2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

News Labeling as Early as Possible: Real or Fake?

Maryam Ramezani*†, Mina Rafiei*‡, Soroush Omranpour§ and Hamid R. Rabiee¶
Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
Email: †maryam.ramezani@sharif.edu, ‡mrafiei@ce.sharif.edu, §omranpour@ce.sharif.edu, ¶rabiee@sharif.edu
*Both authors contributed equally to this research.

***Abstract*—Differentiating between real and fake news propagation through online social networks is an important issue in many applications. The time gap between the news release time and detection of its label is a significant step towards broadcasting the real information and avoiding the fake. Therefore, one of the challenging tasks in this area is to identify fake and real news in early stages of propagation. However, there is a trade-off between minimizing the time gap and maximizing accuracy. Despite recent efforts in detection of fake news, there has been no significant work that explicitly incorporates early detection in its model. The proposed method utilizes recurrent neural networks with a novel loss function, and a new stopping rule. Experiments on real datasets demonstrate the effectiveness of our model both in terms of early labelling and accuracy, compared to the state of the art baseline and models.**

***Index Terms*—Early news labeling, Online Social Networks, Recurrent Neural Networks, Fake News**

I. Introduction

With increased usage of online social networks, social interaction of users are increasingly growing. Users of social networks share ideas, thoughts, feelings, knowledge, news and advertisement via actions such as posting, commenting, liking and reposting. In recent years, various works have been done in detecting the truth of information propagated in social networks. One of the main future directions in fake news research is early fake news detection [1]. Despite various fake news detection methods, there has been no significant work focusing on earliness, and all of prior works have confined earliness just in evaluating the performance of their algorithms. In this paper, we focus on the early labeling of news. We represent a news dissemination by a time series sequence. At each time step, we estimate the probability of labeling the stream. Clearly, with more incoming data, more accurate detection can be achieved. Our aim is to label the stream with high probability of being fake or real, as early as possible.

II. News Early Classification

A *news event* is a widespread news that can be tracked by specific topic and keywords on social networks. Let $E = \{e_1, e_2, e_3, ..., e_N\}$ be the set of different news events disseminated daily among users. Each news event e_i is provided by the set of messages m_is that social users post and share over the time. In the same way, $e_i(t) = \{m_{i1}, m_{i2}, ..., m_{it}\}$ is the streaming set of a news after t steps of spreading. Let τ be the maximum length of the streaming of a news, then e_i is defined as $e_i(\tau)$. Then, the news labeling problem on each e_i is learning a binary classification $f(e_i) = \ell$ such that $\ell \in \{0, 1\}$ where 0 means real label and 1 means fake. Here, we aim to learn an early classifier function f_Θ that minimizes a supervised loss. In the real time classification step for a message stream $e_i(t)$, the learned model will assign a label ℓ with probability P till the time step t, or it waits for more incoming messages. To this end, we define the following function and learn the parameters Θ. Let $S = \{D, W\}$ be the corresponding *Detection* and *Waiting* states, then $f_\Theta(e_i(t)) = (S, \ell, P)$. It is clear that when $S = W$, the ℓ and P are null.

Model Description: The overall architecture of NEC (News Early Classification) is illustrated in Figure 1.

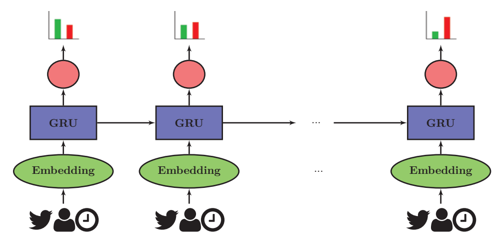


Fig. 1: Model Architecture of NEC.

Formally, X_t is given to GRU (Gated Recurrent Unit), an improved version of RNN, to obtain the hidden state h_t as output. Then h_t will fed into a linear layer for reduction to two dimensions ($\hat{y}_t$), and generating the probability of binary classification by using a soft-max activation function. Here, X_t contains the information of a message m_{it} in stream time series of event e_i, represented by $X_t = (C_t, U_t, DT_t)$, where C, U and DT are input features representing message context, user information and message post time, respectively:

Message context (C): NEC utilized a Doc2Vec embedding [2] on fake and real news to construct feature vectors C_t^F and C_t^R for textual analysis and then $C_t = concatenate(C_t^R, C_t^F)$.

User information (U): User social profile is the simplest and most accessible information that we can have about the users in a social network. A set of eight available features of a user

ASONAM '19, August 27-30, 2019, Vancouver, BC, Canada

ACM ISBN 978-1-4503-6868-1/19/08
http://dx.doi.org/10.1145/3341161.3342957

profile are utilized same as [3].

Message post time (DT): At time step t, if the post time of message is T_t, $DT_t = T_t - T_0$ can be considered as a sign of news diffusion speed.

Training Step: For a focussed learning on earliness, we modify the cross entropy by weighting the loss at each time step $1 \preceq t \preceq \tau$ with $\gamma(t)$ as Equation 1, where y is the ground truth label of the stream sequence of a news and $P(\hat{y}_t|X_{1:t}, \Theta)$ is the probability of our model detected label for time step t. The coefficient $\gamma(t)$ is used as a trade off between two inconsistent objectives including accuracy and earliness. Very high accuracy is not expected at the early steps of diffusion, because of limited amount of information available from the first stage.

$$\mathbb{L}(\Theta|X, Y) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{\tau} -\gamma(t)[y^{(i)}logP(\hat{y_t}^{(i)}|X_{1:t}^{(i)}, \Theta) \\ +(1-y^{(i)})log(1-P(\hat{y_t}^{(i)}|X_{1:t}^{(i)}, \Theta))] \tag{1}$$

However, at later points in the time series sequence, classification should be more accurate. So we define $\gamma(t) = -log(\frac{t}{\tau})$.

Classification Step: The early accurate detection point in different news streams may vary. Therefore, we utilize a stopping rule to evaluate if an accurate detection point is reached. To this end, at time step t with probability of predicted label $\hat{y}$, the expression is defined as:

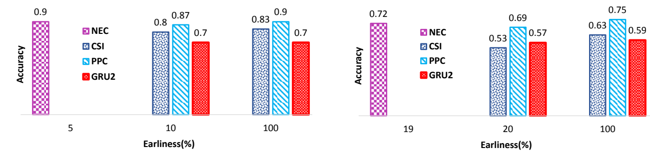
$$[P(\hat{y}_t) \succ P(\hat{y}_{t-1})] \wedge [P(\hat{y}_t) \succ \alpha] \wedge [P(\hat{y}_t) - P(\hat{y}_{t-1}) \prec \gamma(t)] \tag{2}$$

The first clause indicates that whether the sequence is at consistent point where the detected label probability will not decrease by observing more data during next time steps. The second clause defines a minimum value α on probability for labeling the news. The third clause prevents the model from requesting more information in order to increase the classification accuracy when observed time series is becoming undesirably long which can negatively effect the earliness. The classification algorithm starts as the first message is posted. At each time step t, the output is checked in the stopping rule expression (Equation 2). If the expression evaluates as true, the state is set to D and loop of detection terminates with an output label and if the expression evaluates to false, the state is set to W, and loop will be continued for new incoming sequence containing the next message.

III. Experimental Evaluation

We utilized two real-world datasets [4]. Sina Weibo with 2746818 users, 2351 real and 2313 fake events. Twitter with 381541 users, 493 real and 498 fake events. The correctness of event labeling is evaluated with the metrics such as accuracy, precision, recall and F-measure. Also, the model ability to label the events, as early as possible, is evaluated through the $Earliness = \frac{1}{N}\sum_{i=1}^{N}\frac{t_i}{\tau_i}$ metric. where t_i denotes the stopping time step, and τ_i is the event maximum length.

We compared our model with the following models: GRU-2 [4] as a baseline model. CSI [5] as a state of the art model in accuracy by utilizing both the user and text sequences. PPC [3] as a state of the art model using user sequence.



(a) Sina Weibo (b) Twitter

Fig. 2: Accuracy comparison in earliness point and 100% of sequence

In Figure 2, the accuracy against earliness of NEC is compared to other models. As shown, NEC stops the labeling at about 5 time steps for Sina Weibo and 20 time steps for Twitter datasets, while other models use the whole message stream to achieve their labeling. Despite this significant advantage, NEC has a close competition with other models in term of accuracy. The correctness of labeling results is shown in Table I. For a fair comparison between models, we set the input length of baseline models same as our model stopping point. As illustrated, the NEC model outperforms previous models in terms of news labeling correctness.

TABLE I: NEC results compared to competitive models on Sina Weibo and Twitter Dataset

	Sina Weibo						Twitter					
	Fake			Real			Fake			Real		
model	R	P	F	R	P	F	R	P	F	R	P	F
NEC	**0.93**	**0.87**	**0.90**	**0.86**	0.92	**0.89**	0.93	0.64	**0.76**	0.53	**0.90**	0.67
PPC	1	0.78	0.88	0.77	**1**	0.87	0.71	**0.70**	0.71	**0.68**	0.69	**0.69**
GRU2	0.73	0.67	0.70	0.65	0.71	0.68	0.55	0.59	0.57	0.60	0.57	0.59
CSI	0.92	0.73	0.81	0.67	0.90	0.77	**0.98**	0.53	0.68	0.11	0.86	0.20

IV. Conclusions and Future Work

In this paper, we introduced a new real time early news labeling method called NEC. Experiments on real datasets demonstrates that NEC outperforms the competitive methods in term of accuracy while detecting in an earlier stage. As the future works, we may evaluate the performance of adding an attention mechanism to the current model. Considering multi-modal data for early fake news detection is an interesting direction for this work.

References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *SIGKDD Explor. Newsl.*, vol. 19, no. 1, sep 2017.

[2] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the International Conference on International Conference on Machine Learning - Volume 32*, ser. ICML'14. JMLR.org, 2014, pp. II–1188–II–1196.

[3] Y. Liu and Y.-f. B. Wu, "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," in *Proceedings of the Conference on Artificial Intelligence*, 2018, pp. 354–361.

[4] 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," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2016, pp. 3818–3824.

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