Looking Beyond Content: Modeling and Detection of Fake News from a Social Context Perspective

Kenan Xiao
Auburn University
kzx0010@auburn.edu

Longwei Wang
Auburn University
lzw0070@auburn.edu

Ashish Gupta
Auburn University
azg0074@auburn.edu

Xiao Qin
Auburn University
xqin@auburn.edu

Abstract

The widespread fake news on social media has boosted the demand for reliable fake news detection techniques. Such dissemination of fake news can influence public opinions and society. More recently, a growing number of methods for detecting fake news have been proposed. However, most of these approaches have significant limitations in timely detection of fake news. To facilitate early detection of fake news, we propose a unique framework FNEPP (Fake News Engagement and Propagation Path) *from a social context perspective, which explicitly combines news contents, user engagements, user characteristics, and the news propagation path as composite features of two collaborative modules. The engagement module captures news contents and user engagements, while the propagation path module learns global and local patterns of user characteristics and news dissemination patterns. Experimental results on two real-world datasets demonstrate the effectiveness and efficiency of the proposed* FNEPP *framework.*

1. Introduction

Nowadays, people prefer searching and consuming news via social media platforms rather than traditional news venues. According to a Pew Research Center survey conducted between August 31 and September 7, 2020, slightly over half of U.S. adults (53%) claim they read news from social media "often" or "sometimes".[1] Social media, of course, is a double-edged sword in terms of news consumption and distribution. Generally speaking, the quality of news written on social media is not on par with that of news published through traditional sources.

Growing evidence indicate that fake news can impose negative impacts on both individuals and society. First of all, individuals may be duped by fake news

and adopt wrong opinions [1, 2]. Second, fake news is intended to potentially alter people's reactions to legitimate news. Third, widespread dissemination of fake news has a potential to undermine the entire news ecosystem's credibility. As a result, it is crucial and demanding to swiftly identify fake news on social media. Fake news is purposefully designed to deceive readers and; therefore, it is non-trivial to detect fake news solely by scanning news content. In fact, concentrating on news content published on on social media becomes inadequate because news does not exist independently in the form of articles [3]. In order to develop effective and accurate fake news detection systems, we advocate for a diversity of supplementary information gleaned from social media to facilitate fake-news detection.

A typical news propagation pattern is shown in Figure 1. To fully characterize the news ecosystem on social media, we propose to model fake news from four perspectives captures from social media data. The most intrinsic characteristic is the text of news articles. Content-based approaches (1) either determine if a news title coheres with its news or (2) measure quality of the writings. Efforts in automating text assessment have advanced machine learning algorithms that categorize news content as legitimate or fake based on hand-crafted and data-specific textual properties [4, 5, 6, 7, 8, 9]. The development of these cutting-edge detection schemes is challenging because the linguistic properties of fake news are still not fully unraveled. Furthermore, various types of fake news, topics, and media platforms have distinctive linguistic properties.

The second driving force behind this study is the user engagements that reflect responses from news engaged users on social media. According to specialists, fake news frequently contains biased and aggressive language that is designed to construct clickbaits or cause confusions [10, 11]. The New York Times, for example, reported individuals benefiting from the publication of online fake news; the more provocative, a higher response will result in bigger financial benefit [12].
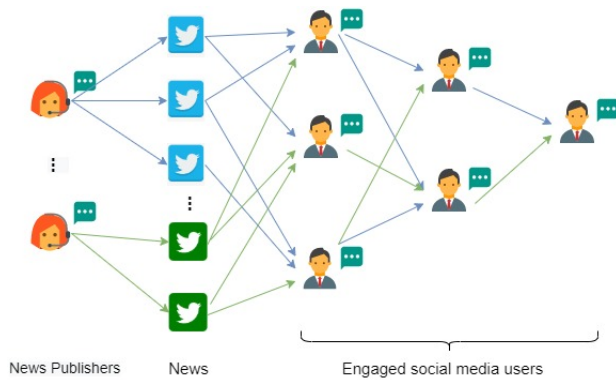
---

[1] https://www.journalism.org/2021/01/12/news-use-across-social-media-platforms-in-2020/

HℓCSS

Figure 1. A typical news propagation pattern on social media

The third aspect motivating our research lies in user characteristics. Spreaders of fake news can post misleading comments as fake news propagate. In comparison to user comments, user characteristics require strenuous effort to manipulate. Efforts in fake news detection by utilizing a series of user characteristics have been investigated in a handful of studies [13, 14, 15]. One notable weakness of those techniques is the lack of consideration of the most significant types of characteristic to detect fake news and whether or not one or more features are unavailable or insufficient in the early period of news dissemination impact the efficacy of these techniques.

The final intriguing aspect is the news dissemination path. A recent study suggests that fake news propagates differently from real news even at the early stages of spreading [16]. Fake news propagates significantly further, faster, deeper, and broader than real news in various categories of information [17]. Thus, news dissemination patterns are valuable features in discerning fake news from legitimate ones. Recent research has studied characteristics taken from propagation paths or networks utilizing temporal-structure to detect false news (see, for example, [18, 19, 20, 21]).

Our work focuses on improving fake news detection systems for social media. Our novel technique is centered around modeling the four social-context perspectives of fake news, namely, fake news content, user engagements, user characteristics, and news propagation path. We frame two collaborative modules, the engagement module and the propagation path module, to accommodate the four social-context characteristics. The two modules are combined seamlessly as the Fake News Engagement and Propagation Path (*FNEPP*) framework. The engagement module is designed as a Recurrent Neural

Network (RNN) [22], which takes in representations of news content and user engagement information. We construct the propagation path module as two cooperative Neural Networks (RNN and CNN [23]), which receive user characteristics features and propagation path information. The two modules are integrated to optimize the fake news detection task jointly.

Experiments on two real-world datasets reveal that the *FNEPP* framework outperforms the existing models in terms of accuracy and other evaluation metrics thanks to concurrently modeling the four social-context perspectives of fake news. The results of early fake news detection also demonstrate that the proposed *FNEPP* framework has the advantage of accurately detecting fake news in the early stage of its dissemination.

In a nutshell, we offer the following three major contributions in this study

(1) We present a principled way for concurrently modeling the four perspectives of fake news posted on social media.

(2) We offer a unique framework *FNEPP* that seamlessly integrates the four characteristics of fake news ecosystems within two collaborative modules.

(3) The experiments driven by two real-world datasets confirm *FNEPP*'s effectiveness and efficiency while retaining the benefit of early detection of fake news.

2. Related Work

The definition of "fake news" has been well studied, but there is no universal definition of "fake news". The "fake news" concept is highly related to the "false news" [17], "satire news" [24], rumor [25], disinformation [26], misinformation [27]. We refer to the definition of "fake news" from Zhou *et al* [28]. That is, fake news is intentionally false news published by a news outlet. Most existing techniques on fake news detection are categorized into two types, namely, content-based approaches (see Section 2.1) and social context-based approaches (see Section 2.2). A content-based approach aims to classify news based on the content of information to be verified, whereas a social context-based scheme utilizes rich secondary information user responses, user characteristics, and the pattern of news propagation through social media to identify fake news.

2.1. Content-based Fake News Detection

By developing a collection of linguistic cues that are instructive of content's truthfulness, cue and feature-based approaches are deployed to discern fake news from real news. Driscoll [29] applied the scientific content analysis (SCAN) scheme incorporating cues related to deception detection. Zhou *et al.* [30] constructed a cue set with 14 linguistic-based cues, which are effective for deception detection. Later research has examined more refined hand-crafted cue sets that are more specifically focused on the challenge of detecting fake news. Rubin *et al.* [31] evaluated a variety of textual features, including the frequency of punctuation marks and text sentiment. Zhao *et al.* [32] offered a variety of regular expressions to capture patterns of inquiry and correction in social media posts. The lack of generalization and task-specific traits, however, limit the usage of cue and feature-based approaches.

Unlike the aforementioned cue and feature-based approaches, linguistic analysis-based methods require no task-specific, hand-engineered cue sets. The most powerful technique of linguistic analysis for detecting fake news is based on n-grams [33, 34, 35]. Part-of-Speech (POS) tags were employed to extract the linguistic characteristics of fake news text [35]. Feng *et al.* [36] examined the use of Probabilistic Context-Free Grammars (PCFG) to encode deeper syntactic features for deception detection. Zhang *et al.* [37] proposed an analytic approach to identify fake news by two phases, that is, fake topics detection and fake events detection.

Deep learning approaches have made substantial progress in text mining and comprehension, especially the ability to learn effective representations. Not surprisingly, existing deep learning approaches devised for fake news detection usually include convolutional neural networks (CNN) and recurrent neural networks (RNN) [38, 39, 40, 41].

2.2. Social Contextual-based Fake News Detection

Early attempts in fake news detection involve hand-engineered features, including propagation pattern features, temporal pattern features, and text-based and user-related features. For example, Castillo *et al.* [13] constructed a feature set to embrace user-based features, text-based features, propagation-based features, and applied a decision tree model to classify fake news. Variants of the above features are comprised of other network-based features that are somewhat extended or tailored to an appropriate context, such as geographic locations [14] or temporal features [21]. Those

approaches usually lack generality while demanding tedious human efforts.

Research has been conducted in utilizing news propagation patterns and structures for fake news detection. Ma *et al.* [20] compared the similarity between propagation trees using tree kernels to detect fake news. A similar strategy was adopted in [19] with random walk graph kernel over propagation trees. Jin *et al.* [18] established a mathematical model named *SEIZ* to model a way of sharing news on social media among people.

Discrepancies in the temporal dynamics of user engagements for news articles are beneficial for detecting false news. Previous work leverages recurrent neural networks to capture temporal patterns [42, 6]. For instance, Ruchansky *et al.* [42] partitioned a sequence of engagements into discrete time intervals with a desired level of abstraction. In another study, Ma *et al.* [6] proposed to sample engagements at regular intervals from the time series to capture temporal differences.

User text responses and user analysis have been explored in the realm of fake news detection. User responses can be highly revealing in terms of discovering fake news. The textual response feature is represented using TF-IDF features as well as doc2vec word embeddings in [42]. Chen *et al.* [43] focused on the collected textual information by LSTM architecture coupled with an attention mechanism. Such a strategy allows for collecting typical fake news words and phrases and the visualization of which part of the text is indicative of truth or deception. When it comes to spreading fake news, news consumers might act as sources or proponents of misinformation. As such, prior studies [44, 45, 46, 47] incorporated user features to enhance the overall performance of fake-news detection systems.

3. Problem Formulation

This section introduces the set of notations and formalize the fake news detection task.

We assume that a series of fake news interactions occur across a time interval $[0, T]$. Our goal is to detect fake news early after it starts to spread on social media. Therefore, we should promptly detect fake news within a short time period (T is a small value). In what follows, our detection model consists of four vital sets, namely, article set A, user set U, engagement set E, and propagation path set P. In what follows, we formally define these four sets based on descriptions provided by [46]. These are subsequently utilized for defining the problem statement of this study.

(1) $A = \{a_1, a_2, ..., a_i, ..., a_{|A|}\}$ is a set of news

articles to be classified as fake or legitimate news.

(2) $U = \{u_1, u_2, ..., u_j, ..., u_{|U|}\}$ is a set of social media users, where each user u_j engaged in spread a news article in set A.

(3) $E = \{e_1, e_2, ..., e_k, ..., e_{|E|}\}$ is a set of engagements. Each e_k is essentially represented as a 3-tuple, (a_i, u_j, t), where user u_j retweets or comments about the news article a_i at time t.

(4) $P = \{p_{a_1}, p_{a_2}, ..., p_{a_i}, ..., p_{a_{|A|}}\}$ is a set of news propagation path. Each propagation path p_{a_i} is associated with news article a_i. Propagation path is naturally denoted as a multivariate time series $p_{a_i} = \{..., (\mathbf{x_{u_j}}, t_{u_j}), ...\}$, where $\mathbf{x_{u_j}}$ is a vector representation of user u_j who engages with news article a_i and $t_{u_j} \in [0, T]$.

With the above notation in place, we formally formalize the problem of detecting fake news from social context in *Definition 1*.

Definition 1. *Social context-based fake news detection.* Given a set of news articles A, a set of social media users U, a set of engagements E, and a set of news propagation path P, social context-based fake news detection is defined as a binary classification problem to predict a label $\hat{y_{a_i}} \in \{0, 1\}$ for news article a_i, where $\hat{y_{a_i}} = 1$ indicates a_i is fake, while $\hat{y_{a_i}} = 0$ indicates a_i is legitimate.

The "fake news early detection" is referred to as the fact that the news is published on a news outlet but has not propagated deeply on social media and is still worthy of intervention. Recall that our goal is to pinpoint fake news in an early stage of news dissemination. In the above formal problem statement, the performance of a fake news detection system is closely related to parameter T. As such, we undertake an empirical study to delve in the correlation between parameter T and detection performance. Please refer to Section 5.3 for the results with respect to early fake news detection.

4. Fake News Engagement and Propagation Path (*FNEPP*) Framework

In this section, we describe the details of our proposed framework, *FNEPP*. *FNEPP* mainly consists of two modules, namely engagement module and propagation path module, that collectively capture news contents, user engagements, user characteristics, and news propagation paths. The engagement module, which is inspired by the study from Ruchansky *et al.* [42], is dedicated to capturing the most efficient representations of user engagements and news

articles. The propagation path module is responsible for capturing the news propagation path along with user characteristics. We draw the idea from Liu *et al.* [44] for constructing the propagation path module. The details of the proposed framework are shown in Figure 2. The engagement module extracts a temporal representation of news articles using a Recurrent Neural Network (more accurately, an Long Short-Term Memory (LSTM) model). The user engagements are represented as vectors and fed into the LSTM to produce a final representation vector $\mathbf{e_{a_i}}$ for the engagement module. The propagation path module utilizes the vector representations of user characteristics to construct propagation paths as multivariate time series. Recurrent Neural Networks (more precisely, GRUs) and Convolutional Neural Networks(CNN) [23] extract the global and local propagation patterns, respectively.

4.1. Engagement Module

The objective of the engagement module is to capture the pattern of users' temporal interactions with a news article a_i in accordance with the occurrence and distribution. Explicitly, the module can capture both the number of engaged users of a_i and the pattern of those interactions over time. Additionally, the textual data associated with the interactions, such as the content of users' retweets, are combined in the module.

We employ a Recurrent Neural Network (RNN) as the basis for the engagement module since RNNs can effectively combine diverse information sources and catch temporal patterns within data. Notably, we choose the LSTM model because of its capability of processing variable-length input and its tendency for capturing long-term dependencies [48]. The critical component of this module is the selection of engagement feature vector $\mathbf{x_{e_k}}$ shown in Figure 2, which serves as the input to the LSTM cell.

The engagement feature vector $\mathbf{x_{e_k}}$ essentially consists of four parts and can be represented as the following vector:

$$\mathbf{x_{e_k}} = (\mathbf{x_u}, \mathbf{x_a}, \Delta t, n) \tag{1}$$

The first part $\mathbf{x_u}$ aims to model the engaged users. We create a binary incidence matrix representing the news articles that a particular user has interacted with. The binary index matrix is high-dimensional and sparse because the number of social media users is much larger than the number of news spreading over social media. Therefore, we employ the Singular value decomposition (SVD) for a binary incidence matrix to obtain lower-dimensional representation for engaged
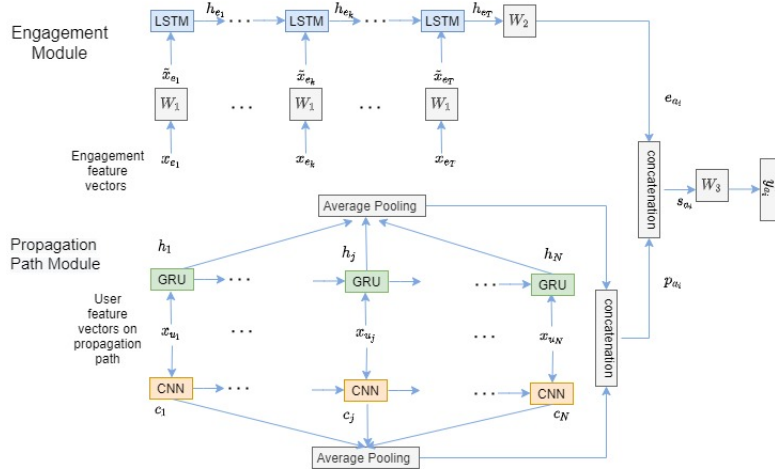
Figure 2. The architecture of FNEPP

users. The second part $\mathbf{x_a}$ is capable of capturing the text of each engagement. In order to prevent hand-crafted textual features, we apply the *doc2vec* [49] embeddings on the text of each engagement. Since we want to capture the occurrence and distribution of engagements over time, we introduce two variables (1) the number of engagements n and (2) the time interval between two consecutive engagements Δt. Notably, the four parts forming the engagement vector $\mathbf{x_{e_k}}$ are from various sources of information, which make the LSTM extremely suitable to serve as the model basis for the engagement module.

As shown in Figure 2, we add an embedding layer right after the raw input vector $\mathbf{x_{e_k}}$. Because the input features are constructed from different sources, it is not an advisable practice to feed the input vector $\mathbf{x_{e_k}}$ directly into the LSTM unit. The embedding layer is a fully connected layer, which transforms the raw input vector $\mathbf{x_{e_k}}$ to $\tilde{\mathbf{x}}_{\mathbf{e_k}}$ by the following formula:

$$\tilde{\mathbf{x}}_{\mathbf{e_k}} = tanh(W_1 \mathbf{x_{e_k}} + \mathbf{b_1}) \qquad (2)$$

where W_1 is the fixed weight matrix and $\mathbf{b_1}$ is the fixed bias vector for all $\mathbf{x_{e_k}}$. The transformed vector $\tilde{\mathbf{x}}_{\mathbf{e_k}}$ is supplied into the LSTM as the input. The last hidden state vector $\mathbf{h_{e_T}}$ is fed into a fully connected layer to obtain the final vector representation $\mathbf{e_{a_i}}$ for news article a_i in the engagement module.

$$\mathbf{e_{a_i}} = tanh(W_2 \mathbf{h_{e_T}} + \mathbf{b_2}) \qquad (3)$$

4.2. Propagation Path Module

The primary task of propagation path module is to assess each user through their profiles and other available information on social media and learn representations to discern bogus propagation patterns from real ones. Our work focuses on the textual fake news, and the propagation of the fake news are not restricted to the textual content. We employ the RNN-based sub-module and CNN-based sub-module to construct the propagation path module.

The propagation path for a particular news article a_i is naturally represented as a multivariate time series as follows:

$$p_{a_i} = \{..., (\mathbf{x_{u_j}}, t_{u_j}), ...\} \qquad (4)$$

where $\mathbf{x_{u_j}}$ is a vector representation of user u_j who engaged with news article a_i and $t_{u_j} \in [0, T]$. $\mathbf{x_{u_j}}$ is constructed by extracting the user characteristics from their social media profile and relevant information. Further technical details on constructing $\mathbf{x_{u_j}}$ will be demonstrated in the Section 5.2.

We consider the interactions that happened within a time interval $[0, T]$ after the news article created on social media. The number of interactions on the propagation path may vary for different news articles. Therefore, to unify a fixed-length propagation path, we propose the following transformation. We assume that the length of the transformed propagation path is N.

- Case 1: If the length of p_{a_j} is not smaller than N, we keep the first N tuples of p_{a_j} as final propagation path sequence $\tilde{p_{a_j}}$.

- Case 2: If p_{a_j} contains less than N tuples, we randomly sample $(N - |p_{a_j}|)$ times and concatenate the sampled tuples to achieve the final propagation path sequence $\tilde{p_{a_j}}$ with length N.

The fixed-length propagation path is represented as follows:

$$\tilde{p_{a_j}} = \{(\mathbf{x_{u_1}}, t_1), ..., (\mathbf{x_{u_j}}, t_j), ..., (\mathbf{x_{u_N}}, t_N)\} \quad (5)$$

For our propagation path module, we only consider the relative time order of each user. Thus, we sort the user feature vectors in $\tilde{p_{a_j}}$ based on an ascending time order and omit the time in the tuple afterward. We rewrite $\tilde{p_{a_j}}$ as follows:

$$\tilde{p_{a_j}} = \{\mathbf{x_1}, ..., \mathbf{x_n}, ..., \mathbf{x_N}\} \quad (6)$$

where $\tilde{p_{a_j}}$ is an ordered sequence according to the engagement time with news article a_j.

Local Propagation Path Representation: Convolutional Neural Networks (CNN) are particularly suited for capturing local variations and representations. We propose to use 1D CNN to learn a vector representation for each propagation path $\tilde{p_{a_j}}$. We assume the user feature vector $\mathbf{x_n} \in \mathbb{R}^l$. We stack the user feature vectors into a user feature matrix $X \in \mathbb{R}^{N \times l}$. 1D CNN is applied on h successive users with a filter $W_c \in \mathbb{R}^{h \times l}$. Each convolution operation produces a scalar feature c_n as follows:

$$c_n = ReLU(W_c X_{n:n+h-1} + b_c) \quad (7)$$

where $X_{n:n-h+1}$ is the subset of h consecutive user feature vectors. $b_c \in \mathbb{R}$ is a scalar bias term. We repeat the above convolution operations with m filters and obtain a feature vector $\mathbf{c_n} \in \mathbb{R}^m$. We apply the convolution procedure for all subsets of consecutive h user feature vectors and obtain a sequence of feature $\{\mathbf{c_1}, ... \mathbf{c_{N-h+1}}\}$. The average pooling is utilized to produce the final vector representation $\mathbf{p_C}$ for local propagation path representation.

$$\mathbf{p_C} = \frac{1}{N} \sum_{n=1}^{N-h+1} \mathbf{c_n} \quad (8)$$

Global Propagation Path Representation: In order to capture the global patterns of propagation path, we propose to utilize Gated Recurrent Unit (GRU) to learn vector representations for transformed propagation paths. A GRU unit takes a user feature vector $\mathbf{x_n}$ and produces the hidden state representation $\mathbf{h_n}$ based on the following formulations, which is adopted from [50].

$$\begin{aligned} \mathbf{z_n} &= \sigma(U_z \mathbf{x_n} + W_z \mathbf{h_{n-1}}) \\ \mathbf{r_n} &= \sigma(U_r \mathbf{x_n} + W_r \mathbf{h_{n-1}}) \\ \tilde{\mathbf{h}}_n &= tanh(U_h \mathbf{x_n} + \mathbf{h_{n-1}} \odot W_h \mathbf{r_n}) \\ \mathbf{h_n} &= (1 - \mathbf{z_n}) \odot \mathbf{h_{n-1}} + \mathbf{z_n} \odot \tilde{\mathbf{h}}_n \end{aligned} \quad (9)$$

where $U_z, U_r, U_h, W_z, W_r, W_h$ are weight matrices, and $\odot$ denotes the element-wise vector multiplication. The detailed description of GRU model can be found in [50]. We apply average pooling over all hidden states produced by GRU and obtain our vector representation $\mathbf{p_R}$ for global propagation paths as follows:

$$\mathbf{p_R} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{h_n} \quad (10)$$

We concatenate the $\mathbf{p_C}, \mathbf{p_R}$ and obtain $\mathbf{p_{a_i}}$ as our final vector representation of propagation path module.

4.3. Integration

As described earlier, the engagement module combines the news article and user engagements to capture the engagement patterns while the propagation path module incorporates the user characteristics from social media and propagation path to discern the fake news dissemination patterns from real ones. On behalf of accommodating representations from two modules, we concatenate $\mathbf{e_{a_i}}, \mathbf{p_{a_i}}$ together as the vector $\mathbf{s_{a_i}}$. $\mathbf{s_{a_i}}$ is utilized as input to a fully connected layer to predict whether the news article a_i is fake or not.

$$\hat{y_{a_i}} = \sigma(W_3^T \mathbf{s_{a_i}} + \mathbf{b_3}) \quad (11)$$

where W_3 is the weight matrix shown in Figure 2, $\mathbf{b_3}$ is the bias term. We apply the cross-entropy loss function for training our model.

The advantage of this integration is that it unifies the two modules to form a more accurate prediction. In addition, the model learns distinctive patterns between fake news and real news by jointly training the engagement module and propagation path module simultaneously.

5. Experiments

In this section, we systematically evaluate our proposed fake news detection framework using two real-world datasets. We compare our solution with the existing baseline schemes coupled with the state-of-the-art models in various evaluation metrics.

5.1. Data

To make a fair comparison, we conduct the experiments on two real-world social media datasets that were also used in previous research, PHEME [51], and WEIBO [6]. The PHEME dataset and WEIBO dataset contain breaking news and each news associates with a set of user engagements. The PHEME dataset consists

	PHEME	WEIBO
# Users	37,175	2,746,818
# News Stories	5,802	4,664
# Real News	3,830	2,351
# Fake News	1,972	2,313

Table 1. Statistics of the datasets

Features	Type
length of user name	integer
length of user description	integer
follower counts	integer
friends counts	integer
favorites counts	integer
status counts	integer
user verified status	boolean
geo enabled status	boolean

Table 2. User characteristics for constructing x_{u_j}

of 5,802 newsworthy stories, which are annotated by the journalists and crowdsourcing. The WEIBO dataset contains 4,664 fact-checked news, and the dataset is annotated by the Sina community management center. The profiles of the engaged users are also available in each dataset, which provides the convenience of constructing propagation paths for modeling purpose. Both datasets are randomly split into the training set (80%), testing set (10%), and validation set (10%). A summary of key statistics is described in Table 1.

5.2. Experimental Setup

Before explaining the major findings, we articulate the specific features within each dataset. Next, we introduce the hyperparameters for training our model. The alternative models, serving as competitors to our model, are briefly outlined at the end of this subsection.

Features: The engagement module essentially judiciously extracts an engagement vector for each news article $x_{e_k} = (x_u, x_a, \Delta t, n)$. Feature vector x_u is constructed by the SVD decomposition with a rank 10 for the PHEME and WEIBO datasets. In order to apply *doc2vec* to obtain textual feature x_a, we perform text segmentation on the WEIBO dataset. The embedding dimension is set to 100 for both datasets, which result in x_a with 100 dimensions. The dimension of the engagement vector x_{e_k} is 112. For the propagation path module, we construct user feature vector x_{u_j} for each engaged user with the following features listed in Table 2. It is note worthy that the user feature vectors are derived from the common user characteristics available in the two tested datasets.

The choice of the characteristics summarized Table 2 unravels the legitimacy of social media users to some extent. For instance, social disrupters tend to post and spread fake news via zombie accounts on social media. Characteristics like follower counts, friend counts, and user verified status can help in discerning potential zombie accounts.

Hyperparameters: For the propagation path module, the GRU units' output dimension is 32. The size of the CNN filter is set to three, and we employ 32 CNN filters to extract local propagation path representations. For the engagement module, the hidden dimension of LSTM is set to 50. For the training purpose, we apply Adam optimizer with a learning rate 0.0001.

Comparison Models: We compare our proposed model against the following four alternative models found in the literature.

> DTR: A decision-tree-based ranking algorithm identifies fake news using query terms. [32].

> SVM-TS: A linear support vector machine classification model utilizes time series to simulate the temporal change of social context characteristics [52].

> GRU: A rumor detection model advocates RNNs and GRU for long-term representation learning of relevant posts [6].

> CSI: An effective recurrent encoder aggregates user features, news content, and user–news engagements [42].

5.3. Results and Analysis

Methods	Class	Accu.	Prec.	Rec.	F_1
DTR	R	0.562	0.549	0.704	0.617
	F		0.588	0.421	0.491
SVM-TS	R	0.651	0.642	0.686	0.663
	F		0.663	0.617	0.639
GRU	R	0.722	0.734	0.712	0.723
	F		0.722	0.733	0.728
CSI	R	0.742	0.743	0.728	0.736
	F		0.735	0.750	0.743
FNEPP	R	0.780	0.789	0.764	0.776
	F		0.771	0.794	0.783

Table 3. Comparison results from the PHEME dataset ("F: " fake news; "R": real news)

Overall Comparison. Tables 3 and 4 demonstrate the performance of all the compared models detecting

Methods	Class	Accu.	Prec.	Rec.	F_1
DTR	R	0.732	0.726	0.749	0.737
	F		0.738	0.715	0.726
SVM-TS	R	0.857	0.878	0.830	0.857
	F		0.839	0.885	0.861
GRU	R	0.892	0.922	0.864	0.893
	F		0.876	0.926	0.901
CSI	R	0.905	0.895	0.907	0.901
	F		0.915	0.909	0.912
FNEPP	R	0.919	0.907	0.915	0.911
	F		0.928	0.921	0.925

Table 4. Comparison results from the WEIBO dataset ("F: " fake news; "R": real news)

fake news from the PHEME and WEIBO datasets. The results indicate that our proposed *FNEPP* outperforms all the competitive models in almost every evaluation metric. For example, when it comes to the PHEME dataset, *FNEPP* boosts the accuracy of DTR, SVM-TS, GRU, and CSI by 22%, 13%, 6%, and 3.8%, respectively. Similarly, *FNEPP* has a clear edge over the four alternative methods in terms of fake news detection accuracy and precision on the WEIBO dataset.

Detailed Analysis. DTR and SVM-TS deliver poor performances on fake news detection because both methods are solely reliant on hand-crafted features. DTR's subpar performance is attributed by the insufficient coverage of patterns described by regular expressions. SVM-TS performs relatively better than DTR thanks to the incorporation of temporal information. The poor detection accuracy of DTR and SVM-TS indicates that hand-crafted features are inadequate for encoding semantic information of news content. Unlike our *FNEPP*, DTR and SVM-TS fail to capture complex feature interactions - key players in fake news detection. GRU integrates the temporal linguistic features, thereby being superior to DTR and SVM-TS. CSI detects fake news from a social context perspective that combines the user engagements and their responses in the model. The performance of GRU suggests that deep learning models can learn semantic representations while enhancing feature interactions. CSI performs slightly better than GRU in the two datasets, implying that the detection accuracy is improved with the provision of valuable social context. Compared to CSI, *FNEPP* incorporates user characteristics along with propagation path, which significantly improves the accuracy of fake news detection. In a nutshell, our results confirm that our *FNEPP* boosts the overall performance of the fake news detection system by the virtue of modeling four perspectives of the social context.

Fake News Early Detection. The "fake news early detection" is referred to as the fact that the news is published on a news outlet but has not propagated deeply on social media and is still worthy of intervention. One of the crucial aspects of fake news detection is to identify fake news in an early dissemination stage on social media. Early alerts essentially prevent further spreading of fake news harassing social media users' opinions. In order to evaluate the performance of early fake news detection, we compare multiple methods by varying time interval $[0, T]$. Accuracy performance is attained by progressively adding data up to a checkpoint T while maintaining the desired time interval. Figure 3 plots the accuracy of all the competitors as a function of parameter T. All the methods enjoy accuracy improvement over time. More importantly, our model exhibits a clear advantage over the alternatives at an early stage. Our model swiftly learns to detect fake news, achieving good performances using less than 8-hour data. In particular, the results on the WEIBO dataset reveal that *FNEPP* outperforms GRU, SVM-TS, and DTR using the 4-hour data. Not surprisingly, similar trends are observed from the PHEME dataset. After 8 hours of spreading fake news, *FNEPP* dramatically surpasses all the compared models. These results confirm the advantage of our model over the existing solutions in terms of early fake news detection.

6. Conclusion and Future Directions

To model fake news detection from a social context perspective, we developed a novel fake news detection framework called *FNEPP* that seamlessly incorporates news contents, user engagements, user characteristics, and propagation paths using two cooperative modules. We conducted extensive experiments driven by two real-world datasets to shed light on the effectiveness of *FNEPP*. We demonstrated the capability of capturing distinctive temporal patterns between fake and real news. The promising results unfold the high efficiency of *FNEPP* in the realm of detecting fake news on social media at an early stage.

This study focuses on data sourced from social media platforms such as Twitter and WEIBO, which are predominantly unmoderated and free-flowing. Future studies can focus on moderated fake news data, such as in fake news sites that exist in edited or semi-edited forms. The size of the dataset utilized in the study is relatively small. Future studies will benefit from increased dataset size. Also, emotions, which could potentially play a key role in fake news detection, have not been studied and are one of the modeling
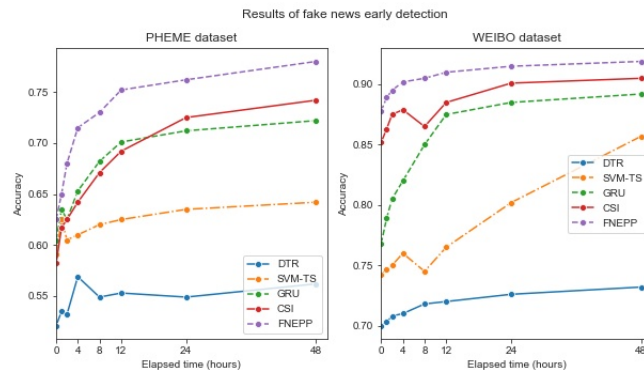
Results of fake news early detection

Figure 3. Results of fake news early detection

limitations. Future studies could investigate the significance of emotional signals in fake news detection. This study predominantly focuses on misinformation detection. However, broader management and governance approaches need to be taken for better control and mitigation of misinformation. Future studies can look at this problem from a disaster control perspective.

References

[1] B. Nyhan and J. Reifler, "When corrections fail: The persistence of political misperceptions," *Political Behavior*, vol. 32, no. 2, pp. 303–330, 2010.

[2] C. Paul and M. Matthews, "The russian "firehose of falsehood" propaganda model," *Rand Corporation*, pp. 2–7, 2016.

[3] Y. Ren, B. Wang, J. Zhang, and Y. Chang, "Adversarial active learning based heterogeneous graph neural network for fake news detection," in *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 452–461, IEEE, 2020.

[4] W. Ferreira and A. Vlachos, "Emergent: a novel data-set for stance classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pp. 1163–1168, 2016.

[5] A. Gupta, P. Kumaraguru, C. Castillo, and P. Meier, "Tweetcred: Real-time credibility assessment of content on twitter," in *International Conference on Social Informatics*, pp. 228–243, Springer, 2014.

[6] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," 2016.

[7] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in *Proceedings of the 5th international workshop on adversarial information retrieval on the web*, pp. 41–48, 2009.

[8] D. M. Markowitz and J. T. Hancock, "Linguistic traces of a scientific fraud: The case of diederik stapel," *PloS one*, vol. 9, no. 8, p. e105937, 2014.

[9] V. L. Rubin, Y. Chen, and N. K. Conroy, "Deception detection for news: three types of fakes," *Proceedings of the Association for Information Science and Technology*, vol. 52, no. 1, pp. 1–4, 2015.

[10] Y. Chen, N. J. Conroy, and V. L. Rubin, "Misleading online content: recognizing clickbait as" false news"," in *Proceedings of the 2015 ACM on workshop on multimodal deception detection*, pp. 15–19, 2015.

[11] V. L. Rubin, "Deception detection and rumor debunking for social media," *The SAGE Handbook of Social Media Research Methods*, pp. 342–364, 2017.

[12] S. Maheshwari, "How fake news goes viral: A case study," *The New York Times*, vol. 20, 2016.

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

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

[15] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pp. 430–435, IEEE, 2018.

[16] Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li, J. Wu, and S. Havlin, "Fake news propagates differently from real news even at early stages of spreading," *EPJ Data Science*, vol. 9, no. 1, p. 7, 2020.

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

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

[19] K. Wu, S. Yang, and K. Q. Zhu, "False rumors detection on sina weibo by propagation structures," in *2015 IEEE 31st international conference on data engineering*, pp. 651–662, IEEE, 2015.

[20] J. Ma, W. Gao, and K.-F. Wong, "Detect rumors in microblog posts using propagation structure via kernel learning," Association for Computational Linguistics, 2017.

[21] S. Kwon, M. Cha, and K. Jung, "Rumor detection over varying time windows," *PloS one*, vol. 12, no. 1, p. e0168344, 2017.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," tech. rep., California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[24] E. C. Tandoc Jr, Z. W. Lim, and R. Ling, "Defining "fake news" a typology of scholarly definitions," *Digital journalism*, vol. 6, no. 2, pp. 137–153, 2018.

[25] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, pp. 1–36, 2018.

[26] N. Kshetri and J. Voas, "The economics of "fake news"," *IT Professional*, vol. 19, no. 6, pp. 8–12, 2017.

[27] J. Corner, "Fake news, post-truth and media–political change," 2017.

[28] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.

[29] L. N. Driscoll, "A validity assessment of written statements from suspects in criminal investigations using the scan technique," *Police Stud.: Int'l Rev. Police Dev.*, vol. 17, p. 77, 1994.

[30] L. Zhou, J. K. Burgoon, J. F. Nunamaker, and D. Twitchell, "Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communications," *Group decision and negotiation*, vol. 13, no. 1, pp. 81–106, 2004.

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

[32] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proceedings of the 24th international conference on world wide web*, pp. 1395–1405, 2015.

[33] R. Mihalcea and C. Strapparava, "The lie detector: Explorations in the automatic recognition of deceptive language," in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 309–312, 2009.

[34] M. Ott, C. Cardie, and J. T. Hancock, "Negative deceptive opinion spam," in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*, pp. 497–501, 2013.

[35] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," *arXiv preprint arXiv:1107.4557*, 2011.

[36] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 171–175, 2012.

[37] C. Zhang, A. Gupta, C. Kauten, A. V. Deokar, and X. Qin, "Detecting fake news for reducing misinformation risks using analytics approaches," *European Journal of Operational Research*, vol. 279, no. 3, pp. 1036–1052, 2019.

[38] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648*, 2017.

[39] F. Qian, C. Gong, K. Sharma, and Y. Liu, "Neural user response generator: Fake news detection with collective user intelligence.," in *IJCAI*, vol. 18, pp. 3834–3840, 2018.

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

[41] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 647–653, 2017.

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

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

[44] Y. Liu and Y.-F. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, 2018.

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

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

[47] Y. Dou, K. Shu, C. Xia, P. S. Yu, and L. Sun, "User preference-aware fake news detection," *arXiv preprint arXiv:2104.12259*, 2021.

[48] M. Hüsken and P. Stagge, "Recurrent neural networks for time series classification," *Neurocomputing*, vol. 50, pp. 223–235, 2003.

[49] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *International conference on machine learning*, pp. 1188–1196, PMLR, 2014.

[50] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[51] A. Zubiaga, M. Liakata, and R. Procter, "Exploiting context for rumour detection in social media," in *International Conference on Social Informatics*, pp. 109–123, Springer, 2017.

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