据来源: https://www.datafountain.cn/competitions/422.

3.2 实验设置

本文设置实验参数如下,学习率为 5×10^{-5},

Bert 层数为 768,头数为 12.

3.3　实验对比

本文采用的基准模型有: (i) TextCNN, Yoon Kim [15] 在 2014 年提出的 TextCNN,将卷积神经网络 CNN 应用到虚假新闻分类任务; (ii) TextRCNN, Lai Siwei 等 [16] 于 2015 年提出的 TextRCNN 模型,该模型广泛应用于文本分类任务; (iii) FastText,2016 年由 Facebook 提出的一种比较迅速的词向量和文本分类方法 [17] ,将整篇文档的词及 n-gram 向量叠加平均得到文档向量,然后使用文档向量做 softmax 多分类; (iv) LSTM,最基础的 LSTM 模型,用于虚假新闻分类任务.

实验设置常用的评价指标: 正确率(Accuracy)、准确率(Precision)、召回率(Recall)以及 F_1 值. 本文算法与其他基线算法的实验结果如表 2 所示,所有实验结果均使用 5 倍交叉验证.

表 2　本文算法与基线算法对比结果

方法	正确率	准确率	召回率	F_1
TextCNN	0.312	0.342	0.302	0.321
TextRNN	0.328	0.354	0.313	0.332
FastText	0.359	0.361	0.338	0.349
LSTM	0.341	0.359	0.329	0.343
Bert	0.423	0.421	0.405	0.413
合成第 1 幅图像对应的描述文本				
TextCNN/caption	0.311	0.336	0.304	0.319
TextRNN/caption	0.329	0.343	0.319	0.331
FastText/caption	0.363	0.371	0.343	0.356
LSTM/caption	0.357	0.339	0.331	0.335
Bert/caption	0.429	0.411	0.426	0.418
合成所有图像对应的描述文本				
TextCNN/caption	0.304	0.321	0.297	0.309
TextRNN/caption	0.317	0.331	0.313	0.322
FastText/caption	0.359	0.356	0.331	0.343
LSTM/caption	0.347	0.329	0.325	0.327
Bert/caption	0.409	0.413	0.396	0.404
合成知识图谱 3 元组对应的文本				
TextCNN/KG	0.331	0.342	0.317	0.329
TextRNN/KG	0.338	0.374	0.328	0.349
FastText/KG	0.353	0.361	0.323	0.341
LSTM/KG	0.352	0.369	0.342	0.355
Bert/KG	0.436	0.441	0.412	0.426
合成知识图谱和第 1 幅图像描述文本				
Bert/KG/caption	0.431	0.429	0.434	0.432

从表 2 实验结果中可以看出: (i) 本文基于图像描述提取向量拼接的模型与基准的 Bert 模型相比,准确率提高了 0.6%,精确率降低了 1.0%,而召回率上升了 1.5%,F_1 值提高了 0.6%; 基于 TransE 提取向量拼接的模型与基准的 Bert 模型相比,准确率提高了 1.0%,精确率提高了 2.0%,召回率上升了 0.7%,F_1 值提高了 1.4%. (ii) 将 2 个向量进行拼接的模型,对比单一的模型在精度上有了更多的提升,F_1 值可以综合衡量模型的性能,从表 2 可以看出,本文算法在 F_1 值上均高于基线算法. 另外,在性能上,合成原始文本对应的第 41 幅图像对应的图像描述优于所有图像,原因在于过多的图像会导致噪音. 这些实验结果证明本文模型有效地挖掘了虚假新闻中的原始文本语义表示,以及利用知识图谱和图像描述对应的深层次语义表示,从而能够有效地检测出虚假新闻.

4　总结

本文提出了融合知识图谱和图像描述获取的文本语义向量的虚假新闻检测深度学习模型. 在模型的向量生成上,一方面,设计了一个通过知识关系抽取的方式获取额外的文本向量,另一方面,采用图像描述的方式获取对应图片的图片描述的文本向量; 之后再将原文与这 2 个向量拼接,作为最后的 Bert 分类模型中进行判别. 在国际基准汉语虚假新闻语料库上的实验结果表明了模型的有效性.

作为将来工作,拟从以下 2 个方面来改进该模型: (i) 可以考虑引入外部知识图谱对知识抽取的结果进行补充,或者考虑直接使用外部知识图谱对分类器进行优化; (ii) 更深入挖掘图像描述获取的图片描述过程,比如,可以考虑先对图片以及文本进行聚类,之后再对于获取到的图片描述进行加权选择,在聚类中距离相近的图片赋予高权重,距离远的图片赋予低权重等.

5　参考文献

[1] Hunt A, Matthew G. Social media and fake news in the 2016 election [J]. Journal of Economic Perspectives, 2017, 31(2): 211-236.

[2] Shao Chengcheng, Giovanni L C, Onur V, et al. The spread of fake news by social bots [EB/OL]. [2020-05-13]. https://arxiv.org/pdf/1707.07592.pdf.

[3] Liu Xiaomo, Armineh N, Li Quanzhi, et al. Real-time rumor debunking on twitter [EB/OL]. [2020-06-15]. https://dl.acm.org/doi/10.1145/2806416.2806651.

[4] Li Jiwei, Myle O, Claire C, et al. Towards a general rule for identifying deceptive opinion spam [EB/OL]. [2020-06-

18]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.661.946&rep=rep1&type=pdf.

[5] Carlos C, Marcelo M, Barbara P. 2011. Information credibility on twitter [EB/OL]. [2020-06-15]. https://doi.org/10.1145/1963405.1963500.

[6] Sejeong K, Meeyoung C, Kyomin J, et al. Prominent features of rumor propagation in online social media [EB/OL]. [2020-06-19]. https://ieeexplore.ieee.org/document/6729605.

[7] Wu Ke, Yang Song, Kenny Q. False rumors detection on sinaweibo by propagation structures [EB/OL]. [2020-07-15]. https://ieeexplore.ieee.org/document/7113322.

[8] Ma Jing, Gao Wei, Wong K F. Detect rumors in microblog posts using propagation structure via kernel learning [EB/OL]. [2020-07-15]. https://aclanthology.org/P17-1066.pdf.

[9] Yang Shuo, Shu Kai, Wang Suhang, et al. Unsupervised fake news detection on social media: a generative approach [EB/OL]. [2020-07-15]. http://www.public.asu.edu/~skai2/files/aaai_2019_unsupervised.pdf.

[10] Jin Zhiwei, Cao Juan, Guo Han, et al. Multimodal fusion with recurrent neural networks for rumor detection on microblogs [EB/OL]. [2020-07-15]. https://dl.acm.org/doi/10.1145/3123266.3123454.

[11] Wang Yaqing, Ma Fenglong, Jin Zhiwei, et al. Eann: event adversarial neural networks for multi-modal fake news detection [EB/OL]. [2020-07-15]. https://www.pianshen.com/article/57871580780/.

[12] Zlatkova D, Nakov P, Koychev I. Fact-checking meets fauxtography: verifying claims about images [EB/OL]. [2020-07-15]. https://arxiv.org/abs/1908.11722.

[13] Antoine B, Nicolas U, Alberto G D, et al. Translating embeddings for modeling multi relational data [EB/OL]. [2020-07-15]. https://dl.acm.org/doi/10.5555/2999792.2999923.

[14] Vinyals O, Toshev A, Bengio S, et al. Show and tell: a neural image caption generator [EB/OL]. [2020-07-15]. https://ieeexplore.ieee.org/document/7298935.

[15] Yoon Kim. Convolutional neural networks for sentence classification [EB/OL]. [2020-07-15]. https://arxiv.org/abs/1408.5882.

[16] Lai Siwei, Xu Liheng, Liu Kang, et al. Recurrent convolutional neural networks for text classification [EB/OL]. [2020-07-15]. https://ieeexplore.ieee.org/document/8852406.

[17] Armand J, Edouard G, Piotr B, et al. Bag of tricks for efficient text classification [EB/OL]. [2020-07-15]. https://arxiv.org/abs/1607.01759.

[18] Qi Peng, Cao Juan, Yang Tianyun, et al. Exploiting multi-domain visual information for fake news detection [EB/OL]. [2020-07-15]. https://arxiv.org/abs/1908.04472v1.

[19] Li Yuezun, LüSiwei. Exposing deepfake videos by detecting face warping artifacts [EB/OL]. [2020-07-15]. https://arxiv.org/abs/1811.00656v3.

[20] Cao Juan, Sheng Qiang, Qi Peng, et al. False news detection on social media [EB/OL]. [2020-07-15]. https://arxiv.org/abs/1908.10818.

[21] 刘知远,张乐,涂存超,等. 中文社交媒体谣言统计语义分析[J]. 中国科学:信息科学,2015,45(12):1536-1546.

[22] 王志宏,过弋. 微博谣言事件自动检测研究[J]. 中文信息学报,2019,33(6):132-140.

The Fake News Detection Based on Knowledge Graph and Image Description

CHEN Kaiyang, XU Fan*, WANG Mingwen

(School of Computer Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: Traditional fake news detection methods mainly use image statistics and image distributed representation features without analyzing the deep semantic meaning of the text expressed behind the image. Based on this observation, the combined model which can integrate knowledge graph and image caption to detect multi-modal fake news is designed. One the one hand, the model can extract triple-style knowledge graph from the texts. On the other hand, the model can generate text description for the images. Meanwhile, the model can successfully integrate the semantic representation of source texts, triples, and image caption. Experimental results on the benchmark Chinese fake news corpus show that the model is significantly better than the representative methods.

Key words: fake news; knowledge graph; image caption; Bert

(责任编辑:冉小晓)