Memory-Guided Multi-View Multi-Domain Fake News Detection

Yongchun Zhu [ID], Qiang Sheng [ID], Juan Cao [ID], Qiong Nan, Kai Shu [ID], Minghui Wu [ID],
Jindong Wang [ID], *Member, IEEE*, and Fuzhen Zhuang [ID]

Abstract—The wide spread of fake news is increasingly threatening both individuals and society. Great efforts have been made for automatic fake news detection on a *single* domain (e.g., politics). However, correlations exist commonly across multiple news domains, and thus it is promising to simultaneously detect fake news of *multiple* domains. Based on our analysis, we pose two challenges in multi-domain fake news detection: 1) *domain shift*, caused by the discrepancy among domains in terms of words, emotions, styles, etc. 2) *domain labeling incompleteness*, stemming from the real-world categorization that only outputs one single domain label, regardless of topic diversity of a news piece. In this paper, we propose a Memory-guided Multi-view Multi-domain Fake News Detection Framework (M^3 FEND) to address these two challenges. We model news pieces from a multi-view perspective, including semantics, emotion, and style. Specifically, we propose a Domain Memory Bank to enrich domain information which could discover potential domain labels based on seen news pieces and model domain characteristics. Then, with enriched domain information as input, a Domain Adapter could adaptively aggregate discriminative information from multiple views for news in various domains. Extensive offline experiments on English and Chinese datasets demonstrate the effectiveness of M^3 FEND, and online tests verify its superiority in practice. Our code is available at https://github.com/ICTMCG/M3FEND.

Index Terms—Fake news detection, multi-domain learning, multi-view learning, memory bank

✦

1 INTRODUCTION

W ITH the rapid growth of web technologies, more and more people rely on online social media for news acquisition. According to the 2021 Pew Research Center survey, 48% of American adults get news from social media "often" or "sometimes" [1]. Meanwhile, the wide spread of fake news on social media has threatened both individuals and society. In the COVID-19 infodemic [2], thousands of fake news pieces have been widely spread around the world [3], which caused social panic [4] and weakened the

- *Yongchun Zhu, Qiang Sheng, Juan Cao, and Qiong Nan are with the Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing 100190, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China. E-mail: {zhuyongchun18s, shengqiang18z, caojuan, nanqiong19z}@ict.ac.cn.*
- *Kai Shu is with the Illinois Institute of Technology, Chicago, IL 60616 USA. E-mail: kshu@iit.edu.*
- *Minghui Wu is with the School of Computer and Computing Science, Zhejiang University City College, Hangzhou 310015, China. E-mail: mhwu@zucc.edu.cn.*
- *Jindong Wang is with Microsoft Research Asia, Beijing 100080, China. E-mail: jindong.wang@microsoft.com.*
- *Fuzhen Zhuang is with the Institute of Artificial Intelligence, SKLSDE, School of Computer Science, Beihang University, Beijing 100191, China. E-mail: zhuangfuzhen@buaa.edu.cn.*

effect of pandemic countermeasures [5]. Under such severe circumstances, automatic detection of fake news has been critical for the sustainable and healthy development of news platforms [6], and great efforts have been made to build automatic fake news detection systems [7], [8], [9], [10].

A real-world news platform releases millions of news pieces in diverse domains every day, e,g., finance, politics, health, as shown in Fig. 1. However, most existing methods only focus on a *single* domain (e.g., politics). In fact, news pieces in different domains are inherently correlated, where news (b) and (c) come from different domains while sharing a similar topic of COVID-19. Intuitively, simultaneously modeling multiple correlated news domains benefits fake news detection, and thus it is valuable to study multi-domain fake news detection in real-world news platforms [11].

In this paper, we first investigate multi-domain news data and find that multi-domain fake news detection empirically faces two challenges (see detailed analysis in Section 2):

(1) Domain Shift Among Multiple News Domains. Various news domains have significant domain discrepancies, e.g., words, emotions, styles. Fig. 1 shows that the topics of Politics and Health Domains are different. In addition, as shown in Fig. 2, we find distributions of the writing styles, word usages, publisher emotions of various domains would be largely different, and the differences among domains are called domain shift [12]. Generally, domain shift could seriously influence the effectiveness of joint training multi-domain data [13], [14], [15]. Thus, it is essential to propose well-designed multi-domain models to alleviate the influence of domain shift.

(2) Domain Labeling Incompleteness for News Pieces. In a real-world news platform, a news piece is released in a single domain (channel). However, a news piece is a mixture
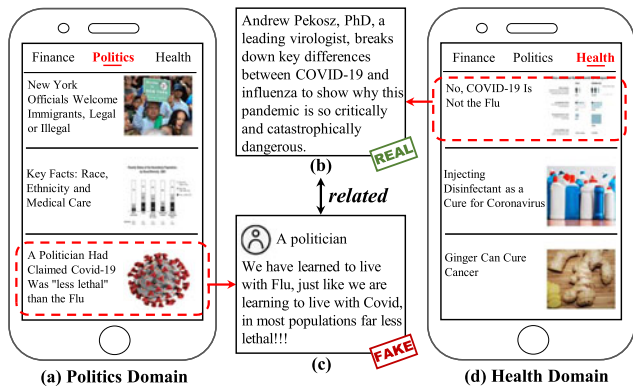
Fig. 1. An example of a real-world news platform with *multiple news domains*. The news distributions vary from domain to domain, leading to the challenge of *domain shift*. However, a news piece is a mixture of diverse elements which makes it relate to multiple news domains, e.g., the political news (c) is also related to the health news (b), leading to the challenge of *domain labeling incompleteness*.

of diverse elements which makes it relate to multiple news domains. As shown in Fig. 1, news (c) is categorized into the Politics Domain, while it also involves a health topic on COVID-19 as shown in news (b). In addition, in Section 2, we find the domain boundary of news domains is not clear, which also indicates a news piece could have multiple domain labels. Since domain labels are useful for multi-domain learning [12], [16], completing the domain labels is important for building an accurate multi-domain fake news detection system.

For multi-domain fake news detection, [17] combined two existing datasets FakeNewsNet [18] and CoAID [19] into a three-domain dataset and proposed a method to preserve domain-specific and domain-shared knowledge. In practice, a news platform consists of far more than three domains, which indicates more challenging domain shift problem, and learning domain-shared knowledge is proved difficult under the challenging scenario [15], [20]. Nan *et al.* [21] proposed a dataset named Weibo21 with nine domains for multi-domain fake news detection and a simple baseline based on Mixture-of-Experts [20], [22]. However, Nan *et al.* [21] ignored the problem of domain labeling incompleteness.

Along this line, we propose a novel Memory-guided Multi-view Multi-domain Fake News Detection Framework (M^3 FEND). Since the distributions of word usage, style, and emotion vary from domain to domain, we first propose three multi-channel networks to model news pieces from semantic view, emotional view, and stylistic view, named SemNet, EmoNet, and StyNet, where each channel can focus on different patterns. Cross-view interactions could capture associations among different views and produce more diverse combinations of views which benefits modeling the domain discrepancy [23], [24], [25], [26], [27]. With the advantage of cross-view interactions, we propose a Multi-head Adaptive Cross-view Interactor to adaptively learn various cross-view interactions. Note that the discriminability of views varies from domain to domain, e.g., the style view is discriminative for Science Domain while not for Entertainment Domain as shown in Section 2. Therefore, we propose a Domain Adapter with domain information as

input to adaptively aggregate discriminative cross-view representations for news in different domains.

To complete domain labels and enrich domain information, we propose a Domain Memory Bank which consists of a Domain Characteristics Memory and multiple Domain Event Memories. Domain Characteristics Memory aims to automatically capture and store information of domain characteristics. In addition, each domain has a Domain Event Memory matrix which records all news released in this domain. Each Domain Event Memory matrix consists of several memory units, and each unit represents a cluster of similar news. Then, we compute the similarity between each Domain Event Memory matrix and a certain news piece. The similarity can denote the distribution of domain labels, and we utilize the similarity distribution to enrich domain information for the news piece. The enriched domain information is utilized to guide the Domain Adapter to aggregate cross-view representations.

M^3 FEND has been successfully deployed in an online fake news detection system that handles millions of news pieces every day, leading to a more trustful online news ecosystem. Our contributions are summarized as follows:

1) We investigate the problem of multi-domain fake news detection and point out two challenges of domain shift and domain labeling incompleteness.
2) To solve the challenges, we propose a novel Memory-guided Multi-view Multi-domain Fake News Detection Framework (M^3 FEND), which can improve the detection performance of most domains.
3) We conduct both offline and online experiments to demonstrate the effectiveness of M^3 FEND. Our code is available at https://github.com/ICTMCG/M3FEND.

2 ANALYSIS

In this section, we first investigate domain shift in multi-domain fake news detection. We analyze domain shift from three views, including word usage, writing style, publisher emotion. In addition, we testify to the problem of domain labeling incompleteness. All analysis is performed based on a Chinese multi-domain fake news detection dataset [21] with nine domains, denoted as Ch-9 in this paper. The statistics of the Ch-9 dataset are shown in Table 2.

2.1 Domain Shift

2.1.1 Word Usage

Intuitively, news pieces published in different domains have various topics and domain-specific word usage [17], [21]. To testify the domain-specific word usage, we plot word clouds of different domains based on word frequency in Fig. 2a. From Fig. 2a, we observe that different news domains have significant differences in the frequently used words.

2.1.2 Writing Style

Fake news publishers utilize particular writing styles to appeal to and persuade a wide scope of consumers. The writing styles of various domains could be largely different. Based on [28] that mines writing style for news on the social

(a) Top 20 words in the nine domains

(b) Style features of fake news

(c) Emotion features of fake news

(d) Style features of real news

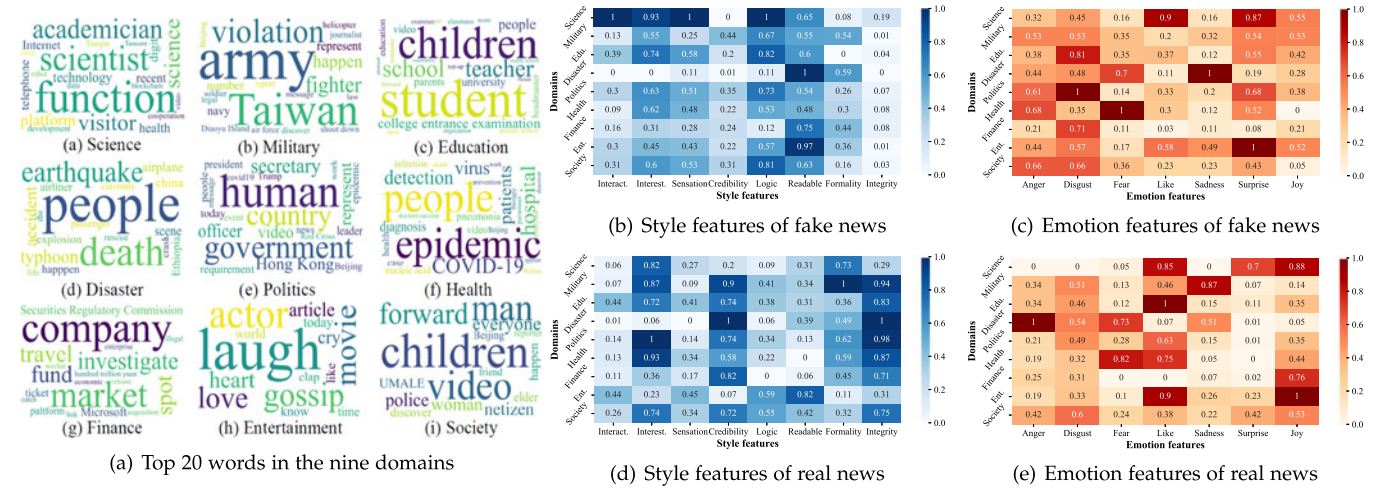(e) Emotion features of real news

Fig. 2. Figure (a) denotes the Top 20 words in the nine domains. Figures (b)-(e) are heatmaps with the average value of different features in the nine domains. Figures (b) and (d) denote style features. Figures (c) and (e) denote emotion features. In fake news (the first row), there are distinct style and emotion features from real news (the second row).

network, we extract eight writing styles including Readability, Logic, Credibility, Formality, Interactivity, Interestingness, Sensation, and Integrity. The average normalized writing style features of fake and real news are shown in Figs. 2b and 2d, respectively. We can find that style features of fake and real news pieces are different, and the discriminative style features of various domains are also different, e.g., the Logic feature is beneficial to detect fake news for Science Domain while not for Finance Domain.

2.1.3 Publisher Emotion

Fake news is more likely to be emotional or inflammatory than real news [6], [29], and emotional signals are useful patterns to detect fake news [30], [31]. According to [31], we evaluate seven kinds of emotions in publisher contents with an affective lexicon ontology database [32], including Anger, Disgust, Fear, Sadness, Surprise, Like, Joy. The emotional features of fake and real news are shown in Figs. 2c and 2e. We can observe that the discriminative emotion features of various domains are different, e.g., fake news publishers of Politics Domain show more disgust, while the publishers of Health Domain show more anger.

2.2 Domain Labeling Incompleteness

Generally, a news piece is manually categorized into a single domain, but the content of a news piece could be related to several domains at the same time. In other words, it is

Fig. 3. Visualizations of unclear domain boundaries with the domain classification task using t-SNE on the Ch-9.

(a) TextCNN

(b) BiGRU

difficult to discriminate domains by news content. To testify this problem, we conduct a supervised domain classification task, which exploits deep models (TextCNN [33], BiGRU [34]) to predict domain labels by news content. First, We visualize in Fig. 3 the network activations of the test set (before classifier) learned by TextCNN and BiGRU learned using t-SNE embeddings [35]. From Fig. 3, we can find that the domain decision boundary is not clear, and many samples of different domains are overlapping, especially Society Domain. For fake news detection without the object of domain classification, it is more difficult to discriminate domain labels. In addition, we evaluate and list the accuracy of domain classification tasks in Table 1, and the unsatisfying accuracy also demonstrates the content of the news piece could be related to several domains. Thus, domain labeling incompleteness is an important issue for multi-domain fake news detection.

3 MODEL

In this section, we introduce the proposed Memory-guided Multi-view Multi-domain Fake News Detection Framework (M^3 FEND) in detail, and the overall structure is shown in Fig. 4. We first give the definitions of multi-domain fake news detection. Then, we detail the components of M^3 FEND: The Multi-view Extractor extracts multi-view representations and adaptively models cross-view interactions. The Domain Memory Bank tackles the problem of domain label incompleteness and enriches domain information. With enriched domain information as input, the Domain Adapter models the domain discrepancy and feeds useful cross-view representations into the Predictor.

3.1 Problem Definition

Let P be a news piece on social media, and the text of the news piece consists of T tokens (words). We adopt RoBERTa [36], [37], a robustly optimized BERT [38] pre-training model to encode tokens of the news content as $\mathcal{T} = \{t_1, \ldots, t_{|\mathcal{T}|}\}$, where $t \in \mathbb{R}^O$ denotes an embedding and O indicates the dimension of embeddings. Inspired by [31] and the above analysis, we extract emotion features from
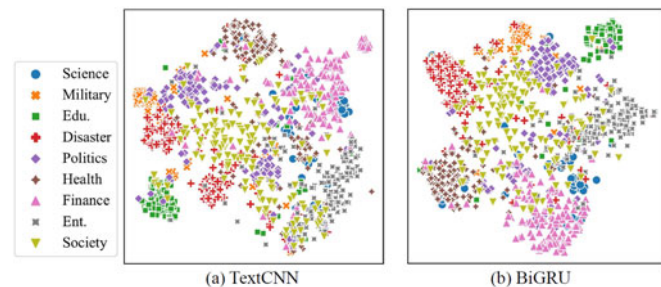
TABLE 1
Accuracy on Ch-9 for Domain Classification

	Science	Military	Edu.	Disaster	Politics	Health	Finance	Ent.	Society
TextCNN	0.42	0.57	0.75	0.74	0.77	0.82	0.83	0.84	0.78
BiGRU	0.56	0.62	0.87	0.69	0.77	0.78	0.84	0.77	0.79

the news piece, including emotion category, emotional intensity, sentiment score, and so on. The emotion features are denoted as $\mathcal{E} = \{e_1, \ldots, e_{|\mathcal{E}|}\}$. Similarly, we extract style features based on [28], and the writing style features of the news piece are denoted as $\mathcal{S} = \{s_1, \ldots, s_{|\mathcal{S}|}\}$. Both emotion and style features are numerical features [39]. Each news piece has a ground-truth label $y \in \{0, 1\}$, where 1 and 0 denote the new piece is fake and real, respectively. In addition, each news piece is manually categorized into a single domain with a domain label $d \in \{Domain_1, \ldots, Domain_N\}$, where N indicates the number of domains. Given a news piece P and a domain label d, multi-domain fake news detection aims to detect whether the news is fake or real.

3.2 Multi-View Extractor

According to the above analysis in Section 2.1, various views can be useful to detect fake news. Thus, first, it is necessary to extract news representations from multiple views. Specifically, we exploit three deep extractors to extract news representations from the semantic view, emotional view, and stylistic view, respectively.

Semantic Network (SemNet). It aims to extract representations from the semantic view with the text content of news pieces. In this paper, TextCNN [33] is employed as SemNet with encoded embeddings $\{t_1, \ldots, t_{|\mathcal{T}|}\}$ as input, which can be formulated as

$$r^{sem} = \text{SemNet}(\{t_1, \ldots, t_{|\mathcal{T}|}\}).\tag{1}$$

Emotion Network (EmoNet). This part focuses on modeling the emotional view with emotional signals $\{e_1, \ldots, e_{|\mathcal{E}|}\}$. In this paper, Multilayer Perceptron (MLP) is adopted as EmoNet to extract emotional representations of news pieces

r^{emo}. Specifically, an EmoNet denoted as

$$r^{emo} = \text{EmoNet}(\{e_1, \ldots, e_{|\mathcal{E}|}\}).\tag{2}$$

Style Network (StyNet). It pays attention to the stylistic view with writing style features $\{s_1, \ldots, s_{|\mathcal{S}|}\}$. MLP is employed as StyNet to extract style representations r^{sty}, which can be formulated as

$$r^{sty} = \text{StyNet}(\{s_1, \ldots, s_{|\mathcal{S}|}\}).\tag{3}$$

Inspired by multi-channel CNN [40], [41], it is beneficial to extract multiple representations for each view with different, learned multi-channel SemNet, EmoNet, and StyNet. Multi-channel extractors allow the model to jointly attend to information from different representation subspaces [40], [42], and different representations could focus on various patterns [43], [44]. Then, we could obtain three groups of representations, each of which corresponds to semantics, emotion, and style views, respectively, denoted as $\{r_i^{sem}\}_{i=1}^{k_{sem}}$, $\{r_i^{emo}\}_{i=1}^{k_{emo}}$, and $\{r_i^{sty}\}_{i=1}^{k_{sty}}$, where k_{sem}, k_{emo}, and k_{sty} indicate the channel number of SemNet, EmoNet, and StyNet, respectively.

Multi-Head Adaptive Cross-View Interactor. Cross-view interactions could capture associations among different views and produce more diverse combinations of views. A direct way is to enumerate the view combinations with all views, which will result in high computational complexity. In addition, such the method cannot model the importance of views in a cross-view interaction, and indiscriminative views may lead to noisy view combinations that degrade model performance. To tackle the problems, we propose an Adaptive Cross-view Interactor to automatically learn cross-view representations, which is formulated as
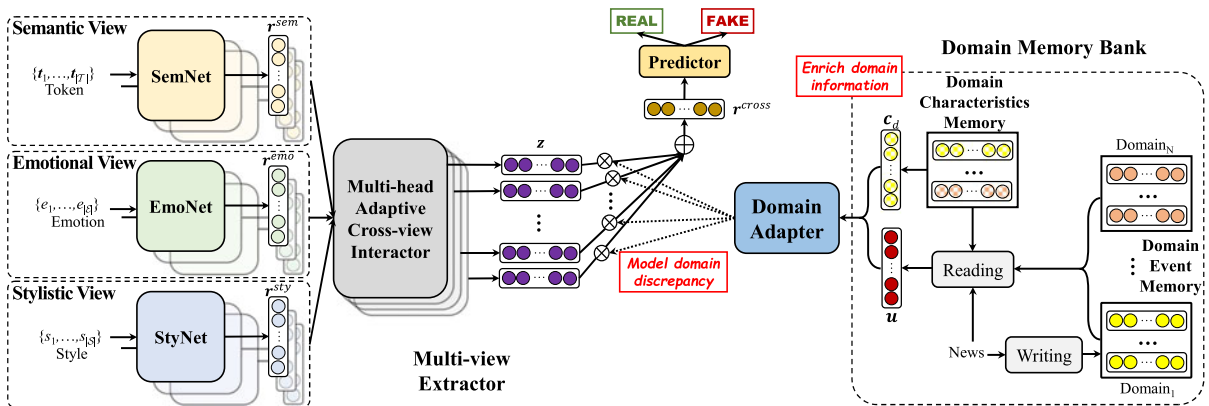
Fig. 4. Overall architecture of the Memory-guided Multi-view Multi-domain Fake News Detection Framework (M³ FEND). The model consists of a Multi-view Extractor, a Domain Memory Bank, a Domain Adapter, and a Predictor. The Multi-view Extractor aims to extract multi-view representations and model cross-view interactions. The Domain Memory Bank stores and provides enriched domain information. The Domain Adapter aggregates discriminative cross-view representations for news in different domains. The Predictor uses the aggregated representations for the final prediction.

$$z = \exp\left[\sum_{i=1}^{k_{sem}} a_i^{sem} \ln r_i^{sem} + \sum_{j=1}^{k_{emo}} a_j^{emo} \ln r_j^{emo} \right.$$
$$\left. + \sum_{q=1}^{k_{sty}} a_q^{sty} \ln r_q^{sty} \right], \tag{4}$$

where a^{sem}, a^{emo}, and a^{sty} indicate the importance of different representations from semantic, emotional, and stylistic views, respectively, which are learnable parameters. z indicates a representation of a cross-view interaction. In addition, the cross-view interaction can be denoted as product operation of various views, and the Equation 4 can be reformulated as

$$z = \prod_{i=1}^{k_{sem}} (r_i^{sem})^{a_i^{sem}} \odot \prod_{j=1}^{k_{emo}} (r_j^{emo})^{a_j^{emo}} \odot \prod_{q=1}^{k_{sty}} (r_q^{sty})^{a_q^{sty}}, \tag{5}$$

where $\odot$ denotes element-wise product operation and $\prod$ indicates continuous element-wise product. An Adaptive Cross-view Interactor can extract representations of a cross-view interaction. However, a single cross-view representation is not discriminative for all domains. For example, the Health Domain may largely rely on the semantic and emotional views, while the Science Domain depends on the semantic and stylistic views. Thus, it is necessary to extract various cross-view representations. Along this line, to model different cross-view interactions, we propose a Multi-head Adaptive Cross-view Interactor with H heads, and each head adaptively learns a kind of cross-view representation. H cross-view representations are denoted as $\{z_i\}_{i=1}^{H}$.

3.3 Domain Memory Bank

Domain Memory Bank aims to complete domain labels and enrich domain information in news pieces, which consists of a Domain Characteristics Memory and a set of Domain Event Memories. The enriched domain information is utilized as the input of the Domain Adapter.

3.3.1 Domain Characteristics Memory

It aims to automatically capture and store domain characteristics. Domain Characteristics Memory can be denoted as $\mathcal{C} = \{c_i\}_{i=1}^{N}$, where c_i represents a memory unit of the ith domain and N denotes the number of domains. All parameters of the Domain Memory $\mathcal{C}$ are randomly initialized. The memory unit c_i is only learned from training samples of the ith domain, so it could be seen as the characteristics representation of the ith domain.

3.3.2 Domain Event Memory

A certain news piece is given a specific domain label d, but the news piece may simultaneously contain information of other domains. To tackle the problem of domain labeling incompleteness, we propose a Domain Event Memory mechanism that aims to discover potential domain labels of news and enrich domain information. The key idea is that a Domain Event Memory matrix records all news released in this domain, and for a certain news piece, we evaluate the similarity between the news and all Domain Event Memory matrices. The similarity can represent the distributions of potential domain labels.

Domain Event Memory of the jth domain is denoted as $\mathcal{M}_j = \{m_i\}_{i=1}^{Q}$, where m represents a memory unit and Q denotes the number of memory units. A memory unit m represents a set of similar news pieces, and all news pieces in a specific news domain can be divided into Q clusters. Each domain has a Domain Event Memory matrix, so there are N Domain Event Memory matrices.

Initialization. We initialize m by clustering similar news pieces with K-means algorithm. A news piece is represented as $n = [\mathcal{G}(\{t_1, \ldots, t_{|\mathcal{T}|}\}); \{e_1, \ldots, e_{|\mathcal{E}|}\}; \{s_1, \ldots, s_{|\mathcal{S}|}\}] \in \mathbb{R}^I$, where $\mathcal{G}(\cdot)$ denotes a learnable attention layer with a mask operation (replacing the padding position as $-\inf$) following [21]. Before training, we obtain representations of all news pieces, and we aggregate news representations into Q clusters using K-means for each domain, respectively. For a specific domain, all centers of clusters are utilized to initialize its memory units.

Reading Operation. This operation aims to evaluate the similarity between a given news piece and all Domain Event Memory matrices. Specifically, for a given news piece, we first find all similar memory units m in a certain Domain Event Memory matrix $\mathcal{M}_j$ and aggregate them into a domain representation

$$o_j = \text{softmax}(nWg(\mathcal{M}_j)/\tau)\mathcal{M}_j, \tag{6}$$

where $g(\cdot)$ denotes transpose function, and $W \in \mathbb{R}^{I \times I}$ is a learnable parameter matrix. We set $\tau = 0.01$ to only find the most similar event cluster. All domain representations are concentrated into a matrix as $\mathcal{D} = [o_1, \ldots, o_N] \in \mathbb{R}^{N \times I}$, where N indicates the number of domains. Then, the similarity distribution can be denoted as

$$v = \text{softmax}(nVg(\mathcal{D})), \tag{7}$$

where $V \in \mathbb{R}^{I \times I}$ is a learnable parameter matrix, and $v \in \mathbb{R}^N$ indicates the similarity distribution.

For a certain news piece, according to the domain label d, we look up the Domain Characteristics Memory $\mathcal{C}$ to obtain an explicit domain representation c_d. Then, with the similarity distribution v, we evaluate an implicit domain representation as: $u = \sum_{i=1}^{o} v_i c_i$, where c_i is the representation of ith domain and v_i denotes the ith element of the similarity distribution vector. The implicit domain representation contain potential domain information. Finally, the implicit representation u and the explicit representation c_d are concentrated into $[c_d, u]$ to represent enriched domain information in the news piece.

Writing Operation. A given domain label d indicates that the news piece contains topics of a certain domain. Thus, we store the news piece in the specific Domain Event Memory $\mathcal{M}_d$. Inspired by neural turning machine (NTM) [45], when writing the Domain Event Memory matrix $\mathcal{M}_d$, it will be erased first before new information is added. We compute the similarity between the news piece and each memory unit as $sim = \text{softmax}(nWg(\mathcal{M}_d)/\tau)$. The erasing vector from the memory unit m_i is denoted as: $erase_i = sim_i \cdot m_i$. Then, an adding vector for the memory unit m_i is denoted as $add_i = sim_i \cdot n$. The overall writing operation can be formulated as

$$m_i = m_i - \beta erase_i + \beta add_i, \tag{8}$$

where β is a parameter to control that the proportions of memory erasing and adding are the same (here, 0.05).

3.3.3 Discussion

In this work, we utilize the Domain Memory Bank to complete the domain labels. Note that domain labeling incompleteness is different from the missing label issue. Domain labeling incompleteness indicates a news piece is given a single domain label, but it could be related to multiple news domains, while the missing label issue denotes some samples have no label. Pseudo label generation has been widely used in semi-supervised scenarios, which trains a classifier to infer the missing labels [46], [47]. Intuitively, such pseudo label generation methods can be exploited to complete domain labels. However, in multi-domain fake news detection, it is difficult to train a satisfying domain classifier as shown in Section 2. Thus, the pseudo labels generated by the domain classifier are unreliable, and the model might be misled by the wrong domain labels. To achieve a similar function and avoid the above issues, we propose the Domain Memory Bank. Different from pseudo label generation methods, our design is not determined by the accuracy of the Domain Memory Bank as it is trained end-to-end with the fake news detection as the final objective. In addition, our method has better transparency because it preserves more event information to help us know how it finds potential domain labels as shown in Table 11.

3.4 Domain Adapter

Due to the existing domain discrepancy, the discriminative cross-view representations of various domains could be different. Thus, we propose a Domain Adapter to model the domain discrepancy. The Domain Adapter takes the enriched domain representation $[c_d, u]$ from the Domain Memory Bank as input to aggregate useful cross-view representations for final prediction. Specifically, the aggregated cross-view representation is formulated as

$$r = \sum_{i=1}^{H} w_i z_i, \quad w = \text{softmax}(f([c_d, u])), \tag{9}$$

where $f(\cdot)$ is a feed-forward network and $w \in \mathbb{R}^H$ is the weight vector representing the importance of different cross-view representations. H is the number of cross-view representations. Note that the weight vector w could help us understand the decision process of a certain news piece depends on which cross-view interactions.

3.5 Predictor

With the aggregated cross-view representation r, we predict the probability of a news piece P being fake with

$$\hat{p} = \text{Sigmoid}(\text{MLP}(r)). \tag{10}$$

All parameters are learnable and can be optimized by minimizing the cross entropy loss with back-propagation

$$\mathcal{L} = -y\log\hat{p} - (1-y)\log(1-\hat{p}). \tag{11}$$

TABLE 2
Data Statistics of Ch-9

Domain	Science	Military	Edu.	Disasters	Politics
#Real	143	121	243	185	306
#Fake	93	222	248	591	546
Total	236	343	491	776	852
Domain	Health	Finance	Ent.	Society	All
#Real	485	959	1,000	1,198	4,640
#Fake	515	362	440	1,471	4,488
Total	1,000	1,321	1,440	2,669	9,128

4 EXPERIMENTS

In this section, we conduct experiments with the aim of answering the following research questions:

RQ1 Does our proposed M^3 FEND outperform other approaches in different datasets?

RQ2 Can M^3 FEND framework improve the performance of a real-world fake news detection system?

RQ3 What are the effects of different views and components in our proposed M^3 FEND?

RQ4 How does M^3 FEND model the domain discrepancy and find potential domain labels? How sensitive are the hyperparameters?

4.1 Experimental Settings
4.1.1 Datasets

We evaluate M^3 FEND with baselines on both English and Chinese datasets of multi-domain fake news detection tasks. The statistics of the Ch-9 and En-3 datasets are shown in Tables 2 and 3, respectively.

English Dataset. Following [17], we combine FakeNews-Net [18] and COVID [48] into an English multi-domain fake news detection dataset with three domains, including GossipCop, PolitiFact, and COVID, which we named as En-3.

Chinese Dataset [21]. It is a Chinese multi-domain fake news detection dataset collected from Sina Weibo with nine domains, including Science, Military, Education, Disaster, Politics, Health, Finance, Entertainment, and Society. We denote the full dataset as Ch-9. In addition, to testify the effectiveness of M^3 FEND under various scenarios, we sample two datasets as Ch-3 and Ch-6. Ch-3 contains the same three domains as the En-3 dataset. Ch-6 contains 6 domains that are related to daily life, including Education, Disaster, Health, Finance, Entertainment, and Society.

To capture multi-view information, we extend the two datasets with emotion and style features. All emotion features are listed in Table 4. The extraction process of emotion features follows the work of Zhang *et al.* [31] with their official codes.[1] In this paper, we follow Yang *et al.* [28] to extract Style features, including eight high-level features:

- *Readability* measures the clarity of the news, evaluated as: Readability = - (Sentence_broken + Characters +

---

1. https://github.com/RMSnow/WWW2021

TABLE 3
Data Statistics of En-3

Domain	GossipCop	PolitiFact	COVID	All
#Real	16,804	447	4,750	22,001
#Fake	5,067	379	1,317	6,763
Total	21,871	826	6,067	28,764

TABLE 4
Emotion Features

Feature	Description
Emotional Category	The probabilities that the given text contains certain emotions obtained from publicly available emotion classifiers.
Emotional Lexicon	The overall emotion score that is aggregated from scores of each word and the whole text across all the emotions.
Emotional Intensity	The overall intensity scores which is extracted from the existing emotion dictionaries annotated with similar process as Emotional Lexicon.
Sentiment Score	The degree of the positive or negative polarity of the whole text calculated by using sentiment dictionaries or public toolkits.
Auxiliary Features	The frequency of emoticons, punctuations, sentimental words, personal pronoun, and uppercase letters.

Words + Sentences + Clauses + Average_word length + Professional_words+ LW + RIX + LIX).

- *Logic* determines whether the news are logical and contextually coherent or not. Logic = Forward_reference + Conjs.
- *Credibility* measures the rigor and reliability of the news, computed as: Credibility = @ + Numerals + Official speech + Time + Place + Object - Uncertainty + Image.
- *Formality* is used to measure the writing normative. Formality = Noun + Adj + Prep - Pron - Verb - Adv - Sentence_broken.
- *Interactivity* represents the interaction between news and the readers, computed as: Interactivity = Question_mark + First_pron + Second_pron + Interrogative_pron.
- *Interestingness* is used to measure whether the news will be attractive. Interestingness = Rhetoric + Exclamation mark + Face + Idiom + Adversative + Adj + Image.
- *Sensation* measures the impression that the news leaves on the reader, computed as: Sensation = Sentiment_score + Adv_of_degree + Modal_particle + First_pron + Second_pron + Exclamation_mark + Question_mark.
- *Integrity* is used to measure whether a news piece is complete. Integrity = 2*HasHead + 2*HasImage + 2*HasVideo + 2*HasTag + HasAt + HasUrl.

We just list how to calculate these style features based on other low-level features, and more details can be found in [28]. Both low- and high-level style features are exploited to extract information of the stylistic view in this paper.

4.1.2 Baselines

We categorize our baselines into three groups. The first group is single-domain methods that separately train models for each domain, including:

- BiGRU [34] is a widely used baseline in many existing works of fake news detection for text encoding. We implement a one-layer BiGRU with a hidden size of 300.
- TextCNN [33] is a popular text encoder. We implement TextCNN with 5 kernels. The 5 kernels with the same 64 channels have different steps of 1, 2, 3, 5, and 10.
- RoBERTa [36], [37] is a robustly optimized BERT [38] pre-training model. We utilize RoBERTa to encode tokens of news content and feed the extracted average embedding into an MLP to obtain the final prediction. The two kinds of RoBERTa [36], [37] are exploited for Chinese and English datasets, respectively.

The second group is the mixed-domain baselines which combine all domains into a single domain. BiGRU [34], TextCNN [33], and RoBERTa [36], [37] of this group have the same implementation as the first group. Another two baselines of this groups are:

- StyleLSTM [49] exploits a BiLSTM to extract news representation from content. Then, it feeds the representation and style features into an MLP to obtain the final prediction. The utilized style features in this paper are the same as StyleLSTM.
- DualEmo [31] extracts news representation with a BiGRU. It exploits both the representation and emotion features to predict fake or real. The utilized emotion features in this paper are the same as DualEmo.

The third group is well-designed multi-domain methods, including:

- EANN [50] consists of three components: feature extractor, event discriminator, and fake news detector. It aims to learn event-invariant representations. In this paper, we modify EANN to learn domain-invariant representations following [17].
- MMoE [20] is a popular multi-domain model that shares a mixture-of-experts (MoE) across various domains, and each domain has its specific head. In this paper, both experts and heads of MMoE are MLPs.
- MoSE [51] is a recent multi-domain model that replaces the experts of MMoE with LSTM.
- EDDFN [17] is a multi-domain fake news detection model which preserves domain-specific and domain-shared knowledge. All domain-specific and domain-shared modules are MLPs. The concentrated domain-specific and domain-shared representations are fed into a classifier to obtain the final prediction.
- MDFEND [21] is the latest multi-domain fake new detection model which utilizes a Domain Gate to select useful experts of MoE.

TABLE 5
Results on the En-3 Dataset

| | Method | Gossip. | Polit. | COVID | overall | | |
					F1	Acc	AUC
single	BiGRU	0.7666	0.7722	0.8885	0.7958	0.8668	0.8840
	TextCNN	0.7786	0.8011	0.9040	0.8079	0.8692	0.9023
	RoBERTa	0.7810	0.8583	0.9288	0.8184	0.8802	0.9108
mixed	BiGRU	0.7479	0.7339	0.7448	0.7501	0.8321	0.8504
	TextCNN	0.7519	0.7040	0.8322	0.7679	0.8362	0.8674
	RoBERTa	0.7823	0.7967	0.9014	0.8101	0.8744	0.9058
	StyleLSTM	0.8007	0.7937	0.9252	0.8285	0.8826	0.9250
	DualEmo	0.8056	0.7868	0.9019	0.8270	0.8818	0.9251
multi	EANN	0.7937	0.7558	0.8836	0.8123	0.8743	0.9053
	MMoE	0.8022	0.8477	<u>0.9379</u>	0.8361	0.8920	<u>0.9265</u>
	MoSE	0.7981	**0.8576**	0.9326	0.8318	0.8885	0.9252
	EDDFN	0.8067	<u>0.8505</u>	0.9306	0.8378	0.8912	0.9263
	MDFEND	<u>0.8080</u>	0.8473	0.9331	<u>0.8390</u>	<u>0.8936</u>	0.9237
	M^3 FEND	**0.8237****	0.8478	**0.9392**	**0.8517****	**0.8977***	**0.9342***

* ($p \leq 0.05$) and ** ($p \leq 0.005$) indicate paired t-test of M^3 FEND versus the best baseline.

4.1.3 Experimental Details

Following [21], [52], all datasets are randomly divided into train/validation/test sets in the ratio of 6:2:2 with keep the domain distribution in each set. We do not perform any dataset-specific tuning except early stopping on validation sets. For all methods, the initial learning rate for the Adam [53] optimizer is tuned by grid searches from 1e-6 to 1e-2. The number of heads and channels is tuned by grid searches from 1 to 10. The number of memory units Q is searched from 5 to 50. The mini-batch size is 64. To ensure a fair comparison, all BiGRU and BiLSTM have one layer with a hidden size of 300. All TextCNN have 5 kernels with steps in {1, 2, 3, 5, 10} and 64 channels. For all MLPs in these methods, the dimension of the hidden layer is set to 384, and ReLU activation function is employed. The maximum sequence length of English and Chinese datasets is set as 300 and 170, respectively. We record average results over ten runs. We report accuracy (Acc), macro F1 score (F1), and Area Under ROC (AUC).

4.2 Offline Results (RQ1)

In this section, we conduct offline experiments on both English and Chinese datasets. The results of En-3 are shown in Table 5, and results of Ch-3, -6, and -9 are listed in Tables 6, 7, and 8 (F1 for each domain and F1, Acc, and AUC for overall performance), respectively. Bold and underlined results indicate the best and second best, respectively. From these results, we have several findings:

(1) On most tasks, the mixed-domain methods outperform the single-domain approaches, which demonstrates jointly training data of multiple domains is helpful to improve not only the overall performance of multiple domains but also the performance of each domain. Meanwhile, we observe that BiGRU, TextCNN, and RoBERTa of the second group have worse performance than the single-domain group in Table 5, and this phenomenon could be caused by serious domain conflict of the En-3 dataset, which shows the necessity of well-designed multi-domain models to alleviate domain conflict.

TABLE 6
Results on the Ch-3 Dataset

| | Method | Politics | Health | Ent. | overall | | |
					F1	Acc	AUC
single	BiGRU	0.8469	0.8335	0.7913	0.8402	0.8411	0.9213
	TextCNN	0.8514	0.9041	0.8423	0.8846	0.8850	0.9521
	RoBERTa	0.8137	0.8924	0.8434	0.8691	0.8697	0.9420
mixed	BiGRU	0.8384	0.8577	0.8687	0.8733	0.8741	0.9402
	TextCNN	0.8579	0.8716	0.8683	0.8833	0.8838	0.9493
	RoBERTa	0.8300	0.8955	0.8862	0.8911	0.8915	0.9566
	StyleLSTM	0.8298	0.8924	0.8896	0.8912	0.8917	0.9564
	DualEmo	0.8362	0.8968	0.9020	0.8977	0.8980	0.9605
multi	EANN	0.8405	0.9189	0.8974	0.9038	0.9042	0.9644
	MMoE	**0.8779**	0.9215	0.8800	0.9048	0.9052	0.9629
	MoSE	0.8564	0.9023	0.8872	0.8978	0.8985	0.9572
	EDDFN	0.8440	0.9235	0.8748	0.8965	0.8970	0.9614
	MDFEND	0.8555	<u>0.9419</u>	<u>0.9103</u>	<u>0.9205</u>	<u>0.9208</u>	<u>0.9750</u>
M^3 FEND		<u>0.8618</u>	**0.9479***	**0.9304****	**0.9308****	**0.9311****	**0.9759**

* ($p \leq 0.05$) and ** ($p \leq 0.005$) indicate paired t-test of M^3 FEND versus the best baseline.

TABLE 7
Results on the Ch-6 Dataset

| | Method | Edu. | Disaster | Health | Finance | Ent. | Society | overall | | |
								F1	Acc	AUC
single	BiGRU	0.7697	0.7191	0.8451	0.8247	0.8026	0.8015	0.8266	0.8270	0.8979
	TextCNN	0.7805	0.4388	0.9012	0.7671	0.7930	0.8654	0.8494	0.8499	0.9195
	RoBERTa	0.8175	0.7584	0.8909	0.8498	0.8549	0.8304	0.8576	0.8580	0.9288
mixed	BiGRU	0.8253	0.7938	0.8626	0.8254	0.8604	0.8206	0.8491	0.8501	0.9249
	TextCNN	0.8593	0.8240	0.8832	0.8646	0.8659	0.8641	0.8776	0.8783	0.9483
	RoBERTa	0.8664	0.8515	0.9100	0.8700	0.8872	0.8634	0.8872	0.8877	0.9494
	StyleLSTM	0.8565	0.8374	0.9080	0.8766	0.8957	0.8546	0.8844	0.8851	0.9489
	DualEmo	0.8472	0.8352	0.9055	0.8951	0.9043	0.8642	0.8904	0.8909	0.9579
multi	EANN	0.8613	0.8657	0.9150	0.8621	0.8871	0.8791	0.8919	0.8925	0.9605
	MMoE	0.8625	0.8777	0.9260	0.8546	0.8882	0.8655	0.8894	0.8900	0.9563
	MoSE	0.8569	0.8588	0.9118	0.8639	0.8904	0.8757	0.8913	0.8918	0.9533
	EDDFN	0.8780	0.8734	0.9280	0.8456	0.8819	0.8716	0.8917	0.8921	0.9544
	MDFEND	0.8826	0.8781	0.9430	0.8749	0.9095	0.8940	0.9093	0.9097	0.9694
M^3 FEND		**0.8836**	**0.8824**	**0.9515***	**0.8997***	**0.9296****	**0.9043****	**0.9208****	**0.9211****	**0.9762***

* $(p \leq 0.05)$ and ** $(p \leq 0.005)$ indicate paired t-test of M^3 FEND versus the best baseline.

TABLE 8
Results on the Ch-9 Dataset

| | Method | Science | Military | Edu. | Disaster | Politics | Health | Finance | Ent. | Society | overall | | |
											F1	Acc	AUC
single	BiGRU	0.5175	0.3365	0.7416	0.7293	0.8588	0.8373	0.8137	0.7992	0.7918	0.8103	0.8103	0.8902
	TextCNN	0.4074	0.3365	0.8059	0.4388	0.8482	0.8819	0.8215	0.7973	0.8615	0.8369	0.8370	0.9094
	RoBERTa	0.7463	0.7369	0.8146	0.7547	0.8044	0.8873	0.8361	0.8513	0.8300	0.8477	0.8477	0.9226
mixed	BiGRU	0.7269	0.8724	0.8138	0.7935	0.8356	0.8868	0.8291	0.8629	0.8485	0.8595	0.8598	0.9309
	TextCNN	0.7254	0.8839	0.8362	0.8222	0.8561	0.8768	0.8638	0.8456	0.8540	0.8686	0.8687	0.9381
	RoBERTa	0.7777	0.9072	0.8331	0.8512	0.8366	0.9090	0.8735	0.8769	0.8577	0.8795	0.8797	0.9451
	StyleLSTM	0.7729	0.9187	0.8341	0.8532	0.8487	0.9084	0.8802	0.8846	0.8552	0.8820	0.8821	0.9471
	DualEmo	0.8323	0.9026	0.8362	0.8396	0.8455	0.8905	**0.9053**	0.8944	0.8569	0.8846	0.8846	0.9541
multi	EANN	0.8225	0.9274	0.8624	0.8666	0.8705	0.9150	0.8710	0.8957	0.8877	0.8975	0.8977	0.9610
	MMoE	**0.8755**	0.9112	0.8706	0.8770	0.8620	0.9364	0.8567	0.8886	0.8750	0.8947	0.8948	0.9547
	MoSE	0.8502	0.8858	0.8815	0.8672	0.8808	0.9179	0.8672	0.8913	0.8729	0.8939	0.8940	0.9543
	EDDFN	0.8186	0.9137	0.8676	0.8786	0.8478	0.9379	0.8636	0.8832	0.8689	0.8919	0.8919	0.9528
	MDFEND	0.8301	0.9389	0.8917	**0.9003**	**0.8865**	0.9400	0.8951	0.9066	0.8980	0.9137	0.9138	0.9708
M^3 FEND		0.8292	**0.9506****	0.8998	0.8896	0.8825	0.9460	0.9009	**0.9315****	**0.9089****	**0.9216****	**0.9216****	**0.9750***

* $(p \leq 0.05)$ and ** $(p \leq 0.005)$ indicate paired t-test of M^3 FEND versus the best baseline.

(2) Then, we find that the multi-domain group outperforms the mixed-domain group, which validates that well-designed multi-domain models are important. The reason could be that the multi-domain models with a well-designed sharing structure could alleviate the domain conflict.

(3) StyleLSTM and DualEmo achieve better results than BiGRU, TextCNN, and RoBERTa, which testifies introducing more views is beneficial for multi-domain fake news detection. Especially, DualEmo is extremely effective for Entertainment and Finance domains and even outperforms most multi-domain methods.

(4) With the results of the t-test, we find that M^3 FEND outperforms the best baseline significantly in most tasks, which demonstrates that M^3 FEND is an effective solution to improve not only overall detection performance but also the performance of the specific domains. The main reason is that M^3 FEND enriches domain information and explicitly

models various domain discrepancies by aggregating useful cross-view interactions for different domains.

(5) We observe MDFEND, and M^3 FEND outperforms EANN, MMoE, MoSE, and EDDFN on most tasks. EANN directly learns a shared network with adversarial training. MMoE and MoSE utilized a shared bottom and multiple separate heads for different domains. EDDFN learns both domain-specific and domain-shared subnetworks. We find that all of them take hard sharing mechanisms to learn shared knowledge from all domains. However, the existing studies [15], [54] found it is hard to learn the common knowledge with a shared structure from too many domains. Different from these methods, MDFEND and M^3 FEND exploit soft sharing mechanisms to aggregate beneficial shared knowledge for fake news detection.

(6) Both MDFEND and M^3 FEND exploit soft sharing mechanisms, but M^3 FEND achieves better results on most tasks. The improvements come from multiple aspects: 1) M^3

TABLE 9
Relative Improvement Over the Online Baseline

Improvement on	SPAUC	AUC	F1
EANN	2.12%	0.67%	0.33%
EDDFN	-0.37%	-2.02%	-3.34%
MDFEND	2.82%	0.74%	1.85%
M³ FEND	**5.50%**	**2.89%**	**4.49%**

FEND contains a Domain Memory Bank to discover potential distributions of domain labels, while MDFEND assumes the given domain label is exact and complete. However, we find that domain labeling incompleteness is an important issue for multi-domain fake news detection in Section 2; 2) MDFEND only extracts semantic information, while M^3 FEND simultaneously models semantic, emotional, and stylistic views and adaptively captures cross-view information.

4.3 Online Tests (RQ2)

The M^3 FEND framework has already been deployed in our online fake news detection system (http://www.newsverify.com/) which handles millions of news pieces every day. To verify the real benefits of M^3 FEND brings to our system, we conduct online testing experiments within one week. Different from offline datasets, the online test set is highly skewed (real versus fake, roughly 300:1). Online data is collected by the system, and the time intervals of training and test sets do not intersect. Precisely, we deploy M^3 FEND and several competitive baselines (EANN, EDDFN, and MDFEND). The online baseline is the mixed-domain RoBERTa model. In real-world scenarios, the number of fake news is much lower than real news, which means that we should detect fake news without misclassifying real news as possible. In other words, the task is improving the True Positive Rate (TPR) on the basis of low False Positive Rate (FPR). Thus, beyond AUC and F1, following [55], [56], we adopt standardized partial AUC (SPAUC $_{\text{FPR} \leq 0.1}$). Due to the company regulations, we cannot detail the online data and the absolute results, so we report the relative improvement over the online baseline RoBERTa. The online results in Table 9 demonstrate that the proposed M^3 FEND achieves a satisfying improvement on all metrics against the baselines.

4.4 Ablation Study (RQ3)

In this section, we analyze the effects of different views and components in our proposed M^3 FEND and conduct an ablation study on the four datasets with the overall F1 score shown in Table 10. First, we conduct experiments to verify the contributions of different views and introduce three kinds of models, w/o SemView, w/o EmoView, and w/o StyView, which remove the semantic, emotional, and stylistic views from M^3 FEND, respectively. We find that all views are beneficial for fake news detection, especially the semantic view, which is the core of most existing methods [52], [57]. Since the emotion and style features are manually extracted from the textual content, these features are usually utilized as auxiliary information for semantic view modeling [31], [49]. In addition, we observe that the emotional view is more effective than the stylistic view. The reason could be that the emotion features include both publisher and social characteristics [31], while the style features only represent the publisher preference [28].

TABLE 10
Results of Ablation Study

	Ch-3	Ch-6	Ch-9	En-3
M³ FEND	**0.9308**	**0.9208**	**0.9216**	**0.8517**
w/o SemView	0.8202	0.8161	0.8249	0.6573
w/o EmoView	0.9195	0.9136	0.9147	0.8403
w/o StyView	0.9255	0.9178	0.9177	0.8472
w/o Interactor	0.9217	0.9169	0.9173	0.8398
w/o Memory	0.9237	0.9182	0.9176	0.8501
w/o Adapter	0.9172	0.9169	0.9157	0.8367

Furthermore, to testify the effectiveness of each component in M^3 FEND, we introduce three kinds of M^3 FEND, (1) w/o Interactor: remove Multi-head Adaptive Cross-view Interactor. The performance drop demonstrates modeling cross-view interaction could capture more information. (2) w/o Memory: remove the Domain Event Memory. The performance decrease on four datasets indicates enriching domain information is useful. (3) w/o Adapter: replace Domain Adapter by average operation. M^3 FEND w/o Adapter obtains the worse results, which demonstrates the aggregation process of discriminative representations is necessary. Note that the Adapter takes two parts from the Domain Event Memory and Domain Characteristics Memory as input. Thus, that M^3 FEND w/o Adapter achieves worse results than M^3 FEND w/o Memory demonstrates the implicit representation from the Domain Characteristics Memory is useful. In addition, we observe the Adapter is more effective than the Interactor, which shows that selecting informative representations is more effective than adaptively learning cross-view information.

4.5 Analysis (RQ4)

4.5.1 Effectiveness of Domain Discrepancy Modeling

In this section, we conduct experiments on Ch-6 to demonstrate M^3 FEND can model the domain discrepancy. Due to domain discrepancy, discriminative cross-view representations vary from domain to domain, and the proposed M^3 FEND can find useful cross-view representations for different domains. For brevity, we set the number H of Interactor heads as 4 and the channel number k of multi-view extractors as 1, and remove the implicit domain representations. First, we visualize the absolute adaptive weight a in Equation (4) for four cross views, as shown in Fig. 5, we see that the cross-view representations derived from M^3 FEND have quite differences on the combination of semantic, emotional, and stylistic views, which diversifies cross-view representations and facilitates the modeling of domain discrepancy.

Then, the domain adapter aggregates useful cross-view representations for each domain, and we visualize the importance w of each cross-view representation for six domains in Fig. 6. We see the discriminative cross-view representations vary from domain to domain, which demonstrates the M^3 FEND can model the domain discrepancy. Such importance distributions are indeed in line with our intuition. For example, the Entertainment news is often sensational, which provokes strong emotion of audience. We see the cross view 3 which largely depends on the emotional view is valued the most by Entertainment Domain.
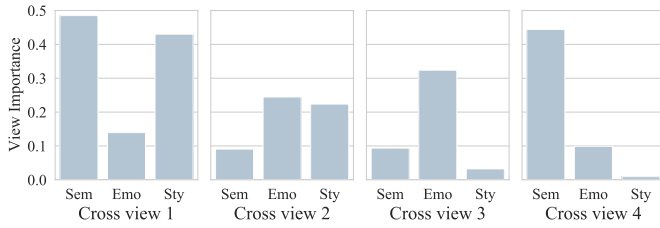
Fig. 5. Each figure indicates importances of different views in a cross-view interaction.

Fig. 6. Various importances of four cross-view interactions for different domains.

4.5.2 Effectiveness of Domain Label Completion

To probe how the Domain Event Memory contributes to discovering potential domain labels, we present a case study. Recall that a unit of a Domain Event Memory represents an event set, and each news piece is only categorized into one unit. For the target political news piece in Table 11, we find the most related memory units of each Domain Event Memory and show a representative example from the event sets of them. The predicted distribution of domain labels (similarity distribution v) indicates that the Domain Event Memory can effectively capture the distributions of potential domain labels.

4.5.3 Hyperparameter Sensitivity

We test the sensitivity of multiple hyperparameters based on the Ch-6 dataset, including #Channel of SemNet k_{sem}, #Channel of EmoNet k_{emo}, #Channel of StyNet k_{sty}, #Head of Interactor H, and the hyperparameter of the memory mechanism β. As shown in Fig. 7, with various hyperparameters, M^3 FEND can achieve satisfying performance. We can observe that even the worst setting of M^3 FEND in Figs. 7a, 7b, 7c, and 7d can get F1 = 0.9149 which is better than the best baseline MDFEND (F1:0.9093) on the Ch-6 dataset, which demonstrates our model are insensitive to the hyperparameters k_{sem}, k_{emo}, k_{sty}, and H. From Fig. 7e, we can observe that the performance first increases and then decreases rapidly as β varies and demonstrates a bell-shaped curve. A big β indicates that the memory module could quickly forget historical samples and focus on most recent samples, which leads to overfitting on recent samples. unsatisfying performance. On the contrary, a small β could lead to underfitting on recent samples. Thus, we need to choose a suitable β for the Domain Event Memory.
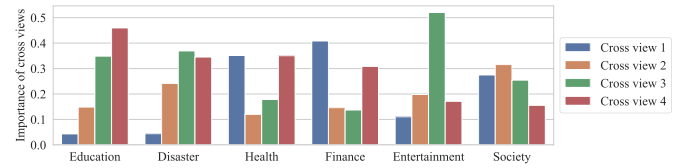
5 RELATED WORK

In this section, we will introduce the related work on Fake News Detection and Multi-domain Learning.

Fake News Detection. Researchers have investigated fake news detection which aims at automatically classifying a news piece as real or fake for a long time. Existing methods can be generally grouped into two clusters: content-based and social-context-based fake news detection [6], [58].

Content-based models mainly rely on news content features and existing factual sources to classify fake news [6], [58]. Some methods focus on extracting textual representations [9], [10], [57], [59], [60]. In addition, visual features of news have been shown to be an important indicator for fake news detection [50], [61], [62]. As fake news publishers tend to use inflammatory and emotional expressions to draw reader's attention for a wide dissemination, style [39], [49], [63] and emotion [30], [31] are useful patterns for fake news detection. Some methods [64], [65], [66], [67] exploit existing factual sources to detect fake news.

Social-context-based models exploit relevant user social engagements to detect fake news [6]. Propagation networks have been testified their effectiveness for fake news detection [68], [69], [70]. In addition, user profile [71], [72] and crowd feedbacks [52], [73] are also important patterns to detect fake news.

Our work mainly falls into the first group, which utilizes semantic, style, emotion features to detect fake news. In addition, most of existing works focus on a single specific domain, e.g., politics [52], [63], health [48], [65], [74]. Our work focuses on multi-domain fake news detection [17], [21]. Silva *et al.* [17] proposed EDDFN that adopts multiple domain-specific subnetworks and a domain-shared subnetwork, and models common knowledge from *all* domains with the domain-shared subnetwork. However, studies have shown that it is hard to simultaneously model the common knowledge of many domains [15], [54]. Thus,

TABLE 11
A Case of the Distribution of Predicted Domain Label

| Target News		Trump nearly fainted during his speech and cancelled his subsequent trip. A symptom of COVID-19?
Domain	Similarity v	Representative Example
Science	0.02	NASA used the Nuclear Spectroscopy Telescope to photo the spiral galaxy 1068 in the Cetus.
Military	0.04	U.S. sends 35 medical ships.
Edu.	0.01	A student admitted to Harvard University.
Disaster	0.02	The US "World Journal" reported a five-level fire in a restaurant.
Politics	0.33	US deaths from COVID-19 exceed 100k.
Health	0.21	The animal experiment of Oxford's COVID-19 vaccine failed.
Finance	0.12	Pfizer's stocking price rose 15%, boosted by the company's COVID-19 vaccine news.
Ent.	0.09	10 more people tested positive for COVID-19 in Italian Serie A.
Society	0.16	A COVID-19 carrier refused security check at the airport.
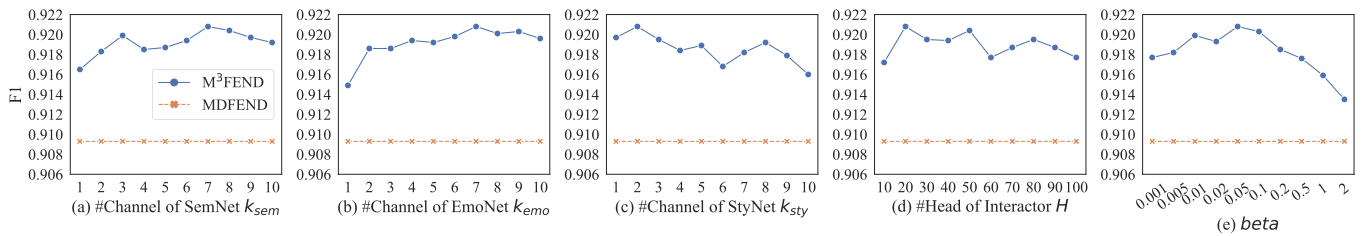
Fig. 7. Performance (F1) of M^3 FEND with various hyperparameters.

EDDFN is unlikely to work well for a real-world news platform that has a large and increasing domain set. Different from EDDFN which uses a hard sharing mechanism, our M^3 FEND adopts a soft sharing mechanism that aggregates the specific shared knowledge of different domains. Nan *et al.* [21] built a multi-domain fake news detection dataset containing news in nine domains and a simple baseline MDFEND which also takes the soft sharing strategy and utilizes a domain gate to aggregate multiple semantic representations extracted by a Mixture-of-Experts structure [20], [22]. However, MDFEND holds an assumption that the provided single domain label is *complete*, which is not always true due to the complex semantic property of news pieces. To tackle the problem of domain labeling incompleteness, our proposed M^3 FEND utilizes the Domain Memory Bank to complete domain labels and enrich domain information in news pieces. In addition, we explore the mechanism of knowledge sharing on not only the semantic view but also emotional and stylistic views, and we further explore the adaptive cross-view knowledge for fake news detection.

Multi-Domain Learning. In the real world, data usually come from various domains. Multi-domain learning aims to simultaneously model multiple domains and improve overall performance, which falls into the area of transfer learning [12]. Some existing methods focus on learning domain-invariant representations [13], [14], and the others model domain relationships [20], [51], [75]. M^3 FEND falls into the second group and outperforms existing methods.

6 CONCLUSION

In this paper, we analyzed two challenges in multi-domain fake news detection, domain shift, and domain labeling incompleteness. Then, to solve the two challenges, we proposed a novel Memory-guided Multi-view Multi-domain Fake News Detection Framework (M^3 FEND). First, we extracted news representations from multiple views and automatically modeled cross-view interactions. To tackle the problem of domain shift, we proposed a Domain Adapter to aggregate cross-view representations for prediction. To solve the second challenge, we proposed a Domain Memory Bank to discover potential domain labels and model domain characteristics. Finally, we demonstrated the superior effectiveness of the proposed M^3 FEND in both offline experiments and online tests, compared to other state-of-the-art approaches.

REFERENCES

[1] E. Shearer and A. Mitchell, "News consumption across social media in 2021," 2021. [Online]. Available: https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/

[2] S. B. Naeem and R. Bhatti, "The COVID-19 'infodemic': A new front for information professionals," *Health Inf. Libraries J.*, vol. 37, no. 3, pp. 233–239, 2020.

[3] Wikipedia, "Misinformation related to the COVID-19 pandemic," 2020. [Online]. Available: https://en.wikipedia.org/wiki/Misinformation_related_to_the_COVID-19_pandemic

[4] Q. Chen, "Coronavirus rumors trigger irrational behaviors among chinese netizens," 2020. [Online]. Available: https://www.globaltimes.cn/content/1178157.shtml

[5] L. Bursztyn, A. Rao, C. P. Roth, and D. H. Yanagizawa-Drott, "Misinformation during a pandemic," Nat. Bureau Econ. Res., Cambridge, MA, Tech. Rep. 27417, 2020.

[6] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newslett.*, vol. 19, no. 1, pp. 22–36, 2017.

[7] X. Zhou *et al.*, "Real-time news certification system on sina weibo," in *Proc. 24th Int. Conf. World Wide Web*, 2015, pp. 983–988.

[8] L. Cui, K. Shu, S. Wang, D. Lee, and H. Liu, "dEFEND: A system for explainable fake news detection," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, 2019, pp. 2961–2964.

[9] Q. Sheng, J. Cao, X. Zhang, R. Li, D. Wang, and Y. Zhu, "Zoom out and observe: News environment perception for fake news detection," in *Proc. 60th Annu. Meeting Assoc. Comput. Linguistics*, 2022, pp. 4543–4556.

[10] Y. Zhu, Q. Sheng, J. Cao, S. Li, D. Wang, and F. Zhuang, "Generalizing to the future: Mitigating entity bias in fake news detection," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022.

[11] Q. Sheng, J. Cao, H. R. Bernard, K. Shu, J. Li, and H. Liu, "Characterizing multi-domain false news and underlying user effects on chinese weibo," *Inf. Process. Manage.*, vol. 59, no. 4, 2022, Art. no. 102959.

[12] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[13] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2096–2030, 2016.

[14] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 1406–1415.

[15] Y. Zhu, F. Zhuang, and D. Wang, "Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources," in *Proc. AAAI Conf. Artif. Intell.*, 2019, pp. 5989–5996.

[16] F. Zhuang *et al.*, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[17] A. Silva, L. Luo, S. Karunasekera, and C. Leckie, "Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 557–565.

[18] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big Data*, vol. 8, no. 3, pp. 171–188, 2020.

[19] L. Cui and D. Lee, "CoAID: COVID-19 healthcare misinformation dataset," 2020, *arXiv:2006.00885*.

[20] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong, and E. H. Chi, "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 1930–1939.

[21] Q. Nan, J. Cao, Y. Zhu, Y. Wang, and J. Li, "MDFEND: Multi-domain fake news detection," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 3343–3347.

[22] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts," *Neural Comput.*, vol. 3, no. 1, pp. 79–87, 1991.

[23] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.

[24] A. Zadeh, P. P. Liang, N. Mazumder, S. Poria, E. Cambria, and L.-P. Morency, "Memory fusion network for multi-view sequential learning," in *Proc. AAAI Conf. Artif. Intell.*, 2018, Art. no. 691.

[25] K. Hassani and A. H. Khasahmadi, "Contrastive multi-view representation learning on graphs," in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 4116–4126.

[26] Y. Wu et al., "Multi-view multi-behavior contrastive learning in recommendation," in *Proc. 27th Int. Conf. Database Syst. Adv. Appl.*, 2022, pp. 166–182.

[27] S. Li, R. Xie, Y. Zhu, X. Ao, F. Zhuang, and Q. He, "User-centric conversational recommendation with multi-aspect user modeling," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2022.

[28] Y. Yang, J. Cao, M. Lu, J. Li, and C.-W. Lin, "How to write high-quality news on social network? Predicting news quality by mining writing style," 2019, *arXiv:1902.00750*.

[29] O. Ajao, D. Bhowmik, and S. Zargari, "Sentiment aware fake news detection on online social networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019, pp. 2507–2511.

[30] A. Giachanou, P. Rosso, and F. Crestani, "Leveraging emotional signals for credibility detection," in *Proc. 42nd Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2019, pp. 877–880.

[31] X. Zhang, J. Cao, X. Li, Q. Sheng, L. Zhong, and K. Shu, "Mining dual emotion for fake news detection," in *Proc. Web Conf.*, 2021, pp. 3465–3476.

[32] L. Xu, H. Lin, Y. Pan, H. Ren, and J. Chen, "Constructing the affective lexicon ontology," *J. China Soc. Sci. Tech. Inf.*, vol. 27, no. 2, pp. 180–185, 2008.

[33] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2014, pp. 1746–1751.

[34] J. Ma et al., "Detecting rumors from microblogs with recurrent neural networks," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 3818–3824.

[35] J. Donahue et al., "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.

[36] Y. Liu et al., "RoBERTa: A robustly optimized bert pretraining approach," 2019, *arXiv:1907.11692*.

[37] Y. Cui et al., "Pre-training with whole word masking for chinese BERT," 2019, *arXiv:1906.08101*.

[38] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2019, pp. 4171–4186.

[39] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. Web Conf.*, 2011, pp. 675–684.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[41] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[42] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[43] C. Tao, W. Wu, C. Xu, W. Hu, D. Zhao, and R. Yan, "Multi-representation fusion network for multi-turn response selection in retrieval-based chatbots," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, 2019, pp. 267–275.

[44] Y. Zhu et al., "Multi-representation adaptation network for cross-domain image classification," *Neural Netw.*, vol. 119, pp. 214–221, 2019.

[45] A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," 2014, *arXiv:1410.5401*.

[46] D. Berthelot, N. Carlini, I. Goodfellow, A. Oliver, N. Papernot, and C. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 5049–5059.

[47] B. Zhang et al., "FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *Proc. 35rd Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 18408–18419.

[48] Y. Li, B. Jiang, K. Shu, and H. Liu, "MM-COVID: A multilingual and multimodal data repository for combating COVID-19 disinformation," 2020, *arXiv:2011.04088*.

[49] P. Przybyla, "Capturing the style of fake news," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 490–497.

[50] Y. Wang et al., "EANN: Event adversarial neural networks for multi-modal fake news detection," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2018, pp. 849–857.

[51] Z. Qin, Y. Cheng, Z. Zhao, Z. Chen, D. Metzler, and J. Qin, "Multitask mixture of sequential experts for user activity streams," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 3083–3091.

[52] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "dEFEND: Explainable fake news detection," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 395–405.

[53] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.

[54] H. Tang, J. Liu, M. Zhao, and X. Gong, "Progressive layered extraction (PLE): A novel multi-task learning (MTL) model for personalized recommendations," in *Proc. 14th ACM Conf. Recommender Syst.*, 2020, pp. 269–278.

[55] D. K. McClish, "Analyzing a portion of the ROC curve," *Med. Decis. Mak.*, vol. 9, no. 3, pp. 190–195, 1989.

[56] Y. Zhu et al., "Modeling users' behavior sequences with hierarchical explainable network for cross-domain fraud detection," in *Proc. Web Conf.*, 2020, pp. 928–938.

[57] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors on twitter by promoting information campaigns with generative adversarial learning," in *Proc. Web Conf.*, 2019, pp. 3049–3055.

[58] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–40, 2020.

[59] H. Karimi and J. Tang, "Learning hierarchical discourse-level structure for fake news detection," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2019, pp. 3432–3442.

[60] G. Kim and Y. Ko, "Effective fake news detection using graph and summarization techniques," *Pattern Recognit. Lett.*, vol. 151, pp. 135–139, 2021.

[61] Y. Wang, F. Ma, H. Wang, K. Jha, and J. Gao, "Multimodal emergent fake news detection via meta neural process networks," in *Proc. 27th ACM SIGKDD Conf. Knowl. Discov. Data Mining*, 2021, pp. 3708–3716.

[62] P. Qi et al., "Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 1212–1220.

[63] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 231–240.

[64] K. Popat, S. Mukherjee, A. Yates, and G. Weikum, "DeClarE: Debunking fake news and false claims using evidence-aware deep learning," in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2018, pp. 22–32.

[65] L. Cui, H. Seo, M. Tabar, F. Ma, S. Wang, and D. Lee, "DETERRENT: Knowledge guided graph attention network for detecting healthcare misinformation," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2020, pp. 492–502.

[66] N. Vo and K. Lee, "Hierarchical multi-head attentive network for evidence-aware fake news detection," in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2021, pp. 965–975.

[67] Q. Sheng, X. Zhang, J. Cao, and L. Zhong, "Integrating pattern- and fact-based fake news detection via model preference learning," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, 2021, pp. 1640–1650.

[68] Y. Liu and Y.-F. B. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, Art. no. 44.

[69] V.-H. Nguyen, K. Sugiyama, P. Nakov, and M.-Y. Kan, "FANG: Leveraging social context for fake news detection using graph representation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 1165–1174.

[70] A. Silva, Y. Han, L. Luo, S. Karunasekera, and C. Leckie, "Propagation2Vec: Embedding partial propagation networks for explainable fake news early detection," *Inf. Process. Manage.*, vol. 58, no. 5, 2021, Art. no. 102618.

[71] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *Proc. IEEE Conf. Multimedia Inf. Process. Retrieval*, 2018, pp. 430–435.

[72] Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User preference-aware fake news detection," in *Proc. 44th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, 2021, pp. 2051–2055.

[73] J. Ma, W. Gao, and K.-F. Wong, "Detect rumor and stance jointly by neural multi-task learning," in *Proc. Companion Proc. Web Conf.*, 2018, pp. 585–593.

[74] X. Zhou, A. Mulay, E. Ferrara, and R. Zafarani, "ReCOVery: A multimodal repository for COVID-19 news credibility research," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, 2020, pp. 3205–3212.

[75] I. Misra, A. Shrivastava, A. Gupta, and M. Hebert, "Cross-stitch networks for multi-task learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3994–4003.

Yongchun Zhu is currently working toward the PhD degree with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. His main research interests include transfer learning, fake news detection and recommender system. He has published more than 20 papers in journals and conference proceedings including KDD, WWW, SIGIR, *IEEE Transactions on Neural Networks and Learning Systems* and so on.

Qiang Sheng received the BE degree from the Beijing University of Posts and Telecommunications, in 2018. He is currently working toward the PhD degree with the Institute of Computing Technology, Chinese Academy of Sciences. His research interests include fake news detection and fact-checking. He has published more than 10 papers in international conferences and journals including ACL, SIGIR, WWW, MM, CIKM, etc.

Juan Cao received the PhD degree from the Institute of Computing Technology, Chinese Academy of Sciences, in 2008. She is a professor with the Institute of Computing Technology, Chinese Academy of Sciences. Her research interests include multimedia content analysis and fake multimedia detection. She has more than 90 publications in international journals and conferences including *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Image Processing*, KDD, WWW, CVPR, etc.

Qiong Nan recieved the BE degree from the Dalian University of Technology, in 2019. She is currently working toward the PhD degree with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. Her research interests lie in fake news detection, transfer learning and multi-task learning.

Kai Shu received the PhD degree in computer science from Arizona State University, in July 2020. He is a Gladwin Development chair assistant professor with the Department of Computer Science, Illinois Institute of Technology (IIT). He was awarded the ASU Fulton Schools of Engineering Dean's Doctoral Dissertation Award, CIDSE Doctoral Fellowship 2015 and 2020. His research lies in machine learning, data mining, social computing, and their applications in disinformation, education, healthcare.

Minghui Wu received the PhD degree in computer science and engineering from Zhejiang University. He is a professor of computer science with Zhejiang University City College. His major interests include artificial intelligence, Big Data, mobile application, and software engineering. He has published more than 100 papers in journals and international conferences including *IEEE Transactions on Knowledge and Data Engineering*, KDD, WWW, etc.

Jindong Wang (Member, IEEE) received the PhD degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. He is currently a researcher as Microsoft Research, Beijing, China. His research interest mainly includes transfer learning, machine learning, data mining, and artificial intelligence. He serves as the reviewers of many journals and conference, including *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *ACM Computing Surveys*, and CHI. He is also a member of ACM, AAAI, and CCF.

Fuzhen Zhuang is a professor in Institute of Artificial Intelligence, Beihang University. His research interests include transfer learning, machine learning, data mining, multi-task learning and recommendation systems. He has published more than 100 papers in the prestigious refereed journals and conference proceedings, such as *Nature Communications*, *IEEE Transactions on Knowledge and Data Engineering*, *Proc. of IEEE*, *IEEE Transactions on Neural Networks and Learning Systems*, *ACM Transactions on Intelligent Systems and Technology*, KDD, WWW, SIGIR, NeurIPS, AAAI, and ICDE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.