An Integrated Multi-Task Model for Fake News Detection

Qing Liao, *Member, IEEE*, Heyan Chai, Hao Han, Xiang Zhang, *Member, IEEE*, Xuan Wang, *Member, IEEE*, Wen Xia, *Member, IEEE*, and Ye Ding, *Member, IEEE*

Abstract—Fake news detection attracts many researchers' attention due to the negative impacts on the society. Most existing fake news detection approaches mainly focus on semantic analysis of news' contents. However, the detection performance will dramatically decrease when the content of news is short. In this paper, we propose a novel *fake news detection multi-task learning (FDML)* model based on the following observations: 1) some certain topics have higher percentages of fake news; and 2) some certain news authors have higher intentions to publish fake news. FDML model investigates the impact of topic labels for the fake news and introduce contextual information of news at the same time to boost the detection performance on the short fake news. Specifically, the FDML model consists of representation learning and multi-task learning parts to train the fake news detection task and the news topic classification task, simultaneously. As far as we know, this is the first fake news detection work that integrates the above two tasks. The experiment results show that the FDML model outperforms state-of-the-art methods on real-world fake news dataset.

Index Terms—Fake news detection, multi-task learning, topic classification

1 Introduction

With the wide dissemination capabilities of the Internet, fake news has been widely spread through online social media. For example, Facebook referrals accounted for 50 percent of total traffic to fake news sites and 20 percent of total traffic to reputable news sites [1]. Moreover, according to Science [2], fake news diffused significantly faster, deeper and more broadly than the truth which brings serious negative impacts to both individuals and society. For instance, a Google engineer was killed by the kidnapping rumours in India [3] and the massive rumours on Facebook and Twitter even influenced the 2016 US presidential election [4]. In order to detect fake news, several fact-checking organizations such as *factcheck.org*, *politifact.com* and *snopes.com* spend huge efforts and costs to hire many domain experts to label the truthfulness of news manually. Thus, automatic fake news detector plays a vital role to detect fake news and reduce the negative effects of the fake news.

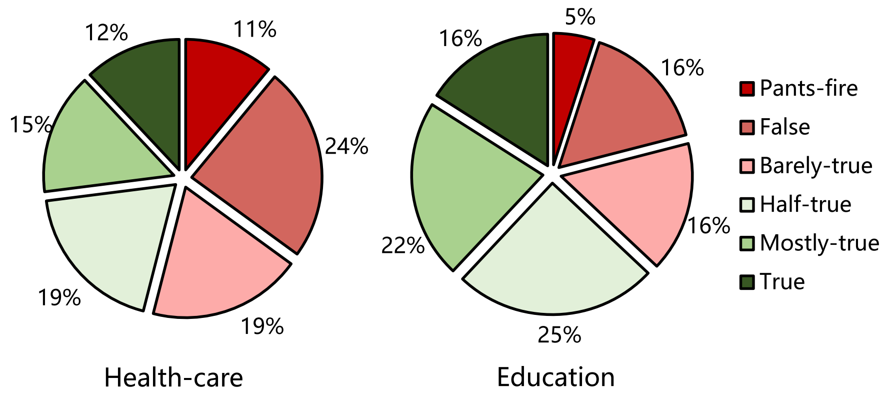
- *Qing Liao is with the Department of Computer Science and Technology, Harbin Institute of Technology(Shenzhen), Shenzhen 518055, China, and also with the Peng Cheng Laboratory, Shenzhen 518055, China. E-mail: liaoqing@hit.edu.cn.*
- *Heyan Chai, Hao Han, Xuan Wang, and Wen Xia are with the Department of Computer Science and Technology, Harbin Institute of Technology(Shenzhen), Shenzhen 518055, China. E-mail: {chaiheyan, hanhao}@stu.hit.edu.cn, wangxuan@cs.hitsz.edu.cn, xiawen@hit.edu.cn.*
- *Xiang Zhang is with the Department of Computer Science and Technology, Harbin Institute of Technology(Shenzhen), Shenzhen 518055, China, and also with the College of Computer, National University of Defence Technology, Changsha 410073, China. E-mail: zhangxiang08@nudt.edu.cn.*
- *Ye Ding is with the School of Cyberspace Security, Dongguan University of Technology, Dongguan 523808, China. E-mail: dingye@dgut.edu.cn.*

Manuscript received 11 June 2019; revised 27 Nov. 2020; accepted 22 Jan. 2021. Date of publication 28 Jan. 2021; date of current version 6 Oct. 2022.
(Corresponding author: Ye Ding.)
Recommended for acceptance by Y. Zhang.
Digital Object Identifier no. 10.1109/TKDE.2021.3054993

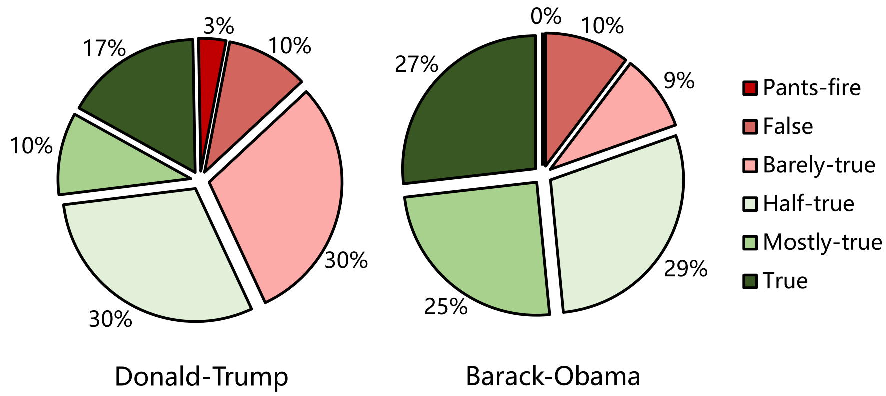
Recently, automatic fake news detection attracts a large number of researchers [5], [6], [7]. Early fake news detection methods [8], [9], [10] often design comprehensive sets of handcrafted features based on news contents, user profiles and propagation paths of news, and then train classifiers to discriminate the truthfulness of news. However, it is difficult to design all-encompassing features since fake news is usually created across different types of topics, writing styles and social media platforms [11]. Hence, deep neural network approaches [12], [13], [14], [15], [16], [17], [18] have been proposed to automatically learn discriminate patterns from news contents and propagation paths. Unfortunately, previous works increase the performances of fake news detection model, but drop dramatically when the news contents are short [15].

Nowadays, Facebook, Twitter, Reddit and Tumblr become important sources of news and thousands of users read news from above social media platforms. For example, there are 66 percent of Facebook users, 59 percent Twitter users and nearly 70 percent Reddit users get news on Facebook, Twitter and Reddit, respectively [19]. In order to distinguish fake news with short contents on social media, we investigate the one of largest fake news dataset which called LIAR dataset as shown in Fig. 1. First, according to Fig. 1a, 63 percent of *education* news tend to be true while only 37 percent of *health-care* news tend to be true. Furthermore, there are 11 percent *health-care* news are *pants-fire* which is more than twice to *education* news (5 percent). Second, according to Fig. 1b, Barack Obama published 81 percent true *economic* news while Donald Trump only published 57 percent true *economic* news. Furthermore, Barack Obama did not published any *pants-fire economic* news, but Donald Trump published 3 percent of them. At the last, we summarize the two observations on social media platform as below:

- *Different topics have different credibility distributions and some certain topics have higher percentages of fake news.*



(a) Credibility distribution of different topics



(b) Credibility distribution of the topic *economic* for different authors

Fig. 1. Credibility distributions of different topics and authors.

For example, *immigration* and *crime* are the main topics of online disinformation in Austria and Germany [20].

- *Different authors also have different credibility distributions and some certain news authors have a higher probability of publishing fake news.* For example, different U.S. Presidents have different credibility distributions of the same topic.

Based on the above two observations, we assume that the news topic and contextual information such as author profiles have certain impacts of fake news detection task. In this work, we propose a novel *fake news detection multi-task learning (FDML)*, which trains the fake news detection task and the news topic classification task simultaneously through multi-task learning, by leveraging the correlations among the news topics, the credibility distributions of authors, and the truthfulness of news at the same time. Moreover, we propose a novel news graph (N-Graph) method to obtain richer news representation via preserving relationships among news and design a simple dynamic weighting strategy to automatically balance the importance of multiple tasks to achieve better performance of both fake news detection task and news topic classification task at the same time. The main contributions of this paper lie in the following aspects:

- We propose a novel *fake news detection multi-task learning (FDML)* model that incorporates topic information and contextual information to enhance the performance on both the fake news detection task and the news topic classification task, simultaneously. To the best of our knowledge, this is the first work that aims to tackle fake news detection task and news topic classification task together in a unified approach through multi-task learning.
- We propose a novel news graph (N-Graph) approach in the representation part to capture relationships of the news based on textual information and contextual information, simultaneously. Furthermore, we design a dynamic weighting strategy to reduce the risk of imbalanced multi-task learning where a single task may occur overfitting across multiple tasks.
- We evaluate the FDML model on one of the largest real-world public fake news dataset and the results show that our approach has the best performances on both tasks in terms of fake news detection task and topic classification task.

The rest of this paper is organized as follows: Section 2 gives a brief review of related works. Section 3 presents the details of the FDML model. Experimental results are presented in Section 4. Finally, Section 5 concludes the study and discusses the future works.

2 Related Work

2.1 Propagation-Based Approach

Using propagation-based fake news detection approaches to mine diffusion patterns of fake news has been studied for many years [21], [22], [23], [24], [25], [26]. For example, Mendoza *et al.* [21] analyzed retweeting topology network and found the differences between rumour diffusion patterns on Twitter and traditional news platforms. Gupta *et al.* [22] proposed a PageRank-like credibility propagation algorithm by encoding the credibilities on three layers of user-tweet-event heterogeneous information network. Kown *et al.* [26] proposed a new method called the Periodic External Shocks (PES) model which combined a set of linguistic features and the network structure together to identify rumours. Chen *et al.* [23] proposed a hybrid approach based on click-baiting cues to detect fake news. Tacchini *et al.* [24] proposed a novel adaptation of boolean crowdsourcing algorithms to show that the diffusion pattern of information can be a useful component of automatic hoax detection systems. However, all these works require a large effort to collect propagation information, especially when diffusion paths change rapidly in social media platforms.

2.2 Stance-Based Approach

Stance-based fake news detection approaches mainly rely on hand-crafted linguistic or embedding features to learn keywords, linguistic cues or writing styles based on news contents [8], [9], [27], [28], [29], [30]. For example, Castillo *et al.* [8] designed a series of hand-crafted features based on linguistic features to determine the truthfulness of Twitter tweets. Yang *et al.* [10] extracted an extensive set of features from micro blog data to train a classifier that automatically detects fake news. Feng *et al.* [9] utilized a wide range of linguistic features such as n-gram, part-of-speech tags and production rules based on the probabilistic context-free grammar to detect deception. Qazivinan *et al.* [27] proposed a general framework based on content-based, network-based, and micro-blog-specific memes to retrieve rumours tweets that match a more general query. Rubin *et al.* [29] provided a conceptual overview of satire and humor, elaborating and illustrating the unique features of satirical news to detect potentially misleading news. Liu *et al.* [31] and Zubiaga *et al.* [32] collected contains many stories and relevant tweets for each story. Conroy *et al.* [30] promised an innovative hybrid approach which combines linguistic cues and network-based behavior data to find fake news. Unfortunately, it is difficult to design all-encompassing features since fake news is usually created across different types of topics. Moreover, the above works cannot capture useful features to achieve a satisfactory performance when the size of the news is not long enough.

2.3 Deep Learning Based Approach

Given the powerful automatic feature learning ability of deep learning, many works focus on detecting fake news by using deep neural networks in recent years. For example, Ma *et al.* [12] proposed a deep neural network based on RNN to capture the temporal and textual features from rumour posts. Liu and Wu [33] utilized recurrent and convolutional networks to detect fake news through propagation path classification. Singhania [34] *et al.* proposed a three-level hierarchical attention network to extract textual features from news content. Ma *et al.* [35] focused on sentiment analysis and jointly optimized rumour detection and stance classification task based RNN. Shu *et al.* [36] proposed a new explainable detection framework for fake news, which exploits both news contents and user comments to explain why the piece of news is fake. Yang *et al.* [37] proposed an unsupervised approach to detect fake news with binary categories by exploiting specific users' engagements on social media platforms. Ruchansky *et al.* [38] proposed a hybrid deep network with multiple branches for high-level representation from contextual and textual information. Nguyen *et al.* [39] proposed a deep network with Markov Random Field (MRF) to incorporate the correlations across the news articles. However, most of previous works collect a large amount on comments of news as an important factor to propose binary classification fake news detection models.

Moreover, there are several works [14], [15], [40], [41] mainly focus on fine-grained short fake news detection. For example, Wang [15] proposed a hybrid convolutional neural networks fake news detection framework with six fake levels on one of the largest short fake news benchmark LIAR dataset. Long *et al.* [40] employed contextual information such as author profiles as attention factors and proposed a hybrid LSTM model to detect fake news with multiple fake labels. Karimi *et al.* [14] applied a deep neural network to learn features from multi-perspective information and then combined these features by attention mechanism. However, none of the existing work investigated the correlations between topics and the truthfulness of news.

Inspired by the success of multi-task learning on data mining applications [42], [43], [44], [45], [46]. We consider the news topic information from the contextual information and capture the relationship between news topic and news truthfulness to jointly learn the fine-grained fake news detection task and news classification task together via multi-task learning.

3 Methodology

In this section, we will show the details of the proposed FDML model. First, we give the definitions of the fake news detection problem and topic classification problem. Then we show the detailed design of the FDML model. Details of each component of the FDML model are introduced in the followed subsections.

3.1 Problem Definition

Before introducing the problem definition, let us first introduce some terminologies used in this paper.

Definition 3.1 (Textual Information). *Let textual information $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$ be a news set that containing $|\mathcal{D}|$ news. Textual information means the text of the news.*

Definition 3.2 (Contextual Information). *Let contextual information $\mathcal{C} = \{R_1, R_2, \ldots, R_{|\mathcal{C}|}\}$ be a contextual set that containing $|\mathcal{C}|$ types of side information of the news.*

Definition 3.3 (News Label). *Each piece of news is associated with labels $y = \{y^t, y^f\}$ for topic label and fake news label.*

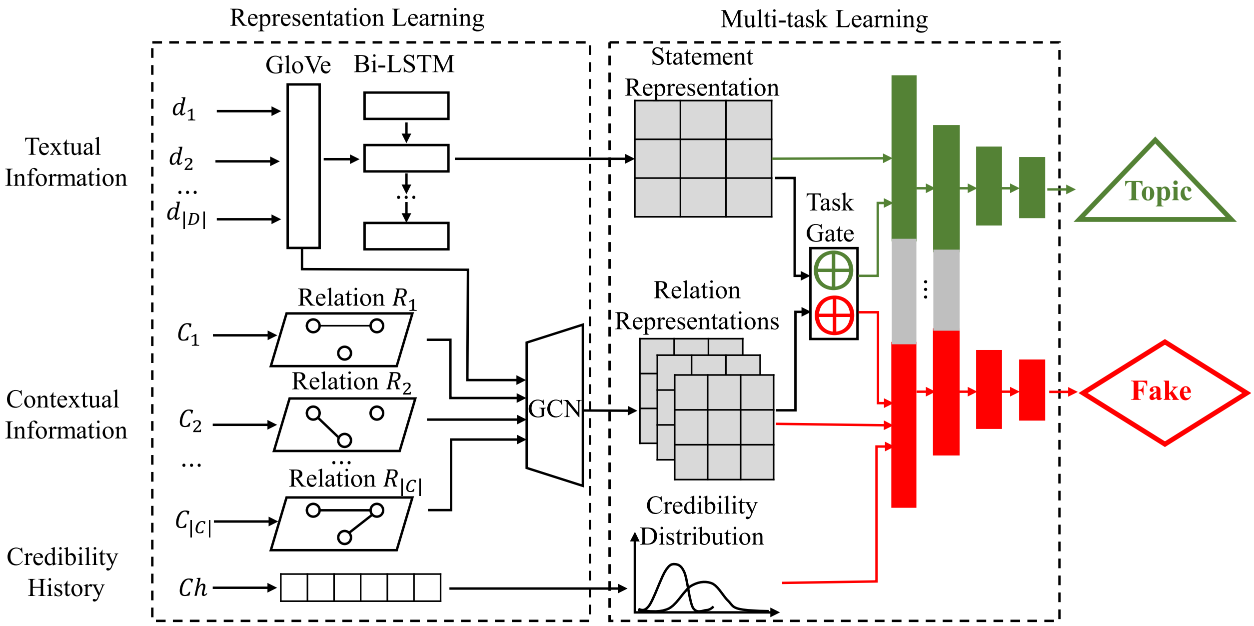


Fig. 2. The framework of FDML. The representation learning part learns statement representation and relation representations from textual information and contextual information. The multi-task learning integrates different types of representations via task gate and dynamically balances the importance of topic classification task and fake news detection task.

Definition 3.4 (Topic Classification). *Given a set of news $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$, predict topic label $\hat{y}^t$ for each piece of unlabelled news d such that $\hat{y}^t \approx y^t$.*

Definition 3.5 (Fake News Detection). *Given a set of news $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$, predict truthfulness label $\hat{y}^f$ for each piece of unlabelled news d such that $\hat{y}^f \approx y^f$.*

Definition 3.6 (Multi-Task Fake News Detection). *Given a set of news $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{D}|}\}$, find a model $\mathcal{F}$ that predicts the the topic label $\hat{y}^t$ and the truthfulness label $\hat{y}^f$ of the news d simultaneously, such that $\mathcal{F}(d) = \{\hat{y}^t, \hat{y}^f\} \approx \{y^t, y^f\}$.*

As defined in Definitions 3.1 and 3.2, a piece of news consists of two kinds of contents: 1) textual information, for example, “Virginia governor Ralph Northam defies calls to resign over racist photo”; and 2) contextual information, for example, the news is published in “Massachusetts”, and the speaker of the news is “Willard Mitt Romney”, who is a “republican”. The “location of speech”, “Speaker name” and “political party” are three types of contextual information.

According to Definition 3.5, the objective of FDML is to jointly train the fake news detection task and the topic classification task, simultaneously.

3.2 Model Overview

Fig. 2 shows the framework of FDML model. FDML consists of two parts: representation learning part and multi-task learning part. For the representation learning part, we propose a novel news graph (N-Graph) approach to preserve relationships of news based on textual information and contextual information, simultaneously. For the multi-task learning part, we design a simple dynamic weighting strategy to balance the importance of multiple tasks. In the following subsections, we will introduce two learning parts of FDML respectively.

3.3 Representation Learning

Several short fake news detection models [15], [38] cooperate contextual data to provide more representations of news to enhance detection accuracy. However, most existing approaches mainly regard contextual data as normal features that ignore the relationships between news. A straightway to preserve relationships among news is using similarity-based methods such as cosine similarity, euclidean distance to calculate differences between news. The above similarity-based methods ignore effects of semantic information of news under different relationship spaces. In the representation learning part of FDML, we propose a novel news graph (N-Graph) method to integrate semantics information and relation subspaces to calculate differences among news. Fig. 3 shows the details of method N-Graph.

First, we build a prototype of relation R_i according to ith type contextual information. $P_k^{R_i}$ is a relation subspace in relation type R_i which is defined as

$$P_k^{R_i} = \frac{1}{|N_k^{R_i}|} \sum_{j=1}^{|N_k^{R_i}|} d_j, \tag{1}$$

where k is a number of relation subspaces in relation type R_i. For example, R_i is political party from contextual C_i, and there is $k = 24$ parties in the experimental datatset such like democratic party, republican party and libertarian party. Hence, there is twenty-four $P_k^{R_i}$ subspaces in relation type R_i. d_j is the embedding vector of jth news via embedding tools GloVe [47]. $|N_k^{R_i}|$ is the number of news which belongs to subspace k in R_i.

Then, we project semantic information of news into subspaces k in relation R_i so that N-Graph can capture more accurate differences between news under different relation subspaces. $S_k(d_m, d_n)$ denotes similarities between news d_m and news d_n in subspace k of relation R_i, which is defined as

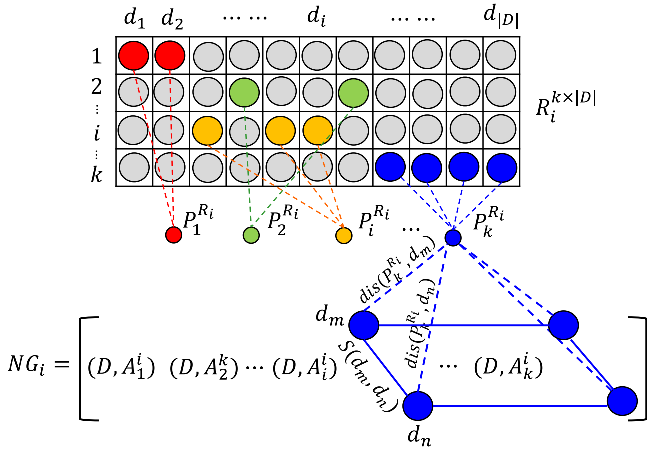


Fig. 3. News graph(N-Graph) preserves relationship among news under different subspace. It projects news semantics information of news into relation subspace $P_k^{R_i}$ to calculate differences between news d_m and news d_n.

$$S_k(d_m, d_n) = \left|dis(P_k^{R_i}, d_m) - dis(P_k^{R_i}, d_n)\right| \cdot \left\|P_k^{R_i}\right\|_2, \tag{2}$$

where $dis(\cdot)$ calculates distances between subspace $P_k^{R_i}$ and news d based on semantics information via euclidean distance or cosine similarity methods. We empirically choose cosine similarity as $dis(\cdot)$ in this work. Both d_m and d_n belong to same relation subspace k in relation type R_i. Furthermore, $\|P_k^{R_i}\|_2$ is the L_2 norm of $P_k^{R_i}$.

We build an adjacent matrix A_i^k of subspace k in relation type R_i according to $S_k(d_m, d_n)$

$$A_k^i = \begin{cases} S_k(d_m, d_n), & \text{if } d_m \text{ and } d_n \text{ belong to space } k \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where $A^i = \{A_1^i, A_2^i, \dots A_k^i\}$ and a news graph $NG_i = (D, A^i)$ preserves relationships among news in relation R_i. A set of news D means nodes in news graph which contain semantics information of news and A^i means edges between news which contains relationship among news in relation type R_i. There is $|C|$ news graphs built by N-Graph method according to contextual information $C = \{R_1, R_2, \dots, R_{|C|}\}$.

In this work, we choose A_k^i as a novel adjacent matrix in graph convolutional network (GCN) [48] to learn different relation representations under different subspaces. Each hidden layer in relation R_i is defined as

$$H_{l+1}^i = f(H_l^i, A^i), \tag{4}$$

where l is the index of layers and A^i is the adjacent matrix of news under relation type R_i. The non-linear function $f(H_l^i, A^i)$ is defined as

$$f(H_l^i, A^i) = \sigma(\widehat{D^i}^{\frac{1}{2}}\widehat{A^i}\widehat{D^i}^{-\frac{1}{2}}H_l^i W_l), \tag{5}$$

where W_l is a weight matrix for the lth hidden layer and $\sigma(\cdot)$ is a non-linear activation function. Adjacent matrix $\widehat{A^i} = A^i + I$, where I is the identity matrix and $\widehat{D^i_{ii}} = \sum_j \widehat{A^i}_{ij}$.

Furthermore, we choose Bi-LSTM [49] to learn statement representation based on textual information of news, due to Bi-LSTM can capture richer semantic features among words in the short texts from forward and backward direction simultaneously [50]. In summary, the representation learning part learns relation representations and statement representations, simultaneously. Then multi-task learning part receives and selectively integrates representations depends on different tasks.

3.4 Multi-Task Learning

In a traditional multi-task learning model, the weight of each task is fixed during the entire learning process, which may cause the imbalance learning problem: the network pays more efforts to train the task with lager weight during the entire learning process and may ignore the importance of another task with a smaller weight. In the multi-task learning part of FDML, we consider the imbalance learning problem to propose a simple dynamic weighting strategy where the weight of each task is dynamically adjusted in each iteration. Specifically, if task A converges or reaches satisfactory performance before task B, then the strategy will assign a smaller weight to task A and a larger weight to task B during future training. The advantage of dynamic weighting is that it reduces the risk of a single task may occur over training across multiple tasks and can search suitable weights of each task automatically.

First, we introduce a task gate to select and fuse relevant features by considering the relationship between features and tasks. We train several task *Mask* matrices to selectively integrate statement representation and relation representations based on attention mechanism for different tasks so that the model can filter irrelevant features of the task to reduce noises and the number of training weights.

The integrated representations of news for topic classification Z_t and fake news detection Z_f are defined as

$$Z_t = concat(f_1(NG) \odot Mask_t, f_2(D)) \tag{6}$$

$$Z_f = concat(f_1(NG) \odot Mask_f, f_2(D), Ch), \tag{7}$$

where $f_1(NG)$ denotes multiple relation representations via GCN and $f_2(D)$ is statement representation of news via Bi-LSTM. $concat$ is a concatenation operator to generate new integrated representations based on relation and statement. For fake news representation, we fuse the credibility history Ch due to considering different authors have different credibility distributions which may influence the truthfulness of the same topic according to observation 2. We train mask matrices $Mask_t$ and $Mask_f$ in task gate to decide which relation representations should be activated in topic classification task and fake news detection task, respectively.

Then we choose an attention mechanism to capture relationships between features and features so that the model can fuse selected features to obtain richer knowledge. Calculating attention scores of features for each news based on different tasks according to [51]

$$u_{fi} = \tanh(W_{fi} Z_{fi} + b_{fi}) \tag{8}$$

$$a_{fi} = \frac{\exp(u_{fi})}{\sum_i \exp(u_{fi})}, \tag{9}$$

where u_{fi} is a word vector of ith news and a_{fi} is a normalized importance weight of each feature for ith news in fake news detection, respectively. Similarly, u_{ti} and a_{ti} are word vector and importance weight vector of the ith news in topic classification task, respectively.

We predict the truthfulness label of news via a softmax classifier, respectively

$$\hat{y}^f = \mathrm{Softmax}(W_2^f \cdot ReLU(W_1^f a_f Z_f) + b^f), \tag{10}$$

where W_1^f and W_2^f are weight matrices of fake news detection task. b^f is the corresponding bias. The news topic label $\hat{y}^t$ is defined similarly as Eq. (10). The final loss function across fake news detection and topic classification task is defined as the cross-entropy error

$$\mathcal{L}(\Theta) = w_f \sum_{d\in D} y_d^f \log \hat{y}_d^f + w_t \sum_{d\in D} y_d^t \log \hat{y}_d^t + \lambda\|\Theta\|_2, \tag{11}$$

where θ_s is the share parameter of multiple task learning, θ_f and θ_d are local parameters corresponding to fake news detection and topic classification task, respectively. $\Theta = \{\theta_s, \theta_f, \theta_t\}$ represents the total parameters, and λ is the trade-off coefficient of L2 regularizer. The $\lambda\|\Theta\|_2$ is to reduce the complexity and overfitting of the model. y_d^f is ground-truth truthfulness label of d news' truthfulness label and $\hat{y}_d^f$ is prediction distribution of d news' truthfulness label. In our work, both y_d^f and $\hat{y}_d^f$ are six-length vectors due to there are six truthfulness labels *pants-fire*, *false*, *barely-true*, *half-true*, *mostly-true* and *true*. Similar, y_d^t and $\hat{y}_d^t$ are ground-truth topic label and prediction distribution of topic label. Furthermore, we dynamically adjust fake news task weight w_f and topic classification weight w_d as below:

$$Speed_f^l = \mathcal{L}_f^l \Big/ \mathcal{L}_f^{l-1}, \quad Speed_t^l = \mathcal{L}_t^l / \mathcal{L}_t^{l-1} \tag{12}$$

$$w_f^l = \frac{Speed_f^l}{Speed_f^l + Speed_t^l}, \quad w_t^l = \frac{Speed_t^l}{Speed_f^l + Speed_t^l}, \tag{13}$$

where l is iteration epoch. $\mathcal{L}_f = \sum_{d\in D} y_d^f \log \hat{y}_d^f$ and $\mathcal{L}_t = \sum_{d\in D} y_d^f \log \hat{y}_d^f$ are loss function of fake news detection and topic classification task, respectively. $Speed_f^l$ and $Speed_t^l$ denote the loss ratio for tasks at l iteration. The FDML model tries to slow down task A's speed of updating if it converges faster than task B so that FDML can reduce the risk of overfitting of task A and take more time to learn of task B.

4 Experiments

In this section, we first introduce the dataset and corresponding settings used in the experiments, then we compare the performance of the FDML model with the state-of-the-art fake news detection approaches and news topic classification approaches, respectively. At last, we discuss the effectiveness of multi-task learning and elaborate on the detailed benefits of combining fake new detection task and topic classification task together.

4.1 Dataset

We conduct our experiments using the LIAR dataset, which is one of the largest real-world public fake news benchmark for short fake news detection [11], [14], [15]. The dataset contains 12,836 manually labeled short news from 3,318 public speakers. All the news within the dataset are short news with an average 17.9 tokens and each news contains rich contextual information. These news come from various contexts/venues, such as tweets, TV ads, debates, etc. Each news is evaluated by *politifact.com* editors and the website provides a detailed verdict report of each fake news. The fine-grained truthfulness labels of the dataset are: *pants-fire*, *false*, *barely-true*, *half-true*, *mostly-true* and *true*.

TABLE 1
An Example of the LIAR Dataset

Statement	"When Mitt Romney was governor of Massachusetts, we didnt just slow the rate of growth of our government, we actually cut it."
Speaker Name	Willard Mitt Romney
Current Job Position	Former Governor
Home State	Detroit, Michigan
Political Party	Republican
Location of Speech	Massachusetts
Credit History	$(34, 32, 58, 33, 19, 33)$
Topic Label	History
Truthfulness Label	False

Table 1 gives an example of the LIAR dataset, each news contains two labels including topic label and truthfulness label. Besides the statement of news, the LIAR dataset also contains many contextual information, including 1) speaker name, 2) current job position, 3) home state, 4) political party, 5) location of speech and 6) credit history of the speaker. In particular, credit history is the historical distribution of the truthfulness labels for the speaker. According to Table 1, the credit history vector of speaker Willard Mitt Romney is (34,32,58,33,19,33), which corresponds to counts of news at *pants-fire*, *false*, *barely-true*, *half-true*, *mostly-true* and *true*. The first term 33 means that Willard Mitt Romney has posted 34 *pants-fire* news and the second term 32 means that he has posted 32 *false* news.

There are 144 news topics in the LIAR dataset, in this work, in order to clearly show a comparison between the FDML model and existing works, we choose top 24 of the most frequent topics as well as an *others* topic containing the rest of the topics as shown in Table 2. The reason we choose top 24 is because they consist of around 80 percent of the entire dataset. To clearly show the scalability of the

TABLE 2
Top 24 News Topics Adopted in the Experiments

Economy	Health-care
Candidates-biography	Education
Elections	Federal-budget
Crime	Taxes
Immigration	Foreign-policy
Energy	Abortion
State-budget	Jobs
Guns	Campaign-finance
Children	Deficit
Congress	Corrections-and-updates
Environment	History
Job-accomplishments	Corporations

TABLE 3
Fake News Detection Performance of Five Detection Models of Six Truthfulness Labels on the LIAR Dataset

Fake News Detectoion Model	Label	Precision	Recall	F1 Score	Accuracy	Macro-F1
Hybrid-CNN [15]	Pants-fire	0.613	0.084	0.147	0.302	0.269
	False	0.323	0.376	0.347		
	Barely-True	0.257	0.231	0.243		
	Half-True	0.312	0.383	0.344		
	Mostly-True	0.306	0.423	0.355		
	True	0.257	0.136	0.178		
LSTM-Attention [39]	Pants-fire	0.514	0.346	0.413	0.390	0.402
	False	0.386	0.576	0.462		
	Barely-True	0.375	0.396	0.385		
	Half-True	0.418	0.428	0.423		
	Mostly-True	0.432	0.424	0.428		
	True	0.373	0.165	0.229		
MMFD [14]	Pants-fire	**0.710**	0.393	0.506	0.422	0.418
	False	0.453	0.479	0.466		
	Barely-True	**0.586**	0.290	0.388		
	Half-True	0.322	**0.636**	0.427		
	Mostly-True	0.425	0.480	0.451		
	True	0.881	0.159	0.269		
Memory-Network [40]	Pants-fire	0.661	0.509	0.575	0.452	0.449
	False	0.421	0.614	0.500		
	Barely-True	0.483	**0.401**	0.438		
	Half-True	0.412	0.525	0.462		
	Mostly-True	0.421	0.483	0.450		
	True	**0.916**	0.156	0.267		
FDML(Ours)	Pants-fire	0.662	**0.554**	**0.604**	**0**.508	**0.516**
	False	**0.445**	**0.644**	**0.526**		
	Barely-True	0.540	0.383	**0.448**		
	Half-True	**0.500**	0.401	**0.511**		
	Mostly-True	**0.493**	**0.530**	**0.511**		
	True	0.567	**0.564**	**0.565**		

FDML model, we have conducted experiments upon different numbers of topics from 2 to 25.

As a benchmark, the LIAR dataset is pre-configured with three subsets: training set containing 80 percent of the entire dataset; testing set containing 10 percent of the entire dataset; and validation set containing 10 percent of the entire dataset.

4.2 Experiment Settings

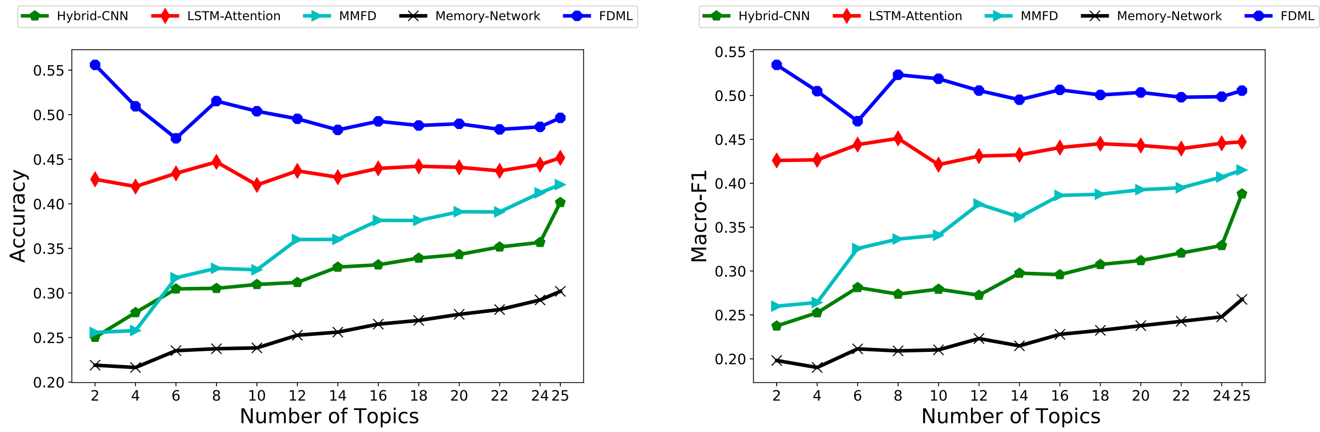
We use the percentages of correctly classified fake news and news topics as the accuracy of the fake news detection task and the topic classification task, respectively. Considering accuracy reflects the classification accuracy of the whole model. The precision and recall reflect the accuracy of positive and negative samples of each label. Hence, in this work, we choose accuracy and Macro-F1 to evaluate performances of all comparison models. To reduce the influence of initial random seeds, we calculate the average accuracy with 10 trials. We use 300-dimensional GloVe tools [47] to initialize word vector. The hidden unit in Bi-LSTM is set to 400 and the hidden unit in GCN is set to [128, 32]. The coefficient of L_2 regularizer λ is set to 0.0001. All parameters of FDML are trained by Adam optimizer with a weight decay strategy for 150 epochs. There are total of four fully connected layers where each task contains two fully connected layers such as ReLu and Softmax to do the classification task as most deep learning works settings.

4.3 Performance of Fake News Detection

To evaluate the fake news detection performance on the LIAR dataset, we compare the FDML model with the following latest fake news detection methods:

- *Hybrid-CNN* [15]: a hybrid CNN that integrates text and contextual information together to detect fake news.
- *LSTM-Attention* [40]: a hybrid LSTM that weighs the importance of words by attention to detect fake news.
- *Memory-Network* [41]: a memory network that uses contextual information as attention factor to detect fake news.
- *MMFD* [14]: a multi-source multi-class fake news detection model which introduces additional information from other sources.

Table 3 shows the detailed fake news detection results in terms of precision, recall, accuracy and F1 score. It is clearly that FDML has about 6 percent improvements in six truthfulness labels of fake news detection in terms of average accuracy and Macro-F1. More specifically, some models such as Hybrid-CNN, LSTM-Attension and Memory-Network have imbalance problem between precision score and recall score. For example, Hybrid-CNN has 0.613 precision score but only 0.084 recall score in *Pants-fire* label. And Memory-Network has 0.916 precision score but only 0.156 recall score in *True* label which seriously impacts the performances of F1 score and accuracy. Therefore, existing comparison methods difficult to achieve satisfied detection performance within all truthfulness labels. On the other hand, FDML has more balanced scores between precision and recall in all the truthfulness labels, such as the performance of True label. The worst comparison methods, Hybrid-CNN only has 0.302 accuracy and 0.269 Macro-F1 score, because it is the first work of fake news detection work in LIAR benchmark and it just simply combines basic CNN and LSTM models together to extract features.



(a) Accuracy of Fake News Detecion

(b) Macro-F1 of Fake News Detecion

Fig. 4. The fake news detection performance of comparison models upon different number of topics.

Fig. 4 shows the accuracy and Macro-F1 score of fake news detection of five comparison methods under different number of topics. More topics are chosen, a larger subset of the original LIAR dataset is constructed. As we can see, all the methods including FDML perform better on a larger subset. Moreover, FDML always performs better than all comparison methods despite of the number of topics. Fig. 4 indicates that FDML has a satisfactory scalability upon the number of topics.

4.4 Performance of News Topic Classification

To evaluate the news topic classification performance of FDML, we compare the FDML model with the following well-known topic classification models on the LIAR dataset:

- *RF* [52]: the random forest classifier using term frequency (TF) and inverse document frequency (IDF) as features.
- *Bi-LSTM* [49]: the bidirectional LSTM for text classification.
- *CNN* [53]: the convolutional neural network for text classification.
- *SWEM* [54]: a word embedding model can be applied in short text classification.
- *LEAM* [55]: a label embedding attentive model can be applied in short text classification which aims to give higher weights to related words.
- *Text-GCN* [56]: a graph convolutional networks for short text classification.
- *TMN* [57]: a topic memory networks for short text that uses topic memory mechanism to encode latent topic representations indicative of class labels.

Table 4 shows average accuracy and Macro-F1 scores from 2 to 25 topics of eight topic classification models in LIAR dataset. As the Table 4 shows, FDML has the highest accuracy and Marco-F1 scores comparing with seven topic classification models on LIAR datasets, which illustrates the effectiveness of the proposed multi-task learning method on the task topic classification. RF is one of the well-known basic models for topic classification which has the lowest accuracy and Marco-F1 score. Topic classification models such as Bi-LSTM, SWEM, LEAM, Text-GCN and TMN can be applied for short text which have similar results in LIAR dataset. However, there is still a gap between FDML, because FDML can not only effectively integrate the contextual information to learn the pattern of news topics, but also cooperate the relationship between the news topics and the truthfulness of news together to provide more useful information to topic classification task.

Fig. 5 shows the trends of topic classification performance of all comparison models upon different numbers of topics in LIAR dataset. The accuracy and Macro-F1 score of all models generally drop when the number of topics increases, which is reasonable due to the difficulty of topic classification rises upon a large number of topics. It is clear that FDML outperforms the other topic classification methods in most of the cases.

4.5 Effectiveness of Multi-Task Learning

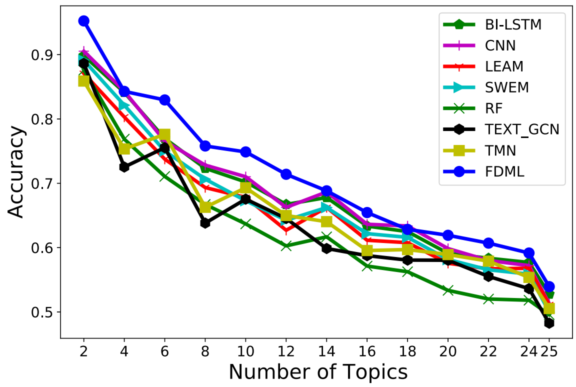
To further show the effectiveness of multi-task learning and explore the relationship between the truthfulness of the news and its corresponding topics, we modify the FDML model to propose two heuristic models:

- *FDML*: the model jointly learns fake news detection task and topic classification task, simultaneously.
- *FDML-*: by removing the topic classification component in FDML, which can only detect fake news.
- *FDML**: by removing the fake news detection component in FDML, which can only classify topics.

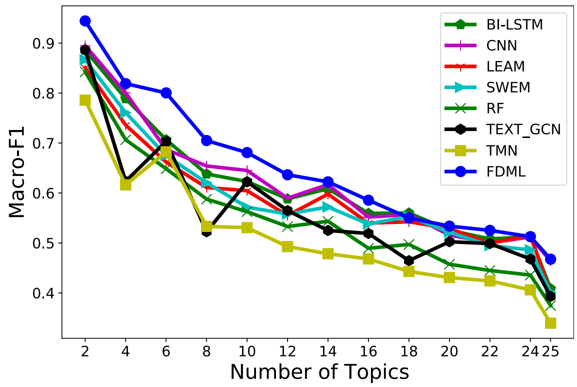
Fig. 6a illustrates the comparison of fake news detection performance of FDML and FDML- of accuracy score. We

TABLE 4
The Topic Classification Performance of Comparison Methods in LIAR Dataset

Topic Classification Model	Accuracy	Marco-F1
RF [52]	0.622	0.548
Bi-LSTM [49]	0.678	0.609
CNN [53]	0.680	0.615
SWEM [54]	0.661	0.586
LEAM [55]	0.655	0.588
Text-GCN [56]	0.634	0.561
TMN [57]	0.650	0.510
FDML(Ours)	**0.706**	**0.645**

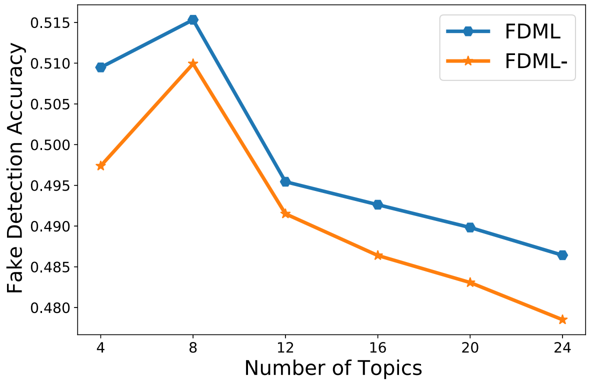


(a) Accuracy of Topic Classification

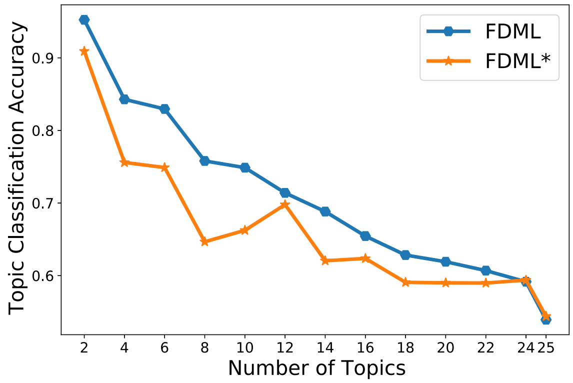


(b) Macro-F1 of Topic Classification

Fig. 5. The news topic classification performance of comparison models upon different number of topics in LIAR dataset.



(a) Fake news detection performance of FDML and FDML- upon different number of topics.



(b) Topic classification performance of FDML and FDML* upon different number of topics.

Fig. 6. The performance of FDML, FDML- and FDML* of two tasks.

Fig. 7. An education news published by Rick Scott [58].

can find that FDML outperforms FDML- despite of the number of topics. It indicates that FDML can effectively utilize the topic information to assist the detection of fake news through multi-task learning.

Fig. 6b illustrates the comparison of topic classification performance of FDML and FDML* of accuracy score. Although the topic classification task becomes difficult with the number of topic increases, the FDML still has the best performance comparing with FDML* in all the cases. It indicates that jointly learning two tasks brings more benefits than single task learning.

In order to get an intuitive understanding of the benefits of jointly performing news topic classification task and fake news detection task, we choose Fig. 7 as a case study to show the effectiveness of FMDL in real-world. According to Fig. 7, it is an education news published by Rick Scott in 2013, says his budget provides "the highest state funding level in history" for education. This news is manually labeled as *half-true* according to the experts in *politifact.com*. Actually, Scott is referring only to the state government's specific contribution to overall spending, which ignores the fact that there are also local and federal investments that fund state schools [58].

Table 5 shows the FDML and FDML- prediction results of *Fig. 7*. It is clear that: 1) for fake news detection, FDML

TABLE 5
Prediction Results of FDML and FDML- for the Education News Published by Rick Scott

	Fake News Detection	Topic Classification
FDML-	*mostly-true*	-
FDML	*half-true*	*education*
Ground Truth	*half-true*	*education*

TABLE 6
Fake News Detection Performance of the FDML Model With/Without Authors' Information

Model	Without author information		With author information		Gap on Accuracy	Gap on Macro-F1
	Accuracy	Macro-F1	Accuracy	Macro-F1		
Hybrid-CNN [15]	0.235	0.183	0.302	0.269	+0.067	+0.086
LSTM-Attention [40]	0.247	0.225	0.390	0.402	+0.143	+0.177
MMFD [14]	0.219	0.271	0.422	0.418	+0.203	+0.147
Memory-Network [41]	0.220	0.266	0.452	0.449	+0.232	+0.183
FDML Model	0.259	0.252	0.508	0.516	**+0.249**	**+0.264**

predicts the truthfulness correctly while FDML- fails; and 2) for news topic classification, FDML predicts the correct topic of the news. Looking into the statement itself, it is very difficult to identify the truthfulness of the news statement only based on text contents without additional information as shown above, especially when the speaker Rick Scott has the highest percentage of *mostly-true* (more than 50 percent) according to his credit history under the *education* topic. Hence, it is reasonable for FDML- not classifying correctly. However, as illustrated in Section 3, by utilizing the hidden information between topic classification and fake news detection, it is possible to distinguish the news as *half-true* via FDML. This concludes the effectiveness of using multi-task learning in FDML.

We find that the author information is very important in the news detection task, especially for the short-text news. Because contents of short news contain limited information which is difficult to obtain enough representations during the learning and leads to bad performance of the model. We conduct some supplementary experiments to illustrate the importance of author information as shown in Table 6. We can find that performances of all the detection models decrease when we remove the author information. That is the reason why the most existed fake news detection models try to introduce author information to help the detection task. Furthermore, we can find that the FDML model achieves higher improvements with author information comparing with other methods due to the FDML model can capture richer representations from author information.

4.6 Effects of Authors Information

In order to investigate the effects of author information in the fake news detection task, we illustrate the importance of author information as shown in Table 6. We can find that the performances of all the detection models decrease when we remove the author information. That is the reason why most existed fake news detection models try to introduce author information to help the detection task, because contents of short news contain limited information which is difficult to obtain enough representations during the model learning. Furthermore, we can find that the FDML model achieves higher improvements with author information comparing with other methods due to the FDML model can capture richer representations from author information.

5 Conclusion

In this paper, we propose a novel fake news detection multi-task learning (FDML) model which is the first work to improve performances of short fake news detection and topic classification, simultaneously. The FDML model is designed based on the observations that 1) news with certain topics have high probabilities to be classified as fake news; and 2) some authors have high probability to publish fake news. In the representation learning part, FDML proposes a novel news graph (N-Graph) to learn statement representation and relation representations from textual and contextual information. In the multi-task learning part, FDML introduce a task gate to selective integrate representations based on different tasks and design a dynamic weight strategy to balance the importance between two tasks. We choose eleven comparison methods in one of the largest short fake news benchmarks to show that FDML has the best performances in terms of fake news detection task and topic classification task.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under grants No.2017YFB0202201, the National Natural Science Foundation of China under grant No.U1711261 and 61976051, the Guangdong Major Project of Basic and Applied Basic Research under grant No.2019B030302002, and the Basic Research Project of Shenzhen under grant No. JCYJ20180306174743727 and No. JCYJ20190806143405318.

References

[1] N. Shavit, "Data on Facebook's fake news problem," *Jumpshot*, 2016. [Online]. Available: https://www.jumpshot.com/data-facebooksfake-news-problem/

[2] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[3] A. Verma, "Google engineer beaten to death, 3 critical in indian lynch mob attack fueled by 'kidnapping' rumor," *Thomson Reuters*, 2018. [Online]. Available: https://www.rt.com/news/433297-india-men-lynched-kidnappers-rumor/

[4] D. DiFranzo and M. J. K. Gloria, "Filter bubbles and fake news," *ACM Crossroads*, vol. 23, no. 3, pp. 32–35, 2017.

[5] K. Shu, D. Mahudeswaran, and H. Liu, "FakeNewsTracker: A tool for fake news collection, detection, and visualization," *Comput. Math. Org. Theory*, vol. 25, no. 1, pp. 60–71, 2019.

[6] S. Ghosh and C. Shah, "Toward automatic fake news classification," in *Proc. 52nd Hawaii Int. Conf. Syst. Sci.*, 2019, pp. 1–10.

[7] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 3391–3401.

[8] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 675–684.

[9] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proc. 50th Annu. Meeting Assoc. Comput. Linguistics*, 2012, pp. 171–175.

[10] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on Sina Weibo," in *Proc. ACM Workshop Mining Data Semantics*, 2012, Art. no. 13.

[11] 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.

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

[13] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal fusion with recurrent neural networks for rumor detection on microblogs,” in *Proc. ACM Multimedia Conf.*, 2017, pp. 795–816.

[14] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, “Multi-source multi-class fake news detection,” in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 1546–1557.

[15] W. Y. Wang, “”Liar, liar pants on fire”: A new benchmark dataset for fake news detection,” in *Proc. 55th Annu. Meeting Assoc. Comput. Linguistics*, 2017, pp. 422–426.

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

[17] J. Zhang, L. Cui, Y. Fu, and F. B. Gouza, “Fake news detection with deep diffusive network model,” 2018, *arXiv: 1805.08751*.

[18] X. Liu *et al.*, “Late fusion incomplete multi-view clustering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2410–2423, Oct. 2019.

[19] J. Gottfried and E. Shearer, “News use across social media platforms 2016,” *Journalism*, 2016. [Online]. Available: https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

[20] E. Humprecht, “Where ‘fake news’ flourishes: A comparison across four western democracies,” *Inf. Commun. Soc.*, vol. 21, pp. 1–16, 2018.

[21] M. Mendoza, B. Poblete, and C. Castillo, “Twitter under crisis: Can we trust what we RT?” in *Proc. 1st Workshop Soc. Media Analytics*, 2010, pp. 71–79.

[22] M. Gupta, P. Zhao, and J. Han, “Evaluating event credibility on Twitter,” in *Proc. SIAM Int. Conf. Data Mining*, 2012, pp. 153–164.

[23] Y. Chen, N. J. Conroy, and V. L. Rubin, “Misleading online content: Recognizing clickbait as false news,” in *Proc. ACM Workshop Multimodal Deception Detection*, 2015, pp. 15–19.

[24] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, “Some like it hoax: Automated fake news detection in social networks,” 2017, *arXiv: 1704.07506*.

[25] Q. Liao, W. Wang, Y. Han, and Q. Zhang, “Analyzing the influential people in Sina Weibo dataset,” in *Proc. IEEE Global Commun. Conf.*, 2013, pp. 3066–3071.

[26] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, “Prominent features of rumor propagation in online social media,” in *Proc. IEEE 13th Int. Conf. Data Mining*, 2013, pp. 1103–1108.

[27] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei, “Rumor has it: Identifying misinformation in microblogs,” in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2011, pp. 1589–1599.

[28] Q. Liao and Q. Zhang, “Efficient rank-one residue approximation method for graph regularized non-negative matrix factorization,” in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases*, 2013, pp. 242–255.

[29] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, “Fake news or truth? Using satirical cues to detect potentially misleading news,” in *Proc. 2nd Workshop Comput. Approaches Deception Detection*, 2016, pp. 7–17.

[30] N. J. Conroy, V. L. Rubin, and Y. Chen, “Automatic deception detection: Methods for finding fake news,” in *Proc. 78th ASIS&T Annu. Meeting: Inf. Sci. Impact: Res. Community*, 2015, Art. no. 82.

[31] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah, “Real-time rumor debunking on Twitter,” in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.*, 2015, pp. 1867–1870.

[32] A. Zubiaga, G. W. S. Hoi, M. Liakata, R. Procter, and P. Tolmie, “Analysing how people orient to and spread rumours in social media by looking at conversational threads,” 2015, *arXiv:1511.07487*.

[33] Y. Liu and Y. 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, pp. 354–361.

[34] S. Singhania, N. Fernandez, and S. Rao, “3HAN: A deep neural network for fake news detection,” in *Proc. 24th Int. Conf. Neural Inf. Process.*, 2017, pp. 572–581.

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

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

[37] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, “Unsupervised fake news detection on social media: A generative approach,” in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 5644–5651.

[38] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A hybrid deep model for fake news detection,” in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 797–806.

[39] D. M. Nguyen, T. H. Do, A. R. Calderbank, and N. Deligiannis, “Fake news detection using deep Markov random fields,” in *Proc. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2019, pp. 1391–1400.

[40] Y. Long, Q. Lu, R. Xiang, M. Li, and C. Huang, “Fake news detection through multi-perspective speaker profiles,” in *Proc. 8th Int. Joint Conf. Natural Lang. Process.*, 2017, pp. 252–256.

[41] T. T. Pham, “A study on deep learning for fake news detection,” 2018.

[42] Q. Liao, L. Jiang, X. Wang, C. Zhang, and Y. Ding, “Cancer classification with multi-task deep learning,” in *Proc. Int. Conf. Secur. Pattern Anal. Cybern.*, 2017, pp. 76–81.

[43] Q. Liao, Y. Ding, Z. L. Jiang, X. Wang, C. Zhang, and Q. Zhang, “Multi-task deep convolutional neural network for cancer diagnosis,” *Neurocomputing*, vol. 348, pp. 66–73, 2019.

[44] Q. Liao, N. Guan, and Q. Zhang, “Logdet divergence based sparse non-negative matrix factorization for stable representation,” in *Proc. IEEE Int. Conf. Data Mining*, 2015, pp. 871–876.

[45] P. Liu, X. Qiu, and X. Huang, “Recurrent neural network for text classification with multi-task learning,” in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2873–2879.

[46] X. Liu *et al.*, “Absent multiple kernel learning algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1303–1316, Jun. 2020.

[47] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global vectors for word representation,” in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2014, pp. 1532–1543.

[48] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” 2016, *arXiv:1609.02907*.

[49] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[50] S. Cao *et al.*, “Knowledge guided short-text classification for healthcare applications,” in *Proc. IEEE Int. Conf. Data Mining*, 2017, pp. 31–40.

[51] Z. Yang, D. Yang, C. Dyer, X. He, A. J. Smola, and E. H. Hovy, “Hierarchical attention networks for document classification,” in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2016, pp. 1480–1489.

[52] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

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

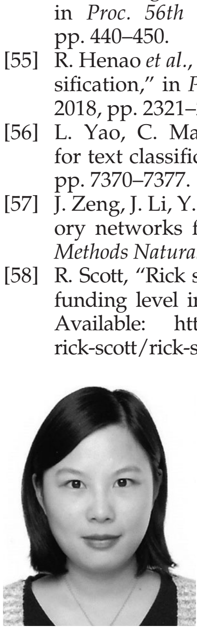
[54] R. Henao *et al.*, “Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms,” in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 440–450.

[55] R. Henao *et al.*, “Joint embedding of words and labels for text classification,” in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics*, 2018, pp. 2321–2331.

[56] L. Yao, C. Mao, and Y. Luo, “Graph convolutional networks for text classification,” in *Proc. 33rd AAAI Conf. Artif. Intell.*, 2019, pp. 7370–7377.

[57] J. Zeng, J. Li, Y. Song, C. Gao, M. R. Lyu, and I. King, “Topic memory networks for short text classification,” in *Proc. Conf. Empir. Methods Natural Lang. Process.*, 2018, pp. 3120–3131.

[58] R. Scott, “Rick scott says his budget proposal has “the highest state funding level in history” for education,” *PolitiFact*, 2013. [Online]. Available: https://www.politifact.com/factchecks/2013/feb/06/rick-scott/rick-scott-says-his-budget-proposal-hashighest-st/



Qing Liao (Member, IEEE) received the PhD degree from the Hong Kong University of Science and Technology, Hong Kong, in 2016. She is currently an associate professor with the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), China. Her research interests include data mining, artificial intelligence, information security, etc.

Heyan Chai received the MEng degree in computer science and technology from the Department of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China, in 2020. He is currently working toward the PhD degree in the Department of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China. His research interests include data mining, multi-task learning, and few-shot learning.

Hao Han received the BEng degree in software engineering from the Department of Computer Science and Technology, Anhui University, Hefei, China, in 2017. He is currently working toward the MEng degree in the Department of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China. His research interests include deep learning and misinformation detection.

Xiang Zhang (Member, IEEE) received the MS and PhD degrees from the National University of Defense Technology, Changsha, China, in 2010 and 2015, respectively. He is currently a research assistant with the Institute for Quantum Information & State Key Laboratory of High Performance Computing, College of Computer, National University of Defense Technology. His current research interests include computer vision and machine learning.

Xuan Wang (Member, IEEE) received the MS and PhD degrees in computer sciences from the Harbin Institute of Technology, Harbin, China, in 1994 and 1997, respectively. He is a professor and PhD supervisor with the Computer Application Research Center, School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), China. His main research interests include artificial intelligence, computer vision, computer network security, and computational linguistics.

Wen Xia (Member, IEEE) received the PhD degree in computer science from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2014. He is currently an associate professor with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen. His research interests include data reduction, storage systems, cloud storage, etc. He has published more than 30 papers in major journals and international conferences including *IEEE Transactions on Computers*, *IEEE Transactions on Parallel and Distributed Systems*, USENIX ATC, USENIX FAST, INFOCOM, IFIP Performance, IEEE DCC, MSST, IPDPS, HotStorage, etc. He is a member of the ACM, CCF, and USENIX.

Ye Ding (Member, IEEE) received the PhD degree in computer science and engineering from the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, in 2014 supervised by Prof. Lionel M. Ni. He is currently an associate professor with the Dongguan University of Technology. His research interests include spatial-temporal data analytics and machine learning.