

基于多模态学习的虚假新闻检测研究

刘华玲¹⁺, 陈尚辉¹, 曹世杰¹, 朱建亮¹, 任青青²

1. 上海对外经贸大学 统计与信息学院 商务大数据实验中心, 上海 201620

2. 哥伦比亚大学 傅氏基金工程与应用科学学院, 纽约 10027

+ 通信作者 E-mail: liuhl@suibe.edu.cn

摘要: 社交媒体在给人们带来便利的同时, 也成为虚假新闻恣意传播的渠道, 如果不及时发现遏止, 极易引发群众恐慌, 激起社会动荡。因此, 探索准确高效的虚假新闻检测技术具有极高的理论价值和现实意义。对虚假新闻相关检测技术做了全面综述。首先, 对多模态虚假新闻的相关概念进行了整理和归纳, 并分析了单模态和多模态新闻数据集的变化趋势。其次, 介绍了基于机器学习和深度学习的单模态虚假新闻检测技术, 这些技术在虚假新闻检测领域已被广泛应用, 而由于虚假新闻通常包含多种数据表现形式, 这些传统的单模态技术无法充分挖掘虚假新闻的深层逻辑, 因此无法有效地应对多模态虚假新闻数据带来的挑战。针对此问题, 对近些年来先进的多模态虚假新闻检测技术进行了整理, 从多流架构和图架构的角度归纳和论述了这些多模态检测的技术方法, 探讨了这些技术的思想理念与潜在缺陷。最后, 分析了目前虚假新闻检测研究领域存在的困难和瓶颈, 并由此给出未来的研究方向。

关键词: 虚假新闻检测; 多模态学习; 深度学习; 社交网络

文献标志码: A **中图分类号:** TP391

Survey of Fake News Detection with Multi-model Learning

LIU Hualing¹⁺, CHEN Shanghui¹, CAO Shijie¹, ZHU Jianliang¹, REN Qingqing²

1. Experimental Center for Business Big Data, School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai 201620, China

2. Fu Foundation School of Engineering and Applied Science, Columbia University in the City of New York, New York 10027, USA

Abstract: While social media brings convenience to people, it has also become a channel for the arbitrary spread of fake news. If not detected and stopped in time, it is easy to cause public panic and social unrest. Therefore, exploring accurate and efficient fake news detection technology has high theoretical value and practical significance. This paper provides a comprehensive overview of the related fake news detection techniques. Firstly, the relevant concepts of multi-modal fake news are sorted and summarized, and the trend of changes in single-modal and multi-modal news datasets is analyzed. Secondly, this paper introduces single-modal fake news detection techniques based on machine learning and deep learning, which have been widely used in the field of fake news detection. However, traditional single-modal techniques cannot fully explore the deep logic of fake news because fake news usually contains multiple data presentations. Thus, they are unable to effectively deal with the challenges brought by multi-modal

基金项目: 国家社科基金重大项目(21ZDA105); 上海哲学社会科学规划课题(2018BJB023)。

This work was supported by the Major Project of National Social Science Foundation of China (21ZDA105), and the Philosophy and Social Science Planning Project of Shanghai (2018BJB023).

收稿日期: 2023-01-31 **修回日期:** 2023-05-10

fake news data. To address this issue, this paper summarizes and discusses advanced multi-modal fake news detection techniques from the perspectives of multi-stream and graph architectures, and explores their concepts and potential drawbacks. Finally, this paper analyzes the difficulties and bottlenecks in the current research on fake news detection and provides future research directions based on these analyses.

Key words: fake news detection; multi-model learning; deep learning; social network

社交平台的信息传播具有低成本、高效率、实时便捷等特点,这些便利为新闻在社区广泛传播提供了可能,然而,信息发布和扩散的同时也导致了社交网络上虚假新闻的恣意横行。据2019年CHEQ和巴尔的摩大学的经济研究报道^[1],全球每年因虚假新闻造成的损失高达780亿美元。2020年7月,江苏南京一小区发生外卖被盜事件,据警方了解,该偷盜居民涉嫌多次盜窃,目前已被刑拘。事发后三天内,众多网络媒体发布新闻,称当事人为考研大学生,报道中还出现了“为供其深造,家中其他3个兄弟姐妹辍学”等说法。7月20日下午,警方发布通报:嫌疑人李某某大学毕业已两年,目前有固定收入,其偷外卖的原因,是一次外卖被人拿走后,产生了报复心理。目前,嫌疑人李某某已被取保候审。李某某父母和大姐在老家务农,二姐、三姐分别在北京、海南工作。换言之,“考研大学生”这一身份是虚假信息,李某某的家庭并不贫困,偷外卖也并非为了维持生活,如图1(a)、图1(b)所示。不良媒体通过散播这些假新闻激起群众的同情,以此获取流量、关注,直到官方辟谣,这些虚假新闻才得以遏止。由此可见,虚假新闻已经成为大量不良媒体获取非法利益的工具,它们的存在会加强人们之间的不信任关系,造成不良的社会影响。因此,探索准确高效的虚假新闻检测方法尤为重要。对于虚假新闻,新闻文字源于图片的恶意编造,其描述的内容必然与图像真实内容存在冲突,即模态之间存在语义不一致性,如果单从图片或者文字角度分析,这种语义不一致性很难被模型识别,容易导致模型分类错误,因此,从多模态的角

度探索虚假新闻检测技术很有必要。

纵观这些年关于虚假新闻检测的综述文章,很少有从多模态角度来分析的。早期研究者们致力于寻找和构建人工特征来表示新闻内容,这时的综述内容大多是关于这些特征的归纳整理^[2-3],后来,随着深度学习技术的发展,学者们将研究重心放在了这种自动化特征提取技术上,其中涌现了大批基于深度学习的虚假新闻检测文章,近些年来,一部分学者对这些方法进行了总结^[4-5]。然而,这些文章的研究角度存在局限,并没有考虑到虚假新闻中的其他模态。有研究发现^[6-7],新闻的视觉内容是能误导读者的关键因素。此外,新闻社交图中蕴含的虚假新闻传播信息是检测取得成功的重要因素^[8],因此从多模态的视角分析新闻很有必要。针对此,本文详尽地梳理了以往虚假新闻检测领域的一些工作,从单模态到多模态的角度对该领域做全面的整理和综述。本文的贡献如下:

(1)详尽地从单模态到多模态角度对虚假新闻检测领域相关技术做了归纳和整理;

(2)将基于新闻社交图的检测技术作为一种特殊的多模态处理方法,并对其最新技术的研究现状做了补充和完善;

(3)梳理了现有虚假新闻检测技术存在的研究瓶颈,并给出了未来研究方向。

1 相关概念

1.1 多模态虚假新闻检测问题的定义

在虚假新闻检测任务中,一般包含以下几个步

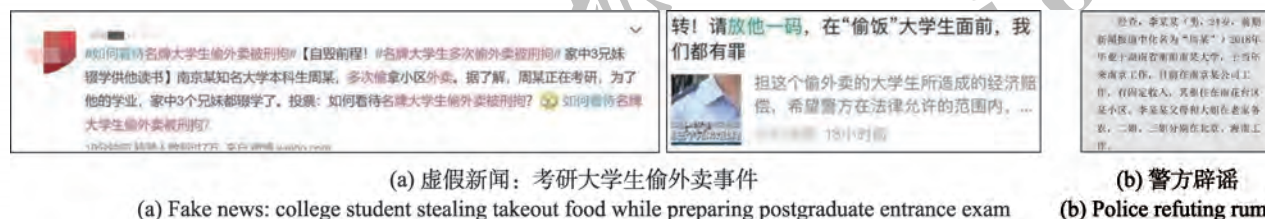


图1 虚假新闻与辟谣案例

Fig.1 Case of fake news and refuting rumors

骤:分析新闻数据的表现形式;构建数学物理模型挖掘新闻的潜在逻辑结构;训练模型并更新模型参数;在新数据集上预测虚假新闻。虚假新闻检测问题可以看作一种分类任务,其定义如下:

假设新闻数据集 $\{x_1, x_2, \dots, x_n\} \in D$ 及类标签 $\{y_1, y_2, \dots, y_n\} \in L$, 对于每条新闻数据 x_i , 学习一个映射函数 $f(x_i)$, 将数据 x_i 映射到类标签 y_i 上, 即: $f(x_i) \rightarrow y_i$, 其中, 模型的输入表示一条事件新闻, 模型的输出通常表示该新闻属于其类标签的概率值。

对于多模态新闻数据, $x_i = [x_i^1, x_i^2, \dots, x_i^m]$, 其中, x_i^j 表示数据 x_i 的第 j 个模态向量, m 表示数据 x_i 中包含的模态个数, 通常情况下 m 为大于等于 2 的值。在本文中, 新闻的模态不仅指新闻的内容(如图片、文章、视频等), 在新闻关系网络图中, 除新闻内容之外, 新闻与新闻、新闻与用户等发生的交互行为也可以看成特殊的模态, 记为行为模态。

1.2 虚假新闻数据集

近些年来, 研究者和媒体平台对新闻数据集进行了收集整理, 这些数据集在标签信息、模态内容、事件类别等多方面存在较大差异。本节主要对近些年研究中常用的单模态和多模态虚假新闻开源数据集进行介绍, 数据集汇总如表 1 所示。

单模态虚假新闻数据集只包含新闻文本内容,

如 LIAR、TSHP-17、FakeCovid 等, 从整体来看, 单模态新闻数据集较大, 划分类别更多, 其主要原因是原始数据存在信息不全, 通常需要更细粒度地划分数据集和不断扩充数据集大小才能带来模型检测性能的提升。相比单模态数据集, 多模态数据集采集工作量较大, 数据量通常较小, 如 Weibo A、MediaEval_Dataset 等数据集大小在十万左右, 数据集中至少包含新闻文本和图片原始数据, 此外一些多模态数据集为新闻传播树, 如 Twitter15、Twitter16 等, 新闻传播数不仅可以记录新闻内容信息, 同时也保留了新闻关系网络中行为信息, 每棵树表示一个新闻事件, 树中包含新闻的相关用户和新闻内容节点, 以及它们之间的交互行为。

以上是近些年研究中常用的经典数据集, 本文收集的大部分工作是在这些数据集上进行的。

1.3 评价指标介绍

虚假新闻检测实质上是一个分类问题, 因此常采用准确率(Accuracy, Ac)、精确率(Precision, P)、召回率(Recall, R)、F1 分数($F1$ -score, $F1$)对结果进行评价。

(1) Ac

Ac 用于评估检测模型的整体性能, 表示分类正确的样本数占总体样本数的比值, 分类正确的样本

表 1 虚假新闻检测数据集汇总

Table 1 Summary of fake news detection datasets

数据集名称	数据集格式	数据集大小	数据集标签	数据集形式
Weibo A ^[9]	文本&图片	7.7×10^3	谣言、非谣言	社交媒体数据
Weibo B ^[10]	文本&图片	3.84×10^4	谣言、非谣言	社交媒体数据
MediaEval_Dataset ^[11]	文本&图片	1.56×10^4	谣言、非谣言	社交媒体数据
Fake_video ^[12]	视频	380	真实、错误	社交媒体数据
Twitter15 ^[13]	传播树	1 381 棵树、 2.7×10^5 用户	未验证的、真实、错误、非谣言	社交媒体数据
Twitter16 ^[13]	传播树	1 181 棵树、 1.7×10^5 用户	未验证的、真实、错误、非谣言	社交媒体数据
HealthStory ^[14]	传播树	1 690 棵树	真实、错误	医疗新闻
MM-COVID ^[15]	传播树	2 263 棵树	真实、错误	新冠疫情新闻
PHEME ^[16]	文本&图片	5.8×10^3	谣言、非谣言	社交媒体数据
FakeNewsNet ^[17]	文本&图片	2.21×10^4	真实、错误	社交媒体数据
LIAR ^[18]	文本	1.28×10^3	谎言、错误、勉强是真、半真半假、大部分是真、完全是真	政治宣言
TSHP-17 ^[19]	文本	1.05×10^3	谎言、错误、勉强是真、半真半假、大部分是真、完全是真	政治宣言
Ahmed2017 ^[20]	文本	2.52×10^4	真实、错误	新闻文章
FAKES ^[21]	文本	804	真实、错误	战争新闻
FakeCovid ^[22]	文本	1.28×10^4	无标签、真实、半真半假等 18 种	新冠疫情新闻
NELA-GT-2018 ^[23]	文本	7.1×10^3	可靠、不可靠、无标签、半真半假	社交媒体数据
NELA-GT-2021 ^[24]	文本	1.8×10^6	可靠、不可靠、无标签、半真半假	社交媒体数据

数越多,模型整体的检测性能越好。

(2) P

P 表示分类器预测的正类样本占实际正类样本的比重。正类样本通常为真实类别样本,在本文中指代虚假新闻, P 反映了模型正确预测虚假新闻的能力。

(3) R

R 表示被分类器正确预测的正类样本所占比例,模型的 R 越高,正类样本越不容易被误分。

(4) $F1$

最大化 P 和 R 是构建分类模型的重要标准之一,其通常用 $F1$ 来衡量, $F1$ 表示 P 和 R 的调和平均值,较高的 $F1$ 可以确保分类模型的 P 和 R 也都较高。

2 单模态虚假新闻检测技术

2.1 人工检测技术

如今,各大社交网络平台仍然依赖人工识别的方法鉴别真假新闻,在互联网中传播的大部分虚假新闻还是依赖用户的及时发现和检举才能得到遏止,一些主流媒体平台如 Twitter、Facebook、新浪微博等会通过雇佣各个领域专家或众包的方式鉴别平台上发布的虚假新闻^[25]。人工虚假新闻检测的优势在于其极高的准确率、灵活性,然而它的缺点也特别明显。首先,互联网上每分钟都会接收成千上万条举报新闻,人工检测会花费昂贵的人力成本,而且人工造成的时间滞后性会导致假新闻造成的损失进一步增加;其次,虚假新闻涉及生活的方方面面,人的知识很难考虑周全,偶尔会出现误判的情况;最后,虚假新闻的伪造手段也在不断进步,人们越来越难以识别新闻中的虚假信息。

随着人工智能技术的进步,许多自动化虚假新闻检测技术被提出,其中这些较早被提出的技术可以统称为传统的虚假新闻检测技术,其可以分为两种主流的方法:基于机器学习的方法和基于神经网络的方法。

2.2 单模态机器学习检测技术

单模态机器学习检测技术的核心在于如何构建和筛选好的特征表示新闻文本信息,在新闻文本数据中,最常用的特征主要包含两种:新闻语义特征和新闻传播特征,细分如表2所示。

真实新闻和虚假新闻的语义风格有明显的差异,虚假新闻通常携带较少的转述动词,语句中有大量的独特动词、脏话和情绪性词语,而且单词较短而

表2 单模态虚假新闻检测技术常用特征

Table 2 Common characteristics of single-modal fake news detection techniques

类别	特征
新闻语义特征	词汇特征
	句法特征
	语义风格特征
新闻传播特征	用户画像
	新闻回复评论
	传播特征

句子较长。因此可以从词汇、句法以及语篇三个语言层面来提取新闻的语义特征。词汇级别的风格表示主要是评估词汇的频率统计^[26]。词频通常分为绝对频率和相对频率,通过计算新闻词的绝对频率可以衡量相同主题下文章关键词特征的相似性,相对频率则可以用于分析新闻作者的偏爱词汇与新闻文章的风格特征。新闻的句法分析方法有浅层句法分析和深层句法分析两种。常用的浅层句法分析方法有基于词性标签的编码方法^[27],在其基础上,深层句法分析方法则会利用新闻文章的补充信息(如文章上下文、词频等)深入挖掘浅层句法特征^[28]。语篇层面上,语篇的风格特征研究新闻内容中的心理语言属性,比如情绪。其主要方法是通过构建情感词典提取新闻文章的情感基调^[29],此外,语篇中通常会使用修辞结构理论作为衡量故事连贯性的标准和假新闻的指标^[30-31]。此外,为了丰富新闻语义特征,学者们会将多种语义特征组合在一起分析,Pérez-Rosas 等人^[29]利用文本的不同语言特征,如语法特征、标点符号、心理学特征等进行组合,采用线性支持向量机(support vector machine, SVM)分类器进行新闻分类,最终获得了74%的分类准确率。然而人工提取的特征泛化性能很难得到保证,Sharma 等人^[32]的研究表明,在跨主题、领域的假新闻检测任务中,难以归纳出人工语言特征。因此后来学者将新闻传播特征与新闻语义特征结合。

据研究,真实新闻与虚假新闻的传播结构有很大不同,相比真实新闻,虚假新闻传播速度更快,范围更广。新闻的传播特征包含多方面,主要涉及用户画像信息、新闻评论、时间因素等。在社交媒体平台中,一些恶意用户会通过社交机器人传播虚假新闻^[33],社交机器人用户一般具有鲜明的特征,与正常用户相比,它们的用户信息通常简洁且不完善,因此,通过建立用户画像特征可以帮助检测出社交机器人发布的虚假新闻。常用的用户画像特征包含用

户的注册年龄、关注者数量、用户的发文数量、用户的可信度和可靠性等^[34-35]。新闻评论体现了新闻受众的情绪(消极、积极等)或是立场(支持、反对、怀疑等),这些评论者的态度和立场对评判新闻信息的真实性具有重要价值。研究者发现^[36-37],通过研究评论内容可信度信息和评论用户的立场观点,可以有效提高虚假新闻检测模型的性能。此外,实验证明,还有一些传播特征如新闻的转发量、新闻的传播时间等^[38]也十分重要。这些传播特征通常为检测的辅助信息,因此,大多为冗余信息,通常需要花费大量时间对数据进行筛选加工。

然而,人工特征的构建需要在数据处理阶段耗费大量的时间,构建的特征通常鲁棒性较差,只在特定的数据集上有好的表现。后来,随着深度学习的兴起,深度学习的自动特征提取能力和模型的高检测准确率完美避开了手工特征提取的缺陷,因此研究者们开始将研究重心转移到深度学习领域。

2.3 单模态深度学习检测技术

相较于机器学习,深度学习方法在单模态虚假新闻检测领域有更好的表现^[39-42]。深度学习将新闻单词特征映射为像素矩阵或张量,并随模型迭代自动更新单词特征,这种基于神经网络的语言模型要明显优于传统的机器学习模型,如N-gram^[43]。

多种研究方法被设计来提高假新闻检测的质量。用户的情感倾向是检测假新闻的有力指标之一^[44]。Ma等人^[45]的研究表明,基于生成对抗网络(generative adversarial network, GAN)的检测方法,通过文本生成器与假新闻鉴别器之间的对抗学习以改进文本的特征表达。Szegedy等人^[46]提出卷积神经网络(convolutional neural network, CNN)来进行目标检测,证实了转向稀疏结构通常能提高检测的质量。何韩森等人^[47]提出了一种基于特征聚合的假新闻内容检测方法CCNN(center cluster neural network),在提升模型检测性能的同时确保其泛化能力。

随着Transformer的提出,各个自然语言任务开始被这种大规模预训练模型取代。孙尉超等人^[48]结合ALBERT(a lite bidirectional encoder representations from transformers)和LSTM(long short-term memory)模型以挖掘文本的深层次特征。Kaliyar等人^[49]提出了FakeBERT虚假新闻检测框架,该框架将不同大小核的卷积块并行,再与BERT模型相结合来提取数据特征,最终模型的准确率高达98.9%。

此外,由于现有数据集数量太少,难以训练出泛

化性较好的模型,有学者探究使用知识图谱学习利用外部知识提取更一般性的新闻特征。Mayank等人^[50]通过RoBERTa(robustly optimized BERT pretraining approach)模型对新闻标题数据进行命名实体识别,并将实体与开源的Wikidata知识图连接,在复数空间中提取实体和关系特征,该方法只利用了新闻标题数据,最终在两个数据集上分别获得了88%和78%的F1值。

虽然基于深度学习的单模态检测模型可以非常准确地识别虚假新闻,但优异的模型性能往往需要大量的训练数据集做支撑,这使得单模态深度学习检测技术往往需要花费大量人力、物力在数据收集工作上,同时,庞大的数据量也会导致模型训练成本太大,模型的每次更新迭代都需要耗费大量时间,因此需要探索新的研究角度来避免这些问题。在多模态学习中,构建多模态模型可以充分学习原始数据中存在的模态关联信息,减少因数据量产生的成本问题。随着多模态学习越来越火热,虚假新闻检测领域的研究者们开始尝试使用多模态的关联信息来提升虚假新闻检测的性能。

3 多模态虚假新闻检测技术

不同形式的信息源可以看成不同的模态^[51],新闻是典型的多模态数据,书面报道的新闻通常包含图片和文本两种模态信息,短视频新闻至少包含图像、音频和字幕等多模态信息,新闻社交图中包含新闻内容以及新闻行为等多种模态信息。

多模态虚假新闻检测技术的关键是如何构建模型框架学习新闻数据的多模态信息,以提升虚假新闻检测性能。总结至今提出的一些文章,大致可以划分为两类:基于流形式的多模态虚假新闻检测技术和基于图形式的多模态虚假新闻检测技术。

3.1 基于流形式的虚假新闻检测技术

3.1.1 基于单流架构的技术

单流架构指在模型输入之前,不同模态数据的初级特征会通过拼接、函数映射等方式进行数据融合,得到的多模态特征内部中各个模态的信息是独立的,而多模态信息需要在后续模型中学习。最具代表性的是基于Transformer架构的多模态模型,如ViLT(vision-and-language transformer)^[52]、MBT(multi-modal bottleneck transformer)^[53]等,各模态的数据会预处理为序列化数据,例如,文本会转化为多个token组成的序列,图片会转化为多个不重叠的图片patch

序列,音频数据会先转化为频谱图,最终组成多个不重叠的频谱图 patch 序列,多个模态的特征最终会进行拼接,构成模型的多模态输入特征,单流架构框架如图2所示。

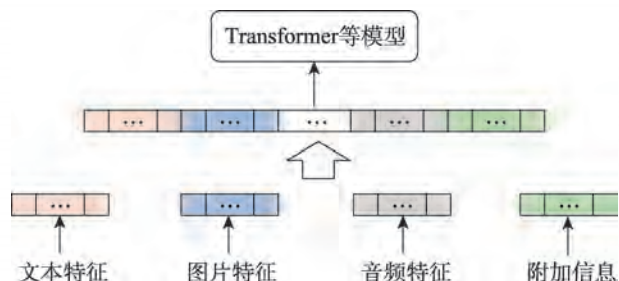


图2 单流架构图

Fig.2 Single-stream architecture diagram

目前,单流架构模型在视频分类、情感分析、图像生成等多模态领域中得以广泛应用,单流模型具有结构简单、容易实现、高准确率等优势,在虚假新闻检测领域中,是一个极具潜力的研究方向。但参考目前的一些研究,其也存在一些缺陷:

(1)在网络训练时需要花费更多的迭代次数才能获得好的多模态表示;

(2)由于模型的输入特征通常是多个模态特征拼接而成,模型有较高的计算复杂度;

(3)单流模型的学习需要大量的训练数据集,而在虚假新闻检测领域中,目前没有足够多可以训练的数据。

3.1.2 基于多流架构的技术

近些年来,关于多模态虚假新闻检测领域,研究者们更常用的是基于多流架构的技术。多流架构是指根据不同模态数据设计不同模型提取模态高级特征,从各个模态高级特征中学习多模态特征并输入下游的分类器中预测各个类别的概率。相比单流架构,多流架构更加灵活,其可以针对不同模态数据单独设计模型提取模态特征。多流框架如图3所示。

Jin 等^[54]最早提出了一个双流的多模态虚假新闻检测框架,其中一支使用 LSTM 模型提取新闻文本和社交文本特征,另一分支通过 VGG-19 深度卷积网络提取新闻图片特征,最终将两支的特征串联融合并通过全连接层进行二分类预测。通过对比实验发现,相比单模态技术,多模态的检测方法在各项评价指标上都表现优异。后来,一系列学者在此框架的基础上做了以下一些改进:

(1)选取新的单模态特征提取模型(如 BERT、

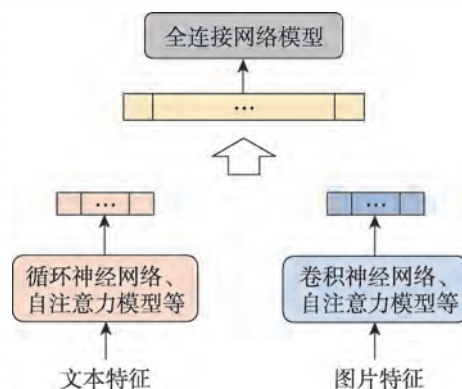


图3 多流架构图

Fig.3 Multi-stream architecture diagram

ResNet等)替代传统模型,提升原始数据的表征能力;

(2)学习多模态信息的互补性,对多模态特征融合模块进行改进,比如使用注意力机制、外积等方法融合多模态特征;

(3)学习多模态信息的一致性,研究者们引入了多模态对齐、跨模态相似性学习等思想;

(4)其他领域的思想,如对抗学习、引入外部关联知识等方法。

单模态特征的质量决定了多模态特征的表示能力,在初期,学者们关于多流框架中的工作主要是尝试一些新的模型替代传统的单模态特征提取模型。Singhal 等人^[55]在双流的多模态虚假新闻检测框架中使用 BERT 模型替代传统的循环神经网络模型,最终在 Twitter 和 Weibo 数据集上分别取得了 77.8% 和 89.2% 的分类准确率。后来,他们又提出用 XLNet 代替 BERT 模型^[56],模型性能得到进一步提升。陈志毅等人^[57]针对多流框架中的新闻社交数据,使用 DeepFM 算法提取新闻的社会特征,并与文本特征和图像特征拼接融合成多模态特征。引入新特征提取模型的方式是简单有效的,但这种提升局限于模态本身,对于多模态特征学习并没有太大帮助,而要充分利用多模态信息,需要学习模态之间的互补性和一致性。

多模态信息的互补性通常是通过多模态融合方法实现的,最简便的方法是采用串联的方式融合多模态特征,这种方法虽然容易实现,但其也存在以下两个问题:一方面,特征的简单拼接会导致数据维度增加,产生大量冗余信息;另一方面,多模态信息交互只能依赖后续的网络层,不可避免地会存在多模态信息交互不足的问题。为了解决多模态融合中存在的问题,学者们提出了交叉注意力融合的方法融

合多模态特征,交叉注意力方法可以在模型训练时利用辅助模态为目标模态各个部分信息赋予不同的权重,因此,可以有效地提取不同模态之间的交互特征,交叉注意力结构如图4所示。Song等人^[58]提出了跨模态残差网络,首先使用交叉注意力有选择性地从另一模态提取与目标模态相关的信息,其次通过残差的方法保留目标模态的信息,通过四个数据集上的对比实验证实该方法的优越性。其后,Qian等人^[59]在进行多模态特征融合时提出了多模态上下文注意网络,该网络中使用了两个Transformer块分别学习包含多模态信息的文本特征和图像特征,他们的方法最终在Twitter和PHEME数据集上准确率分别达到了89.7%和88.1%。不同特征的融合方法特点汇总在表3中。此外,还有一些基于交叉注意力的多模态特征融合的方法^[60-61],实验证明这些相似的模块确实提高了虚假新闻模型的检测性能。而且值得注意的是,交叉注意力获得的多模态特征可以一定程度上反映不同信息对检测结果的贡献程度,这也为模型提供了可解释性。

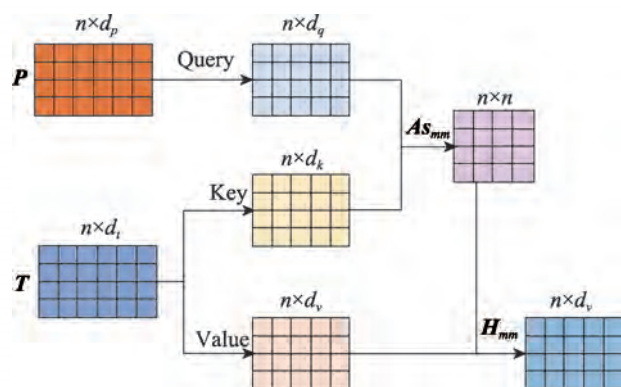


图4 交叉注意力融合

Fig.4 Cross-attention fusion

表3 不同特征融合方法的特点

Table 3 Characteristics of different feature fusion methods

特征融合方法	特点
拼接融合	优点:结构简单易实现 缺陷:无法有效提取多模态交互信息,不同模态的冗余信息叠加在一起
交叉注意力融合	优点:有效的特征交互方法,不改变输出特征维度 缺陷:输出特征中多模态信息不对称、计算复杂度较高
外积融合	优点:可以提取多模态特征相似性信息 缺陷:输出特征维度过大,存在维度爆炸风险

多模态一致性的方法是通过研究不同模态信息差异来发掘假新闻中存在的信息表达方式。Zhou等人^[62]较早提出在新闻内容中,不同模态特征之间的信息差异越大,新闻是杜撰的可能性越大,因此一部分学者开始聚焦于新闻模态语义一致性的研究。Giachanou等人^[63]通过计算新闻图像和新闻标题之间的余弦相似度来提取文本和图片的相似特征。Müller-Budack等人^[64]从新闻文本和新闻图像中提取实体,并通过命名实体链接利用外部知识计算实体之间的跨模态相似性。Xue等人^[65]认为图像伪造是假新闻产生的重要原因,他们以不同的等级压缩原始图片提取图像深层特征,同时结合文本特征、文本图像的余弦相似度特征发现,假新闻中确实存在严重的图像伪造问题。此外,还有学者借鉴多模态对齐的思想来研究语义一致性。Chen等人^[66]提出了一种跨模态歧义学习的技术,他们首先借鉴了对比学习的思想学习多模态对齐信息,然后结合VAE(variational auto-encoder)和KL(Kullback-Leibler)散度学习跨模态歧义信息,当多模态歧义值较大的时候单模态特征重要性更大,反之,多模态特征重要性更大。最终,该方法分别在Twitter和Weibo数据集上获得了80.6%和84%的分类准确率。研究多模态一致性问题可以为虚假新闻检测提供有关相似性特征、跨模态歧义值等信息。然而,这种方法一方面对数据质量的要求较高,需要保证数据的完整性,另一方面模态之间的相似性计算需要解决不同模态之间的语义鸿沟。

由于多流架构的灵活性,许多其他深度学习领域的思想也被广泛用于该架构中。在这过程中,研究者们试图寻找能表示普遍新闻的一般特征。Wang等人^[67]在多流架构中结合对抗学习的思想,增加了一个新闻事件分类的子任务,诱导模型学习与事件无关的多模态特征,以提高模型的泛化能力。Khatter等人^[68]提出了变分自编码器模型来探索虚假新闻特征的潜在分布。Qian等人^[69]在多流架构中加入了知识图谱模块,利用外部知识提高特征的泛化性能。类似的,Zhang等人^[70]在文本图像特征提取的基础上构建知识编码器从知识图谱中挖掘新闻的背景知识,为了保证模型在未知数据中的预测能力,他们构建事件记忆单元抽取新闻事件之间的共享特征。

基于多流架构的方法汇总在表4中。总而言之,基于多流的多模态虚假新闻检测技术如今已趋于成熟,但是受到新闻数据集的影响,不同学者对于新闻的研究角度会有不同,模型框架的设计通常有较大

表4 基于多流架构的方法汇总

Table 4 Summary of methods based on multi-stream architecture

分类	方法	时间	模型	表现效果(准确率)	框架结构
单模 态特 征提 取	Jin 等人 ^[54]	2017	LSTM、VGG19	Weibo A:0.788 MediaEval_Dataset:0.682	结合 Attention 的 LSTM、预训练的 VGG19 学习文本评论和图像特征,并串联拼接获得多模态特征
	Singhal 等人 ^[55]	2019	BERT、VGG19	Weibo A:0.892 MediaEval_Dataset:0.777	从零训练的 BERT 模型和预训练的 VGG19 学习新闻文本和新闻图像特征,并串联拼接获得多模态特征
	陈志毅等人 ^[57]	2022	DeepFM、text-CNN、VGG19	Weibo A:0.803	深度学习与 FM 算法结合的 DeepFM 学习新闻社交特征, text-CNN、VGG19 学习新闻的文本和图像特征,并将三部分特征串联获得多模态特征
多模 态信 息互 补性	Song 等人 ^[58]	2021	多头注意力、VGG19	Weibo A:0.853 Weibo B:0.869 MediaEval_Dataset:0.741	多头注意力和预训练的 VGG19 学习新闻文本和图像特征,基于双向的交叉注意力机制有选择的从一个模态学习另一模态信息,并结合残差保留原始特征信息
	Qian 等人 ^[59]	2021	BERT、ResNet	Weibo A:0.885 MediaEval_Dataset:0.897 PHEME:0.881	从零训练的 BERT 模型和预训练的 ResNet 模型学习新闻文本和图像特征,构建 Transformer 架构,并将其中的输入输出替换为图片和文本特征,经过多次的注意力和全连接层的计算输出多模态特征
	亓鹏等人 ^[60]	2021	ERNIE、卷积网络	Weibo A:0.895	文本数据包含新闻文本和图片 OCR 后得到的文本,使用预训练的 ERNIE 和卷积模型提取文本和图像特征,将这两部分特征经过注意力计算获得第三部分特征,最终的多模态特征为三部分特征串联
多模 态信 息一 致性	Giachanou 等人 ^[63]	2020	BERT、VGG16	FakeNewsNet:0.768	预训练的 BERT 和预训练的 VGG16 模型提取文本和图像特征,将两个特征通过余弦相似度计算得到模态一致性特征,最终将三特征串联拼接得到多模态特征
	Müller-Budack 等人 ^[64]	2020	知识图谱、卷积网络	—	对新闻文本使用命名实体链接技术从 Wikidata 的知识图谱中获得外部的知识(新闻相关图片),将获得的图片特征与原始图片特征进行相似度计算,相似度的大小决定新闻的真实性
	Chen 等人 ^[66]	2022	BERT、ResNet	Weibo A:0.840 MediaEval_Dataset:0.806	预训练的 BERT 和 ResNet 抽取新闻文本和图像特征,计算跨模态相关性,通过跨模态相关性决定多模态特征的偏重
其他 领域 思想	Wang 等人 ^[67]	2018	Text-CNN、VGG19	Weibo A:0.827 MediaEval_Dataset:0.715	Text-CNN 和预训练的 VGG19 抽取新闻文本和图像特征,通过串联拼接获得多模态特征,在此基础上增加一个事件分类任务,使模型在真假分类和时间分类任务上都有较好的表现
	Khattar 等人 ^[68]	2019	BiLSTM、VGG19	Weibo A:0.824 MediaEval_Dataset:0.745	BiLSTM 和预训练的 VGG19 抽取新闻文本和图像特征,通过串联获得多模态特征,之后使用变分自编码器学习多模态特征分布
	Qian 等人 ^[69]	2021	图卷积网络、VGG19	Weibo A:0.944 MediaEval_Dataset:0.827 PHEME:0.867	预训练的 VGG19 抽取新闻图像特征,将文本和新闻关联信息构建图数据,并使用图卷积网络提取图特征,最后将两部分特征串联得到多模态特征

的差异,如今还需要探索一种通用、稳定、高效的多模态虚假新闻检测框架。此外,多流的框架还存在其他的缺点:

(1)由于多流框架需要对不同模态设计不同的特征提取模型,这不可避免地会丢失模态之间的时间同步信息,目前仍然没有有效的解决方法来处理多模态的时间异步问题;

(2)相对于单流框架来说,多流框架的结构更加复杂,而且还需要研究者对不同模态领域知识和处

理方法有足够的了解。

3.2 基于图形式的虚假新闻检测方法

社会性是新闻的基本特性之一,新闻数据可以表示为新闻和新闻受众互动的社交网络图,新闻社交网络图包含了新闻文章、评论等纯文本数据,也包含了节点、连边等关系型数据,这些不同形式数据组成的图可以看作特殊的多模态数据。本节主要综述基于新闻社交图的虚假新闻检测技术,其大致可以包含两类:基于图机器学习的技术和基于图神经网络

络的技术。

3.2.1 基于图机器学习的技术

虚假信息的传播主要包含三种因素^[71]：一是新闻内容的合理性；二是传播者的个性以及可信度；三是传播网络的同质性。基于以上因素，研究者根据新闻内容和社交信息建立了不同的新闻社交图，如新闻传播树、新闻立场网络等，以探究虚假新闻的传播模式。

传播树代表了在社交媒体上新闻文章的发帖和转发之间的关系。Wu 等人^[72]将消息传播模式描述为树结构的关系，传播树不仅能反映转发者与作者之间的关系，还能反映转发者的即时行为和情感。其次，Ma 等人^[13]分别构建了真新闻和假新闻的消息传播树，利用真新闻和假新闻存在的不同传播模式，计算两棵传播树之间的子结构的相似性，实验证明该方法可以有效帮助检测假新闻。

立场网络的节点表示新闻和帖子，边表示帖子与帖子之间的支持和反对关系。利用立场网络进行虚假新闻检测，即检测与某新闻相关帖子的可信度，可信度越低，代表该新闻是假新闻的可能性越大。在新闻的传播中，有学者发现^[73]，可以通过用户分享的观点、猜测和证据来自我纠正一些不正确的信息。如图 5 所示，图 5(a) 表示虚假新闻的立场网络，图 5(b) 表示真实新闻的立场网络。此外，有学者对假新闻传播树和立场网络进行综合分析。Davoudi 等人^[74]提出了一种包含动态分析、静态分析和结构分析三个结构的检测框架。其分别使用循环神经网络、全连接神经网络和 Node2Vec 学习传播树和立场网络随时间的演化模式、检测结束时传播树和立场网络的整体特征以及传播树和立场网络的结构特征，最终汇总三个结构的输出完成虚假新闻的检测。

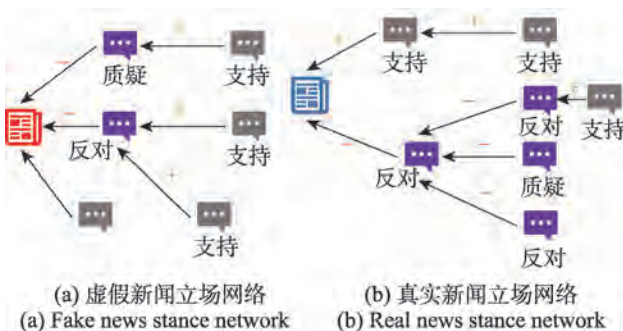


图5 新闻立场网络图

Fig.5 News stance network diagram

在处理图数据中，非负矩阵分解是一种常用的

机器学习处理方法，在处理新闻数据中，非负矩阵分解的方法将新闻词矩阵 X 投影到一个低维的联合潜在语义空间中，用语义空间中两个非负矩阵的内积表示新闻的词关系。

具体来说，给定新闻的词矩阵 $X \in \mathbb{R}^{n \times t}$ ， n 表示新闻中单词的个数， t 表示单词的嵌入维度，NMF (non-negative matrix factorization) 算法尝试寻找两个非负矩阵 $D \in \mathbb{R}^{n \times d}$ 和 $V \in \mathbb{R}^{t \times d}$ ，其中 d 为潜在空间的维度，其大小通常要小于 n 和 t ，目标函数如下：

$$\min_{D, V \geq 0} \|X - DV^T\|_F^2 + \lambda(\|D\|_F^2 + \|V\|_F^2) \quad (1)$$

其中， D 可以表示新闻的低维表示， V 表示单词的低维表示， $\lambda(\|D\|_F^2 + \|V\|_F^2)$ 为正则项，其目的是为了提高模型的泛化能力， λ 为惩罚系数，用来控制正则项的权重。目标函数可以通过广义拉格朗日方法求解更新 D 和 V ，之后用分解后的矩阵优化任务目标，以欺诈新闻分类任务为例，任务目标可以表示为：

$$\min_w \|wD - y\|_2^2 + \lambda \|w\|_2^2 \quad (2)$$

其中， $w \in \mathbb{R}^d$ 是将新闻潜在特征映射到欺诈新闻标签的权重矩阵， $y \in \mathbb{R}^n$ 为欺诈新闻标签向量。

Shu 等人^[75]曾提出了一种基于非负矩阵分解的虚假新闻检测模型，它从新闻的传播网络中构建了社交用户矩阵、新闻内容矩阵、社交用户-新闻行为矩阵和作者-新闻发布关系矩阵。对于社交用户矩阵和新闻内容矩阵，采用了 NMF 方法进行分解，对于后两个矩阵，分别通过用户信用和新闻真实性标签之间的规则和发布者党派标签和新闻真实性性标签之间的规则完成矩阵分解。该方法最终在多个数据集上都取得了不错的成绩。

3.2.2 基于图神经网络的技术

近年来，研究者们借鉴了卷积网络、循环网络和深度自编码器的思想，设计了可以用于处理图数据的神经网络结构——“图神经网络”^[76]。该技术在处理图关系数据时有独特的优势，而虚假新闻的散布和传播是以图形式实现的，图中节点表示与新闻相关的实体信息，而连边表示不同实体之间的联系。新闻社交传播图如图 6 所示。

图结构也可以挖掘虚假新闻之间存在的相似性。Dou 等人^[77]提出了 CARE-GNN (camouflage-resistant graph neural network) 虚假新闻检测框架，他们认为虚假新闻发布者伪造的新闻数据特征往往相似度较高，在图结构上表现为节点之间具有较高的相似性。针对这一目标，他们根据新闻发布者发布的文

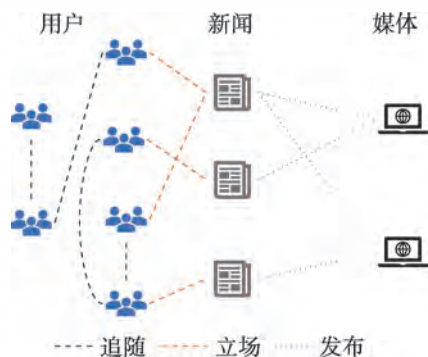


图6 新闻社交传播图

Fig.6 News social communication graph

本信息,结合新闻社交图中节点之间的联系,挖掘图节点之间的相似关系,以此辅助预测社交图上的虚假信息。

图卷积网络(graph convolutional network, GCN)是借用卷积网络的思想处理图数据而提出的一种图神经网络模型,其核心思想是学习一个映射函数,对于图中的一个节点,聚合该节点的特征和邻居节点的特征来生成该节点的新表示。Chandra等人^[78]提出的 SAFER (socially aware fake news detection framework)模型使用GCN来获取具有用户信息的新闻表示,然而他们构建的是同质图网络,会导致信息丢失问题。在此基础上,Wang等人^[79]以新闻文本、图片和知识概念为节点构建异质图,一定程度上缓解了该问题。此外,Bian等人^[80]从新闻的传播深度和散布广度两个角度研究虚假新闻的扩散模式,如图7所示,他们提出了双向图卷积神经网络,从自上而下和自下而上两个方向分别获取虚假新闻传播和散布的模式,最终的实验结果证明该方法的有效性。

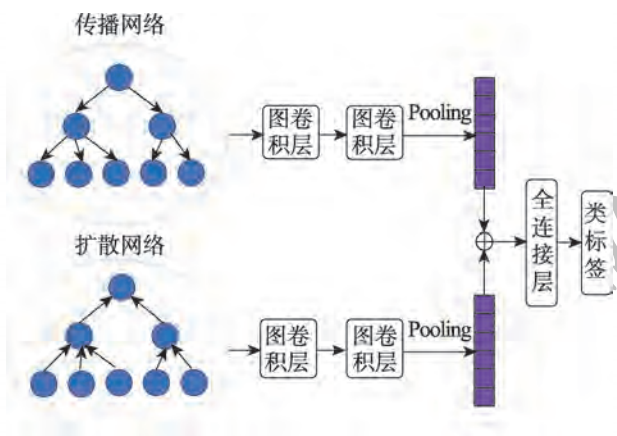


图7 Bi-GCN模型总览图

Fig.7 Model overview diagram of Bi-GCN

图注意力网络(graph attention, GAT)的结构在图卷积网络的基础上,选择使用注意力机制作为聚合方法。Ni等人^[81]以新闻文本和转发用户构建新闻传播图,分别使用文本注意力和图注意力方法从新闻文本中和传播图中寻找涉及虚假信息的关键词和可疑用户,最终在平均准确度上比最先进的技术高了2.5个百分点。Hu等人^[82]构建了一个包含句子、主题和实体的有向异质文本图,并借助图注意力网络学习主题增强的新闻表示以及基于上下文的实体表示,其后又设计了一个实体对比网络,捕获实体与知识库中的实体之间的一致性,最终该模型的F1分数在实验结果上比最佳基准模型高了3个百分点。

此外,利用图深度学习探索时间因素对虚假新闻传播的影响也是一个热门话题,其主要包含两种范式:一种是以图神经网络和循环神经网络(recurrent neural network, RNN)相结合,利用RNN学习新闻图的演变;另一种是将时间因素作为特征放入图神经网络中进行学习。Nguyen等人^[83]提出了一种虚假新闻检测框架FANG(factual news graph),其结构如图8所示,他们将时间戳、用户立场以及GraphSAGE提取的新闻用户特征进行拼接,并使用BiLSTM+Attention模型对拼接特征进行更新,以此来获取新闻社交图随时间变化的信息。Song等人^[84]基于图注意力网络设计了一个时序图注意力机构(temporal graph attention, TGAT),该模型以历史交互信息来生成新闻图的记忆向量,以学习不同时刻图结构的时态特征。

这些引入时间因素的框架可以展现新闻传播图随时间发生的变化,一定程度上增强了检测方法的解释能力。然而,这些框架设计起来较为繁琐,通常需要将检测任务分解为多个子任务,这样导致各个任务之间的错误信息不断累计,最终影响模型性能;此外,这些具有时序信息与背景信息的新闻特征很难通过端到端的方式学习。为了解决这些问题,学者们提出了以元路径为研究角度来避免新闻信息的损失,其中,元路径是指新闻发布到最后传播的节点之间的路径,并使用GRU作为聚合函数来实现对时间信息的端到端学习^[85]。

总的来说,基于图形式的虚假新闻检测方法具有准确率高、灵活性强等优点,可以识别影响虚假信息传播的重要节点,为模型提供了一定的解释能力。但也存在一些问题,如新闻社交图需要事先人为构建,当与新闻相关的实体数量太多时,需要花费

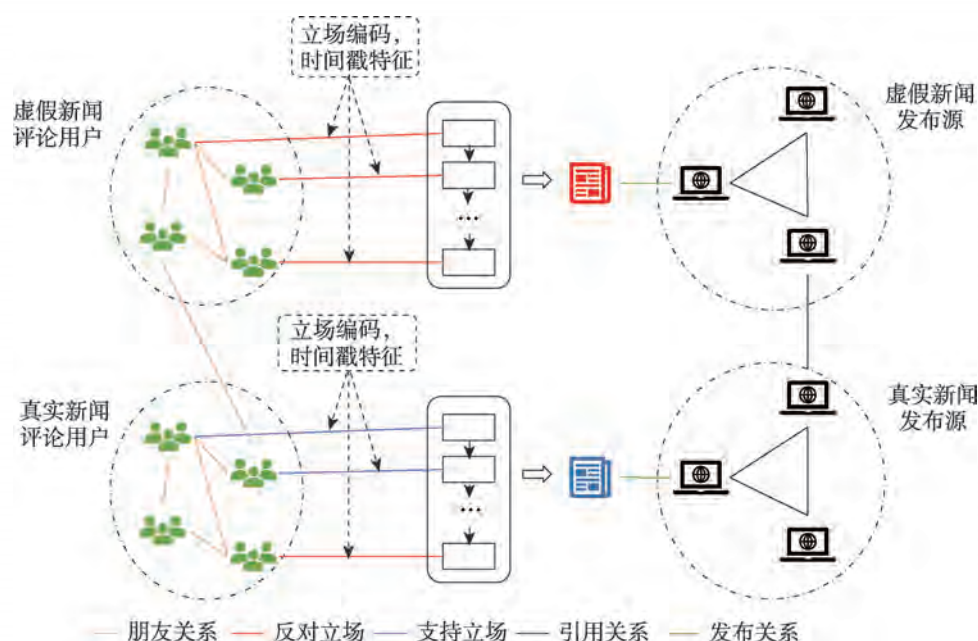


图8 FANG网络结构图

Fig.8 FANG network structure diagram

大量时间,有时还可能错漏关键实体信息;其次,图的训练需要花费大量时间,对硬件的需要较大;此外,涉及时间因素的图检测技术仍然发展不完善。

4 研究与挑战

尽管近些年来虚假新闻检测技术已有了长足的进步,研究中的检测准确率都已媲美甚至超越了人工,但其在面对实际问题时仍然存在不少的缺陷,造成这些缺陷的主要原因有以下几个方面:

(1)数据集问题。虚假新闻数据集通常具有规模小、内容形式不一致等问题,这些问题的存在严重阻碍了该领域的进步,因此,后续研究者一方面可以收集和共享规模更大、内容更全面的数据集,另一方面需要积极开发虚假新闻检测领域的弱监督和无监督方法。

(2)模型问题。受虚假新闻复杂多样性的影响,研究者们试图设计一个万全的模型,能综合考虑新闻的内容属性、社交属性以及多模态属性,但这使得他们设计出的模型框架是非常庞大的,有较高的算法复杂度。

(3)语境问题。现有的关于虚假新闻的语境信息通常只涉及了新闻的评论信息、新闻发布者的基本信息、新闻评论者的立场态度等,而关于新闻更一般的背景,如新闻的报道的事件、人物的背景知识、新闻的传播动态、评论者之间的可信度、社交信息

等,很少有研究从这些角度考虑,因此需要对这些语境信息进行整合,从整体的视角上进行分析。

(4)细粒度问题。现有的虚假新闻检测任务大多被看作二分类任务,但某些情况下新闻的虚假程度是更加值得关心的话题,因此需要对新闻进行更细粒度的划分,如搞笑性、讽刺性、攻击性等虚假新闻。甚至可以建立回归模型对新闻的虚假程度进行打分。

(5)可解释性问题。深度学习方法具有弱解释性,在虚假新闻领域,探究影响虚假新闻传播的因素也是一个极重要的研究领域。因此,怎样构建模型使得保持高检测准确率的同时,模型结果能具有良好的可解释性,这在虚假新闻检测领域也是一个值得探讨的话题。

5 结束语

在互联网时代下,如何在海量的新闻中准确高效地识别虚假信息成为了国际关心的热点话题。经过多年的研究探索,虚假新闻检测技术已经从早期的人工检测发展成如今的自动化检测,基于机器学习的人工特征提取转变为如今的深度学习自动特征提取,对新闻单一对象的检测方法演变为用户特征、文本、图片、视频特征以及传播特征等多模态联合的检测方法。

本文对虚假新闻检测研究相关理论进行了整

理,从单模态到多模态角度对虚假新闻检测数据集与相关技术做了全面的综述,并对现有研究中存在的缺陷做了归纳整理,最后给出该领域存在的问题以及以后的研究方向。本文不仅对后来的学者们有借鉴作用,而且还对专业媒体平台应对虚假新闻冲击提供重要的实际应用价值。

参考文献:

- [1] NGADIRON S, ABD AZIZ A, MOHAMED S S. The spread of COVID-19 fake news on social media and its impact among Malaysians[J]. *International Journal of Law, Government and Communication*, 2021, 6(22): 253-260.
- [2] SHU K, SLIVA A, WANG S, et al. Fake news detection on social media: a data mining perspective[J]. *ACM SIGKDD Explorations Newsletter*, 2017, 19(1): 22-36.
- [3] MANZOOR S I, SINGLA J. Fake news detection using machine learning approaches: a systematic review[C]//*Proceedings of the 3rd International Conference on Trends in Electronics and Informatics*, Tirunelveli, Apr 23-25, 2019. Piscataway: IEEE, 2019: 230-234.
- [4] MRIDHA M F, KEYA A J, HAMID M A, et al. A comprehensive review on fake news detection with deep learning [J]. *IEEE Access*, 2021, 9: 156151-156170.
- [5] KUMAR S, KUMAR S, YADAV P, et al. A survey on analysis of fake news detection techniques[C]//*Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems*, Coimbatore, Mar 25-27, 2021. Piscataway: IEEE, 2021: 894-899.
- [6] ALAM F, CRESCI S, CHAKRABORTY T, et al. A survey on multimodal disinformation detection[J]. *arXiv:2103.12541*, 2021.
- [7] GUIMARÃES N, FIGUEIRA Á, TORGO L. Can fake news detection models maintain the performance through time? A longitudinal evaluation of Twitter publications[J]. *Mathematics*, 2021, 9(22): 2988.
- [8] ALI I, AYUB M N B, SHIVAKUMARA P, et al. Fake news detection techniques on social media: a survey[J]. *Wireless Communications and Mobile Computing*, 2022: 6072084.
- [9] 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*, New York, Jul 9-15, 2016. Menlo Park: AAAI, 2016: 3818-3824.
- [10] CAO J, GUO J, LI X, et al. Automatic rumor detection on microblogs: a survey[J]. *arXiv:1807.03505*, 2018.
- [11] BOIDIDOU C, PAPADOPOULOS S, ZAMPOGLOU M, et al. Detection and visualization of misleading content on Twitter[J]. *International Journal of Multimedia Information Retrieval*, 2018, 7(1): 71-86.
- [12] PAPADOPOULOU O, ZAMPOGLOU M, PAPADOPOULOS S, et al. A corpus of debunked and verified user-generated videos[J]. *Online Information Review*, 2018, 43(1): 72-88.
- [13] 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*, Vancouver, Jul 30-Aug 4, 2017. Stroudsburg: ACL, 2017: 708-717.
- [14] DAI E, SUN Y, WANG S. Ginger cannot cure cancer: battling fake health news with a comprehensive data repository[C]//*Proceedings of the 14th International AAAI Conference on Web and Social Media*, Atlanta, Jun 8-11, 2020. Menlo Park: AAAI, 2020: 853-862.
- [15] LI Y, JIANG B, SHU K, et al. MM-COVID: a multilingual and multimodal data repository for combating COVID-19 disinformation[J]. *arXiv:2011.04088*, 2020.
- [16] ZUBIAGA A, LIAKATA M, PROCTER R. Learning reporting dynamics during breaking news for rumour detection in social media[J]. *arXiv:1610.07363*, 2016.
- [17] SHU K, MAHUESWARAN D, WANG S, et al. Fake-newsnet: 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.
- [18] WANG W Y. "liar, liar pants on fire": a new benchmark dataset for fake news detection[C]//*Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Jul 30-Aug 4, 2017. Stroudsburg: ACL, 2017: 422-426.
- [19] 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, Sep 9-11, 2017. Stroudsburg: ACL, 2017: 2931-2937.
- [20] AHMED H, TRAORE I, SAAD S. Detection of online fake news using n-gram analysis and machine learning techniques[C]//*Proceedings of the 1st International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, Vancouver, Oct 26-28, 2017. Cham: Springer, 2017: 127-138.
- [21] SALEM F K A, FEEL R A, ELBASSUONI S, et al. FA-KES: a fake news dataset around the Syrian war[C]//*Proceedings of the 13th International Conference on Web and Social Media*, Munich, Jun 11-14, 2019. Menlo Park: AAAI, 2019: 573-582.
- [22] SHAHI G K, DIRKSON A, MAJCHRZAK T A. An exploratory study of COVID-19 misinformation on Twitter[J]. *Online Social Networks and Media*, 2021, 22: 100104.
- [23] NØRREGAARD J, HORNE B D, ADALI S. NELA-GT-2018: a large multi-labelled news dataset for the study of misinformation in news articles[C]//*Proceedings of the 13th*

- International Conference on Web and Social Media, Munich, Jun 11-14, 2019. Menlo Park: AAAI, 2019: 630-638.
- [24] GRUPPI M, HORNE B D, ADALI S. NELA-GT-2021: a large multi-labelled news dataset for the study of misinformation in news articles[J]. arXiv:2203.05659, 2022.
- [25] 高玉君, 梁刚, 蒋方婷, 等. 社会网络谣言检测综述[J]. 电子学报, 2020, 48(7): 1421-1435.
- GAO Y J, LIANG G, JIANG F T, et al. Social network rumor detection: a survey[J]. Acta Electronica Sinica, 2020, 48(7): 1421-1435.
- [26] ZHOU X, ZAFARANI R. Network-based fake news detection: a pattern-driven approach[J]. SIGKDD Explorations, 2019, 21(2): 48-60.
- [27] ZHAO Y, ZOBEL J. Searching with style: authorship attribution in classic literature[C]//Proceedings of the 30th Australasian Computer Science Conference, Victoria, Jan 30-Feb 2, 2007. Sydney: Australian Computer Society, 2007: 59-68.
- [28] FENG S, BANERJEE R, CHOI Y. Syntactic stylometry for deception detection[C]//Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Jeju Island, Jul 8-14, 2012. Stroudsburg: ACL, 2012: 171-175.
- [29] PÉREZ-ROSAS V, KLEINBERG B, LEFEVRE A, et al. Automatic detection of fake news[J]. arXiv:1708.07104, 2017.
- [30] RUBIN V L, CONROY N J, CHEN Y. Towards news verification: deception detection methods for news discourse[C]//Proceedings of the 2015 Hawaii International Conference on System Sciences, Hawaii, Jan 5-8, 2015. Washington: IEEE Computer Society, 2015: 5-8.
- [31] KARIMI H, TANG J. Learning hierarchical discourse-level structure for fake news detection[J]. arXiv:1903.07389, 2019.
- [32] SHARMA K, QIAN F, JIANG H, et al. Combating fake news: a survey on identification and mitigation techniques[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(3): 1-42.
- [33] CHU Z, GIANVECCIO S, WANG H, et al. Detecting automation of Twitter accounts: are you a human, bot, or cyborg? [J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(6): 811-824.
- [34] CASTILLO C, MENDOZA M, POBLETE B. Information credibility on Twitter[C]//Proceedings of the 20th International Conference on World Wide Web, Hyderabad, Mar 28-Apr 1, 2011. New York: ACM, 2011: 675-684.
- [35] GUO H, CAO J, ZHANG Y, et al. Rumor detection with hierarchical social attention network[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Oct 22-26, 2018. New York: ACM, 2018: 943-951.
- [36] 段大高, 盖新新, 韩忠明, 等. 基于梯度提升决策树的微博虚假消息检测[J]. 计算机应用, 2018, 38(2): 410-414.
- DUAN D G, GAI X X, HAN Z M, et al. Micro-blog misinformation detection based on gradient boost decision tree [J]. Journal of Computer Applications, 2018, 38(2): 410-414.
- [37] 张仰森, 彭媛媛, 段宇翔, 等. 基于评论异常度的新浪微博谣言识别方法[J]. 自动化学报, 2020, 46(8): 1689-1702.
- ZHANG Y S, PENG Y Y, DUAN Y X, et al. The method of Sina Weibo rumor detecting based on comment abnormality [J]. Acta Automatica Sinica, 2020, 46(8): 1689-1702.
- [38] 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, Melbourne, Oct 18-23, 2015. New York: ACM, 2015: 1751-1754.
- [39] BAHAD P, SAXENA P, KAMAL R. Fake news detection using bi-directional LSTM-recurrent neural network[J]. Procedia Computer Science, 2019, 165: 74-82.
- [40] CHEN T, LI X, YIN H, et al. Call attention to rumors: deep attention based recurrent neural networks for early rumor detection[C]//Proceedings of the 2018 Pacific-Asia Conference on Trends and Applications in Knowledge Discovery and Data Mining, Melbourne, Jun 3-8, 2018. Cham: Springer, 2018: 40-52.
- [41] YU F, LIU Q, WU S, et al. A convolutional approach for misinformation identification[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Aug 19-25, 2017. Menlo Park: AAAI, 2017: 3901-3907.
- [42] QIAN F, GONG C, SHARMA K, et al. Neural user response generator: fake news detection with collective user intelligence[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Jul 13-19, 2018. New York: AAAI Press, 2018: 3834-3840.
- [43] MIKOLOV T, DEORAS A, KOMBRINK S, et al. Empirical evaluation and combination of advanced language modeling techniques[C]//Proceedings of the 12th Annual Conference of the International Speech Communication Association, Florence, Aug 27-31, 2011. Baixas: ISCA, 2011: 605-608.
- [44] DUNGS S, AKER A, FUHR N, et al. Can rumour stance alone predict veracity? [C]//Proceedings of the 27th International Conference on Computational Linguistics, New Mexico, Aug 20-26, 2018. Stroudsburg: ACL, 2018: 3360-3370.
- [45] MA J, GAO W, WONG K F. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning[C]//Proceedings of the 2019 World Wide Web Conference, San Francisco, May 13-17, 2019. New York: ACM, 2019: 3049-3055.
- [46] 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, Boston, Jun 7-12, 2015. Washington: IEEE Computer Society, 2015: 1-9.
- [47] 何韩森, 孙国梓. 基于特征聚合的假新闻内容检测模型 [J]. 计算机应用, 2020, 40(8): 2189-2193.
- HE H S, SUN G Z. Fake news content detection model based on feature aggregation[J]. Journal of Computer Appli-

- cations, 2020, 40(8): 2189-2193.
- [48] 孙尉超, 陈涛. 基于 ALBERT-BiLSTM 模型的微博谣言识别方法研究[J]. 计算机时代, 2020(8): 21-26.
SUN W C, CHEN T. Research on Microblog rumor recognition method based on ALBERT-BiLSTM model[J]. Computer Era, 2020(8): 21-26.
- [49] KALIYAR R K, GOSWAMI A, NARANG P. FakeBERT: fake news detection in social media with a BERT-based deep learning approach[J]. Multimedia Tools and Applications, 2021, 80(8): 11765-11788.
- [50] MAYANK M, SHARMA S, SHARMA R. DEAP-FAKED: knowledge graph based approach for fake news detection [J]. arXiv:2107.10648, 2021.
- [51] MORENCY L P, LIANG P P, ZADEH A. Tutorial on multi-modal machine learning[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Tutorial Abstracts, Seattle, Jul 10-15, 2022. Stroudsburg: ACL, 2022: 33-38.
- [52] KIM W, SON B, KIM I. ViLT: vision-and-language transformer without convolution or region supervision[C]//Proceedings of the 38th International Conference on Machine Learning, Jul 18-24, 2021: 5583-5594.
- [53] NAGRANI A, YANG S, ARNAB A, et al. Attention bottlenecks for multimodal fusion[C]//Advances in Neural Information Processing Systems 34, Dec 6-14, 2021: 14200-14213.
- [54] 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, California, Oct 23-27, 2017. New York: ACM, 2017: 795-816.
- [55] SINGHAL S, SHAH R R, CHAKRABORTY T, et al. SpotFake: a multi-modal framework for fake news detection [C]//Proceedings of the 2019 IEEE 5th International Conference on Multimedia Big Data, Singapore, Sep 11-13, 2019. Piscataway: IEEE, 2019: 39-47.
- [56] SINGHAL S, KABRA A, SHARMA M, et al. SpotFake+: a multimodal framework for fake news detection via transfer learning (student abstract)[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conference, the 10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, Feb 7-12, 2020. Menlo Park: AAAI, 2020: 13915-13916.
- [57] 陈志毅, 隋杰. 基于 DeepFM 和卷积神经网络的集成式多模态谣言检测方法[J]. 计算机科学, 2022, 49(1): 101-107.
CHEN Z Y, SUI J. DeepFM and convolutional neural networks ensembles for multimodal rumor detection[J]. Computer Science, 2022, 49(1): 101-107.
- [58] SONG C, NING N, ZHANG Y, et al. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks[J]. Information Processing & Management, 2021, 58(1): 102437.
- [59] QIAN S, WANG J, HU J, et al. Hierarchical multi-modal contextual attention network for fake news detection[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Jul 11-15, 2021. New York: ACM, 2021: 153-162.
- [60] 亓鹏, 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 P, CAO J, SHENG Q. Semantics-enhanced multi-modal fake news detection[J]. Journal of Computer Research and Development, 2021, 58(7): 1456-1465.
- [61] 威力鑫, 万书振, 唐斌, 等. 基于注意力机制的多模态融合谣言检测方法[J]. 计算机工程与应用, 2022, 58(19): 209-217.
QI L X, WAN S Z, TANG B, et al. Multimodal fusion rumor detection method based on attention mechanism[J]. Computer Engineering and Applications, 2022, 58(19): 209-217.
- [62] ZHOU X, WU J, ZAFARANI R. SAFE: similarity-aware multi-modal fake news detection[C]//LNCS 12085: Proceedings of the 24th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, Singapore, May 11-14, 2020. Cham: Springer, 2020: 354-367.
- [63] GIACHANOU A, ZHANG G, ROSSO P. Multimodal multi-image fake news detection[C]//Proceedings of the 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, Sydney, Oct 6-9, 2020. Piscataway: IEEE, 2020: 647-654.
- [64] MÜLLER-BUDACK E, THEINER J, DIERING S, et al. Multimodal analytics for real-world news using measures of cross-modal entity consistency[C]//Proceedings of the 2020 International Conference on Multimedia Retrieval, Dublin, Jun 8-11, 2020. New York: ACM, 2020: 16-25.
- [65] XUE J, WANG Y, TIAN Y, et al. Detecting fake news by exploring the consistency of multimodal data[J]. Information Processing & Management, 2021, 58(5): 102610.
- [66] CHEN Y, LI D, ZHANG P, et al. Cross-modal ambiguity learning for multimodal fake news detection[C]//Proceedings of the 2022 ACM Web Conference, Lyon, Apr 25-29, 2022. New York: ACM, 2022: 2897-2905.
- [67] 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, Aug 19-23, 2018. New York: ACM, 2018: 849-857.
- [68] KHATTAR D, GOUD J S, GUPTA M, et al. MVAE: multi-modal variational autoencoder for fake news detection[C]//Proceedings of the 2019 World Wide Web Conference, San Francisco, May 13-17, 2019. New York: ACM, 2019: 2915-2921.
- [69] QIAN S, HU J, FANG Q, et al. Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2021, 17(3): 1-23.

- [70] ZHANG H, FANG Q, QIAN S, et al. Multi-modal knowledge-aware event memory network for social media rumor detection[C]//Proceedings of the 27th ACM International Conference on Multimedia, Nice, Oct 21-25, 2019. New York: ACM, 2019: 1942-1951.
- [71] ASWANI R, KAR A K, ILAVARASAN P V. Experience: managing misinformation in social media—insights for policymakers from Twitter analytics[J]. Journal of Data and Information Quality, 2019, 12(1): 1-18.
- [72] WU K, YANG S, ZHU K Q. False rumors detection on Sina Weibo by propagation structures[C]//Proceedings of the 2015 IEEE 31st International Conference on Data Engineering, Seoul, Apr 13-17, 2015. Piscataway: IEEE, 2015: 651-662.
- [73] MA J, GAO W, WONG K F. Rumor detection on Twitter with tree-structured recursive neural networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Jul 15-20, 2018. Stroudsburg: ACL, 2018: 1980-1989.
- [74] DAVOUDI M, MOOSAVI M R, SADREDDINI M H. DSS: a hybrid deep model for fake news detection using propagation tree and stance network[J]. Expert Systems with Applications, 2022, 198: 116635.
- [75] SHU K, WANG S, LIU H. Beyond news contents: the role of social context for fake news detection[C]//Proceedings of the 12th ACM International Conference on Web Search and Data Mining, Melbourne, Feb 11-15, 2019. New York: ACM, 2019: 312-320.
- [76] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [77] DOU Y, LIU Z, SUN L, et al. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Ireland, Oct 19-23, 2020. New York: ACM, 2020: 315-324.
- [78] CHANDRA S, MISHRA P, YANNAKOUDAKIS H, et al. Graph-based modeling of online communities for fake news detection[J]. arXiv:2008.06274, 2020.
- [79] WANG Y, QIAN S, HU J, et al. Fake news detection via knowledge-driven multimodal graph convolutional networks[C]//Proceedings of the 2020 International Conference on Multimedia Retrieval, Dublin, Jun 8-11, 2020. New York: ACM, 2020: 540-547.
- [80] BIAN T, XIAO X, XU T, et al. Rumor detection on social media with bi-directional graph convolutional networks[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conference, the 10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, Feb 7-12, 2020. Menlo Park: AAAI, 2020: 549-556.
- [81] NI S, LI J, KAO H Y. MVAN: multi-view attention networks for fake news detection on social media[J]. IEEE Access, 2021, 9: 106907-106917.
- [82] HU L, YANG T, ZHANG L, et al. Compare to the knowledge: graph neural fake news detection with external knowledge[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Aug 1-6, 2021. Stroudsburg: ACL, 2021: 754-763.
- [83] NGUYEN V H, SUGIYAMA K, NAKOV P, et al. FANG: leveraging social context for fake news detection using graph representation[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Ireland, Oct 19-23, 2020. New York: ACM, 2020: 1165-1174.
- [84] SONG C, SHU K, WU B. Temporally evolving graph neural network for fake news detection[J]. Information Processing & Management, 2021, 58(6): 102712.
- [85] CUI J, KIM K, NA S H, et al. Meta-path-based fake news detection leveraging multi-level social context information[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, Oct 17-21, 2022. New York: ACM, 2022: 325-334.



刘华玲(1964—),女,重庆人,博士,教授,主要研究方向为知识管理与智能决策、机器学习等。
LIU Hualing, born in 1964, Ph.D., professor. Her research interests include knowledge management and intelligent decision making, machine learning, etc.



陈尚辉(1998—),男,江苏连云港人,硕士研究生,主要研究方向为机器学习、人工智能等。
CHEN Shanghui, born in 1998, M.S. candidate. His research interests include machine learning, artificial intelligence, etc.



曹世杰(1997—),男,天津人,硕士研究生,主要研究方向为数据挖掘、机器学习等。
CAO Shijie, born in 1997, M.S. candidate. His research interests include data mining, machine learning, etc.



朱建亮(1999—),男,江西南昌人,硕士研究生,主要研究方向为数据挖掘、机器学习等。
ZHU Jianliang, born in 1999, M.S. candidate. His research interests include data mining, machine learning, etc.



任青青(1997—),女,上海人,硕士,主要研究方向为数据挖掘、管理科学等。
REN Qingqing, born in 1997, M.S. Her research interests include data mining, management science, etc.