Journal Pre-proof

Deep learning for fake news detection: A comprehensive survey

LinMei Hu, SiQi Wei, Ziwang Zhao, Bin Wu

PII: S2666-6510(22)00013-4
DOI: https://doi.org/10.1016/j.aiopen.2022.09.001
Reference: AIOPEN 37

To appear in: *AI Open*

Received date : 17 December 2021
Revised date : 8 August 2022
Accepted date : 13 September 2022

Please cite this article as: L. Hu, S. Wei, Z. Zhao et al., Deep learning for fake news detection: A comprehensive survey. *AI Open* (2022), doi: https://doi.org/10.1016/j.aiopen.2022.09.001.

Deep Learning for Fake News Detection: A Comprehensive Survey

LinMei Hu[a], SiQi Wei[b], Ziwang Zhao[b] and Bin Wu[b,*]

[a]Beijing Institute of Technology, Beijing 100081, China
[b]School of Computer Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

ARTICLE INFO

Keywords:
Fake News Detection
Deep Learning

ABSTRACT

The information age enables people to obtain news online through various channels, yet in the meanwhile making false news spread at unprecedented speed. Fake news exerts detrimental effects for it impairs social stability and public trust, which calls for increasing demand for fake news detection (FND). As deep learning (DL) achieves tremendous success in various domains, it has also been leveraged in FND tasks and surpasses traditional machine learning based methods, yielding state-of-the-art performance. In this survey, we present a complete review and analysis of existing DL based FND methods that focus on various features such as news content, social context, and external knowledge. We review the methods under the lines of supervised, weakly supervised, and unsupervised methods. For each line, we systematically survey the representative methods utilizing different features. Then, we introduce several commonly used FND datasets and give a quantitative analysis of the performance of the DL based FND methods over these datasets. Finally, we analyze the remaining limitations of current approaches and highlight some promising future directions.

1. Introduction

The Internet and social media platforms have become indispensable ways for people to obtain news in their daily life. Because they allow the news to be disseminated rapidly and freely, public viewers have unhindered access to consult whenever and wherever they want. And till August 2018, over 68 percent of Americans acquire their news through social media[1]. However, the deficiency of censorship from authority renders the quality of news spread on the far lower than by traditional way. The online information ecosystem is extremely noisy and fraught with disinformation and fake news.

Fake news refers to the news that is fabricated to deceive people, which exerts negative effects on individuals as well as the whole society. It misleads people with false or biased stories for self-serving purposes, seriously affecting public opinion and social stability. For example, during the COVID-19 pandemic, a mix of real and fake information about the outbreak is so overwhelming that the World Health Organization[2] called it an "information epidemic." In the first three months of 2020, about 6 000 people were hospitalized worldwide because of coronavirus misinformation, and researchers said at least 800 people may have died due to misinformation related to COVID-19.[3]

Fake news detection aims to identify the fake news automatically. Existing traditional ML based FND methods require feature engineering. According to the features that the models utilized, those methods can be broadly divided into three categories: linguistic features, temporal-structural features and the hybrid features. *Linguistic feature* based approaches detect fake news based on the text content (Castillo et al., 2011). For example, Castillo et al. (2011) employed a variety of linguistic features such as special characters, emoticon symbols, sentimental words, ect. Popat (2017) investigated language style features such as assertive verbs and factive verbs. In addition, some methods explore the *temporal-structural* feature (e.g., the propagation feature) based on social networks to detect fake news (Jin et al., 2013). For instance, Wu et al. (2015) proposed SVM to learn high-order propagation patterns. Sampson et al. (2016) employed implicit links between conversation fragments to properly classify emergent conversations. *Hybrid* approaches that combine different types of features to detect fake news are also proposed. For example, Sun et al. (2013) combined news content, user, and multimedia features. Ma et al. (2015) combined the temporal variations of content-based, user-based, and diffusion-based features as the news propagation evolves. Though these traditional ML methods achieve promising results, they rely heavily on laborious feature engineering.

Deep learning based fake news detection. With the success of deep learning in various domains, DL based FND methods have been proposed and attracted significant attention recently. Firstly, deep learning can avoid feature engineering and take full advantages of its strong expressive power to model the features of input news. For example, Ma et al. (2016) modeled the sequential relationship between news posts utilizing recurrent neural networks. Yu et al. (2017) utilized convolutional neural networks to represent high-level semantic relationships between news posts. Bian et al. (2020) leveraged a Graph Neural Networks (GCNs) with both top-down and bottom-up directed graph of rumor spreading to learn its propagation and dispersion patterns. Khattar et al. (2019) proposed the multi-modal Variational Autoencoder (MVAE) to extract the hidden multi-modal representations of multimedia news.

*Corresponding author
hulinmei1991@gmail.com (L. Hu); Weisiqi@bupt.edu.cn (S. Wei); zhaoziwang@bupt.edu.cn (Z. Zhao); wubin@bupt.edu.cn (B. Wu)
ORCID(s):

[1]https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/.
[2]https://www.worldometers.info/coronavirus/
[3]https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time

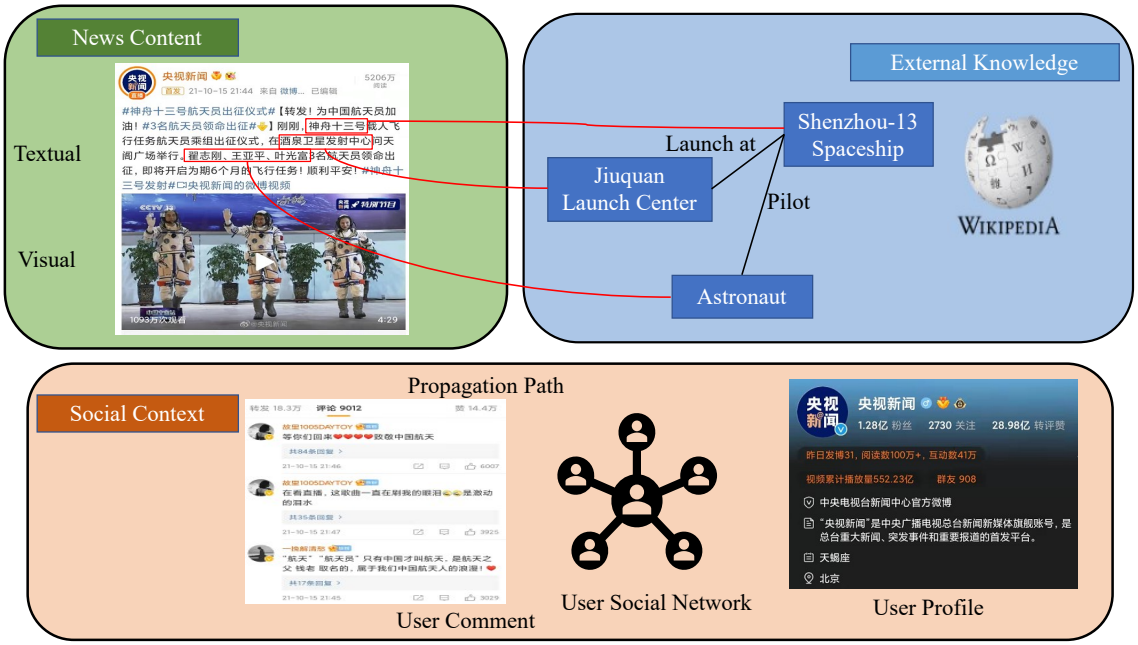


Figure 1: Information used to handle fake news detection tasks.

Motivations. Although there have been several surveys on FND, most of them divide the existing research from the feature perspective, for example, Zhou and Zafarani (2018) categorized the approaches for detecting fake news into the four categories listed below: external knowledge-based detection methods, style-based detection methods, propagation-based detection methods, and credibility-based detection methods. Similarly, Shu et al. (2017) divided the detection methods into those based on news content and those based on social context. These previous surveys provide insights and facilitate the research. Recently, weakly supervised as well as unsupervised methods have attracted attention to resolve problems of limited labeled data. It is an important concern so as to be applied to real-world situation, because large scale labeled examples are either too expensive or unavailable due to privacy concerns. Furthermore, we also notice that the methods attach different emphasis on different feature information according to the manner of learning (supervised, weakly supervised and unsupervised). This prompts us to write this survey with a novel perspective to systematically review and summarize the current status of deep learning techniques for fake news detection.

Overall, the contributions of this survey can be summarized as follows.

- ***A comprehensive review.*** We conduct a comprehensive survey to present a thorough overview and analysis of DL-based FND methods along the three lines: supervised, semi-supervised, and unsupervised learning.
- ***A quantitative analysis.*** We present a quantitative analysis on the performance of the DL based FND methods on a variety of data sets, so that researchers can learn from it.
- ***Some future directions.*** We discuss the remaining limitations of existing FND methods and point out possible future directions.

The remainder of this survey is organized as follows. We first retrospect the task definition of FND in Section 2. Then we explain our new taxonomy of DL based FND methods in Section 3. We present a comprehensive survey along with the three categories in Sections 4, and 5 respectively. Section 6 introduces several commonly used FND datasets and presents a quantitative performance analysis of the DL based FND methods, and in Section 7 we finally give a conclusion as well as some promising directions.

2. Problem Definition

In the FND task, we define the output space $\mathcal{Y} = \{0, 1\}$ indicating whether the news is fake or not. The input space $\mathcal{X}$ is relatively complex, comprising information not only from the news originally carries, such as the news content, but also from social context and external knowledge like the knowledge base. See in table 1.

Formally, let $\mathcal{D} = \{x_i, y_i\}_{i=1}^{n}$ denote a set of n news with annotated labels, with $\mathcal{X} = \{x\}_{i=1}^{n}$ denoting the news pieces and $\mathcal{Y} = \{y_i\}_{i=1}^{n} \subset \{0, 1\}^n$ denoting the corresponding labels of whether the news is fake or not. Note that, since manually annotating the data is expensive and time consuming, there are large-scale unlabeled news in real world.

Table 1
The available information for different FND methods.

Methods	News Content	Social Context	External Knowledge
Supervised	multi-modal information (textual, visual)	user credibility propagation pattern	knowledge graph pre-trained language model
Weakly/Un Supervised	single modal information (textual)	user engagement network(friends, interaction, dissemination)	pre-trained language models probabilistic knowledge

Thus, facing the scenarios with few and no labeled data, weakly supervised and even unsupervised methods are in need. Formally, given the news data $\mathcal{X}$, fake news detection models aim to learn a mapping function $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Y}$.

3. A Taxonomy of DL-based Methods

The previous survey (Zhou and Zafarani, 2018) divided the fake news detection methods from the view of features. Note that limited labeled data has been attracting increasingly more researchers to combat fake news with few or no labeled data. We propose to categorize the fake news detection methods as supervised, weakly supervised, and unsupervised. Figure 2 shows the classification taxonomy of the fake news detection methods based on deep learning in this paper. Moreover, it's worth noting that disparate methods along these aforementioned three dimensions focus on different features, which we will elaborate on in subsequent subsections.

- Supervised methods: Supervised methods learn with labeled data. The main concern is: how the DL models utilize these rich feature information. Therefore, we further divide these methods into: news content based, social context based, and external knowledge based. Besides, we particularize solutions to integrate different or even heterogeneous feature information in section 4.4.
- Weakly/Un- supervised methods: Weakly supervised methods assume only limited labeled data is available in the learning stage. The common solution is to derive weak supervision information from available information. According to the way of getting weak supervision, we divide the semi-supervised methods into weak content supervision and weak social supervision. Unsupervised methods learn with totally unlabeled data. Some researchers resort to generative methods or utilize probabilistic knowledge. We will particularize them in section 5.

4. Supervised Methods

Supervised methods tend to learn from various features with labeled examples. In order to carry out better performance, amounts of techniques such as the introduction of multi-modal information, external knowledge, and integration strategy have been explored. According to the features utilized, we further divide them as follows.

4.1. News Content-based Methods

News content means explicit information the news originally carries, such as the text of articles or images attached to it. Generally, news content-based methods utilize these textual/visual features. Table 2 compares different news content-based methods, which we will elaborate on in subsequent subsections.

4.1.1. Single Modality

Textual features can be classified as generic features and latent features. The former are often used within a traditional ML framework, which describe textual content from linguistic levels: lexicon, syntax, discourse, semantic, ect. Previous work has summarized them into a detailed table (Zhou and Zafarani, 2020). Latent textual features refer to news text embedding. The text embedding can be derived at word, sentence, and document levels. In that, a news article can be represented by latent vectors, which can either be utilized as the input for classifiers (like SVMs) right away or subsequently integrated into neural network structures.

Recurrent neural networks (RNNs) are highly capable of modeling sequential data. (Ma et al., 2016) utilized recurrent neural networks (RNN) as the basis and captured the relevant information of the event over time through learning its hidden layer representation. Chen et al. (2019b) proposed an Attention-Residual network called ARC which can capture long-range dependency by attention mechanism. Meanwhile, the convolution neural network is applied to select important components and local features. Ma et al. (2019) proposed a GAN-based model to obtain low-frequency while strong representations of fake news. The model designed a generator based on GRU to produce controversial instances that makes the distribution of tweets' opinions complicated, and a RNN-based discriminator to identify the effective features from hard samples generated by the generator. Although the above RNN-based models have achieved good results, they are biased towards the latest elements of the input sequence, while key features do not necessarily appear at the rear part of an input sequence. To address the issue with RNN-based models, Yu et al. (2017) proposed a method for fake news detection based on a convolutional neural network (CNN). The model can extract essential features from an input sequence and form relationships among relevant features at a high level. Vaibhav and Hovy (2019) modeled each news article in the dataset as a graph and reformulated the fake news detection task as a graph classification task, where the nodes represent the sentences of the article and the edges represent the semantic similarity between a pair of sentences. They used two widely applicated graph neural networks, GCN and GAT, to generate graph embedding and use the embedding to classify the fake news. Inspired

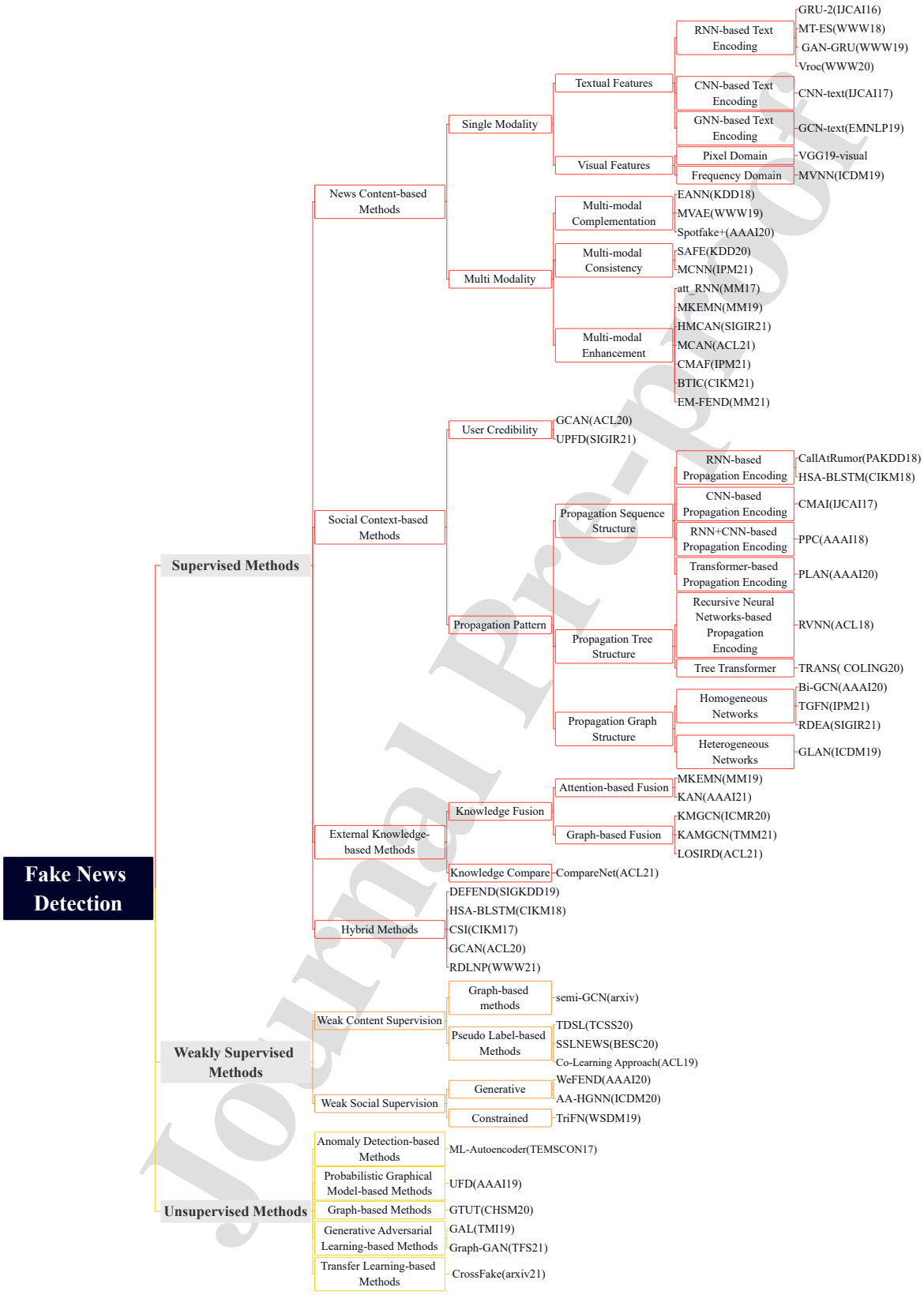


Figure 2: Taxonomy of fake news detection methods.

Table 2
Comparisons of different news content-based methods.

Input	Method	Backbone			Multi-modal Correlations	
		Text Encoder	Image Encoder	Multi-modal Fusion	Multi-modal Consistency	Multi-modal Enhancement
Text-only Method	GRU-2(Ma et al., 2016)	GRU	–	–	–	–
	CNN-text(Yu et al., 2017)	CNN	–	–	–	–
	GCN-text(Vaibhav and Hovy, 2019)	GCN	–	–	–	–
	Vroc(Cheng et al., 2020)	LSTM + VAE	–	–	–	–
Image-only Method	VGG19-visual (Simonyan and Zisserman)	–	Pixel Domain CNN	–	–	–
	MVNN(Qi et al., 2019)	–	Pixel Domain CNN + Frequency Domain CNN	–	–	–
Multi-modal Method	EANN(Wang et al., 2018b)	Text-CNN	VGG-19	Concat	–	–
	metaFEND(Wang et al., 2021)	Text-CNN	VGG-19	Concat	–	–
	MVAE(Khattar et al., 2019)	Bi-LSTM	VGG-19	Concat	–	–
	MSRD (Liu Jinshuo, 2020)	LSTM	VGG-19 + OCR	Concat	–	–
	SpotFake(Singhal et al., 2019)	BERT	VGG-19	Concat	–	–
	SpotFake+(Singhal et al., 2020)	XLNet	VGG-19	Concat	–	–
	SAFE(Singhal et al., 2020)	Text-CNN	Image2sentence + Text-CNN	Cross-modal Similarity Capture	Text-Image Caption	–
	MCNN(Xue et al., 2021)	BERT + Bi-GRU	ResNet50	Attention + Similarity	Text-Vision feature	–
	FNR(Ghorbanpour et al., 2021)	BERT	ViT	Concat + Similarity	Text-Vision Feature	–
	att_RNN(Jin et al., 2017)	Bi-LSTM	VGG19	Attention	–	Text– >Vision Feature
	MKEMN(Zhang et al., 2019)	Bi-GRU	VGG19	Attention + Multi-channel CNN	–	Text– >Vision Feature
	HMCAN(Qian et al., 2021b)	BERT	ResNet	Hearchical Multi-modal Contextual Attention Network	–	Text< – >Vision Feature
	CARMN(Song et al., 2021a)	BERT	VGG19	Co-attention Transformer + Multi-channel CNN	–	Text< – >Vision Feature
	MCAN(Wu et al., 2021b)	BERT	Pixel Domain CNN + Frequency Domain CNN	Co-attention Networks	–	Text< – >Vision Feature
	EMAF (Li et al., 2021b)	BERT	Faster-RCNN	Capsule	–	Text< – >Vision Object Feature
	BTIC(Zhang et al., 2021a)	BERT	Resnet	Multi–head Transformer	–	Text< – >Vision Patch Feature
	EM-FEND(Qi et al., 2021)	BERT	VGG19 + Vision Entity Detector	Co-attention Transformer	Text-Vision Entity	Text< – >Vision Feature, Text< – >Vision Entity

by multi-task learning, Wu et al. (2019) designed a multi-task learning model that employs a fake news detection task and a stance classification task to optimize the shared layer simultaneously, resulting in enhanced news representations. The sharing layer filters and selects shared feature flows between the tasks of fake news and stance detection using a gate mechanism and an attention mechanism. Cheng et al. (2020) proposed an LSTM-based variational autoencoder model to extract latent representations of tweet-level text. Besides, in order to *detect fake news as early as possible*, some existing researchers assumed that multifaceted information is not available before a news article has already become popular. These works utilized text-only features on purpose. For example, Qian et al. (2018) proposed a model to generate user feedback based on text, which was then used in the classification process along with word-level and sentence-level information from real articles to address the lack of user reviews as an auxiliary source of information in early detection. Giachanou et al. (2019) considered the role of emotional signals and proposed an LSTM model that incorporates emotional signals obtained from the text of the claims in order to distinguish between real and fake news.

Visual features : With the development of multimedia, the news of social networks contains not only text information but also images, videos and other visual information that involves rich semantics. Due to the heterogeneity between textual and visual information, it is difficult for textual feature-based approaches to capture visual information. Many existing scholars have proposed using visual features to detect fake news, which we will elaborate on in subsequent subsections.

Some early researches utilize basic statistical features of the attached images such as the number of attached images (Wu et al., 2015; Yang et al., 2012), image popularity and image type (Jin et al., 2016) to help detect fake news. For those tampered images which are digitally modified, Boididou et al. (2015) extracted advanced forensics features of image and combined them with post-based and user-based features to detect fake news. However, these statistical features are insufficient for describing the complicated distributions of visual information in fake news.

As a typical deep learning-based model, convolutional neural networks can effectively capture the visual features in news images. Sermanet et al. (2014); Sharif Razavian et al. (2014) demonstrated that features extracted from the pre-trained CNN could be considered a generic image representation for diverse visual recognition tasks.

Inspired by the capability of CNN, several existing efforts (Wang et al., 2018b; Khattar et al., 2019) obtained generic visual representations using a pre-trained deep CNN such as VGG19(Simonyan and Zisserman). To better utilize task-relevant information like the intrinsic features of fake-news images, Qi et al. (2019) proposed a multi-domain visual neural framework that combined frequency domain and pixel domain visual information to distinguish real news from false one by visual characteristics. Fake news images may be of bad quality, appearing in the frequency domain. The proposed model automatically captures image quality features in the frequency domain using a convolutional neural network(CNN) and automatically extracts image semantic features in the pixel domain using a CNN-RNN.

4.1.2. Multi Modality

Textual and visual features are efficacious in fake news detection tasks, respectively. News in existing social networks often contains both textual and visual information. It is a natural idea to combine them for better performance. We will illustrate multi modal-based methods along three axes according to the different multi-modal perspectives they adopt to facilitate FND.

Multi-modal Complementation. Some studies consider visual information as a complement to fake news texts. They used a text encoder to extract text features and a visual encoder to extract visual features, and simply concatenated them as the feature of the news. The framework of multi-modal complementation that commonly used in fake news detection obtains image features from pre-trained VGG19Simonyan and Zisserman first, and then concatenate these visual features simply with textual features. Under this framework, Wang et al. (2018b) improved the generalization ability of the model through introducing event classification

as an extra task to guide the process of learning event-invariant multi-modal features. After that, Wang et al. (2021) proposed a method to detect multi-modal fake news on emergent events through meta neural process. Dhruv et al. Khattar et al. (2019) modified this general complementation framework into a multimodal variational autoencoder to obtain multi-modal representations which are utilized for fake news detection. With the development of the current *pre-trained model*, Singhal et al. (2019, 2020) first introduced pre-trained language models such as BERT and XLnet to encode text features and then complemented them with visual features. Despite the success achieved by these works, they all fail to consider the complex cross-modal correlations contained in the fake news, which restricts the effectiveness of multi-modal content detection.

Multi-modal Consistency. Irrelevant images are characteristics of multi-modal fake news. Therefore, some works have paid attention to measuring the multi-modal consistency in detection.

Zhou et al. (2020b) utilized the image captioning model to transform images into sentences, and then assessed the sentence similarity between the original news text and the produced image captions to calculate multimodal inconsistency. However, due to the disparities between the image caption model's training dataset and the actual news corpus, the model's performance is limited. Xue et al. (2021) projected both visual and textual features into a mutual feature space using weight sharing encoders, and then calculated the similarity of transformed multi-modal features; however, it remains challenging to capture multi-modal inconsistency due to the semantic gap between visual and textual features. Inspired by the excellent performance of the transformer in visual representation, Ghorbanpour et al. (2021) proposed a Fake News Revealer (FNR) method that utilized vision transformer (Dosovitskiy et al., 2020) and BERT(Devlin et al., 2019) to extract image features and text features separately. And then, FNR used the contrastive loss to contrastive loss to determine image and text similarity.

Multi-modal Enhancement. Rather than directly modeling images in a coarse-grained manner when fusing information, news text and images are related in high-level semantics, whose aligned parts usually indicate the important features of news. Therefore, some works focus on extracting features in images and texts and conducting mutual enhancement to detect fake news better.

Jin et al. (2017) presented an att-RNN model to use RNN with an attention mechanism to combine text, image, and social context information for rumor detection. Zhang et al. (2019) used multi-channel CNN with the attention mechanism to fuse multi-modal information. They focused on the unidirectional enhancement of multi-modal content, highlighting the essential image regions under textual guidance. Song et al. (2021a) modeled the bidirectional enhancement between images and text using the co-attention transformer. Similarly, Qian et al. (2021b) proposed the hierarchical Multi-modal Contextual Attention Network (HMCAN) architecture to jointly models the multi-modal context information and hierarchical semantics of text in a deep, unified framework for fake news detection. The multi-modal context of each news is modeled by the multi-modal contextual attention network. Wu et al. (2021b) proposed a Multi-modal Co-Attention Network (MCAN), which extracted spatial-domain and frequency-domain features from the image and textual features from the text. MCAN also developed a novel fusion approach with multiple co-attention layers to learn inter-modality relations, which fused visual features first and then the textual features. The fused representation obtained from the last co-attention layer was used for fake news detection. Wang et al. (2020a) utilized GCN to model the relationship between words and image-extracted objects. Equally, Li et al. (2021b) adopted entity-centric cross-modal interaction, which can reserve semantic integrity and capture the details of multi-modal entities. Specifically, they designed an alignment module with the improved dynamic routing algorithm and introduced a fusion module based on the comparison, the former aligned and captured the essential entities, and the latter compared and aggregated entity-centric features. Zhang et al. (2021a) proposed a novel method for COVID-19 fake news detection. It uses a BERT-based multi-modal model to encode text and visual information, which captures the interactions between text and image. It further adopts contrastive learning in order to better learn multi-modal representations using the past articles that report similar events. Qi et al. (2021) imported the visual entities to improve the understanding of the high-level semantics related to news in images, as well as model the inconsistency and mutual enhancement of multi-modal entities.

To sum up, there is three valuable inductive bias when considering text-image correlations in multi-modal fake news detection task: The images contribute additional information to the original text, which calls for multi-modal complement. Text and images with inconsistent elements are a possible signal for the multimodal detection of fake news. Text and images enhance each other by spotting the essential features.

4.2. Social Context-based Methods

In some cases, it is not satisfying to spot fake news merely from news content because fake news pieces were written intentionally to confound the public. Social media platforms' networking and interconnecting characteristics give extra information, namely, the social context features. It represents the user engagements and social behaviors of users on social media. Methods can be separated further into credibility-based and propagation-based categories.

4.2.1. User Credibility

Credibility-based techniques aim at utilizing users' reliability to help identify fake news. Normally, the credibility information can be collected either from the explicit description of the user or by analyzing relationships between news

Table 3
Comparisons of different propagation-based methods.

Propagation Pattern Structure	Method	News Encoder	Comment/Retweets Encoder	Propagation Pattern Encoder
Sequence Structure	ML-GRU(Ma et al., 2016)	TF-IDF	TF-IDF	RNN
	CallAtRumor(Chen et al., 2018a)	TF-IDF	TF-IDF	RNN+Attention
	HSA-BLSTM(Guo et al., 2018)	Bi-LSTM+Attention	Bi-LSTM+Attention	Hierarchical Bi-LSTM
	CAMI(Yu et al., 2017)	MLP	MLP	CNN
	PPC(Liu and Wu, 2018)	MLP	MLP	CNN+RNN
	dEFEND(Shu et al., 2019a)	RNN	RNN	Co-attention
	PLAN(Khoo et al., 2020)	Transformer	Transformer	Transformer
Tree Structure	RvNN(Ma et al., 2018b)	MLP	MLP	Recursive Neural Networks
	TRANS(Ma and Gao, 2020)	Transformer	Transformer	Tree Transformer
Graph Structure	Bi-GCN(Bian et al., 2020)	MLP	MLP	GCN
	VAE-GCN(Lin et al., 2020)	TF-IDF	TF-IDF	GCN+VAE
	TGNF(Song et al., 2021b)	Average of Word Embedding	Average of Word Embedding	Dynamic Graph Neural Network
	RDEA(He et al., 2021)	MLP	MLP	GCN+Graph Contrastive Learning
	GLAN(Yuan et al., 2019)	Word Embedding+CNN	Word Embedding+CNN	Heterogeneous Graph Neural Network

articles and other components such as users, publishers and posts.

Shu et al. (2018) provided a systematic study of the relationship between users' information and the credibility of news. Later, Shu et al. (2019c) studied the problem of understanding and exploiting user profiles for fake news detection and proved its effectiveness. Dong et al. (2018) utilized an attetion-based Bi-GRU to extract news content information, a deep neural network to extract user information, and an attention mechanism to fuse text information and user information for improved fake news detection. Jiang et al. (2019) utilized attributed network representation learning to explore possible user conrrelations in the friendship network based on user attributes and reconstruct news-user network, so as to enhance the news and user embeddings in the news propagation network. Lu and Li (2020) constructed all users involved in social interaction into a fully connected graph and used graph neural network(GNN) methods to model users to detect fake news. Emotion also plays an essential role in detecting fake news online. Zhang et al. (2021b) proposed BERT-EMO to represent publisher emotion and social emotion and also considered the relation between publisher emotion and social emotion(dual emotion) to expose the distinctive emotional signals for detecting fake news.

Although models could attain good representations of users with the help of deep neural networks, the limitations are mainly manifested in privacy issues. In other words, many users are unwilling to show accurate personal information, which leads to the introduction of noisy information.

4.2.2. Propagation Pattern

Fake news often spreads differently from real news on social networks. Propagation-based methods aim to model the propagation path of news and analyze the difference between real news dissemination and fake news dissemination to identify fake news. Table 3 shows the comparisons of different propagation-based methods. We will detail them in subsequent subsections.

Some recent approaches use temporal-linguistic features extracted from a sequence of user comments for fake news detection. They model the news dissemination process as a **sequence structure**. Ma et al. (2016) utilized recurrent neural networks on a sequence of user comments for fake news detection. As an extension, Chen et al. (2018a) incorporated a soft-attention mechanism. Yu et al. (2017) proposed a convolutional neural network (CNN)-based model for fake news detection. Liu and Wu (2018) modeled news propagation in social networks as a time series and utilized CNN and RNN to capture both global and local information on this time series in order to detect fake news. In recent years, the transformer(Vaswani et al., 2017) has exhibited improved performance on a variety of NLP tasks, including machine translation, sentence representation(Devlin et al., 2019), and conversation generation(Tao et al., 2018). Khoo et al. (2020) designed a transformer-based fake news detection model PLAN. PLAN used the transformer to capture the dependency between any two posts, regardless of the reposting relationship between the posts.

Although sequence structure modeling achieves good results, sequence structure has difficulty capturing the structural relationship between news retweets and comments. In order to better learn the structural information between news and its retweeted or comment sequences. Some researchers try to model the news dissemination process as a **tree structure.** Ma et al. (2018b) used a recursive neural network to model top-down and bottom-up propagation trees for news dissemination. Because of their inherently recursive character, propagation trees may be used to speed up the learning of posts representation by inserting hidden indicative signals in the structure to recognize rumors more accurately. Inspired by the idea of enhancing the structured objects representation power using transformer, Ma and Gao (2020) proposed a tree transformer-based fake news detection model, which uses a self-attention mechanism to model post-level semantic interactions inside and among sub-trees. To enhance the robustness of the model, Ma et al. (2021) proposed a GAN-style approach, where a transformer-based generator was designed to produce uncertain or conflicting voices, further polarizing the original conversation thread to pressurize the discriminator to learn more substantial rumor indicative features from the augmented, more challenging examples.

However, the efficiency of the above methods is too low to understand the features of the propagation structure, and the global structure features of rumor dispersion are ignored. Other researchers tried to model the news dissemination process as a **graph structure**, thereby turning the fake news detection problem into a graph classification problem. Both homogeneous and heterogeneous graphs are used to model the propagation structure of news.

Homogeneous networks are composed of nodes and edges of a single type. (Zhou and Zafarani, 2019). According to the survey (Vosoughi et al., 2018), fake news spreads more quickly, further, and broadly than the real news. Bian et al. (2020) proposed a bi-directional graph convolutional network to learn both propagation and dispersion characteristics of fake news. Specifically, it applied a top-down directed graph convolutional network that considers the causal features of rumor spreading and a bottom-up oppositely directed graph convolutional network that captures the structural features from rumor dispersion. Inspired by the success of the autoencoder model in the field of learning latent information, Lin et al. (2020) proposed AE-GCN and VAE-GCN, which used GAE (Graph AutoEncoder) and its variant Variational GAE (VGAE) to learn graph structure information on rumor detection. To detect fake news as early as possible, Silva et al. (2021) proposed training an autoencoder to learn the embedding of the entire propagation network based on the partial network. They proved that the predicted embedding of the entire propagation network might yield improved results for the early detection of fake news. Rather than centering on static networks and assuming that the entire information propagation network structure can be accessed before conducting learning algorithms, Song et al. (2021b) proposed a dynamic graph-based detection framework to model the temporal evolution process of news in the real-world as the graph evolution under the view of continuous time. Inspired by contrastive learning Chen et al. (2020); He et al. (2021) proposed a graph contrastive learning model that integrated three augmentation strategies to extract meaningful rumor propagation patterns and capture intrinsic representations of user engagement. GNN-based encoder and contrastive learning mechanism allowed the proposed model to dig information from the complex rumor propagation structure and learn mutual information between propagation event and its augmentation.

Methods based on the homogeneous graph can only model a specific type of node or edge while are not able to integrate the information of multiple nodes or multiple relationships. However, news dissemination often involves multiple types of node information, such as users and comments, which calls for introducing a heterogeneous graph neural network to fuse information of multiple nodes or edges.

Heterogeneous networks are consisted of multiple types of nodes or edges (Zhou and Zafarani, 2019). Zhang et al. (2018) used the information of news articles, creators and subjects to build a heterogeneous information network and proposed a deep diffusion network which can learn the representations of news articles, creators and subjects at the same time. Huang et al. (2019) presented a meta-path-based heterogeneous graph attention network to encode the global semantic relations among user behavior. In order to increase the robustness of the model, Yang et al. (2020) first employed a particular graph adversarial learning framework to learn more characteristic structure features. Nguyen et al. (2020) proposed an inductive heterogeneous graph neuarl network for effectively modeling the news artical, retwitter and users. The work mentioned above ignores the local information and only takes into account the propagation graph's global information. However, a propagation graph's local information commonly involves rich semantic information. Yuan et al. (2019) constructed a heterogeneous graph using source tweets, retweets, and users. They then proposed a global-local attention network (GLAN) that can concurrently encode local semantic and global structural information. They first used an attention mechanism to fuse news-related retweets with the original news to obtain a local news representation, and then model the global relations between all source tweets, retweets, and users to obtain a global news representation.

In a nutshell, algorithms based on social context exploit user profile information and propagation information to detect fake news and have achieved good results when the information is sufficient. However, it is not easy to obtain users' personal information out of the concern for privacy protection. Furthermore, obtaining complete propagation information is infeasible in the early stage of news dissemination. So from the angle of early detection, social context-based methods are not conducive enough to be applicable in the real world.

4.3. External Knowledge-based Methods

Most above methods rely heavily on linguistic and semantic features from the news content or social context features. However, they fail to effectively exploit external knowledge, which could help determine whether the news document is trusted. News content is highly condensed and comprised of many entity mentions. Due to issues with aliases, abbreviations, and alternative spellings, it is not always possible to understand news text content directly. Thus, several research studies introduce external knowledge to help improve the performance of fake news detection, whose effectiveness has also been analyzed in (Ahmed et al., 2019). Table 4 compares different external knowledge-based methods, which we will elaborate on in subsequent subsections.

A typical source of prior knowledge is knowledge graphs (KGs), which describe entities and their relationships as a graph. Particularly, KGs include information obtained from numerous fields. KGs define the categories and relationships between entities (Paulheim, 2017).

The critical issue of external knowledge-based methods is: how to obtain the background knowledge of the news content; and fuse the background information and knowledge concepts.

Table 4
Comparisons of different external knowledge-based methods.

Method	Form of Knowledge	Knowledge Fusion	Knowledge Compare
MKEMN(Zhang et al., 2019)	Knowledge Graph	Attention	–
KAN(Dun et al., 2021)	Knowledge Graph	Knowledge Aware Attention	–
KMGCN(Wang et al., 2020a)	Knowledge Graph	GCN	–
KAMAGCN(Qian et al., 2021a)	Knowledge Graph	GAT	–
LOSIRD(Li et al., 2021a)	Wikipedia Information Corpus	GCN	–
CompareNet(Hu et al., 2021)	Knowledge Graph+Entity Descriptions	–	Entity Comparison Network

Knowledge Retrieval. The process of knowledge retrieval is to find a related concept set C, given a post text. It is composed of two steps: entity linking and entity conceptualization. Formally, given a post p_i, we first retrieve a set of concepts C_t for each entity t in p_i from the knowledge graph. We can use the entity linking methods like Rel-Norm (Le and Titov, 2018), Link Detector (Milne and Witten, 2008) and EDEL (Kolitsas et al., 2018) to link the entity mentions M in the text to the corresponding entities T in the knowledge graph. For entity conceptualization, we utilize the current knowledge graph, such as YAGO (Suchanek et al., 2008) and Probase (Wu et al., 2012) for retrieval.

Knowledge Fusion. After getting the related knowledge concepts in the Knowledge Graph, models need to fuse the information of news text and background knowledge concepts to obtain the representation of each post and detect fake news. The existing methods to fuse the information of news text and external knowledge can be broadly divided into two mainstreams.

Attention-based approaches fuse the information of text and external knowledge concepts by attention mechanism. Zhang et al. (2019) proposed a multi-modal knowledge perception network and event storage network as the basis of detection. Among them, the perception network fused the information of external knowledge from the knowledge graph to supplement the semantic representation of news content by attention; the storage network extracted the invariant features of the event and stored them in the global memory for subsequent retrieval. Similarly, Dun et al. (2021) strove to integrate knowledge of entities and entity contexts from the knowledge graph for detect fake news. Firstly, they identified entities mentioned in news content and aligned them with those in the knowledge graph. Then, the entities and their contexts were used as an external knowledge to provide supplementary information. Finally, they designed a Knowledge-aware attention mechanism to measure the importance of knowledge. Particularly, there are two components to the knowledge-aware attention mechanisms. First, they calculated the semantic similarity between news contents and their related entities using News towards Entities ($N-E$) attention, where each entity is given a weight to signify its importance. Second, they created News towards Entities and Entity contexts ($N-E2C$) attention with the goal of integrating entity contexts in order to provide the entity context weight based on the associated entity.

Graph-based approaches use graph topology structure for information fusion. Specifically, these methods construct external knowledge entities and text entities into heterogeneous graphs for every news and use graph neural network methods for information fusion. This way, the fake news detection problem is converted into graph classification to identify fake news.

Wang et al. (2020a) proposed a knowledge-driven multimodal graph convolutional neural network, which combined entity linking technology and target recognition technology to model text contend, knowledge concepts, and images into a unified framework. Qian et al. (2021a) proposed a Knowledge-aware Multi-modal Adaptive Graph Convolutional Network (KMAGCN) which models posts as graphs to obtain the long-range semantic representations, and fused the texts, knowledge concepts, and images into a unified framework. Wu et al. (2021a) constructed a credential-based multi-relation knowledge graph from existing fake news corpora and designed a new framework to enhance semantic embeddings with structured knowledge to predict the trustworthiness of a news article. Li et al. (2021a) proposed a model called LOSIRD, which contains two functions: evidence providing function and rumor identifying function. First, LOSIRD mined appropriate evidence sentences and classified them by automatically checking the veracity of the relationship of the given claim and its evidence from about 5 million Wikipedia documents. LOSIRD constructed a heterogeneous graph to describe the relationship between the claim and the evidence and used Graphsage (Hamilton et al., 2017) to aggregate the information from the external knowledge evidence to detect fake news.

The above knowledge fusion-based approaches all incorporate external knowledge into the news, thus making the model have a better news understanding ability. However, the above approaches ignore the fact that the content in fake news often contradicts external knowledge. Hu et al. (2021) proposed an end-to-end graph neural model called CompareNet, which compared the news to the knowledge base (KB) through entities for fake news detection. Specifically, they first constructed a directed heterogeneous document graph for each news incorporating topics and entities, then developed a heterogeneous graph attention network for learning the topic-enriched news representation and contextual entity representations that encoded the semantics of the news content. Finally, the contextual entity representations

were compared to the corresponding KB-based entity representations through a carefully designed entity comparison network to capture the consistency between the news content and KB. It is worth noting that this method was based on the assumption that the content of real news should be consistent with objective knowledge, while the content of fake news may be contrary to objective knowledge. Afterward, fake news can be debunked by comparing the entity representation in the knowledge graph with the representation of mention in the fake news.

The existing authoritative knowledge bases(e.g., Freebase, Wikidata, DBpedia, Google's Knowledge Graph) contain millions of entities and statements, facilitating the detecting approaches to check the fact of the news by knowledge retrieval, fusion, and completion. Thanks to the reliable extra information source, external knowledge-based methods can detect the truthfulness of the news and provide interpretability to some extent.

4.4. Hybrid Methods

The news content, social context information, and external knowledge involved in the news are helpful for the detection of fake news. Therefore, many studies have tried to use the hybrid method to achieve better results.

The users' interactive information, such as their likes and comments on news articles, conveys their opinions regarding the event. Several studies have attempted to detect fake news by combining news content and user-interactive information. Della Vedova et al. (2018) fused news content with user like behaviour to detect fake news. The study is based on an intuitive assumption that news that receives a higher number of likes tends to have a higher level of veracity. Volkova and Jang (2018) proposed to extract audience attitudes and mental states from news comments and combine them with news content to identify fake news. Similarly, Shu et al. (2019a) utilized the co-attention network to combine the content information of news with user comment information for fake news detection. They also sought the top-k importance of news content sentences and user comments to provide interpretability for fake news detection.

The credibility of news authors also plays a significant role in the detection of fake news. Some studies have attempted to mine user credibility using user profiles and social relationships. Additionally, they combine information from user credibility and news content to detect fake news. Ruchansky et al. (2017) utilized the user profiles of the news author to determine their reliability. Additionally, they combined it with news text information to identify fake news. Guo et al. (2018) used an attention mechanism to fuse news author's profile, news content and the users' social relationships information for fake news detection. Shu et al. (2019b) modeled the publication relationship between news publishers and news, the distribution relationship between news and users, and the social relationship between users to obtain news representations integrating publisher, content, and audience for detecting fake news. Lu and Li (2020) combined news content, user information, and comments to detect fake news using the co-attention method. Dou et al. (2021) analyzed the influence of user preferences on the detection of fake news and presented a method called User Preference-aware Fake Detection (UPFD). UPFD mined the past posts of news publishers to obtain a representation of their preferences and combined the intrinsic preferences of news publishers with the external propagation of news in order to detect fake news.

External knowledge involved in the news can assist readers in comprehending the news. Some studies have attempted to detect fake news by using external knowledge about news, news content, and social context. Wang et al. (2020a) presented a multimodal graph convolutional neural network based on external knowledge for detecting fake news by fusing textual information of news, visual information, and external knowledge information. Sun et al. (2022) not only evaluated the external knowledge associated with the news content, but also the external knowledge associated with the news commentary. They designed a Dual Dynamic Graph Convolutional Network for Social Media Rumor Detection (DDGCN). DDGCN employs the external knowledge involved in news and comments to improve the comprehension of content and comments, and it models the news propagation process using dynamic graph neural networks, which has produced positive results.

5. Weakly/Un-supervised Methods

Supervised methods have shown promising results. However, they suffer from one conspicuous limitation: requiring a reliably laebeld dataset to train the model, which is often complex, time-consuming, costly to procure, or unavailable due to privacy or data access constraints. Even worse, this limitation is exacerbated under the FND setting because of the dynamic nature of news, for annotated news may soon become outdated and cannot represent the news articles on events that newly emerge. Thus, some researchers explore the weakly supervised or unsupervised methods to detect fake news.

5.1. Weakly Supervised Methods

Weakly supervised learning (Zhou, 2018) is a promising solution to employ deep learning models for fake news detection, which learns from experience containing only weak supervision (such as incomplete, inexact, inaccurate or noisy supervised information). Traditionally, according to whether the oracle or human intervention (subject matter expert) is leveraged, this can be further classified into:

Semi-supervised learning. It refers to learning from a small number of labeled samples and (usually a large number of) unlabeled samples.

Active learning. It selects informative unlabeled data to query an oracle for output y.

To the best of our knowledge, most extant work in FND choose the semi-supervised way. According to the source of social media data to derive weak supervision, we classify the semi-supervised way as weak content supervision and weak social supervision.

5.1.1. Weak Content Supervision

Weak content supervision methods use partially labeled news text data as input, leveraging the linguistic information of labeled news articles, simultaneously exploring the hidden patterns in unlabelled data. Weak content supervision methods can be roughly divided into Graph-based methods and Pseudo label-based methods.

Graph-based methods use the similarity between news texts to model the labeled news and the unlabeled news into a single graph. The underlying constraint is: nodes in the graph that are close to each other often have the same label. They converted the semi-supervised fake news detection problem into a node classification problem, and the problems of scarce labeled data are addressed through propagating known labels on a graph to determine unknown labels. Guacho et al. (2018) leveraged tensor-based article embeddings, which produce concise representations of articles concerning their spatial context. Similarly, Benamira et al. (2019) proposed to use graph neural networks to classify fake news. They employed word embeddings to construct hidden representations of news items in a lower dimensional space, and then a graph-based representation method may capture the contextual relationships and similarities across articles, and finally, missing labels can be inferred by graph learning techniques. Hu et al. (2019) proposed to learn the news node representations via graph embedding and used multi-depth GCN blocks to capture multi-scale neighbor information combined by attention mechanism. Meel and Vishwakarma (2021a) came up with a semi-supervised fake news detection technique based on graph convolutional network. The recommended architecture comprises three essential components: collecting word embeddings from the news articles in datasets utilizing GloVe, building similarity graph using Word Mover's Distance (WMD) and applying GCN for binary classification of news articles in a semi-supervised paradigm.

Pseudo label-based methods utilize the labeled data to train a supervised model, then apply the supervised model to label the unlabeled data with pseudo labels, and finally use the actual labels and pseudo labels to train the unsupervised model or retrain the supervised model. Dong et al. (2019) proposed a model called two-path-convolutional neural networks, in which one path adopts supervised learning, and the other path adopts unsupervised learning for FND. Specifically, the supervised method needs labeled data to train by cross-entropy loss. The trained supervised model provides pseudo-labels for unlabeled data, then the unsupervised path is trained using the mean squared error loss between the pseudo-label from the supervised path and the predicted result from the unsupervised path. Then, the weighted sum of the two losses is used as the total loss to train the overall network architecture jointly. Victor (2020) proposed an SSDL pipeline, which used attention RNN models to detect fake news in semi-supervised setting. It employed two paths to obtain supervised loss (i.e., cross-entropy) and unsupervised loss (i.e., mean squared error) respectively. Then these two losses are jointly optimized for model training. Mansouri et al. (2020) proposed a method based on a semi-supervised learning framework, targeting both labeled and unlabeled data. In this method, various features of text and image data were first extracted using CNN. Then, Linear Discrimination Analysis (LDA) was used to predict the classes of unclassified data. Inspired by the semi-supervised learning method, temporal ensembling, Laine and Aila (2017); Meel and Vishwakarma (2021b) proposed a convolutional neural network semi-supervised framework built on the self-ensembling concept to take leverage of the linguistic information of annotated news articles, at the same time explore the hidden patterns in unlabelled data. That modal used previous epochs outputs of network-in-training to provide pseudo-labels for unlabeled data. The uniqueness of the framework is that it ensembled all the outputs of previous training epochs of the neural network and used them as an unsupervised target for comparing them with the current output prediction of unlabelled articles. Li et al. (2021c) proposed a self-learning mechanism to perform semi-supervised fake news detection. They proposed a confidence network layer for evaluating the confidence of pseudo labels and retrained the model using pseudo labels with high confidence and the original label data.

The above methods all use a trained model to provide pseudo-labels for unlabeled data, and use the actual labels and pseudo-labels to train the other model or retrain the model. However, they ignore that data can be classified from different views, so different classifiers can be trained from different views. Qiao et al. (2018) proposed the framework include two co-training networks to classify images from different views and achieve good performance. Inspired by the co-training model, Helwe et al. (2019) proposed deep co-learning, which is a semi-supervised deep learning method to assess the credibility of Arabic blogs. In this approach, a small labeled dataset was used to train multiple weak deep neural network classifiers, and then these classifiers which are based on different view of data was used to classify unlabeled data. The predictions of each classifier was used to train the other classifiers in the semi-supervised method.

5.1.2. Weak Social Supervision

Compared with news content, social context information has unique properties that make it suitable for deriving weak supervision. On the one hand, information carried by news content is limited, while the quantity of information produced by users (e.g., comments, behaviors) is not proscribed, so it provides abundant sources to get useful features. On the other hand, when there is no explicit labeled data for models to optimize, social context information (e.g.users' attitudes towards news as well as their credibility) becomes an essential auxiliary reference to judge the veracity of news claims (Shu et al., 2020b).

Users' personal information and characteristics could be captured for FND, such as the age of registration, the number of followers/fans, and the number of tweets published by users. The injected constraint of weak supervision is that the

authors of fake and true news may create distinct groups with distinct traits that user profile signals might illustrate.

Posts from users' comments or positions are also essential in inferring news authenticity. The underlying injected constraint is that the credibility of a news is highly correlated to the credibility of previous related news published by that news publisher.

Network. There are various types of networks (such as friendship networks, dissemination networks, and interaction networks.), and each reflects certain characteristics that distinguish fake news from the true one (Shu and Liu, 2019). For example, news dissemination involves the real-time participation of a large number of users on social media. Fake news might burgeon quickly, whereas real news shows a steady trend (Shu and Liu, 2019). Interaction networks show the connections between various entities, such as publishers, news articles, and readers, which could be used to extract network features of relevant entities and predict news credibility according to their association.

There are two manners to utilize weak social supervision. **Generative,** it means to generate weak labels, then directly learn with weak labels. **Constrained,** it means to represent weak supervision based on social media data as constraints.

The following work (Shu et al., 2019b) is a typical work to leverage social context information as constraints. The rules to derive constraints are as follows: Firstly, related users are more likely to share similar news articles. Secondly, politically biased publishers are more probably to create fake stories. Thirdly, users with low trustworthiness are more likely to propagate fake news. In this way, the label distribution is estimated by injecting constraints into the heterogeneous network embedding framework for learning the news representations: on the one hand, for publishing relationship, the presentation of news must take into consideration the publisher's political leaning; on the other hand, for the spreading relationship, constraining that the news presentation and user representation are close to each other if the news is fake and the user is less-credible, and vice versa.

Besides, some works (Konkobo et al., 2020; Yuan et al., 2020) focus on addressing the problem of early detection based on the framework of weak social supervision. Specifically, Konkobo et al. (2020) first built a model to extract users' opinions expressed in comments, then they used CredRank algorithm to evaluate users' credibility and built a small network of users involved in the spread of given news. Yuan et al. (2020) built a heterogeneous information network composed of news publishers, news content, and news users, and presented an early fake news detection approach Multi-head Structure-aware Attention Network (SMAN). SMAN treats the credibility of news creators and readers as weakly supervised signals. They used user credibility as well as news text information for early fake news detection.

5.2. Unsupervised Methods

In search of an alternative to supervised methods, more scholars are considering detecting fake news in an unsupervised manner.

5.2.1. Anomaly Detection-based Methods

According to the principles of social psychology and social communication dynamics (Kwon et al., 2013), the user behaviors of posting rumors will diverge from those of posting genuine facts. In order to exploit such differences to help detect rumors, some researchers treat rumor posts as anomalies in social networks, and **regard fake news detection as anomaly detection.** This method uses unsupervised anomaly detection, which collects regular posts from users' posting history and projects them into the deep embedding space using unsupervised coding methods. In the testing phase, the models calculate the distance between the data in the test set and the regular posts, and select the outlier with a more considerable distance as the rumor. Chen et al. (2016) viewed it as an anomaly detection task. They built a user behavior model based on the recent microblogs posted by the user within a short period. Specifically, they extracted the features of both rumors and non-rumors posted by the same user within a time window and used a PCA-like method to preserve significant features of these posts and ranked them according to their deviation degree. Zhang et al. (2017) modeled users' normal behaviors through their recent post sets and viewed rumors as anomalies. A multi-layer structured autoencoder-based model was proposed to detect rumors on open social networks automatically. Chen et al. (2018b) proposed a combination of RNN and variant AE to learn the typical behaviors of individual users. The errors between outputs and inputs of the model are used to describe the deviation degree of posts and are compared with the self-adapting thresholds to determine whether it is a rumor. Chen et al. (2018b) judged the truth of the news according to the posting behaviors of users. The proposed model leveraged AE to learn the hidden representations of the post and comments of news, and can be used to measure the authenticity of news when its reconstruction error converges. If the model's reconstruction error reaches a particular level, the post will be deemed fake.

The above work has achieved good performance in unsupervised scenarios. However, all of them need to collect the user's previous behavior data in advance, which is not easy to achieve.

5.2.2. Probabilistic Graphical Model-based Methods

The probabilistic graph-based approach treats the problem of fake news detection as a probabilistic problem. Yang et al. (2019) treated the truthfulness of news and the credibility of users as hidden variables. It considers that the truthfulness of news is more related to the credibility of users. They first mined the information of user comments so as to obtain the users' views on the news. After that, they built a Bayesian probabilistic graphical model to capture the complete generative process of news truth and users'

views. Meanwhile, they propose a collapsed Gibbs sampling approach to solve the inference problem.

5.2.3. Graph-based Methods

Unsupervised graph-based approaches for detecting fake news exploit the property that close nodes in a network tend to have similar labels. (Gangireddy et al., 2020) presented an unsupervised graph-based approach for detecting fake news called GTUT. Using news as nodes and similarities between news as edges, they developed a graph of news. It begins with the identification of a seed set of fake and true articles by utilizing high-level observations on inter-user behavior in fake news propagation. GTUT then utilizes the similarity between news nodes and extends the labels to all news in the graph. Utilizing textual information from news articles and social background information, GTUT achieved success. GTUT is ineffective in situations where social context is not always present.

5.2.4. Generative Adversarial Learning-based Methods

The generative adversarial learning (GAL) (Chen et al., 2019a) was suitable for achieving adaptive learning for unsupervised scenarios. The GAL contained two components: a generator and a discriminator (Qiu et al., 2019). The former generated unknown samples and the latter discriminated whether the generated samples were close to real ones. The adversarial training between them was expected to produce optimal outputs. Guo et al. (2021) proposed a graph embedding-based generative adversarial network (Graph-GAN) model. First of all, it constructed fine-grained feature spaces via graph-level encoding. Furthermore, it introduced continuous adversarial training between a generator and a discriminator for unsupervised decoding, which can actively learn the rules of feature spaces. The two-stage scheme not only solved fuzzy rumor detection under unsupervised scenarios but also improved the robustness of the unsupervised training.

5.2.5. Transfer Learning-based Methods

The goal of transfer learning is to obtain generalizable knowledge by using the richly labeled source domain data and transfer this generalizable knowledge to the target task, thereby helping to train the target task. In fake news detection tasks, English documents often have many annotated data, but other minor languages have less annotated data. The model is trained on the English dataset to obtain generalized prior knowledge and transferred to the task of detecting fake news in other languages. Du et al. (2021) proposed a deep learning framework named CrossFake which trained BERT on labeled English news datasets and could detect most of the unlabeled Chinese fake news after translation. Tian et al. (2021) proposed a zero-shot cross-lingual transfer learning framework to construct a fake news detection system that does not require any annotated data for a new language. This system is cross-lingual because it can detect rumors in two languages based on only one model. It first fine-tuned a multi-lingual pre-trained language model (e.g., multi-lingual BERT) for fake news detection using annotated data of a source language (e.g., English), and then used this model to classify fake news in another target language (zero-shot prediction) to create "silver" rumor labels of the target language. These silver labels are used to fine-tune the multi-lingual model further to adapt it to the target language. It is worthwhile noting that under the unsupervised setting where ground-truth labels are unavailable, methods attend more to users and their social media engagements. This way, models can capture some implicit information as referential judgment hints.

6. Experiment

In this section, we first introduce some relevant datasets. Then we pick typical methods and give descriptions of them. Finally, we give analysises of these methods.

6.1. Dataset

We summarize representative datasets in the field of fake news as follows.

BuzzFeedNews (Silverman et al., 2016) consists of complete news samples released on Facebook by 9 news agencies from September 19 to 23 and September 26 and 27, over a week before the 2016 U.S. election. Each post and its linked articles were verified by five BuzzFeed journalists one by one. It contains 1627 articles, including 826 mainstream articles, 356 left-wing articles and 545 right-wing articles.

BuzzFace (Santia and Williams, 2018) is organized by adding news-related Facebook comments to the BuzzFeed dataset. The dataset consists of 2,263 news and 1.6 million comments.

LIAR (Wang, 2017) comprises 12,836 real-world news gathered from PolitiFact. Each news is labeled with six-grade truthfulness: true, false, half-true, part-true, barely-true, and mostly-true. It also includes the information about subjects, party, context, and speakers.

CREDBANK (Mitra and Gilbert, 2015) is a large crowd-sourced data set comprised of sixty million tweets over 96 days starting from October 2015. The tweets pertain to almost a thousand news events. Each event's authenticity is rated by 30 Amazon Mechanical Turk annotators.

FacebookHoax (Tacchini et al., 2017) contains information about the posts from Facebook pages associated with scientific news (non-hoax) and conspiracy pages (hoax), gathered using the Facebook API. The data collection includes 15,500 postings from 32 pages (14 conspiracy and 18 scientific) that have received over 2,300,000 likes.

Twitter15 and Twitter16 (Ma et al., 2017) respectively contains 1,381 and 1,181 propagation trees. In each data set, the tree structure contains a collection of source tweets, as well as their propagation threads, such as responses and retweets. Each tree is categorized as non-rumor, false rumor, real rumor, or unverified rumor. Each event label in Twitter15 and Twitter16 is tagged with the article's veracity tag from rumor busting websites (e.g., snopes.com, Emergent.info).

Ma-Twitter (Ma et al., 2016) is collected on Twitter [4]. In the Ma-Twitter dataset, 498 rumors are collected from Snopes[5], a real-time rumor debunking website. It also contains 494 normal events from Snopes and two public datasets. Each post in the dataset contains Twitter content, replies and comments, and the publisher's profile.

Media-Twitter (Boididou et al., 2014) has two parts: the development set which contains about 9,000 fake tweets and 6,000 real tweets from 17 events, and the test set which contains about 2,000 tweets from another 35 events. Therefore, the tweets in two sets cover different events. The posts in the Media-Twitter dataset contain tweets with text, image/video attachments, and additional social contextual information.

Ma-Weibo (Ma et al., 2016) was collected on the Chinese social media site Sina Weibo[6]. The known rumors are collected from the Sina community management center [7] reporting a variety of misinformation. The Weibo API can capture the original messages and all their repost/reply messages given an event. The real news is gathered by collecting posts from regular threads that are not labeled as rumors.

Media-Weibo (Jin et al., 2017) is first presented for the multi-modal fake news detection task. Each post includes text content, user profile and additional images these three parts. The fake news having been verified is collected from the official fake news exposing system of Sina Weibo which is similar to Twitter. The period of the data is from May 2012 to January 2016. Jin et al. The true news is obtained from Xinhua News Agency, a reputable Chinese news agency. The low-quality and duplicated images are removed according to the data pre-processing methods in previous research. Each post in the dataset contains a tweet id, text, and image.

Weibo-20 (Zhang et al., 2021b) is a Chinese fake news detection dataset which is an expanded version of Media-Weibo(Jin et al., 2017). It keeps the two-class setting (i.e., fake or real for each news piece). For fake news, It retains the 1,355 fake news pieces on Media-Weibo and collects news pieces varified as fake officially by Weibo Community Management Center from April 2014 to Nov 2018. For real news, it retains the 2,351 real news pieces of Media-Weibo and gathers 850 unique new pieces in the same period as the fake news. The newly-collected real news pieces are real news verified by NewsVerify, which focuses on discovering and verifying suspicious news pieces on Weibo. Weibo-20 contains 3,161 fake news pieces and 3,201 real news pieces.

Weibo21 (Nan et al., 2021) is a multi-domain fake news dataset in Chinese with domain label annotated. Both fake and real news are collected from Sina Weibo from December 2014 to March 2021. In terms of fake data, Weibo21 collects news pieces that are officially judged as misinformation by Weibo Community Management Center. For real data, Weibo21 gathers real news pieces in the same period as the fake news, which has been verified by NewsVerify[8] (a platform that discovers and verifies suspicious news pieces on Weibo). In the dataset, each news piece contains news content, the image of the news, timestamp, users' comments, and domain label. Weibo21 involves news belonging to the following 9 domains: Science, Military, Education, Disasters, Politics, Health, Finance, Entertainment, and Society.

MCG-FNeWS (Cao et al., 2019) is a multimodal fake news detection dataset. Each news item in the dataset includes both text and an accompanying image. The dataset includes Sina Weibo news from May 2012 to November 2018.

NewsBag (Jindal et al., 2020) is a multimodal dataset for detecting fake news. Its training set includes 200,000 real news items and 15,000 fake news items, with the real news stories originating from The Wall Street Journal [9] and the fake news stories from The Onion [10]. The test set includes 11,000 real news pieces and 18,000 fake news stories, with the real news sourced from The Real News[11] and the fake news sourced from The Poke[12].

Ti-CNN (Yang et al., 2018) is a multimodal dataset for detecting fake news. The dataset contains a total of 20,015 news items, of which 11,941 are fake and 8,074 are true. Each news article in the dataset includes a title, text, image, and author information. The fake news items are from Kaggle[13], whereas the real news items are from the New York Times[14] and the Washington Post[15].

PolitiFact[16] is a famous non-profit fact-checking website in the United States that provides political statements and reports. News in PolitiFact data set was published from May 2002 to July 2018. Domain experts offer ground truth labels (false or real) for news items in the dataset. The news after fact-checking from PolitiFact mainly are the statements or news articles posted by the politicians (Congress members, White House staff, lobbyists) and political groups. For these news articles, PolitiFact will provide the original contents, fact-checking results, and comprehensive fact-checking reports on the website. The platform will group them into different subjects based on content and topic. The subject has a brief description. The fact-checking results can demonstrate the credibility of corresponding news articles True (2149), Mostly True (2676), Half True (2765), Mostly False (2539), False (2601), Pants on Fire (1322). We take the labels Pants on fire, False, Mostly False as fake news and take True, Mostly True, Half True as real news. Thus we can get 6465 fake news and 7590 real news. The fact-checking results will be taken as the ground truth in experiments.

[4] http://www.twitter.com
[5] http://www.snopes.com
[6] http://www.weibo.com
[7] http://service.account.weibo.com
[8] https://www.newsverify.com/
[9] www.wsj.com
[10] www.theonion.com
[11] www.therealnews.com
[12] www.thepoke.co.uk
[13] www.kaggle.com/mrisdal/fake-news
[14] www.nytimes.com
[15] www.washingtonpost.com
[16] https://www.politifact.com/

GossipCop[17] is a fact checking website. News in GossipCop dataset was published from July 2000 to December 2018. Domain experts give the ground truth labels for news items in the dataset, ensuring the quality of news tags.

FakeNewsNet (Shu et al., 2020a) contains news from the fact-checking websites BuzzFeed[18] and PolitiFact. The dataset contains news content, user information, and retweets. The dataset contains a total of 23,196 news articles and 69,733 retweets.

PHEME (Zubiaga et al., 2017) is constituted by tweets from the Twitter platform. In addition, it collected from five breaking news sources, each of which had a collection of tweets. Each piece of tweet contains texts as well as images.

WeChat (Wang et al., 2020b) is a semi-supervised fake news detection dataset. The database contains both news articles and user comments. The dataset comprises news from the social media platform WeChat[19] between March 2018 and October 2018. Only a small portion of the dataset's news items are labeled by the experts of the WeChat team, while the most are left unlabeled. However, whether or not the news is labeled, they all include comment information.

Fakeddit (Nakamura et al., 2020) is sourced from multiple subreddits of the Reddit platform. This dataset gathers rich information, including textual sentences, images, and social context information. In addition, the Fakeddit dataset provides 2-way, 3-way, and 6-way labels for each sample, which enables researchers to develop a more fine-grained fake news detection model. The 2-way categorization establishes that whether the news is real or false. The 3-way categorization evaluates if the news is entirely real, fake and contains true text, or fake and contains fake content. Instead of a basic binary or trinary classification, the 6-way classification was developed to distinguish different types of fake news. The six categorization labels are described in detail below: True, Satire/Parody, Deceptive Content, Fake Content, False Connection, Manipulated Content.

FakeHealth (Dai et al., 2020) is a dataset for detecting health-related fake news. The data in the dataset originate from the website of Health News Review[20]. The dataset contains news content, users' comment , and social network for people.

CoAID (Cui and Lee, 2020) is a dataset for detecting fake news connected to COVID-19. The dataset includes news articles, user comments, and user data. The dataset includes 4,251 news articles, 296,000 user comments, and actual news labels.

FakeCovid (Shahi and Nandini, 2020) is a multi-lingual cross-domain dataset collecting 5182 articles circulated in 105 countries from 92 fact-checkers. The articles are manually annotated into 11 categories of fact-checked news. 40.8% articles are in English.

ReCOVery (Zhou et al., 2020a) is a COVID-19 correlated multimodal fake news detection dataset collected from NewsGaurd[21] website. The dataset contains news-related images, textual content, and social context. This dataset contains 2,029 news articles and 1,40820 tweets.

MM-COVID (Li et al., 2020) is a COVID-19 related multilingual fake news detection dataset. The dataset contains news content, social posts, and spatial information about the news. This dataset contains news from six distinct languages: English, Spanish, Portuguese, Hindi, French, and Italian. The dataset contains 3,981 fake news items and 7,192 real news items.

CHECKED (Yang et al., 2021) is a Chinese COVID-19 related fake news detection dataset. The dataset consists of 2104 news articles from the social media platform Sina Weibo. The dataset contains textual news content, images, propagation information, and lables.

Cross-lingual COVID-19 (Du et al., 2021) is a cross-lingual COVID-19 dataset which contains both English and Chinese news about COVID-19. The dataset is divided into two parts: the training and test sets. The training set contains 2840 reports in English, and 49.43% of the news is fake news. The test set contains 200 Chinese news, 43% of which are fake news.

SLN (Rubin et al., 2016) contains 360 news stories, which collectively include the topics covered by national newspapers in the United States and Canada. The dataset was constructed in two separate sets. The first set was collected from The Onion and The Beaverton[22], two satirical news websites, and The Toronto Star[23] and The New York Times, two legitimate news organizations. The second set, also divided between satirical and legitimate news articles, was drawn from 6 legitimate and 6 satirical North American online news sources.

LUN (Rashkin et al., 2017) is a multi-category fake news detection dataset. The dataset contains news from PolitiFact and its sister sites (PunditFact, etc.). News in the dataset can be classified into six categories: True, Mostly True, Half True, Mostly False, False, and Pants-on-Fire. Additionally, there is a variant with two classes that considers the top three true ratings to be true and the bottom three to be false.

GermanFakeNC (Vogel and Jiang, 2019) is a German-language fake news detection dataset. The dataset consists of 490 news articles. The ground truth labels were retrieved from German well-known news media websites. This dataset only contains text information for news, without social information.

We summarize the public datasets that have been used, as shown in Table 5, whose data is collected from Sina Weibo, Twitter, and other social platforms, as well as fact-checking sites (e.g., Emergent, BuzzFeedWeb, LIAR, FakeNewsNet).

6.2. Methods Overview

In this section, we briefly review the representative methods again.

[17] https://www.gossipcop.com/
[18] www.buzzfeed.com
[19] https://www.wechat.com/
[20] HealthNewsReview.org
[21] www.newsguardtech.com
[22] www.beavertonoregon.gov
[23] www.thestar.com

Table 5
Summary of datasets of fake news detection.

Dataset	News Content				Social Context		Spatiotemporal Information		Labels
	Multi-lingual	Text	Visual	User Profile	Repost&Response	Network	Spatial	Temporal	
BuzzFeedNews		√							4
LIAR		√							6
CREDBANK		√		√			√	√	5
BUzzFace		√			√			√	4
FacebookHoax		√		√	√				2
FakeNewsNet		√	√	√	√	√	√	√	2
FakeCovid	√	√					√	√	2
ReCOVery		√	√	√	√	√	√	√	2
CoAID		√	√		√			√	2
MM-COVID	√	√	√	√	√	√	√	√	2
Cross-lingual COVID-19	√	√		√				√	2
CHECKED		√	√	√	√		√	√	2
FakeHealth		√	√	√	√	√	√	√	2
Fakeddit		√	√		√		√	√	2,3,6
PHEME	√	√	√	√	√		√	√	2
WeChat		√			√				2
Ma-Twitter		√		√		√	√	√	2
Media-Twitter		√	√	√	√	√	√	√	2
Twitter15 Twitter16		√		√	√			√	4
Ma-Weibo		√		√		√	√	√	2
Media-Weibo		√	√	√	√	√	√	√	2
Weibo-20		√	√	√	√	√	√	√	2
Weibo-21		√	√		√			√	2
GermanFakeNC		√							2
TI-CNN		√	√	√					2
NewsBag		√	√						2
MCG-FNeWS		√	√						2
SLN		√							2
LUN		√							2,6

6.2.1. News Content-based Methods

The following are some single modal methods.

SVM-TS (Ma et al., 2015) is a method that employs heuristic rules and a linear SVM classifier to detect fake news.

GRU-2 (Ma et al., 2016) uses 2-layer GRU to learn the hidden representations of the news and a classification head to detect fake news.

CNN-text (Yu et al., 2017) is a convolutional neural network to learn the news feature for fake news identification. The model uses CNN to extract the textual information of the news, and then uses a fully-connected layer for classification.

HAN (Yang et al., 2016) is a hierarchical attention neural network framework for detecting fake news based on news content. It encodes news content using word-level attention of each sentence and sentence level-attention of each artical.

GCN-text (Vaibhav and Hovy, 2019) is an approach for detecting fake news based on graph neural networks. He models the relationships between sentences in the news using graph structures and transforms the task of detecting fake news into a graph classification problem.

BERT-text (Qi et al., 2021) utilizes a pre-trained BERT to acquire the representation of the news article, and uses a fully connected layer for classification.

GAN-GRU (Ma et al., 2019) is an adversarial training-based approach. The generator is used to produce uncertain or conflicting voices and complicates the original conversational, which threads to pressurize the discriminator to learn stronger rumor indicative representations from the augmented, more challenging examples.

VRoC (Cheng et al., 2020) is a tweet-level text-based novel rumor classification system based on variational autoencoders. VRoC realizes all four tasks include rumor detect, rumor track, stance classification, and veracity classification in the rumor classification system.

VGG19-visual (Khattar et al., 2019) is an approach for identifying fake images. It utilizes a pre-trained VGG19 model to extract image information for classification.

The following are some multi-modal methods.

att_RNN-S att_RNN (Jin et al., 2017) is an attention RNN-based multimodal fake news detection approach. It employs an LSTM model to extract text and social context information, and then employs an attention mechanism to fuse image information and text information for detecting fake news. In order to make a fair comparison, the authors created a variant called att-RNN- that deleted the part tackling with social context features.

EANN (Wang et al., 2018b) is a multimodal fake news detection method based on adversarial training. It consists of three components: a feature extractor for extracting visual and textual information from the news, an event discriminator for extracting event-irrelevant information from the news, and a classifier for fake news detection.

MVAE (Khattar et al., 2019) is a multimodal method for detecting fake news that is based on an autoencoder. It makes use of a multimodal variational autoencoder to extract

modal shared features and then classes the news based on this information.

MSRD (Liu Jinshuo, 2020) takes into account the text embedded in the image. It employs OCR techniques to extract text from images. It combines the visual information of the image, text information embedded inside the image, and news text information for detecting fake news.

SAFE (Zhou et al., 2020b) evaluates the effect of similarities between news images and text on the veracity of the news. It translates a news image into a sentence and assesses its similarity to the news text as a crucial clue for detecting fake news.

SpotFake (Singhal et al., 2019) identifies fake news by concatenating the text features acquired from a pre-trained BERT model with the image features obtained from a VGG19 model.

SpotFake+ (Singhal et al., 2020) encodes image and text information of news using pre-trained VGG19 and XLNET, then concat them as news representations for classification.

CARMN (Song et al., 2021a) presents a cross-modal attention residual network to fuse multi-modal features.

MCAN (Wu et al., 2021b) utilizes three different sub-networks to extract features from the text, spatial domain, and frequency domain, respectively. Then the three features are deeply fused by stacking co-attention layers inspired by human behavior. When people read news with images, images and text are read once or multiple times and continuously fused in the brain.

HMCAN (Qian et al., 2021b) is a hierarchical multi-modal contextual attention network, which jointly models the multi-modal context information and hierarchical semantics of text in a deep and unified model to detect fake news. In detail, HMCAN employs BERT and ResNet to learn better text and image representations. Then, HMCAN feeds the obtained image and text representations into the multi-modal contextual attention network to blend the inter-modality and intra-modality relations. Finally, HMCAN designs a hierarchical encoding network in order to obtain the rich hierarchical semantics to detect fake news.

EM-FEND (Qi et al., 2021) proposes a novel entity-enhanced multi-modal fusion framework, which simultaneously models different cross-modal correlations to detect diverse multi-modal fake news. It utilizes visual entity information and textual entity information to extract higher-order semantic relationships from news articles.

6.2.2. Social Context-based Methods

ML-GRU (Ma et al., 2016) uses a multi-layer GRU network to model the microblog event as a variable-length time series, which is effective for the early detection of rumors.

CallAtRumor (Chen et al., 2018a) presents an LSTM model to automatically identify rumors. By using the standard attention mechanism at the word level, this method could detect rumors effectively, which is effective for early detection of rumors.

HSA-BLSTM (Guo et al., 2018) is a kind of hierarchical neural network combined with social information. At first, a hierarchical Bi-LSTM model is built TO represent the text information. Then, the social contexts are then added to the network via an attention technique.

MT-ES (Ma et al., 2018a) is a multi-task architecture based on RNN, which is used to capture the shared features from stance prediction and rumor detection these two subtasks.

CMAI Yu et al. (2017) proposed a CNN-based approach for rumor detection. A two-layer convolutional neural network receives the event posts as input to generate a representation of the event. To obtain the classification outcome, the model feeds the event representation into the MLP layer.

PPC (Liu and Wu, 2018) is a model to detect fake news by combining the propagation path classification of recurrent network and convolutional network.

CSI (Ruchansky et al., 2017) employs the LSTM network to extract textual information from news and comments, and combines it with user credibility scores in order to detect fake news.

dEFEND (Shu et al., 2019a) models news content and news comments using a deep co-attention method to detect fake news. dEFEND simultaneously chooses the most significant news sentences and comments for interpretability.

RvNN (Ma et al., 2018b) models the dissemination of news in social networks using a tree structure and employs a recursive neural network to model news propagation trees.

Undirected GCN (Song et al., 2021b): Graph convolutional neural network(GCN) is the most commonly model which can capture the high order neighborhoods information. In this study, the rumor propagation network is viewed as an undirected graph. Specifically, the rumor propagation network is first fed into the GCN to obtain the embedding representation and then fed into the fully connected network to obtain classification results.

Undirected GAT (Song et al., 2021b) : Graph attention networks (GAT) employs the attention mechanism to provide distinct weights to neighboring nodes in order to aggregate neighbor information. This study models the news propagation network with undirected graphs. And it uses GAT to model the distribution network of news in social networks and input the embedding to the fully connected layer to identify fake news.

Bi-GCN (Bian et al., 2020) considers the behavior of rumor propagation and dissemination in social networks. Top-down and bottom-up graph architectures are used to simulate propagation and dissemination behavior, respectively. It leverages not only a GCN considering rumor spreading from top to down, but also a GCN with an oppositely rumor diffusion direction to obtain the structures of rumor dispersion. In addition, each layer of GCN incorporates information from the source post to enhance the effect of rumor sources.

AE-GCN, VAE-GCN (Lin et al., 2020): These two models model the propagation of news in social networks using graph structures. They transform the identification of fake news into a graph classification task. They use graph

autoencoders and variational graph autoencoders, respectively, to encode graph structures. The distinction between AE-GCN and VAE-GCN is whether the distribution limit of the latent variable is based on a Gaussian distribution.

GCAN (Lu and Li, 2020) is an approach for identifying fake news based on user information, comment information, and news content. GCAN consists of five components: (1) user feature extraction, (2) new story encoding, (3) user propagation representation, (4) dual co-attention between source tweet and users' propagation, (5) fake news detection.

GLAN (Yuan et al., 2019) is a social context-based method for detecting fake news that jointly models the local information and global relationships of news propagation. GLAN first aggregates the textual information of news and comments to produce a local representation of news. GLAN then combines news, comments, and users to build a heterogeneous graph of news propagation and encodes this graph to produce a global representation of the news for fake news detection.

PLAN (Khoo et al., 2020) models long-distance interactions between tweets with a transformer network. There exist ts of PLAN: (1) a structure-aware self-attention model (StA-PLAN) that takes into account the tree structure in transformer, and (2) a hierarchical token and post-level attention model (StA-HiTPLAN) that obtains a sentence representation via token-level self-attention.

HDGCN (Kang et al., 2021) uses the connection between multiple news, such as their relevance in time, content, topic, and source, to conduct fake news detection. They construct a heterogeneous graph with multiple types of nodes and edges, to integrate various information of multiple news. A Heterogenous Deep Convolutional Network (HDGCN) is proposed to learn the node representations.

TGNF (Song et al., 2021b) is a dynamic graph-based approach for detecting fake news. It models the temporal and structural information of news propagation in social networks using dynamic graphs. Specifically, TGNF can model temporal evolution patterns of the news as the graph evolving via continuous-time dynamic diffusion networks.

RDLNP (Lao et al., 2021) is a rumor detection method which is based on dual propagation: linear and non-linear propagation. This approach captures the tree-aware representations from non-linear diffusion patterns and pays attention to the sequential property of linear temporal interactions.

TriFN (Shu et al., 2019b) utilizes both user-news interactions and publisher-news relations for learning news feature representations to detect fake news.

UPFD (Dou et al., 2021) has three major components. First, given a news piece, UPFD crawls the historical posts of the users engaged in the news to learn user endogenous preferences. UPFD implicitly extracts the preferences of engaged users by encoding historical posts using text representation learning techniques. The news textual data is encoded using the same approach. Second, to leverage user exogenous context, UPFD builds the news propagation graph according to its engagement information on social media platforms (e.g., retweets on Twitter). Third, UPFD devises a hierarchical information fusion process to fuse the user endogenous preference and exogenous context.

BERT-EMO (Zhang et al., 2021b) proposes the feature set, dual emotion features, to comprehensively represent dual emotion and the relationship between the two kinds of emotions, and exhibit how to plug it into the fake news detectors as a complement and enhancement.

6.2.3. External knowledge-based Methods

B-TransE (Pan et al., 2018) combines positive and negative single models for fake news detection, based on news content and knowledge graphs.

KCNN (Wang et al., 2018a) takes external knowledge of the entities covered into consideration for a better understanding of the news. kCNN first uses entity extraction methods to mine the news's entities. Following this, the entity linking technique is utilized to link the entities to their appropriate nodes in the knowledge network and capture their contextual information. Lastly, CNN is employed to simultaneously mine news text information, entity information, and entity context information.

KAN (Dun et al., 2021) is a fake news detection method based on external knowledge. KAN first links entity mentions in text to external knowledge graphs using knowledge retrieval techniques. KAN then creates a subtle attention technique to combine news content and knowledge information for detecting fake news.

KMGCN (Wang et al., 2020a) models semantic representations by combining text contend, background knowledge, and visual information into a unified framework for fake news identification. KMGCN constructs a graph taking textual information, visual information, and external knowledge as nodes, and employs a well-designed graph convolutional network for learning the graph representations.

KMAGCN (Qian et al., 2021a) models fake news as a graph structure where text, images and knowledge concepts are used as nodes in the graph. KMAGCN designs an end-to-end graph neural network to represent this graph. The standard KMAGCN uses word2vec to capture information about the nodes in the graph.

MKN & MKEMN (Zhang et al., 2019) learns the post representation by taking the word embedding, visual embedding, and knowledge embedding of the post as multiple stacked channels while explicitly keeping their alignment relationships. Event Memory Network (EMN) builds an external shared memory during training to capture event-independent latent topic information (event-invariant features). MKEMN processes event posts to obtain multi-modal knowledge-aware representation and the event-invariant features for rumor detection.

6.2.4. Weakly supervised Methods

TDSL (Dong et al., 2020) is a semi-supervised approach for detecting fake news. TDSL offers supervised and unsupervised CNN paths to mine the textual content of news articles. The supervised path is trained by minimizing the

cross-entropy loss between label and model predictions. The unsupervised path is trained with the mean squared loss of the text representation derived from the two-path CNN.

SSDL (Victor, 2020) is a kind of attention RNN-based model for detecting fake news at semi-supervised setting. The pipeline utilizes two paths to create unsupervised loss (mean squared error) and supervised loss (cross-entropy), separately. The training is then completed by minimizing these two losses.

SSLNews (Konkobo et al., 2020) is a kind of semi-supervised learning model to detect fake news on social media as early as possible. Using semi-supervised learning, SSLNews can deal with the massive amount of unlabeled data on social media. SSLNews first built a model to extract users' opinions expressed in comments, then SSLNews used CredRank Algorithm to evaluate users' credibility and built a small network of users involved in the spread of given news. The outputs of these three steps serve as inputs of the news classifier. SSLNews comprises three networks: a shared CNN, an unsupervised CNN, and a supervised CNN, similar to TDSL.

MWSS (Shu et al., 2020b) makes use of a variety of weak signals from various sources, including user and content interactions (also known as weak social supervision). Deep neural networks are trained in a meta-learning framework using a combined limited quantity of clean data and weak signals from social interactions. This allows it to evaluate the quality of various weak examples.

WeFEND and WeFEND- (Wang et al., 2020b) exploits users' comments as weak supervision to expand training data for fake news detection. WeFEND includes three components: the annotator, the reinforced selector, and the fake news detector. The annotator aims to automatically generate weak labels to unlabeled news according to user posts. The reinforced selector chooses high-quality samples from the weakly labeled data by reinforcement learning. The fake news detector identifies the news. WeFEND- is a variant that removing the data selector to validate the effectiveness of the selector.

SRLF (Yuan et al., 2021) first finetunes the pre-trained BERT on a small labeled dataset and then leverages this model to annotate weak stance labels for users' comment data. Then, Stance-aware Reinforcement Learning Framework (SRLF) chooses high-quality labeled stance data. The stance selection and rumor detection tasks are optimized simultaneously for mutually enhancing both tasks.

SMAN (Yuan et al., 2020) is a structure-aware multi-attention network. It combines news content with the heterogeneous graphs among publishers and users. It jointly optimizes the fake news detection task and user credibility prediction.

AA-HGNN (Ren et al., 2020) utilizes a hierarchical attention mechanism to address the heterogeneity of News-HIN and learns textual and structural information simultaneously. An active learning framework is applied in AA-HGNN to deal with the paucity of labeled data. The HGAT-based selector is trained in an adversarial manner to query high-value candidates for the active learning setting.

6.2.5. Unsupervised Methods

Majority Voting (Yin et al., 2008) estimates each news using the most frequent verified user's viewpoint.

TruthFinder (Yin et al., 2008) is an unsupervised approach for detecting fake news. He utilized conflicting correlations between tweets to judge the veracity of each news item in an iterative way.

UFD (Yang et al., 2019) is an effective collapsed Gibbs sampling method for inferring the veracity of news and the reliability of users in the absence of labeled data. UFD employs a probabilistic graphical model to model the news veracity and user reliability. A collapsed Gibbs sampling strategy is presented as a solution to the inference problem.

GTUT (Gangireddy et al., 2020) is a graph-based approach for unsupervised fake news detection. GTUT is founded on the premise that neighboring nodes in a network usually share a similar label. GTUT first selects a seed set of true and fake news based on social contextual information, then employs a graph neural network technique to transmit information about the graph's nodes, and finally generates labels for the unlabeled data.

6.3. Experimental analysis

Table 6 presents the performance of representative fake news detection methods on datasets introduced in Section 6.1. To ensure correctness, we collect the experimental results from their original papers. To ensure fairness, we only compare the results of experiments with the same dataset and the same data pre-processing conditions. The accuracy scores metric is commonly used by fake news detection methods to report performance. Therefore, we show accuracy scores for all data sets.

In the following section, we compare the detection accuracy of different fake news detection methods under typical data sets and give an analysis.

6.3.1. News Content-based Methods

Single-modal Methods

1. Among all models on most datasets, SVM-TS performs the lowest, showing that hand-crafted features are ineffective in detecting fake news.
2. CNN-text and GRU-2, which are based on deep learning, surpass SVM-TS, which is based on machine learning. This implies that the deep learning technique can successfully identify significant characteristics in news content, but the traditional machine learning strategy is less effective and depends heavily on feature engineering.

Table 6
The performance of representative fake news detection methods on various datasets taken from their original papers.

	Dataset																			
Model	Ma-Weibo	Media-Weibo	Weibo-20	Media-Twitter	Ma-Twitter	PHEME	PolitiFact	GossipCop	Twitter15	Twitter16	FakeNewsNet	BuzzFeed	Wechat	LIAR	Fakeddit	HealthStory	HealthRelease	TI-CNN	MCG-FNeWS	NewsBag
Single-modal News Content-based Methods																				
SVM-TS(CIKM15)	–	0.64	–	0.529	–	0.639	0.597	0.498	–	–	–	–	–	–	–	0.64	0.657	–	–	0.51
GRU-2(IJCAI16)	–	0.702	–	0.634	–	0.742	–	–	–	–	–	–	0.733	72.44	0.839	0.691	0.665	0.864	–	–
CNN-text(IJCAI17)	–	0.74	–	0.549	–	0.779	0.649	0.775	–	–	–	–	0.747	–	0.834	0.742	0.67	–	–	0.5
GCN-text(EMNLP19)	–	0.81	–	0.703	–	0.828	–	–	–	–	–	–	–	–	–	–	–	–	–	–
HAN(NAACL16)	–	–	–	–	–	–	0.837	0.742	–	–	–	–	–	–	–	–	–	–	–	–
GAN-GRU(WWW19)	–	–	–	0.863	–	0.781	–	–	–	–	–	–	–	–	–	–	–	–	–	–
VRoC(WWW20)	–	–	–	–	–	0.876	–	–	–	–	–	–	–	–	–	–	–	–	–	–
BERT-text(MM21)	–	0.83	0.9	–	–	–	0.7104	0.8576	–	–	–	–	–	–	0.886	–	–	0.873	–	–
VGG19-visual(WWW19)	–	0.73	–	0.596	–	–	0.649	0.775	–	–	–	–	–	–	0.7355	–	–	0.758	0.743	–
Multi-modal News Content-based Methods																				
att_RNN-S(MM17)	–	0.772	–	0.664	–	–	–	–	–	–	–	–	–	–	–	–	–	0.899	0.783	–
att_RNN(MM17)	–	0.788	–	0.682	–	0.85	0.769	0.743	–	–	–	–	–	–	–	–	–	–	–	–
EANN-(KDD18)	–	0.782	–	0.648	–	0.681	–	–	–	–	–	–	–	–	–	–	–	–	–	–
EANN(KDD18)	–	0.827	–	0.715	–	–	–	–	–	–	–	–	0.767	–	–	–	–	0.855	0.823	–
MVAE(WWW19)	–	0.824	–	0.745	–	0.852	–	–	–	–	–	–	–	–	–	–	–	0.908	0.876	0.71
MSRD(JCRD20)	–	0.749	–	0.685	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SAFE(KDD20)	–	0.763	–	0.766	–	0.811	0.874	0.838	–	–	–	–	–	–	–	–	–	0.922	0.924	–
SpotFake(BigMM19)	–	0.869	–	0.771	–	0.823	–	–	–	–	–	–	–	–	–	–	–	–	–	–
SpotFake+(AAAI20)	–	0.869	–	0.79	–	0.8	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CARMN(IPM21)	–	0.865	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MCAN(ACL21)	–	0.899	–	0.809	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
HMCAN(SIGIR21)	–	0.885	–	0.897	–	0.881	–	–	–	–	–	–	–	–	–	–	–	–	–	–
EM-FEND(MM21)	–	0.904	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
External Knowledge-based Methods																				
B-TransE	–	–	–	–	–	0.72	0.7694	0.7394	–	–	–	–	–	–	–	–	–	–	–	–
KCNN(WWW18)	–	–	–	–	–	0.7265	0.7827	0.7491	–	–	–	–	–	–	–	–	–	–	–	–
KAN(AAAI21)	–	–	–	–	–	0.783	0.8586	0.7766	–	–	–	–	–	–	–	–	–	–	–	–
KMGCN(ICMR20)	–	0.8863	–	–	–	0.8756	–	–	–	–	–	–	–	–	–	–	–	–	–	–
KMGCN-NoKD(ICMR20)	–	0.8799	–	–	–	0.869	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MKEMN(MM19)	–	–	–	0.866	–	0.816	–	–	–	–	–	–	–	–	–	–	–	–	–	–
MKN(MM19)	–	–	–	0.841	–	0.808	–	–	–	–	–	–	–	–	–	–	–	0.889	–	–
KMAGCN(TOMMCCAP21)	–	0.849	–	0.787	–	0.864	–	–	–	–	–	–	–	–	–	–	–	–	–	–
KMAGCN$_{BERT}$(TOMMCCAP21)	–	0.922	–	0.804	–	0.865	–	–	–	–	–	–	–	–	–	–	–	–	–	–
KMAGCN$_{XLNET}$(TOMMCCAP21)	–	0.944	–	0.827	–	0.867	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Social Context-based Methods																				
ML-GRUIJCAI16	0.862	–	0.839	–	0.784	–	–	–	0.5547	0.6612	–	–	–	–	–	–	–	–	–	–
CMAI(IJCAI17)	0.933	–	–	–	0.777	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
CallAtRumors(PAKDD17)	0.887	–	–	–	0.804	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
HSA-BLSTM(CIKM18)	0.943	–	0.913	–	0.844	–	0.846	0.753	–	–	–	–	–	–	–	–	–	–	–	–
MT-ES(WWW18)	–	–	–	–	–	–	–	–	0.848	0.864	–	–	–	–	–	–	–	–	–	–
CSI(CIKM17)	0.953	–	–	–	0.892	–	0.827	0.772	0.6987	0.6987	–	–	–	–	–	–	–	–	–	–
dEFEND(KDD19)	–	–	–	–	–	–	0.904	0.808	0.7383	0.7383	–	–	–	–	–	–	–	–	–	–
PPC(AAAI18)	0.916	–	–	–	–	–	–	–	0.842	0.863	–	–	–	–	–	–	–	–	–	–
GCAN(ACL20)	–	–	–	–	–	–	–	–	0.8767	0.9084	–	–	–	–	–	–	–	–	–	–
RvNN(ACL18)	0.908	–	–	–	–	0.7865	–	–	0.723	0.737	0.828	–	–	–	–	–	–	–	–	–
AE-GCN(DSAA20)	0.942	–	–	–	–	–	–	–	0.851	0.881	–	–	–	–	0.892	–	–	–	–	–
VAE-GCN(DSAA20)	0.944	–	–	–	–	–	–	–	0.856	0.868	–	–	–	–	0.881	–	–	–	–	–
Undirected GCN	0.892	–	–	–	–	–	0.836	0.8943	0.841	0.852	–	–	–	–	–	–	–	–	–	–
Undirected GAT	0.881	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Bi-GCN(AAAI20)	0.961	–	–	–	0.864	0.8678	–	–	0.886	0.88	0.889	–	–	–	–	–	–	–	–	–
GLAN(ICDM20)	0.946	–	–	–	–	–	–	–	0.905	0.902	–	–	–	–	–	–	–	–	–	–
HDGCN(PAKDD21)	0.961	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
TGNF(IPM21)	0.968	–	–	–	0.923	–	–	–	–	–	0.935	–	–	–	–	–	–	–	–	–
RDLNP(WWW21)	–	–	–	–	–	0.8867	–	–	–	–	–	–	–	–	0.906	–	–	–	–	–
PLAN(AAAI20)	–	–	–	–	–	–	–	–	0.845	0.874	–	–	–	–	–	–	–	–	–	–
StA-PLAN (AAAI20)	–	–	–	–	–	–	–	–	0.852	0.868	–	–	–	–	–	–	–	–	–	–
TRABS(COLING20)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
UPFD(SIGIR21)	–	–	–	–	–	–	0.8462	0.9723	–	–	–	–	–	–	–	–	–	–	–	–
BERT-EMO(WWW21)	0.908	–	0.932	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Weakly Supervised Methods																				
WeFEND(AAAI20)	–	–	–	–	–	–	–	–	–	–	–	–	0.824	–	–	–	–	–	–	–
WeFEND-(AAAI20)	–	–	–	–	–	–	–	–	–	–	–	–	0.807	–	–	–	–	–	–	–
SRLF(IJCNN21)	–	–	–	–	–	–	–	–	0.89	0.886	–	–	–	–	–	–	–	–	–	–
SMAN(COLING20)	–	–	–	–	0.951	–	–	–	0.929	0.935	–	–	–	–	–	–	–	–	–	–
TDSL	–	–	–	–	–	0.6064	0.531	0.501	–	–	–	–	–	0.834	–	–	–	–	–	–
SSDL	–	–	–	–	–	0.8478	–	–	–	–	–	–	–	0.8355	–	–	–	–	–	–
SSLNews	–	–	–	–	–	–	0.722	0.704	–	–	–	–	–	–	–	–	–	–	–	–
AA-HGNN(ICDM20)	–	–	–	–	–	–	0.6155	–	–	–	–	0.7351	–	–	–	–	–	–	–	–
HGAT-based classifier(ICDM20)	–	–	–	–	–	–	0.6154	–	–	–	–	0.7022	–	–	–	–	–	–	–	–
CNN-MWSS(ECML-PKDD20)	–	–	–	–	–	–	0.823	–	–	–	–	0.77	–	–	–	–	–	–	–	–
RoBERTa-MWSS(ECML-PKDD20)	–	–	–	–	–	–	0.825	–	–	–	–	0.803	–	–	–	–	–	–	–	–
Unsupervised Methods																				
Majority Voting	–	–	–	–	–	–	0.65	–	–	–	–	0.556	–	0.586	–	–	–	–	–	–
TruthFinder(TKDE08)	–	–	–	–	–	–	0.59	–	–	–	–	0.554	–	0.634	–	–	–	–	–	–
UFD(AAAI20)	–	–	–	–	–	–	0.7	–	–	–	–	0.679	–	0.759	–	–	–	–	–	–
GTUT	–	–	–	–	–	–	0.8	–	–	–	–	–	–	–	–	–	–	–	–	–

3. On the Media-Weibo dataset, the methods based on the single text modality are better than methods based on the single visual modality, which shows that the detection of fake news mainly relies on text clues.
4. HAN performs better than SVM-TS and CNN-text on PolitiFact and GossipCop datasets. It is because that HAN can better capture both syntactic and semantic information in news contents through hierarchical attention for differentiating fake and real news.
5. On the Media-Twitter and PHEME datasets, GAN-GRU performs much better than GRU-2, demonstrating the benefit of adversarial learning.
6. On the PHEME dataset, VRoC outperforms GRU-2, demonstrating the value of the autoencoder structure in the detection of fake news.
7. GCN-text performs better than CNN-text and SVM-TS on Media-Weibo, Media-Twitter, and PHEME datasets. The results demonstrate that the graph structure is capable of capturing word co-occurrences and document-word associations.
8. The pre-trained language model outperforms conventional text modeling techniques like CNN-text and GRU-2. The greater modeling capability of the transformer is one factor contributing to this improvement. On the other hand, it gains from the linguistic expertise acquired by the language model that has been pre-trained using a large pre-trained corpus.

Multi-modal Methods

1. Multi-modal methods are generally better than methods based on single-modal information because multimodal data can provide additional details and more signs for detecting fake news.
2. As a multi-modal model, att_RNN has superior performance than GRU-2. The reason is that it considers the text-related parts of the image for improving fake news detection. If the information on social context is removed, the performance would drop by a certain degree, which means the social context is useful for att_RNN.
3. Without an event discriminator, the form of the proposed model EANN− has a tendency to collect event-specific aspects rather than learning enough common features across events. The complete EANN, greatly increases performance with the aid of the event discriminator, proving the event discriminator's value for enhancing performance.
4. The performance of MVAE outperforms that of single-modal methods, suggesting that the detection of fake news may benefit from the addition of extra visual information. Due to the design of a multi-modal variational autoencoder, which is trained to reconstruct both modalities from the acquired shared representation and hence finds correlations across modalities, the MVAE outperforms several multi-modal approaches like attRNN and EANN.
5. Because MSRD takes into account both the visual information and the embedded text in images, it outperforms att_RNN on the Media-Twitter dataset.
6. SAFE outperforms CNN on most datasets because SAFE jointly uses multi-modal news content and relational information to learn the representation of posts. SAFE recognizes fake news by calculating the similarity between news text and news image and has a good performance in tasks where the text and pictures do not match.
7. Spotfake and Spotfake+ outperform att_RNN, EANN, and MVAE on Media-Twitter and Media-Weibo datasets. This is due to the fact that Spotfake and Spotfake+ employ the pre-trained language models BERT and XLNet to better capture the linguistic characteristics of media articles.
8. MCAN performs better than att_RNN, EANN and MVAE on Media-Twitter and Media-Weibo datasets, which embodies MCAN proposes feature fusion method is indeed better than the simple concatenation method and embodies that both spatial-domain features and frequency-domain features from the image are helpful to detect fake news.
9. HMCAN outperforms att_RNN, EANN, MVAE, SpotFake, and SpotFake+ on the datasets of meida-Twitter and PHEME. The results demonstrate that HMCAN can jointly model multi-modal context information and hierarchical semantics of text in a unified model, which better captures the underlying representation of posts, and improves the performance of fake news detection.
10. EM_FEND outperforms att_RNN, MVAE, SAFE, and SpotFake on the Media-Weibo dataset. It validates that EM-FEND can effectively capture important multi-modal clues that existing works ignore to detect fake news.

6.3.2. Social Context-based Methods

1. ML-GRU, CallAtRummer, and CMAI perform better than methods based on classical machine learning, indicating that deep learning algorithms can automatically capture more complicated patterns than the approaches using hand-crafted features. CallAtRummer outperforms ML-GRU because the typical attention method at the word level can increase accuracy even more.
2. HSA-BLSTM performs better than HAN on PolitiFact and GossipCop datasets, and HSA-BLSTM performs better than ML-GRU and CallATRummer on Ma-Twitter and Ma-Weibo datasets. There are two reasons: first, the hierarchy-based models can effectively learn the representation of each semantic level. Second, incorporating social features into the model helps detect fake news on the social network.
3. CSI performs better than ML-GRU and CallAtRummer on the Ma-Twitter and Ma-Weibo datasets. This is because CSI considers the influence of user credibility on news authenticity.

4. dEFEND outperforms HAN on the PolitiFact and GossipCop datasets because HAN only utilizes news content. However, dEFEND utilizes both news content and user comments, indicating that users' comments can provide a wealth of additional clues for fake news detection. The performance of dEFEND on the PolitiFact and GossipCop datasets is superior to that of CSI because co-attention modeling of news sentences and user comments is critical for fake news identification.
5. GCAN outperforms CSI on Twitter15 and Twitter16, indicating that the dual co-attention mechanism in GCAN is extremely effective. While both GCAN and dEFEND are co-attention-based, GCAN demonstrates superior performance. This is due to the fact that GCAN uses graph structure to represent the relationship between users and may thus reveal more clues between users.
6. PPC performs much better than ML-GRU on Ma-Weibo, Twitter15 and Twitter16 datasets. There are two reasons: 1) ML-GRU only focuses on the text information of news articles and comments, whereas PPC additionally considers user information. 2) PPC utilizes both CNN and RNN to capture the variations of user characteristics.
7. The sequential neural model ML-GRU performs worse than RvNN on the Twitter15 and Twitter16 datasets. Because RVNN employs a tree structure to represent the structural information of news distribution and to better define the patterns of news dissemination. Moreover, ML-GRU disregards structural information.
8. PLAN performs better than RvNN on Twitter15 and Twitter16. This is because that PLAN employs a self-attention mechanism to capture the long-range dependencies between tweets and better comprehend the semantic associations of the news propagation process. Although findings were not significantly different, StA-PLAN performs well on Twitter15 but does not outperform PLAN on Twitter16. Because Twitter15's data set is so much larger than Twitter16's, the structure-aware model can only be used with that platform. To take use of the comprehensive structural information in the structure-aware model, a large dataset could be required.
9. By modeling dependencies between any two tweets, PLAN makes the most of the transformer's representation power, but this may underuse the structural information. In contrast, tree transformer models outperform PLAN on Ma-Twitter and PHEME datasets because they both naturally harness propagation structure and make use of the transformer's powerful representational capabilities.
10. On Ma-Weibo, Twitter15, and Twitter16 datasets, AE-GCN and VAE-GCN outperform RvNN, a model that similarly learns propagation information but is based on recursive neural network. The categorization outcomes mainly rely on the most recent posts, which are insufficiently detailed because RvNN only takes into account leaf nodes and use a max-pooling operator. Additionally, sequential propagation limits the efficiency of GRU units, hence the GCN module outperforms them.
    The performance of AE-GCN and VAE-GCN is better than Undirected GCN, which indicate that the Autoencoder is useful for modeling propogation structure.
11. Bi-GCN outperforms PPC, demonstrating the value of including the dispersion structure in rumor detection. PPC ignores crucial structural aspects of rumor dispersion since RNN and CNN are unable to analyze data with a graph structure.
    Bi-GCN significantly outperforms RvNN. Since RvNN only uses the hidden leaf node feature, it is heavily impacted by the information of the latest posts. However, the latest posts just follow the former posts and are always lacking in information such as comments. Differently, Bi-GCN uses root feature enhancement and thus pay more attention to the information of the source posts.
    Bi-GCN outperforms than Undirected GCN, AE-GCN, and VAE-GCN on Ma-Weibo, Twitter15, and Twitter16 datasets. There are two reasons to explain. Firstly, Bi-GCN uses the root post's feature to enhance, while other methods do not, this indicates that the source posts play an important role in rumor detection. Secondly, Bi-GCN considers the directions of edges in news propagation graphs, but other methods only consider undirected representations. It indicates that both top-down propagation information and bottom-up dispersion information are useful for detecting fake news.
12. On Ma-Weibo, Ma-Twitter, and FakeNewsNet datasets, TGNF beats Bi-GCN, demonstrating the significance of temporal propagation information in detecting the truthfulness of the news.
13. Because RDLNP combines the non-linear diffusion model (NLSL) and linear time-dependent model (LSL) to examine the variety of rumor dissemination, it performs better than Bi-GCN on PHEME dataset. Additionally, in accordance with the property of rumors, the RDLNP simultaneously models the claim's content and social context data, enabling it to fully use additional features.
14. GCAN uses the information of news content, user characteristics, propagation, and structure of social networks. GCAN performs better than ML-GRU because it uses more abundant information than ML-GRU. In the case of consistent input information, GCAN performs better than CSI on Twitter15 and Twitter16, which suggests that the dual co-attention mechanism in GCAN can more effectively fuse news content with social context information. Although both GCAN and dEFEND are based on co-attention,

GCAN significantly improves the performance by learning additional sequential features from the retweet user sequence.

15. On the Twitter15, Twitter16, and Ma-Weibo datasets, GLAN beats PPC and ML-GRU. These findings show how important local semantic and global structural information are for understanding how to distinguish between fake and real news.
16. UPFD performs better than Undirected GCN on Politifac an Gossipcop datasets because Undirected GCN only considers the propagation structure of news in a social network. However, UPFD not only considers the propagation structure but also consider the users' preference. We can see that leveraging the historical posts as user preferences could improve the fake news detection performance.
17. In comparison to the homogeneous graph-based methods Undirected GCN, Undirected GAT, and GLAN, HDGCN performs better. In contrast toUndirected GCN, Undirected GAT, HDGCN takes use of the heterogeneous graph and is able to learn heterogeneous information from various kinds of nodes and edges.
18. Bert-EMO performs better than Bert-text on the Ma-Weibo dataset because Bert-EMO considers the publisher's emotion and social emotion for fake news detection.

6.3.3. External Knowledge-based Methods

1. The methods using news content and external knowledge consistently outperforms those based on news content only. This indicates that incorporating external knowledge can significantly improve the detection performance.
   KAN outperforms KCNN and B-TransE. There are two underlying reasons: 1) KAN utilizes the knowledge-aware attention network, which may minimize ambiguity produced by the entity cited in the news and discover knowledge-level links between news items. 2) KAN leverages the attention network, which can quantify the significance of entity and entity context information and effectively combine them into news representation.
2. KMGCN's performance is better than att_RNN, EANN, MVAE, SAFE, SpotFake, and SpotFake+ on Media-Weibo and PHEME datasets. There are two reasons: 1) KMGCN enhances the semantic information of post text with visual information and knowledge concepts. 2) It benefits from multi-modal graph convolutional networks which better captures non-consecutive phrases and word dependency in post representations. For fairness, the original authors of KMGCN removed knowledge input in KMGCN as KMGCN-NoKD and tested on Media-Twitter and PHEME datasets. The experimental results show that after removing the knowledge input, the experimental effect of KMGCN-NoKD on Media-Twitter and PHEME is weaker than that of KMGCN, which shows that external knowledge does enhance the semantic information of the text. Knowledge information is an essential kind of complementary information for fake news detection.
3. MKEMN has achieved better performance than EANN. There exist two underlying reasons: 1) MKEMN uses the multi-modal knowledge-aware network for post representation, which better captures the multi-channel semantic information. 2) MKN with EMN (a.k.a MKEMN) is better than MKN. MKEMN exploits the event memory network for obtaining the event-invariant features which is helpful for the newly emerged events.
4. KMAGCN$_{BERT}$/KMAGCN$_{XLNET}$ achieves better performance than SpotFake/SpotFake+ on three datasets, demonstrating the effectiveness of KMAGCN for improving the performance of fake news detection. KMAGCN, KMAGCN-BERT, and KMAGCN-XLNET consistently outperform att_RNN, EANN, and MVAE on the three datasets. It demonstrates that KMAGCN better captures the underlying representation of the posts by modeling textual data, knowledge concepts, and image features inside a unified framework.

6.3.4. Weakly Supervised Methods

1. TDSL, SSDL, and SSLNews use only 30% of the labeled data to achieve accuracy comparable to partially supervised learning approaches, suggesting that weakly supervised learning is effective.
2. can achieve good performance with less labels, even outperforming some supervised learning algorithms on specific datasets. This is because WeFEND makes use of annotators to weakly label unlabelled data, effectively using a large amount of unlabelled data. WeFEND's variation without the reinforcement learning selection is less effective than WeFEND. This is because the labels generated by the annotator are noisy, and using them directly will introduce noise into the model training and affect the results. The reinforcement learning selector can filter out the noisy labels and leave valid labels to enhance the model.
3. SRLF achieves significant improvement over some supervised methods such as ML-GRU, RvNN, PPC, and MT-ES on Twitter15 and Twitter16 datasets. The main reason is that SRLF uses a selected stance label as a supervised signal, providing more practical information since users' stances usually can reveal the veracity of rumors. Compared with MT-ES, which uses weak stance labels without selecting, SRLF can select high-quality labeled stance labels for training, avoiding the influence of wrongly labeled data, thus achieving greater performance. The alternate training process can optimize both tasks simultaneously and promote them mutually.
4. SMAN significantly outperforms GLAN on Ma-Weibo, Twitter15, and Twitter16 datasets. Different from GLAN, SMAN not only optimizes the fake news detection task but also predicts the credibility of publishers and users. The results demonstrate that

the credibility of publishers and users is important for differentiating fake and real news.

5. AA-HGNN achieves improvement over some homogeneous graph neural network methods on the PolitiFact and BuzzFeed datasets. A News-HIN integrates all available heterogeneous data in the form of a graph structure. Intuitively, AA-HGNN fully uses New-HIN as training data and achieves better performance.
   The performance of the HGAT-based classifier is worse than that of AA-HGNN on PolitiFact and BuzzFeed datasets. It means the adversarial learning between the classifier and the selector provides an effective active learning query strategy. The queried candidates are highly valued for the classifier, thus the performance can be significantly enhanced.
6. CNN-MWSS and RoBERT-MWSS achieve competitive performance on GossipCop and Politifact datasets, which shows that the weak social supervision signal is useful for fake news detection. Comparing the two encoders, RoBERTa performed better on the GossipCop and PolitiFact datasets than CNN, indicating that the pre-trained language model can capture news content information more effectively and thus obtain better results.

6.3.5. Unsupervised Methods

1. The Majority Voting method performs the lowest on the LIAR and BuzzFeed datasets because it equally aggregates user opinions without taking credibility information into account.
2. UFD achieves better performance than Majority Voting and TruthFinder on LIAR and BuzzFeed datasets. Utilizing user actions (likes, retweets, and responses), as opposed to benchmarks that solely utilize the information in news tweets, may significantly boost speed. UFD performs better on the LIAR dataset than the BuzzFeed dataset, mostly due to the sparser user interactions in BuzzFeed.
3. GTUT outperforms Majority Voting, TruthFinder, and UDF on the PolitiFact and GossipCop datasets, thus proving the validity of the model design.

7. Future Direction

Based on the above review and analysis, we believe there exists much space for further enhancement in this field. In this section, we discuss the emerging trends for further exploration in FND research.

7.1. Emergent Events Fake News Detection

Despite the powerful ability of deep learning models to capture various news features, the algorithms still struggle when it comes to detecting fake news on emerging events. The model trained on past events may not perform satisfactorily due to the domain shift in news events, so new information from emergent events is required to supplement fake news detection models. But doing so necessitates either starting from scratch with a new model or continuing to fine-tune the existing model using newly gathered labeled data, which can be difficult, expensive, and unrealistic for real-world settings.

Moreover, fake news frequently appears at recent events where we hardly ever receive enough posts quickly. We only typically have a small number of related verified posts in the early stages of emergent events. Another significant challenge is to use a small number of verified posts to help the model quickly learn to identify false information about recently occurring events.

Few-shot learning, which aims to use a small set of data instances for quick learning, is one potential solution to the problem mentioned above. The basic concept of meta-learning, a promising research area in few-shot learning, is to use the general knowledge from earlier tasks to aid in learning the new task. However, the effectiveness of current meta-learning methodologies is strongly correlated with a crucial assumption: the tasks are drawn from a similar distribution, and the common global knowledge applies to various tasks. This presumption usually does not apply to the fake news detection problem because news articles on various events usually have different writing styles, content, vocabulary, and even class distributions.

In conclusion, it is crucial and worthwhile to research how to accurately identify fake news using the limited data from emergent news events.

7.2. Multi-domain Fake News Detection

Existing fake news detection methods either completely ignore news domain information for fake news detection within a generic domain dataset (e.g., Twitter15 and Twitter16 (Ma et al., 2017), Ma-Weibo (Ma et al., 2016), Media-Weibo (Jin et al., 2017)), or perform fake news detection within a single domain dataset (e.g., FakeHealth (Dai et al., 2020), MM-COVID (Li et al., 2020)) only. Concerning a specific domain, the domain-specific fake news detection dataset usually contains much more knowledge of this domain than the general-purpose fake news detection dataset. Research has shown a great difference in domain-specific word usage and domain-specific propagation patterns between news from different domains. For example, the health domain-specific fake news detection dataset FakeHealth (Dai et al., 2020) dataset contains more specialized vocabulary in the health domain.

The information about the news domain is essential for identifying fake news because, in reality, news on social media covers various domains. Therefore, it is essential to carry out research on the identification of fake news in situations involving multiple domains. Existing studies, however, have difficulty addressing the issue of cross-domain fake news detection. Generic domain fake news detection models cannot take advantage of domain-specific prior knowledge because they completely ignore the distinctions between domains. It is challenging to directly transfer fake news detection models developed for one domain to another due

to the significant differences in the news across different domains.

To promote the development of multi-domain fake news detection research, Nan et al. (2021) designed a benchmark of fake news dataset for multi-domain fake news detection with domain label annotated, namely Weibo21, which consists of 4,488 fake news and 4,640 real news from 9 different domains. We strongly believe that this direction deserves much deeper exploration by the community.

7.3. Multi-lingual Fake News Detection

The majority of studies to date have focused on detecting fake news in a particular language. However, a lot of the information is disseminated not just among native English speakers but also among speakers of other languages who come from other cultures. It raises a crucial query regarding the applicability of current techniques for identifying fake news.

A large multilingual news corpus is necessary for training a multilingual fake news detection model. To the best of our knowledge, there aren't many datasets for multilingual rumor detection. The PHEME (Zubiaga et al., 2017) dataset includes tweets in both German and English. Cross-lingual COVID-19 (Du et al., 2021) contains COVID-19 news in both English and Chinese, whereas fakeCovid (Shahi and Nandini, 2020) contains COVID-19 news in forty various languages. The COVID-19 news is available in six languages on mm-COVID (Li et al., 2020). Although the available datasets can aid academics in their work on cross-lingual fake news detection, they are limited in the number of languages they include and the quantity of data they contain. There is a requirement for a dataset with a greater quantity of languages and news articles.

It is necessary to detect fake news in texts in different languages. However, there are often some lexical, syntactic and grammatical differences between different languages, and making the model with multi-lingual understanding is one of the critical challenges to solving multi-lingual fake news detection. The English fake news corpus is large, and the model can learn much knowledge to identify fake news from it, but the corpus of other languages is small, so how to transfer the knowledge obtained from the English fake news detection model to the fake news detection model of other languages by using transfer learning method is another challenge.

8. Acknowledgments

The work is supported by the National Natural Science Fundation of China (No. 62276029) and Beijing Academy of Artificial Intelligence (BAAI).

References

Ahmed, S., Hinkelmann, K., Corradini, F., 2019. Combining machine learning with knowledge engineering to detect fake news in social networks-a survey, in: Proceedings of the AAAI Spring Symposium, p. 8.

Benamira, A., Devillers, B., Lesot, E., Ray, A.K., Saadi, M., Malliaros, F.D., 2019. Semi-supervised learning and graph neural networks for fake news detection, in: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 568–569.

Bian, T., Xiao, X., Xu, T., Zhao, P., Huang, W., Rong, Y., Huang, J., 2020. Rumor detection on social media with bi-directional graph convolutional networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 549–556.

Boididou, C., Papadopoulos, S., Dang-Nguyen, D.T., Boato, G., Kompatsiaris, Y., 2015. The certh-unitn participation@ verifying multimedia use 2015., in: MediaEval, pp. 1–3.

Boididou, C., Papadopoulos, S., Kompatsiaris, Y., Schifferes, S., Newman, N., 2014. Challenges of computational verification in social multimedia, in: Proceedings of the 23rd International Conference on World Wide Web, pp. 743–748.

Cao, J., Sheng, Q., Qi, P., Zhong, L., Wang, Y., Zhang, X., 2019. False news detection on social media. arXiv preprint arXiv:1908.10818 .

Castillo, C., Mendoza, M., Poblete, B., 2011. Information credibility on twitter, in: Proceedings of the 20th international conference on World wide web, pp. 675–684.

Chen, T., Kornblith, S., Norouzi, M., Hinton, G., 2020. A simple framework for contrastive learning of visual representations, in: Proceedings of the 2020 International conference on machine learning, pp. 1597–1607.

Chen, T., Li, X., Yin, H., Zhang, J., 2018a. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection, in: Proceedings of the Pacific-Asia conference on knowledge discovery and data mining, pp. 40–52.

Chen, W., Yeo, C.K., Lau, C.T., Lee, B.S., 2016. Behavior deviation: An anomaly detection view of rumor preemption, in: Proceedings of the 7th Annual Information Technology, Electronics and Mobile Communication Conference, pp. 1–7.

Chen, W., Zhang, Y., Yeo, C.K., Lau, C.T., Lee, B.S., 2018b. Unsupervised rumor detection based on users' behaviors using neural networks. Pattern Recognition Letters 105, 226–233.

Chen, X., Lian, C., Wang, L., Deng, H., Fung, S.H., Nie, D., Thung, K.H., Yap, P.T., Gateno, J., Xia, J.J., et al., 2019a. One-shot generative adversarial learning for mri segmentation of craniomaxillofacial bony structures. IEEE transactions on medical imaging 39, 787–796.

Chen, Y., Sui, J., Hu, L., Gong, W., 2019b. Attention-residual network with cnn for rumor detection, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1121–1130.

Cheng, M., Nazarian, S., Bogdan, P., 2020. Vroc: Variational autoencoder-aided multi-task rumor classifier based on text, in: Proceedings of The International World Wide Web Conferences, pp. 2892–2898.

Cui, L., Lee, D., 2020. Coaid: Covid-19 healthcare misinformation dataset. arXiv preprint arXiv:2006.00885 .

Dai, E., Sun, Y., Wang, S., 2020. Ginger cannot cure cancer: Battling fake health news with a comprehensive data repository, in: Proceedings of the International AAAI Conference on Web and Social Media, pp. 853–862.

Della Vedova, M.L., Tacchini, E., Moret, S., Ballarin, G., DiPierro, M., de Alfaro, L., 2018. Automatic online fake news detection combining content and social signals, in: Proceedings of the 22nd Conference of Open Innovations Association, pp. 272–279.

Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. Bert: Pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, pp. 4171–4186.

Dong, M., Yao, L., Wang, X., Benatallah, B., Sheng, Q.Z., Huang, H., 2018. Dual: A deep unified attention model with latent relation representations for fake news detection, in: Proceedings of the International conference on web information systems engineering, pp. 199–209.

Dong, X., Victor, U., Chowdhury, S., Qian, L., 2019. Deep two-path semi-supervised learning for fake news detection. arXiv preprint arXiv:1906.05659 .

Dong, X., Victor, U., Qian, L., 2020. Two-path deep semisupervised learning for timely fake news detection. IEEE Transactions on Computational

Social Systems 7, 1386–1398.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale, in: International Conference on Learning Representations.

Dou, Y., Shu, K., Xia, C., Yu, P., Sun, L., 2021. User preference-aware fake news detection, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2051–2055.

Du, J., Dou, Y., Xia, C., Cui, L., Ma, J., Yu, P.S., 2021. Cross-lingual covid-19 fake news detection. arXiv preprint arXiv:2110.06495 .

Dun, Y., Tu, K., Chen, C., Hou, C., Yuan, X., 2021. Kan: Knowledge-aware attention network for fake news detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 81–89.

Gangireddy, S.C.R., Long, C., Chakraborty, T., 2020. Unsupervised fake news detection: A graph-based approach, in: Proceedings of the 31st ACM Conference on Hypertext and Social Media, pp. 75–83.

Ghorbanpour, F., Ramezani, M., Fazli, M.A., Rabiee, H.R., 2021. Fnr: A similarity and transformer-based approachto detect multi-modal fake-news in social media. arXiv preprint arXiv:2112.01131 .

Giachanou, A., Rosso, P., Crestani, F., 2019. Leveraging emotional signals for credibility detection, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 877–880.

Guacho, G.B., Abdali, S., Shah, N., Papalexakis, E.E., 2018. Semi-supervised content-based detection of misinformation via tensor embeddings, in: Proceedings of the IEEE/ACM international conference on advances in social networks analysis and mining, pp. 322–325.

Guo, H., Cao, J., Zhang, Y., Guo, J., Li, J., 2018. Rumor detection with hierarchical social attention network, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 943–951.

Guo, Z., Yu, K., Jolfaei, A., Bashir, A.K., Almagrabi, A.O., Kumar, N., 2021. A fuzzy detection system for rumors through explainable adaptive learning. IEEE Transactions on Fuzzy Systems 29, 3650–3664.

Hamilton, W.L., Ying, R., Leskovec, J., 2017. Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 1025–1035.

He, Z., Li, C., Zhou, F., Yang, Y., 2021. Rumor detection on social media with event augmentations, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2020–2024.

Helwe, C., Elbassuoni, S., Al Zaatari, A., El-Hajj, W., 2019. Assessing arabic weblog credibility via deep co-learning, in: Proceedings of the Fourth Arabic Natural Language Processing Workshop, pp. 130–136.

Hu, G., Ding, Y., Qi, S., Wang, X., Liao, Q., 2019. Multi-depth graph convolutional networks for fake news detection, in: Proceedings of the Natural Language Processing and Chinese Computing, pp. 698–710.

Hu, L., Yang, T., Zhang, L., Zhong, W., Tang, D., Shi, C., Duan, N., Zhou, M., 2021. Compare to the knowledge: Graph neural fake news detection with external knowledge, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pp. 754–763.

Huang, Q., Zhou, C., Wu, J., Wang, M., Wang, B., 2019. Deep structure learning for rumor detection on twitter, in: Proceedings of the 2019 International Joint Conference on Neural Networks, pp. 1–8.

Jiang, S., Chen, X., Zhang, L., Chen, S., Liu, H., 2019. User-characteristic enhanced model for fake news detection in social media, in: Proceedings of the Natural Language Processing and Chinese Computing, pp. 634–646.

Jin, F., Dougherty, E., Saraf, P., Cao, Y., Ramakrishnan, N., 2013. Epidemiological modeling of news and rumors on twitter, in: Proceedings of the 7th workshop on social network mining and analysis, pp. 1–9.

Jin, Z., Cao, J., Guo, H., Zhang, Y., Luo, J., 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs, in: Proceedings of the 25th ACM international conference on Multimedia, pp. 795–816.

Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q., 2016. Novel visual and statistical image features for microblogs news verification. IEEE transactions on multimedia 19, 598–608.

Jindal, S., Sood, R., Singh, R., Vatsa, M., Chakraborty, T., 2020. Newsbag: A multimodal benchmark dataset for fake news detection, in: Proceedings of the Workshop on Artificial Intelligence Safety, co-located with 34th AAAI Conference on Artificial Intelligence, pp. 138–145.

Kang, Z., Cao, Y., Shang, Y., Liang, T., Tang, H., Tong, L., 2021. Fake news detection with heterogenous deep graph convolutional network., in: Proceedings of the Knowledge Discovery and Data Mining - 25th Pacific-Asia Conference, pp. 408–420.

Khattar, D., Goud, J.S., Gupta, M., Varma, V., 2019. Mvae: Multimodal variational autoencoder for fake news detection, in: Proceedings of the International World Wide Web Conferences, pp. 2915–2921.

Khoo, L.M.S., Chieu, H.L., Qian, Z., Jiang, J., 2020. Interpretable rumor detection in microblogs by attending to user interactions, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8783–8790.

Kolitsas, N., Ganea, O.E., Hofmann, T., 2018. End-to-end neural entity linking, in: Proceedings of the 22nd Conference on Computational Natural Language Learning, pp. 519–529.

Konkobo, P.M., Zhang, R., Huang, S., Minoungou, T.T., Ouedraogo, J.A., Li, L., 2020. A deep learning model for early detection of fake news on social media, in: Proceedings of the 7th International Conference on Behavioural and Social Computing, pp. 1–6.

Kwon, S., Cha, M., Jung, K., Chen, W., Wang, Y., 2013. Aspects of rumor spreading on a microblog network, in: Proceedings of the International Conference on Social Informatics, pp. 299–308.

Laine, S., Aila, T., 2017. Temporal ensembling for semi-supervised learning, in: Proceedings of the 5th International Conference on Learning Representations, pp. 1–13.

Lao, A., Shi, C., Yang, Y., 2021. Rumor detection with field of linear and non-linear propagation, in: Proceedings of the International World Wide Web Conferences, pp. 3178–3187.

Le, P., Titov, I., 2018. Improving entity linking by modeling latent relations between mentions, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 1595–1604.

Li, J., Ni, S., Kao, H.Y., 2021a. Meet the truth: Leverage objective facts and subjective views for interpretable rumor detection, in: Proceedings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 705–715.

Li, P., Sun, X., Yu, H., Tian, Y., Yao, F., Xu, G., 2021b. Entity-oriented multi-modal alignment and fusion network for fake news detection. IEEE Transactions on Multimedia , 1–14.

Li, X., Lu, P., Hu, L., Wang, X., Lu, L., 2021c. A novel self-learning semi-supervised deep learning network to detect fake news on social media. Multimedia Tools and Applications , 1–9.

Li, Y., Jiang, B., Shu, K., Liu, H., 2020. Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation. arXiv preprint arXiv:2011.04088 .

Lin, H., Zhang, X., Fu, X., 2020. A graph convolutional encoder and decoder model for rumor detection, in: Proceedings of the 7th International Conference on Data Science and Advanced Analytics, pp. 300–306.

Liu, Y., Wu, Y.F.B., 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, pp. 354–361.

Liu Jinshuo, Feng Kuo, J.Z.P.D.J.W.L., 2020. Msrd: Multi-modal web rumor detection method. Journal of Computer Research and Development 57, 2328–2336.

Lu, Y.J., Li, C.T., 2020. Gcan: Graph-aware co-attention networks for explainable fake news detection on social media, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 505–514.

Ma, J., Gao, W., 2020. Debunking rumors on twitter with tree transformer, in: Proceedings of the 28th International Conference on Computational Linguistics, pp. 5455–5466.

Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M., 2016. Detecting rumors from microblogs with recurrent neural networks, in: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 3818–3824.

Ma, J., Gao, W., Wei, Z., Lu, Y., Wong, K.F., 2015. Detect rumors using time series of social context information on microblogging websites, in: Proceedings of the 24th ACM international on conference on information and knowledge management, pp. 1751–1754.

Ma, J., Gao, W., Wong, K.F., 2017. Detect rumors in microblog posts using propagation structure via kernel learning, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 708–717.

Ma, J., Gao, W., Wong, K.F., 2018a. Detect rumor and stance jointly by neural multi-task learning, in: proceedings of the the International World Wide Web Conferences, pp. 585–593.

Ma, J., Gao, W., Wong, K.F., 2018b. Rumor detection on twitter with tree-structured recursive neural networks, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 1980–1989.

Ma, J., Gao, W., Wong, K.F., 2019. Detect rumors on twitter by promoting information campaigns with generative adversarial learning, in: Proceedings of the International World Wide Web Conferences, pp. 3049–3055.

Ma, J., Li, J., Gao, W., Yang, Y., Wong, K.F., 2021. Improving rumor detection by promoting information campaigns with transformer-based generative adversarial learning. IEEE Transactions on Knowledge and Data Engineering , 1–14.

Mansouri, R., Naderan-Tahan, M., Rashti, M.J., 2020. A semi-supervised learning method for fake news detection in social media, in: Proceedings of the 28th Iranian Conference on Electrical Engineering, pp. 1–5.

Meel, P., Vishwakarma, D.K., 2021a. Fake news detection using semi-supervised graph convolutional network. arXiv preprint arXiv:2109.13476 .

Meel, P., Vishwakarma, D.K., 2021b. A temporal ensembling based semi-supervised convnet for the detection of fake news articles. Expert Systems with Applications 177, 115002.

Milne, D., Witten, I.H., 2008. Learning to link with wikipedia, in: Proceedings of the 17th ACM conference on Information and knowledge management, pp. 509–518.

Mitra, T., Gilbert, E., 2015. Credbank: A large-scale social media corpus with associated credibility annotations, in: Proceedings of the Ninth International Conference on Web and Social Media, pp. 258–267.

Nakamura, K., Levy, S., Wang, W.Y., 2020. Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection, in: Proceedings of the 12th Language Resources and Evaluation Conference, pp. 6149–6157.

Nan, Q., Cao, J., Zhu, Y., Wang, Y., Li, J., 2021. Mdfend: Multi-domain fake news detection, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 3343–3347.

Nguyen, V.H., Sugiyama, K., Nakov, P., Kan, M.Y., 2020. Fang: Leveraging social context for fake news detection using graph representation, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1165–1174.

Pan, J.Z., Pavlova, S., Li, C., Li, N., Li, Y., Liu, J., 2018. Content based fake news detection using knowledge graphs, in: Proceedings of the International semantic web conference, pp. 669–683.

Paulheim, H., 2017. Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic web 8, 489–508.

Popat, K., 2017. Assessing the credibility of claims on the web, in: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 735–739.

Qi, P., Cao, J., Li, X., Liu, H., Sheng, Q., Mi, X., He, Q., Lv, Y., Guo, C., Yu, Y., 2021. Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues, in: Proceedings of the 29th ACM International Conference on Multimedia, pp. 1212–1220.

Qi, P., Cao, J., Yang, T., Guo, J., Li, J., 2019. Exploiting multi-domain visual information for fake news detection, in: Proceedings of the International Conference on Data Mining, pp. 518–527.

Qian, F., Gong, C., Sharma, K., Liu, Y., 2018. Neural user response generator: Fake news detection with collective user intelligence., in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence,, pp. 3834–3840.

Qian, S., Hu, J., Fang, Q., Xu, C., 2021a. Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. ACM Transactions on Multimedia Computing, Communications, and Applications 17, 1–23.

Qian, S., Wang, J., Hu, J., Fang, Q., Xu, C., 2021b. Hierarchical multi-modal contextual attention network for fake news detection, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 153–162.

Qiao, S., Shen, W., Zhang, Z., Wang, B., Yuille, A., 2018. Deep co-training for semi-supervised image recognition, in: Proceedings of the european conference on computer vision, pp. 135–152.

Qiu, S., Zhao, Y., Jiao, J., Wei, Y., Wei, S., 2019. Referring image segmentation by generative adversarial learning. IEEE Transactions on Multimedia 22, 1333–1344.

Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y., 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking, in: Proceedings of the 2017 conference on empirical methods in natural language processing, pp. 2931–2937.

Ren, Y., Wang, B., Zhang, J., Chang, Y., 2020. Adversarial active learning based heterogeneous graph neural network for fake news detection, in: Proceedings of the 2020 IEEE International Conference on Data Mining, pp. 452–461.

Rubin, V.L., Conroy, N., Chen, Y., Cornwell, S., 2016. Fake news or truth? using satirical cues to detect potentially misleading news, in: Proceedings of the second workshop on computational approaches to deception detection, pp. 7–17.

Ruchansky, N., Seo, S., Liu, Y., 2017. Csi: A hybrid deep model for fake news detection, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 797–806.

Sampson, J., Morstatter, F., Wu, L., Liu, H., 2016. Leveraging the implicit structure within social media for emergent rumor detection, in: Proceedings of the 25th ACM international on conference on information and knowledge management, pp. 2377–2382.

Santia, G.C., Williams, J.R., 2018. Buzzface: A news veracity dataset with facebook user commentary and egos, in: Proceedings of the Twelfth International AAAI Conference on Web and Social Media, pp. 531–541.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2014. Overfeat: Integrated recognition, localization and detection using convolutional networks, in: Proceedings of the 2nd International Conference on Learning Representations, pp. 1–16.

Shahi, G.K., Nandini, D., 2020. Fakecovid–a multilingual cross-domain fact check news dataset for covid-19. arXiv preprint arXiv:2006.11343 .

Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. Cnn features off-the-shelf: an astounding baseline for recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 806–813.

Shu, K., Cui, L., Wang, S., Lee, D., Liu, H., 2019a. defend: Explainable fake news detection, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 395–405.

Shu, K., Liu, H., 2019. Detecting fake news on social media. Synthesis lectures on data mining and knowledge discovery 11, 1–129.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H., 2020a. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Journal on big data 8, 171–188.

Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H., 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter 19, 22–36.

Shu, K., Wang, S., Liu, H., 2018. Understanding user profiles on social media for fake news detection, in: Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval, pp. 430–435.

Shu, K., Wang, S., Liu, H., 2019b. Beyond news contents: The role of social context for fake news detection, in: Proceedings of the twelfth ACM international conference on web search and data mining, pp. 312–320.

Shu, K., Zheng, G., Li, Y., Mukherjee, S., Awadallah, A.H., Ruston, S., Liu, H., 2020b. Early detection of fake news with multi-source weak social supervision, in: Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 650–666.

Shu, K., Zhou, X., Wang, S., Zafarani, R., Liu, H., 2019c. The role of user profiles for fake news detection, in: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp. 436–439.

Silva, A., Han, Y., Luo, L., Karunasekera, S., Leckie, C., 2021. Propagation2vec: Embedding partial propagation networks for explainable fake news early detection. Information Processing & Management 58, 102618.

Silverman, C., Strapagiel, L., Shaban, H., Hall, E., Singer-Vine, J., 2016. Hyperpartisan facebook pages are publishing false and misleading information at an alarming rate. Buzzfeed News 20, 68.

Simonyan, K., Zisserman, A., . Very deep convolutional networks for large-scale image recognition, in: Proceedings of the 3rd International Conference on Learning Representations, pages=1–14, year=2015.

Singhal, S., Kabra, A., Sharma, M., Shah, R.R., Chakraborty, T., Kumaraguru, P., 2020. Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract), in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 13915–13916.

Singhal, S., Shah, R.R., Chakraborty, T., Kumaraguru, P., Satoh, S., 2019. Spotfake: A multi-modal framework for fake news detection, in: Proceedings of the fifth international conference on multimedia big data, pp. 39–47.

Song, C., Ning, N., Zhang, Y., Wu, B., 2021a. A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Information Processing & Management 58, 1–14.

Song, C., Shu, K., Wu, B., 2021b. Temporally evolving graph neural network for fake news detection. Information Processing & Management 58, 102712.

Suchanek, F.M., Kasneci, G., Weikum, G., 2008. Yago: A large ontology from wikipedia and wordnet. Journal of Web Semantics 6, 203–217.

Sun, M., Zhang, X., Zheng, J., Ma, G., 2022. Ddgcn: Dual dynamic graph convolutional networks for rumor detection on social media, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 4611–4619.

Sun, S., Liu, H., He, J., Du, X., 2013. Detecting event rumors on sina weibo automatically, in: Proceedings of the Asia-Pacific web conference, pp. 120–131.

Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S., de Alfaro, L., 2017. Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506 .

Tao, C., Gao, S., Shang, M., Wu, W., Zhao, D., Yan, R., 2018. Get the point of my utterance! learning towards effective responses with multi-head attention mechanism., in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, pp. 4418–4424.

Tian, L., Zhang, X., Lau, J.H., 2021. Rumour detection via zero-shot cross-lingual transfer learning, in: Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 603–618.

Vaibhav, R.M.A., Hovy, E., 2019. Do sentence interactions matter? leveraging sentence level representations for fake news classification, in: Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing, TextGraphs@EMNLP, pp. 134–139.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, in: Proceedings of the neural information processing systems, pp. 5998–6008.

Victor, U., 2020. Robust Semi-Supervised Learning for Fake News Detection. Ph.D. thesis. Ph. D Thesis, Prairie View A&M University, Prairie View, TX, USA, 2020 ....

Vogel, I., Jiang, P., 2019. Fake news detection with the new german dataset "germanfakenc", in: Proceedings of the 23rd International Conference on Theory and Practice of Digital Libraries, pp. 288–295.

Volkova, S., Jang, J.Y., 2018. Misleading or falsification: Inferring deceptive strategies and types in online news and social media, in: Proceedings of the International World Wide Web Conferences 2018, pp. 575–583.

Vosoughi, S., Roy, D., Aral, S., 2018. The spread of true and false news online. Science 359, 1146–1151.

Wang, H., Zhang, F., Xie, X., Guo, M., 2018a. Dkn: Deep knowledge-aware network for news recommendation, in: Proceedings of the International World Wide Web Conferences, pp. 1835–1844.

Wang, W.Y., 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pp. 422–426.

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Su, L., Gao, J., 2018b. Eann: Event adversarial neural networks for multi-modal fake news detection, in: Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining, pp. 849–857.

Wang, Y., Ma, F., Wang, H., Jha, K., Gao, J., 2021. Multimodal emergent fake news detection via meta neural process networks, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 3708–3716.

Wang, Y., Qian, S., Hu, J., Fang, Q., Xu, C., 2020a. Fake news detection via knowledge-driven multimodal graph convolutional networks, in: Proceedings of the 2020 International Conference on Multimedia Retrieval, pp. 540–547.

Wang, Y., Yang, W., Ma, F., Xu, J., Zhong, B., Deng, Q., Gao, J., 2020b. Weak supervision for fake news detection via reinforcement learning, in: Proceedings of the 34th AAAI Conference on Artificial Intelligence, pp. 516–523.

Wu, K., Yang, S., Zhu, K.Q., 2015. False rumors detection on sina weibo by propagation structures, in: Proceedings of the 31st international conference on data engineering, pp. 651–662.

Wu, K., Yuan, X., Ning, Y., 2021a. Incorporating relational knowledge in explainable fake news detection, in: Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 403–415.

Wu, L., Rao, Y., Jin, H., Nazir, A., Sun, L., 2019. Different absorption from the same sharing: Sifted multi-task learning for fake news detection, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 4644–4653.

Wu, W., Li, H., Wang, H., Zhu, K.Q., 2012. Probase: A probabilistic taxonomy for text understanding, in: Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, pp. 481–492.

Wu, Y., Zhan, P., Zhang, Y., Wang, L., Xu, Z., 2021b. Multimodal fusion with co-attention networks for fake news detection, in: Proceedings of the Association for Computational Linguistics, pp. 2560–2569.

Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., Wei, L., 2021. Detecting fake news by exploring the consistency of multimodal data. Information Processing & Management 58, 102610.

Yang, C., Zhou, X., Zafarani, R., 2021. Checked: Chinese covid-19 fake news dataset. Social Network Analysis and Mining 11, 1–8.

Yang, F., Liu, Y., Yu, X., Yang, M., 2012. Automatic detection of rumor on sina weibo, in: Proceedings of the ACM SIGKDD workshop on mining data semantics, pp. 1–7.

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., Liu, H., 2019. Unsupervised fake news detection on social media: A generative approach, in: Proceedings of the AAAI conference on artificial intelligence, pp. 5644–5651.

Yang, X., Lyu, Y., Tian, T., Liu, Y., Liu, Y., Zhang, X., 2020. Rumor detection on social media with graph structured adversarial learning., in: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, pp. 1417–1423.

Yang, Y., Zheng, L., Zhang, J., Cui, Q., Li, Z., Yu, P.S., 2018. Ti-cnn: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749 .

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E., 2016. Hierarchical attention networks for document classification, in: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp. 1480–1489.

Yin, X., Han, J., Philip, S.Y., 2008. Truth discovery with multiple conflicting information providers on the web. IEEE Transactions on Knowledge and Data Engineering 20, 796–808.

Yu, F., Liu, Q., Wu, S., Wang, L., Tan, T., et al., 2017. A convolutional approach for misinformation identification., in: Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 3901–3907.

Yuan, C., Ma, Q., Zhou, W., Han, J., Hu, S., 2019. Jointly embedding the local and global relations of heterogeneous graph for rumor detection, in: Proceedings of the 2019 IEEE International Conference on Data Mining, pp. 796–805.

Yuan, C., Ma, Q., Zhou, W., Han, J., Hu, S., 2020. Early detection of fake news by utilizing the credibility of news, publishers, and users based on weakly supervised learning, in: Proceedings of the 28th International Conference on Computational Linguistics, pp. 5444–5454.

Yuan, C., Qian, W., Ma, Q., Zhou, W., Hu, S., 2021. Srlf: A stance-aware reinforcement learning framework for content-based rumor detection on social media, in: Proceedings of the International Joint Conference on Neural Networks, pp. 1–8.

Zhang, H., Fang, Q., Qian, S., Xu, C., 2019. Multi-modal knowledge-aware event memory network for social media rumor detection, in: Proceedings of the 27th ACM International Conference on Multimedia, pp. 1942–1951.

Zhang, J., Cui, L., Fu, Y., Gouza, F.B., 2018. Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751 .

Zhang, W., Gui, L., He, Y., 2021a. Supervised contrastive learning for multimodal unreliable news detection in covid-19 pandemic, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 3637–3641.

Zhang, X., Cao, J., Li, X., Sheng, Q., Zhong, L., Shu, K., 2021b. Mining dual emotion for fake news detection, in: Proceedings of the International World Wide Web Conferences, pp. 3465–3476.

Zhang, Y., Chen, W., Yeo, C.K., Lau, C.T., Lee, B.S., 2017. Detecting rumors on online social networks using multi-layer autoencoder, in: Proceedings of the 2017 IEEE Technology & Engineering Management Conference, pp. 437–441.

Zhou, X., Mulay, A., Ferrara, E., Zafarani, R., 2020a. Recovery: A multimodal repository for covid-19 news credibility research, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 3205–3212.

Zhou, X., Wu, J., Zafarani, R., 2020b. Safe: Similarity-aware multi-modal fake news detection, in: Proceedings of the 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD 2020, Springer. pp. 354–367.

Zhou, X., Zafarani, R., 2018. Fake news: A survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315 .

Zhou, X., Zafarani, R., 2019. Network-based fake news detection: A pattern-driven approach. ACM SIGKDD explorations newsletter 21, 48–60.

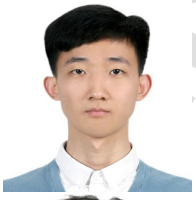
Zhou, X., Zafarani, R., 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys 53, 1–40.

Zhou, Z.H., 2018. A brief introduction to weakly supervised learning. National science review 5, 44–53.

Zubiaga, A., Liakata, M., Procter, R., 2017. Exploiting context for rumour detection in social media, in: Proceedings of the International Conference on Social Informatics, pp. 109–123.

Linmei Hu received her Ph.D degree from Tsinghua University in 2018. She has published more than 30 papers in refereed journals and conferences. Her research interests include Natural Language Processing, Knowledge Graph and Multimedia. E-mail: hulinmei1991@gmail.com.

Siqi Wei received the B.E. and M.E. degrees from NanChang University and LiaoNing University, in 2018 and 2021, respectively. He is currently pursuing the Ph.D. degree in computer science and technology at Beijing University of Posts and Telecommunications, since 2021. His research interests include deep learning, fake news detection, and social network analysis. E-mail: Weisiqi@bupt.edu.cn.



Ziwang Zhao is currently pursuing the Master's Degree in computer science and technology at Beijing University of Posts and Telecommunications, since 2021. His research interests include natural language processing and multimedia. E-mail: zhaoziwang@bupt.edu.cn.

Bin Wu received the B.E. degree from the Beijing University of Posts and Telecommunications, in 1991, and the M.E. and Ph.D. degrees from the Institute of Computing Technology of Chinese Academic of Sciences, in 1998 and 2002, respectively. He joined the Beijing University of Posts and Telecommunications as a Lecturer, in 2002, where he is currently a Professor. He has published more than 100 papers in refereed journals and conferences. His current research interests include social computing, data mining, and complex network. E-mail: wubin@bupt.edu.cn.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: