Contradicted in Reliable, Replicated in Unreliable: Dual-Source Reference for Fake News Early Detection

Yifan Feng[1], **Weimin Li**[*1], **Yue Wang**[1], **Jingchao Wang**[1], **Fangfang Liu**[1], **Zhongming Han**[2],

[1]School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[2]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
{fengyffyf, wmli, wyofficial, static, ffliu}@shu.edu.cn, hanzhongming@btbu.edu.cn

Abstract

Early detection of fake news is crucial to mitigate its negative impact. Current research in fake news detection often utilizes the difference between real and fake news regarding the support degree from reliable sources. However, it has overlooked their different semantic outlier degrees among unreliable source information during the same period. Since fake news often serves idea propaganda, unreliable sources usually publish a lot of information with the same propaganda idea during the same period, making it less likely to be a semantic outlier. To leverage this difference, we propose the Reliable-Unreliable Source Reference (RUSR) Fake News Early Detection Method. RUSR introduces the publication background for detected news, which consists of related news with common main objects of description and slightly earlier publication from both reliable and unreliable sources. Furthermore, we develop a strongly preference-driven support degree evaluation model and a two-hop semantic outlier degree evaluation model, which respectively mitigate the interference of news with weak validation effectiveness and the tightness degree of semantic cluster. The designed redistribution module and expanding range relative time encoding are adopted by both models, respectively optimizing early checkpoint of training and expressing the relevance of news implied by their release time gap. Finally, we present a multi-model mutual benefit and collaboration framework that enables the multi-model mutual benefit of generalization in training and multi-perspective prediction of news authenticity in inference. Experiments on our newly constructed dataset demonstrate the superiority of RUSR.

Introduction

Fake news refers to entirely false news reports (Rastogi and Bansal 2023). These reports influence individuals' perceptions of social (Wu et al. 2023), health (Silva et al. 2021), and other issues. As fake news spreads, it can even weaken social stability and national security (Yin et al. 2024). Early detection involves identifying the truthfulness of news at the beginning of its dissemination (Liu and Wu 2020). Thus, automating the early detection of fake news is of significant practical importance.

Fake news often imitates the content patterns of real news to avoid detection by humans and machines (Hu et al. 2021), making content-only detection methods (Wang et al. 2018; Ma et al. 2016) less effective. To address this problem, some methods incorporate reliable source information as external evidence and improve accuracy by leveraging the difference between real and fake news regarding the support degree from reliable sources (Popat et al. 2018; Vo and Lee 2021; Hu et al. 2021; Zhang et al. 2024). However, current methods ignore their different semantic outlier degrees among unreliable source information during the same period.

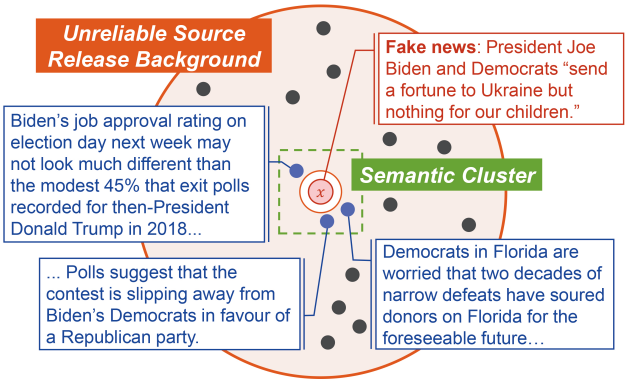

Figure 1: Fake news and its semantic cluster within its unreliable sources publication background which consist of 23 related news items with common main object of description and slightly earlier publication from unreliable sources.

Promoting a specific viewpoint for an interest group is a primary motivation for publishing fake news (Khan et al. 2022). These news may shape political views (Khan et al. 2022), propagandize ideologies (Baptista and Gradim 2021), or change consumer perceptions of products (Domenico et al. 2021). When fake news is published to promote an opinion for an interest group, the fabricator often manipulates unreliable sources to simultaneously disseminate other information supporting that opinion. Therefore, fake news typically has greater semantic cluster among unreliable source information during the same period than real news.

Figure 1 illustrates an example of fake news (2022/11/2) published before the US 2022 mid-term elections and its semantic cluster. This fake news claims that Democrats priori-

*Weimin Li is the corresponding author.

tize foreign aid over the needs of American children. Among its unreliable sources publication background, three news items also tends to support that Democrats are unfit to govern, forming a semantic cluster with the fake news. These news highlight the dissatisfaction with Biden's job performance, the decline of Democratic approval rating and the longstanding losses in Florida of the party.

To leverage this difference, we propose **R**eliable-**U**nreliable **S**ource **R**eference (RUSR) Fake News Early Detection Method, which constructs the proposed publication background of detected news by extracting related news with common main objects of description and slightly earlier publication from both reliable and unreliable sources. Based on this, a support degree evaluation model and a semantic outlier degree evaluation model are developed. They are less disturbed by background news with weak validation effectiveness from reliable sources and by the semantic cluster tightness of detected news within its unreliable source publication background, respectively. Besides, they use a new relative time encoding method, whose encoding reflects the relevance of news based on the release time gap, and employ an innovative redistribution module to optimize early checkpoint of training. Additionally, we present a multi-model mutual benefit and collaboration framework based on the above two models and a content authenticity evaluation model. The framework achieves the multi-model mutual benefit of generalization in training and multi-perspective prediction of news authenticity in inference.

Our major contributions are summarized as follows:

- We propose to apply the different semantic outlier degrees between real and fake news among unreliable source information during the same period to fake news detection.
- We propose RUSR, which analyzes the support degree and semantic outlier degree of detected news based on proposed publication background. Besides, the designed redistribution module and relative time encoding respectively optimize early checkpoint of training and imply news relevance. Finally, the proposed framework achieves multi-model mutual benefit and multi-perspective prediction.
- We propose to construct the dataset comprising verified news and news tagged with source reliability, both published within the same time period. Experiments demonstrate the superiority of our proposal.

Problem Definition

A piece of news is represented as $x = (\text{text}(x), t_x)$, where $\text{text}(x)$ and t_x denote the text and publication time of x, respectively. y_x represents the authenticity label of x, with 0 indicating fake news and 1 indicating real news. The total labeled news, news from reliable sources, and news from unreliable sources collected before t are denoted by $\mathcal{L}_t$, $\mathcal{R}_t$ and $\mathcal{U}_t$, respectively. For any given $l \in \mathcal{L}_t$, there is a large amount of news released shortly before t_l in both $\mathcal{R}_t$ and $\mathcal{U}_t$. Our goal is to develop a fake news detection function using $\mathcal{L}_{t_{\text{test}}}$, $\mathcal{R}_{t_{\text{test}}}$, and $\mathcal{U}_{t_{\text{test}}}$, ensuring good early detection performance for news published at t_{test} and thereafter.

Proposed Method

In this chapter, we introduce the proposed RUSR. First, we describe the framework of RUSR, followed by detailed explanations of each component.

Multi-Model Mutual Benefit and Collaboration Framework

The RUSR includes three models: the strongly preference-driven support degree evaluation model Model_r, the two-hop semantic outlier degree evaluation model Model_u, and the content authenticity evaluation model Model_c. During training, they share the bottom module to form a joint model (JM), enabling multi-model joint learning. During inference, the multi-perspective authenticity scores output by three models are integrated to calculate the final prediction.

Multi-Model Mutual Benefit Training The training set includes all news published before t_{val} in $\mathcal{L}_{t_{\text{test}}}$. t_{val} is earlier than t_{test}. The validation set includes the other part of $\mathcal{L}_{t_{\text{test}}}$. The architecture of JM is shown in Figure 2 and can be expressed as:

$$(v_x^{\text{c}}, v_x^{\text{r}}, v_x^{\text{u}}) = \text{JM}(\text{in}(x); \mathbf{\Theta}), \tag{1}$$

where x represents any news, $\mathbf{\Theta}$ denotes the set of all learnable parameters in JM, $\text{in}(x)$ is $(x, \mathcal{B}_x^{\text{r}}, \mathcal{B}_x^{\text{u}})$. The publication background of x is denoted as $\mathcal{B}_x$, and the news from reliable and unreliable sources in it constitute $\mathcal{B}_x^r$ and $\mathcal{B}_x^u$ respectively. The output is composed of three authenticity scores, which represent the content authenticity of x, the support degree of x from $\mathcal{B}_x^r$, and the semantic outlier degree of x in $\mathcal{B}_x^u$ respectively. $v_x^h \in (0.5, 1]$, $v_x^h \in [0, 0.5)$, and $v_x^h = 0.5$ indicate real, fake, and uncertain news, where $h \in \text{c}, \text{r}, \text{v}$.

JM consists of the news time-content encoding module M_e, the content authenticity evaluation module M_c, the support degree evaluation module M_r, and the outlier degree evaluation module M_u, which are represented as follows:

$$(\boldsymbol{e}_x, \boldsymbol{E}_x^{\text{r}}, \boldsymbol{E}_x^{\text{u}}) = \text{M}_\text{e}(\text{in}(x); \mathbf{\Theta}^{\text{e}}), \tag{2}$$

$$v_x^{\text{c}} = \text{M}_\text{c}(\boldsymbol{e}_x; \mathbf{\Theta}^{\text{c}}), \tag{3}$$

$$v_x^{\text{r}} = \text{M}_\text{r}(\boldsymbol{e}_x, \boldsymbol{E}_x^{\text{r}}; \mathbf{\Theta}^{\text{r}}), \tag{4}$$

$$v_x^{\text{u}} = \text{M}_\text{u}(\boldsymbol{e}_x, \boldsymbol{E}_x^{\text{u}}; \mathbf{\Theta}^{\text{u}}), \tag{5}$$

where $\boldsymbol{e}_x \in \mathbb{R}^{dim}$, $\boldsymbol{E}_x^{\text{r}} \in \mathbb{R}^{|\mathcal{B}_x^{\text{r}}| \times dim}$, and $\boldsymbol{E}_x^{\text{u}} \in \mathbb{R}^{|\mathcal{B}_x^{\text{u}}| \times dim}$ are the time-content features of x and the matrices composed of time-content features of news in $\mathcal{B}_x^{\text{r}}$ and $\mathcal{B}_x^{\text{u}}$, respectively. $dim = 256$. $\mathbf{\Theta}^{\text{e}}$, $\mathbf{\Theta}^{\text{c}}$, $\mathbf{\Theta}^{\text{r}}$, $\mathbf{\Theta}^{\text{u}}$ are sets of modules' learnable parameters.

The sub-models are defined as follows:

$$\text{Model}_h(\text{in}(x); \mathbf{\Theta}^h, \mathbf{\Theta}^{\text{e}}) = v_x^h, \tag{6}$$

where Model_h contains M_e and M_h.

For any training news l, the loss function $L(v_l^h, y_l)$ on single output v_l^h is defined as follows:

$$L(\cdot, \cdot) = \max\left\{-\ln\left(v_l^h \times (-1)^{(y_l+1)} + 1 - y_l + \varepsilon\right), 0\right\}, \tag{7}$$

where ε is a very small positive number. The max function is used to prevent negative loss. The loss on l is the sum of $L(v_l^c, y_l)$, $L(v_l^r, y_l)$ and $L(v_l^u, y_l)$.

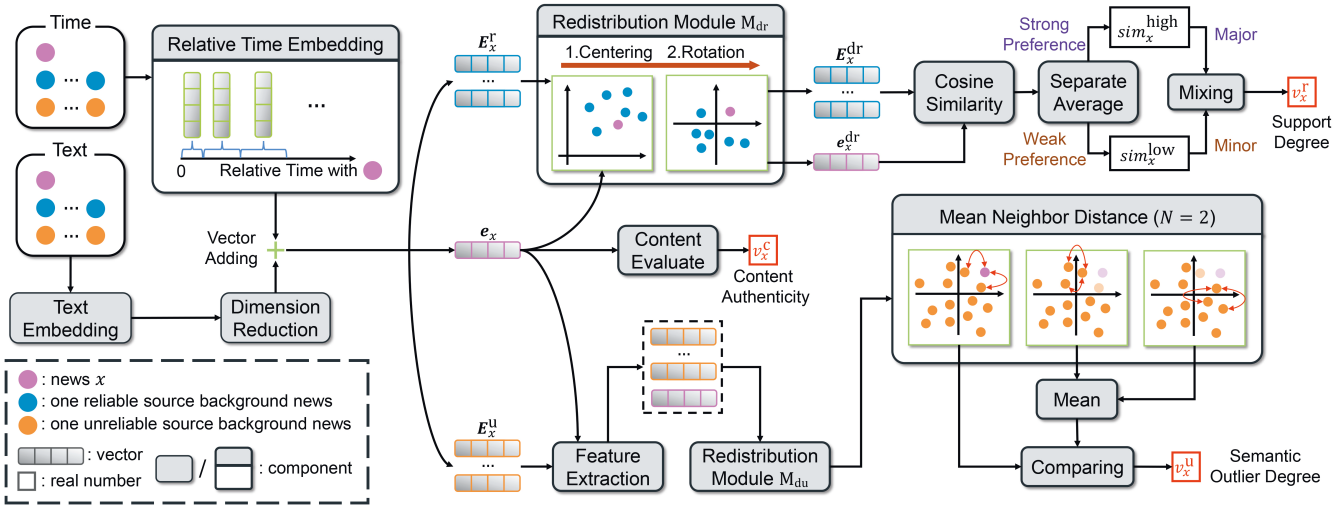

Figure 2: The overall architecture of the proposed JM model.

After the k-th, $k \geq 0$ training epoch, the parameter set of JM is denoted as $\mathbf{\Theta}_k = \mathbf{\Theta}^{\mathrm{e}}_k \cup \mathbf{\Theta}^{\mathrm{c}}_k \cup \mathbf{\Theta}^{\mathrm{r}}_k \cup \mathbf{\Theta}^{\mathrm{u}}_k$, and the accuracies of the three outputs on the validation set are denoted as acc^{c}_k,acc^{r}_k,acc^{u}_k[1].

Multi-Model Collaboration Inference Let acc^h_k reach its maximum when k equals k_h. Consequently, when $\mathbf{\Theta}^h$ and $\mathbf{\Theta}^{\mathrm{e}}$ are $\mathbf{\Theta}^h_{k_h}$ and $\mathbf{\Theta}^{\mathrm{e}}_{k_h}$, respectively, Model_h achieves optimal future generalization.

While detecting news x, for each model Model_h, RUSR calculates $\mathrm{Model}_h(\mathrm{in}(x); \mathbf{\Theta}^h_{k_h}, \mathbf{\Theta}^{\mathrm{e}}_{k_h})$. The real, fake or unknown result leads to increasing the score of real or fake class by $acc^h_{k_h}$, or remaining their scores, respectively. After the calculation of three models, the prediction of RUSR is real news if real class score is larger than fake class score. Otherwise, the prediction is fake news.

The significant differences in effective inputs and structures ensure that three models are unlikely to produce the same prediction for one news, ensuring their effective collaboration.

Publication Background Construction

RUSR selects related news in content and time from $\mathcal{R}_{t_x}$ and $\mathcal{U}_{t_x}$ to construct $\mathcal{B}_x$.

News with Common Main Objects of Description Selecting The content-related news set $\mathcal{C}_x$ for news x is defined as the news in $\mathcal{R}_{t_x}$ and $\mathcal{U}_{t_x}$ that share common main objects of description with x. The method to obtain the main objects of description set $\mathcal{O}_m$ for any news m is as follows:

The *en_core_web_sm* pipeline[2] from the *spacy* package is used to process $\mathrm{text}(m)$. First, the text is converted into a sequence of tokens, where each token is a word or punctuation from the text. Then, each word token is tagged with its part of speech and their base forms are obtained. Finally, the named entities consisting of tokens are identified in $\mathrm{text}(m)$.

For each named entity, the base forms of its tokens are concatenated with space to get the base form of the named entity. After that, we extract one sequence consisting of the base forms of all named entities and another composed of the base forms of noun tokens that are not part of any named entity. We concatenate the two sequences, then analyze the frequency of each unique string in the combined sequence. Ultimately, we select up to five strings with the highest frequency, and form a set with them as $\mathcal{O}_m$.

Temporally Close News Selecting Due to the varying popular topics, news in $\mathcal{C}_x$ that is temporally closer to x is more likely to be related to x. Therefore, we retain up to $N^{\mathrm{B}}_{\max} = 30$ latest news in $\mathcal{C}_x$ to construct $\mathcal{B}_x$. An exception is when x is in the training set, $\mathcal{B}_x$ consists of up to $N^{\mathrm{B}}_{\max}$ random news from the latest $1.1 \times N^{\mathrm{B}}_{\max}$ news in $\mathcal{C}_x$. Thus, $\mathcal{B}_x$ varies in each epoch to enhance generalization.

News Time-Content Encoding Module

For any $m \in \{x\} \cup \mathcal{B}_x$, we first generate its low-dimensional content vector $\boldsymbol{e}^{\mathrm{low}}_m$. Then, we encode the relative time from m to x into a vector $\boldsymbol{e}^{\mathrm{time}}_{d_{m \to x}}$ using expanding range relative time encoding. The sum of these two vectors gives the time-content encoding $\boldsymbol{e}_m$.

Low-Dimensional Content Vector Counting First, the *all-mpnet-base-v2* model[3] from the *sentence-transformers* package is used to generate a 768-dimensional high-dimensional content vector $\boldsymbol{e}^{\mathrm{high}}_m$ based on $\mathrm{text}(m)$. The parameters of this model are not trained and are not in $\mathbf{\Theta}^{\mathrm{e}}$.

For generalization, a two-layer MLP is used to reduce the dimensions of $\boldsymbol{e}^{\mathrm{high}}_m$, resulting in $\boldsymbol{e}^{\mathrm{low}}_m$. The MLP progres-

[1]In this paper, when calculating one metric on a given dataset for a sub-model or a specific output of JM, samples with an output value of 0.5 are excluded.

[2]https://spacy.io/models/en#en_core_web_sm

[3]https://sbert.net/docs/pretrained_models.html

sively compresses the dimensions to 512 and dim dimensions. The first layer uses ReLU as the activation function, while the second layer has no activation function.

Expanding Range Relative Time Encoding Calculating support and outlier degree needs to compute the semantic similarity between x and each background news based on their time-content encodings. To take news semantic relevance implied by the release time gap into account when computing similarity, we include the encoding of the release time gap between m and x in the time-content emcoding of m and make the time encoding reflect their semantic relevance implied by the gap.

Let $d_{m\to x}$ be the number of days in the gap, representing the release time gap. We divide all possible values of $d_{m\to x}$ into 22 segments and define a learnable dim-dimensional parameter vector for each segment. The vector corresponding to the segment where $d_{m\to x}$ falls is $\boldsymbol{e}^{\text{time}}_{d_{m\to x}}$. The segment index $idx_{m\to x}$ of $d_{m\to x}$ is:

$$idx_{m\to x} = \begin{cases} d_{m\to x}, & 0 \le d_{m\to x} < 7, \\ 6 + \lfloor d_{m\to x}/7 \rfloor, & 7 \le d_{m\to x} < 28, \\ 9 + \lfloor d_{m\to x}/28 \rfloor, & 28 \le d_{m\to x} < 28 \times 12, \\ 21, & d_{m\to x} \ge 28 \times 12. \end{cases} \tag{8}$$

The reason for designing expanding range segments is that we want $\boldsymbol{e}^{\text{time}}_{d_{m\to x}}$ to express the semantic relevance between x and m implied by $d_{m\to x}$. The relevance is denoted as the temporal relevance between them. Intuitively, the temporal relevance decreases as $d_{m\to x}$ increases, with the rate of decrease being initially fast and then slower. Therefore, for a positive integer Δd, the smaller the value of $d_{m\to x}$, the greater the difference in temporal relevance between $d_{m\to x}$ and $d_{m\to x} + \Delta d$, and hence, $\boldsymbol{e}^{\text{time}}_{d_{m\to x}}$ and $\boldsymbol{e}^{\text{time}}_{d_{m\to x}+\Delta d}$ should be less similar.

Content Authenticity Evaluation Module

M_c is defined as:

$$\boldsymbol{a}_x = \text{Linear\_Layer}\left(\text{ReLU}\left(\boldsymbol{e}_x\right)\right), \tag{9}$$

$$v^{\text{c}}_x = \frac{e^{a^1_x}}{e^{a^0_x} + e^{a^1_x}}, \tag{10}$$

where Linear_Layer don't have activation function, and $\boldsymbol{a}_x$ is a vector with two dimensions.

Strongly Preference-Driven Support Degree Evaluation Module

M_r first calculates the M_r redistributed time-content vectors for each time-content vector in $\boldsymbol{e}_x$ and $\boldsymbol{E}^{\text{r}}_x$ through the redistribution module M_{dr}, resulting in $\boldsymbol{e}^{\text{dr}}_x \in \mathbb{R}^{dim}$ and $\boldsymbol{E}^{\text{dr}}_x \in \mathbb{R}^{|\mathcal{B}^{\text{r}}_x|\times dim}$, respectively. M_{dr} consists of centering and rotation. During centering, the element-wise mean $\boldsymbol{mean}^{\text{r}}_x$ of the $|\mathcal{B}^r_x| + 1$ vectors from $\boldsymbol{e}_x$ and $\boldsymbol{E}^{\text{r}}_x$ is computed. Then, for any $n \in \{x\} \cup \mathcal{B}^{\text{r}}_x$, the time-content encoding of n minus $\boldsymbol{mean}^{\text{r}}_x$ element-wise resulting in $\boldsymbol{e}^{\text{mid}}_n$. During rotation, the M_r redistributed time-content vector of n is represented as:

$$\boldsymbol{e}^{\text{dr}}_n = \frac{\boldsymbol{e}^{\text{mid}}_n}{\|\boldsymbol{e}^{\text{mid}}_n\|_2} + \boldsymbol{b}^{\text{r}} + \frac{\boldsymbol{e}^{\text{mid}}_n}{\|\boldsymbol{e}^{\text{mid}}_n\|_2} \odot \boldsymbol{m}^{\text{r}}. \tag{11}$$

where $\boldsymbol{m}^{\text{r}}, \boldsymbol{b}^{\text{r}} \in \mathbb{R}^{dim}$ are learnable parameter vectors, and $\odot$ denotes element-wise multiplication. The role of the redistribution modules in M_r and M_u will be described in the experiments.

Next, to assess the degree of semantic similarity or opposition between x and each $m \in \mathcal{B}^{\text{r}}_x$, the cosine similarity between $\boldsymbol{e}^{\text{dr}}_m$ and $\boldsymbol{e}^{\text{dr}}_x$ is calculated and denoted as $sim^{\text{r}}_{m,x}$. If $|sim^{\text{r}}_{m,x}| > 0.7$, it is considered a strong preference similarity, indicating m has strong validation effectiveness. Otherwise, they are weak. The mean of the strong and weak preference similarities are sim^{high}_x and sim^{low}_x, respectively (if strong or weak preference similarity does not exist, the corresponding mean is 0). v^{r}_x is calculated as follows:

$$v^{\text{r}}_x = (0.9 \times sim^{\text{high}}_x + 0.1 \times sim^{\text{low}}_x + 1)/2. \tag{12}$$

where 0.1 suppress the interference of weak validation effectiveness news. Additionally, if $|\mathcal{B}^r_x| = 0$, x is deemed unmanageable for M_r. So v^{r}_x is not calculated in M_r, and 0.5 is set.

Two-Hop Semantic Outlier Degree Evaluation Module

M_u considers the news in $\mathcal{G}_x = \{x\} \cup \mathcal{B}^{\text{u}}_x$ as a whole and calculates the semantic outlier degree of x in $\mathcal{B}^{\text{u}}_x$:

$$v^{\text{u}}_x = (ndif_x - \frac{\sum_{n\in\mathcal{G}^N_x}(ndif^x_n)}{N} + 2)/4, \tag{13}$$

where $\mathcal{G}^N_x$ is the set of the $N = 5$ nearest semantic neighbors of x in $\mathcal{B}^{\text{u}}_x$, and $ndif_x$ and $ndif^x_n$ are defined as:

$$ndif_x = \text{ndif}^N_x(x, \mathcal{G}_x), \tag{14}$$

$$ndif^x_n = \text{ndif}^N_x(n, (\mathcal{B}^{\text{u}}_x \backslash \mathcal{G}^N_x) \cup \{n\}), \tag{15}$$

where $\mathcal{B}^{\text{u}}_x \backslash \mathcal{G}^N_x$ represents the relative complement of $\mathcal{G}^N_x$ in $\mathcal{B}^{\text{u}}_x$. The function $\text{ndif}^N_x(m, \mathcal{S})$ is defined as:

$$\text{ndif}^N_x(m, \mathcal{S}) = \frac{\sum_{n\in\mathcal{S}^N_m} \text{dis}_x(m, n)}{N}. \tag{16}$$

Its domain is $m \in \mathcal{S}$, and $\mathcal{S} \subseteq \mathcal{G}_x$. Here, $\mathcal{S}^N_m$ is the set of the N nearest semantic neighbors of m in S. In M_u, the semantic distance function used for calculating the nearest semantic neighbors and semantic distances is $\text{dis}_x(m, n)$, whose domain is $m, n \in \mathcal{G}_x$ and range is $[0, 2]$.

To avoid the sensitivity of v^{u}_x to the tightness degree of the semantic cluster of x in $\mathcal{G}_x$, we use relative outlier degree instead of $ndif_x/2$ as v^{u}_x. The tightness degree is based on pairwise semantic distances. To ensure the accessibility of smaller values in $[0, 1]$ for v^{u}_x, $\mathcal{G}^N_x$ is excluded when calculating $ndif^x_n$. Otherwise, when $ndif_x$ is small, the pairwise distances in x and the news in $\mathcal{G}^N_x$ are not far generally, making it difficult for $ndif^x_n, n \in \mathcal{G}^N_x$ to reach large values.

To construct the function $\text{dis}_x(m, n)$, for any $m \in \mathcal{G}_x$, calculate the deep time-content feature vector:

$$\tilde{\boldsymbol{e}}_m = \text{Linear\_Layer}\left(\text{ReLU}\left(\boldsymbol{e}_m\right)\right), \tag{17}$$

where $\boldsymbol{e}_m$ is the time-content feature of m, $\tilde{\boldsymbol{e}}_m \in \mathbb{R}^{dim}$ and Linear_Layer don't have activation function. Then, using the

redistribution module $\mathrm{M_{du}}$, we calculate a $\mathrm{M_u}$ redistributed time-content vector for each deep time-content vector in $\tilde{\boldsymbol{e}}_x$ and $\tilde{\boldsymbol{E}}_x^{\mathrm{u}} \in \mathbb{R}^{|\mathcal{B}_x^{\mathrm{u}}| \times dim}$, resulting in $\boldsymbol{e}_x^{\mathrm{du}} \in \mathbb{R}^{dim}$ and $\boldsymbol{E}_x^{\mathrm{du}} \in \mathbb{R}^{|\mathcal{B}_x^{\mathrm{u}}| \times dim}$. $\mathrm{dis}_x(m,n)$ is defined as:

$$\mathrm{dis}_x(m,n) = -\mathrm{cos\_sim}(\boldsymbol{e}_m^{\mathrm{du}}, \boldsymbol{e}_n^{\mathrm{du}}) + 1, \tag{18}$$

where $\boldsymbol{e}_m^{\mathrm{du}}$ and $\boldsymbol{e}_n^{\mathrm{du}}$ are the $\mathrm{M_u}$ redistributed time-content vectors of m and n.

If $|\mathcal{B}_x^{\mathrm{u}}|$ is too small, it is difficult to identify x as real or fake news based on its semantic cluster size in $\mathcal{G}_x$. Therefore, if $|\mathcal{B}_x^{\mathrm{u}}| < M \times N$, x is deemed unmanageable for $\mathrm{M_u}$. So v_x^{u} is not calculated in $\mathrm{M_u}$, and 0.5 is set.

Experiments

Dataset

Existing fake news detection datasets do not include verified news and news with source reliability label published within the same period. Therefore, we constructed a new English dataset. Besides this innovation, the advantages of our dataset are: 1) The verified news is recent (January 2019 to October 2023), reflecting updated forgery methods. 2) News with source reliability label includes summarized text to highlight key points and avoid token limits of text models. 3) It provides publication dates and main objects of description for all news, facilitating correlation analysis. 4) It offers vectors for the text of all news and the names of main objects of description to facilitate research, with the calculation method in method section. The construction details are as follows:

Verified News We scraped verified news texts and publication dates from fact-checking websites Politifact and Snopes. Since Snopes does not have the latter, we used the publication dates of fact-checking articles. For binary classification tasks and balanced categories, we labeled news marked as true and mostly-true on Politifact, and True and Mostly True on Snopes as real news, and news marked as pants-fire on Politifact, and False and Fake on Snopes as fake news, excluding other labels. Totally, the dataset contains 2,602 real news and 3,253 fake news.

News with Source Reliability Label We collected 33,212 reliable source news and 114,600 unreliable source news published from January 2019 to December 2022 from the nela-gt2019-2022 (Gruppi, Horne, and Adalı 2020, 2021, 2022, 2023) datasets, with reliability labels from nela-gt2020. Using the BART model[4], we summarized the body text of each news and removed unexpected characters in the summaries, which we provided as the news texts in our dataset.

Settings

To validate the superiority of RUSR in real-world early detection tasks, we designed four experiments (Q1-Q4) inspired by (Hu et al. 2023) to correspond to four instances of the proposed problem. The four instances set t_{test} as the first day of each quarter in 2022. $\mathcal{L}_t$, $\mathcal{R}_t$, and $\mathcal{U}_t$ are the verified news, reliable source news and unreliable source news published before t in our dataset, respectively. The test set consists of verified news from the corresponding quarter, while the validation and training sets consist of verified news from the preceding quarter and earlier, respectively. Results on the test set are presented.

Baseline Fake news detection baselines include (1) $\mathrm{EANN_T}$ (Wang et al. 2018), which uses the publication years of the news as auxiliary task labels to produce time-invariant features, upgrading TextCNN to pre-trained BERT and removing the image part as (Hu et al. 2023). (2) Emo (Zhang et al. 2021), which detects news based on the sentiment of the publisher, excluding the sentiment of news comments. (3) BERT+FTT (Hu et al. 2023), focusing on the model's performance on high-frequency future news topics. (4) DeClarE (Popat et al. 2018) and MAC (Vo and Lee 2021), which detect based on news content and evidence. Due to the absence of news sources and low variability in evidence sources, the use of these sources is removed. (5) Emo+NEP and DeClarE+NEP (Sheng et al. 2022), which additionally consider the popularity and novelty of detected news in its news environment. (1)(2)(3) are content-only methods.

Implementation Details We use the RAdam optimizer with a learning rate of 0.0002 and a batch size of 64. Due to the relatively small dataset, pre-trained models for all methods are not fine-tuned during training. For each detected news, up to 5 reliable source news published before are retrieved as evidence in the baselines using evidence. Methods using NEP obtain the news environment from news with source reliability label.

Performance Comparison

The experimental results of Q1-Q4 and the average results are reported in Table 1, with the best values in each row highlighted in bold. The experiments indicate that the methods using evidence do not necessarily outperform content-only methods, and adding NEP does not always enhance performance. This may be because introducing evidence or news environment adds complexity to the model structure while supplementing information for detection. As news evolves rapidly over time, splitting the training, validation, and test news by time results in significant distribution difference between training and test news (Hu et al. 2023). More complex models tend to exhibit greater performance drops from training to testing in the condition. RUSR is the best in all average metrics. This improvement is firstly attributed to the support and outlier degree evaluation models, whose generalization stems from the interpretability of considering their output as authenticity, and benefits from the proposed publication background, redistribution module and time embedding. Additionally, the multi-model mutual benefit and collaboration framework plays a crucial role in the effectiveness of RUSR.

Ablation Analysis

To validate the three models and framework, we compare the performance of RUSR with three methods of removing one model in Q1-Q4 and present the average perfor-

[4]https://huggingface.co/facebook/bart-large-cnn

2022	Metric	Emo	EANN$_T$	DeClarE	MAC	DeClarE +NEP	Emo +NEP	BERT +FTT	RUSR
Q1	macF1	0.703	0.650	0.676	0.693	0.682	0.677	0.682	**0.712**
	Acc	0.712	0.651	0.681	0.703	0.686	0.681	0.694	**0.716**
	F1$_{fake}$	**0.754**	0.667	0.718	0.748	0.714	0.714	0.745	0.745
	F1$_{real}$	0.653	0.633	0.633	0.638	0.650	0.640	0.620	**0.680**
Q2	macF1	0.696	0.678	0.649	0.673	0.673	0.640	0.674	**0.706**
	Acc	0.699	0.678	0.653	0.674	0.674	0.640	0.674	**0.708**
	F1$_{fake}$	0.726	0.675	0.685	0.691	0.691	0.635	0.675	**0.729**
	F1$_{real}$	0.667	0.681	0.613	0.655	0.655	0.644	0.672	**0.682**
Q3	macF1	0.700	0.616	0.665	0.621	0.661	0.664	0.666	**0.743**
	Acc	0.705	0.633	0.667	0.638	0.662	0.676	0.676	**0.748**
	F1$_{fake}$	0.737	0.698	0.690	0.701	0.643	0.728	0.724	**0.778**
	F1$_{real}$	0.663	0.533	0.639	0.542	0.679	0.600	0.609	**0.707**
Q4	macF1	0.752	0.668	0.697	0.721	**0.756**	0.696	0.712	0.745
	Acc	0.758	0.686	0.716	0.733	**0.763**	0.703	0.720	**0.763**
	F1$_{fake}$	0.791	0.747	0.773	0.779	0.797	0.743	0.761	**0.812**
	F1$_{real}$	**0.714**	0.589	0.621	0.633	**0.714**	0.650	0.663	0.678
AVG	macF1	0.713	0.653	0.672	0.677	0.693	0.669	0.684	**0.727**
	Acc	0.719	0.662	0.679	0.687	0.696	0.675	0.691	**0.734**
	F1$_{fake}$	0.752	0.697	0.716	0.730	0.711	0.705	0.726	**0.766**
	F1$_{real}$	0.674	0.609	0.627	0.625	0.675	0.634	0.641	**0.687**

Table 1: The performance of the baseline methods and RUSR.

Method	macF1	Acc	F1$_{fake}$	F1$_{real}$
w/o Model$_c$	0.692	0.705	0.752	0.632
w/o Model$_r$	0.710	0.723	**0.768**	0.653
w/o Model$_u$	0.714	0.721	0.754	0.674
w/o time	0.709	0.720	0.764	0.653
w/o M$_{dr}$	0.693	0.701	0.736	0.650
w/o M$_{du}$	0.712	0.717	0.744	0.680
RUSR	**0.727**	**0.734**	0.766	**0.687**

Table 2: Ablation studies of RUSR. The maximum value in a column is bolded.

Model	w/o Model$_c$	w/o Model$_r$	w/o Model$_u$	RUSR
Model$_c$	None	0.710	0.715	**0.721**
Model$_r$	**0.690**	None	0.687	**0.690**
Model$_u$	0.668	0.688	None	**0.690**

Table 3: The impact of single model ablation on other models. The maximum value in a row is bolded.

mance of sub-model predictions and method predictions across four experiments in Table 2 and 3. Table 3 shows the average macF1 for the sub-models. The results indicate: (1) All sub-models are important, as removing any one results in method performance decline. (2) The generalization of other models decreases when a single model is removed, with the exception that the performance of Model$_r$ remains unchanged when Model$_c$ is removed, demonstrating the rationality of the multi-model joint training. (3) The performance of RUSR is superior to any single model, indicating the effectiveness of multi-model collaboration.

For relative time encoding, we compare RUSR with a variant without it across four experiments, with average results in Table 2. Performance drops significantly without this encoding, showing the importance of news relevance implied by the release time gap.

Then we design two variants, each removing one redistribution module, and present the mean results across Q1-Q4 in Table 2. The results show that RUSR outperforms both variants. To analyze the reasons for this optimization, experiments are conducted on Q4. w/o M$_{dr}$ and w/o M$_{dr}^{rotate}$ refer to the removal of M$_{dr}$ and the removal of only the rotation step, respectively. Similarly for w/o M$_{du}$ and w/o M$_{du}^{rotate}$. RUSR, w/o M$_{dr}$ and w/o M$_{dr}^{rotate}$: After the 0th epoch, for each M$_r$-processed real validation news we collect the cosine similarities in M$_r$ and calculate the frequency of similarity falling into each of the four ranges. Finally, frequencies is averaged over all validation news for each range. Similarly for calculating the average frequencies of fake validation news. RUSR, w/o M$_{du}$ and w/o M$_{du}^{rotate}$: We similarly calculate the average frequencies for real and fake validation news, replacing all M$_r$ with M$_u$. The results are shown in Table 4 with the model performance.

The results show that removing the redistribution module leads to performance loss. In inference, M$_r$ is expected to have a similarity distribution towards 1 for real news and towards -1 for fake news. Additionally, M$_u$ is expected to have a similarity distribution far from 1 for real news, and for fake news, N similarities close to 1, with the rest far from 1. Table 4 shows that the average similarity distribution of real and fake news for w/o M$_{dr}$ is dominated by $[0.5, 1]$, deviating significantly from the expected distribution. The inclusion of M$_{dr}$ optimizes both distributions, reducing the deviation. Similarly for w/o M$_{du}$ and M$_{du}$.

Method	Model_r macF1	M_r distribution(%) [-1,-0.5)/[-0.5,0)/[0,0.5)/[0.5,1]	
		real	fake
w/o M_dr	0.711	0/0/0/100	0/0/0/100
w/o $\text{M}_\text{dr}^\text{rotate}$	0.382	12/80/8/0	15/78/7/0
RUSR	0.738	8/80/13/0	12/74/15/0
Method	Model_u macF1	M_u distribution(%) [-1,-0.5)/[-0.5,0)/[0,0.5)/[0.5,1]	
		real	fake
w/o M_du	0.616	0/0/0/100	0/0/0/100
w/o $\text{M}_\text{du}^\text{rotate}$	0.705	1/62/37/0	0/64/36/0
RUSR	0.752	0/56/44/0	0/54/46/0

Table 4: Ablation studies of redistribution module and its rotation part.

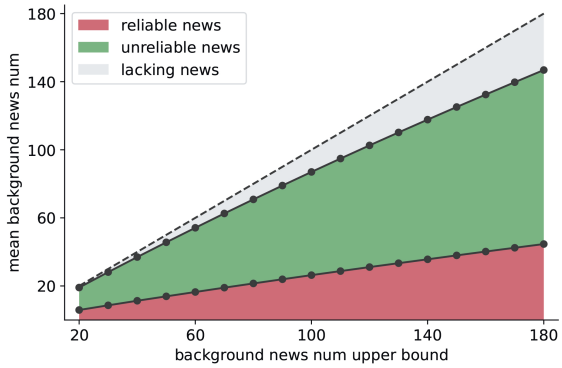

Figure 3: The average number of background news changes with N_max^B. Gray means the difference between N_max^B and the sum of two averages.

The inclusion of the centering step in w/o $\text{M}_\text{dr}^\text{rotate}$ and w/o $\text{M}_\text{du}^\text{rotate}$ also reduces the deviation but causes the feature vectors of detected news and its background news used for similarity calculation to tend to be distributed around the origin, hindering the realization of the expected distribution during inference. The addition of the rotation component eliminates this problem, and thus the results show that both methods have lower performance than RUSR.

Publication Background Discussion

First, we present the average number of reliable and unreliable source background news for verified news published before 2023 as N_max^B changes from 20 to 180 in increments of 10. Common background construction is conducted. As shown in the Figure 3, two average numbers increase significantly with the increasement of N_max^B. To study the impact of N_max^B, Figure 4 shows the average macF1 of the three output modules and RUSR in Q1-Q4 as N_max^B changes. The following conclusions can be drawn: (1) As the proportion of weakly related news in reliable source background increases, the macF1 of M_r remains generally stable, reflecting the rationality of focusing on news with strong validation effectiveness. (2) As N_max^B increases from 110, unlike

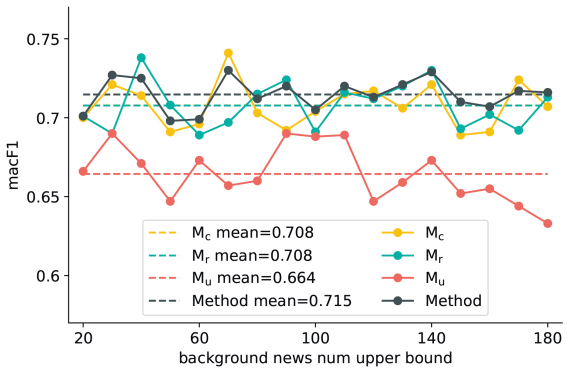

Figure 4: Performance changes as N_max^B increases. For one solid line, the dashed line with the same color indicates the mean of y-coordinates.

the previous relatively stable phase, the performance of M_u declines. This may be due to the fact that, at this stage, each N_max^B increase in an experiment leads to a general rise in unreliable source background news which is not in the period of the detected news. This causes the difference between the semantic cluster size of real and fake news to narrow. (3) The performance of RUSR is higher on average across all N_max^B than the three output modules, indicating the effectiveness of the multi-model collaboration strategy.

Related Work

We categorize fake news detection methods suitable for early detection into two types: **Content-only**: Many methods focus on text modalty (Ma et al. 2016; Ajao, Bhowmik, and Zargari 2019; Przybyla 2020; Karimi and Tang 2019) or image modalty (Jin et al. 2017; Qi et al. 2019), while others incorporate both modalities to enhance detection (Wang et al. 2018; Qi et al. 2021; Chen et al. 2022; Zhou et al. 2023; Wang et al. 2024). **Reference-enhanced**: Most methods use knowledge as supplementary information (Hu et al. 2021; Qian et al. 2021; Popat et al. 2018; Vo and Lee 2021; Zhang et al. 2024; Ma et al. 2019), while (Sheng et al. 2022) enhances detection by perceiving the popularity and novelty of the detected news within its news environment.

Conclusion

In this paper, we introduce RUSR, which calculates the support degree driven by strong preferences and the two-hop semantic outlier degree with the proposed publication background. Besides, the designed redistribution module and expanding range relative time encoding respectively optimize early checkpoint of training and imply news relevance. Ultimately, the proposed framework ensures mutual benefit among multiple models and supports multi-perspective prediction. Experimental results demonstrate the superiority of our proposed method.

Acknowledgements

This work was supported by National Key Research and Development Program of China (No. 2022YFC3302600).

References

Ajao, O.; Bhowmik, D.; and Zargari, S. 2019. Sentiment Aware Fake News Detection on Online Social Networks. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2507–2511.

Baptista, J. P.; and Gradim, A. 2021. "Brave New World" of Fake News: How It Works. *Javnost - The Public*, 28(4): 426–443.

Chen, Y.; Li, D.; Zhang, P.; Sui, J.; Lv, Q.; Tun, L.; and Shang, L. 2022. Cross-modal Ambiguity Learning for Multimodal Fake News Detection. In *Proceedings of the ACM Web Conference 2022*, WWW '22, 2897–2905. New York, NY, USA: Association for Computing Machinery. ISBN 9781450390965.

Domenico, G. D.; Sit, J.; Ishizaka, A.; and Nunan, D. 2021. Fake news, social media and marketing: A systematic review. *Journal of Business Research*, 124: 329–341.

Gruppi, M.; Horne, B. D.; and Adalı, S. 2020. NELA-GT-2019: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv:2003.08444.

Gruppi, M.; Horne, B. D.; and Adalı, S. 2021. NELA-GT-2020: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv:2102.04567.

Gruppi, M.; Horne, B. D.; and Adalı, S. 2022. NELA-GT-2021: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv:2203.05659.

Gruppi, M.; Horne, B. D.; and Adalı, S. 2023. NELA-GT-2022: A Large Multi-Labelled News Dataset for The Study of Misinformation in News Articles. arXiv:2203.05659.

Hu, B.; Sheng, Q.; Cao, J.; Zhu, Y.; Wang, D.; Wang, Z.; and Jin, Z. 2023. Learn over Past, Evolve for Future: Forecasting Temporal Trends for Fake News Detection. In Sitaram, S.; Beigman Klebanov, B.; and Williams, J. D., eds., *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track)*, 116–125. Toronto, Canada: Association for Computational Linguistics.

Hu, L.; Yang, T.; Zhang, L.; Zhong, W.; Tang, D.; Shi, C.; Duan, N.; and Zhou, M. 2021. Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge. In Zong, C.; Xia, F.; Li, W.; and Navigli, R., eds., *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 754–763. Online: Association for Computational Linguistics.

Jin, Z.; Cao, J.; Zhang, Y.; Zhou, J.; and Tian, Q. 2017. Novel Visual and Statistical Image Features for Microblogs News Verification. *IEEE Transactions on Multimedia*, 19(3): 598–608.

Karimi, H.; and Tang, J. 2019. Learning Hierarchical Discourse-level Structure for Fake News Detection. In Burstein, J.; Doran, C.; and Solorio, T., eds., *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 3432–3442. Minneapolis, Minnesota: Association for Computational Linguistics.

Khan, S. A.; Shahzad, K.; Shabbir, O.; and Iqbal, A. 2022. Developing a Framework for Fake News Diffusion Control (FNDC) on Digital Media (DM): A Systematic Review 2010–2022. *Sustainability*, 14(22).

Liu, Y.; and Wu, Y.-F. B. 2020. FNED: A Deep Network for Fake News Early Detection on Social Media. *ACM Trans. Inf. Syst.*, 38(3).

Ma, J.; Gao, W.; Joty, S.; and Wong, K.-F. 2019. Sentence-Level Evidence Embedding for Claim Verification with Hierarchical Attention Networks. In Korhonen, A.; Traum, D.; and Màrquez, L., eds., *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2561–2571. Florence, Italy: Association for Computational Linguistics.

Ma, J.; Gao, W.; Mitra, P.; Kwon, S.; Jansen, B.; Wong, K.; and Cha, M. 2016. Detecting rumors from microblogs with recurrent neural networks. *IJCAI International Joint Conference on Artificial Intelligence*, 2016-January: 3818–3824. 25th International Joint Conference on Artificial Intelligence, IJCAI 2016 ; Conference date: 09-07-2016 Through 15-07-2016.

Popat, K.; Mukherjee, S.; Yates, A.; and Weikum, G. 2018. DeClarE: Debunking Fake News and False Claims using Evidence-Aware Deep Learning. In Riloff, E.; Chiang, D.; Hockenmaier, J.; and Tsujii, J., eds., *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 22–32. Brussels, Belgium: Association for Computational Linguistics.

Przybyla, P. 2020. Capturing the Style of Fake News. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01): 490–497.

Qi, P.; Cao, J.; Li, X.; Liu, H.; Sheng, Q.; Mi, X.; He, Q.; Lv, Y.; Guo, C.; and 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*, MM '21, 1212–1220. New York, NY, USA: Association for Computing Machinery. ISBN 9781450386517.

Qi, P.; Cao, J.; Yang, T.; Guo, J.; and Li, J. 2019. Exploiting Multi-domain Visual Information for Fake News Detection. In *2019 IEEE International Conference on Data Mining (ICDM)*, 518–527.

Qian, S.; Hu, J.; Fang, Q.; and Xu, C. 2021. Knowledge-aware Multi-modal Adaptive Graph Convolutional Networks for Fake News Detection. *ACM Trans. Multimedia Comput. Commun. Appl.*, 17(3).

Rastogi, S.; and Bansal, D. 2023. A review on fake news detection 3T's: typology, time of detection, taxonomies. *International Journal of Information Security*, 22(1): 177–212.

Sheng, Q.; Cao, J.; Zhang, X.; Li, R.; Wang, D.; and Zhu, Y. 2022. Zoom Out and Observe: News Environment Perception for Fake News Detection. In Muresan, S.; Nakov, P.; and Villavicencio, A., eds., *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 4543–4556. Dublin, Ireland: Association for Computational Linguistics.

Silva, A.; Luo, L.; Karunasekera, S.; and Leckie, C. 2021. Embracing Domain Differences in Fake News: Cross-domain Fake News Detection using Multi-modal Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(1): 557–565.

Vo, N.; and Lee, K. 2021. Hierarchical Multi-head Attentive Network for Evidence-aware Fake News Detection. In Merlo, P.; Tiedemann, J.; and Tsarfaty, R., eds., *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 965–975. Online: Association for Computational Linguistics.

Wang, J.; Zhang, H.; Liu, C.; and Yang, X. 2024. Fake News Detection via Multi-scale Semantic Alignment and Cross-modal Attention. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '24, 2406–2410. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704314.

Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. 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*, KDD '18, 849–857. New York, NY, USA: Association for Computing Machinery. ISBN 9781450355520.

Wu, L.; Rao, Y.; Zhang, C.; Zhao, Y.; and Nazir, A. 2023. Category-Controlled Encoder-Decoder for Fake News Detection. *IEEE Transactions on Knowledge and Data Engineering*, 35(2): 1242–1257.

Yin, S.; Zhu, P.; Wu, L.; Gao, C.; and Wang, Z. 2024. GAMC: An Unsupervised Method for Fake News Detection Using Graph Autoencoder with Masking. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1): 347–355.

Zhang, L.; Zhang, X.; Zhou, Z.; Huang, F.; and Li, C. 2024. Reinforced Adaptive Knowledge Learning for Multimodal Fake News Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(15): 16777–16785.

Zhang, X.; Cao, J.; Li, X.; Sheng, Q.; Zhong, L.; and Shu, K. 2021. Mining Dual Emotion for Fake News Detection. In *Proceedings of the Web Conference 2021*, WWW '21, 3465–3476. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.

Zhou, Y.; Yang, Y.; Ying, Q.; Qian, Z.; and Zhang, X. 2023. Multi-modal Fake News Detection on Social Media via Multi-grained Information Fusion. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, ICMR '23, 343–352. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701788.