A Convolutional Approach for Misinformation Identification

Feng Yu[1,3], **Qiang Liu**[1,3], **Shu Wu**[1], **Liang Wang**[1,2,3], **Tieniu Tan**[1,2,3]
[1]Center for Research on Intelligent Perception and Computing
National Laboratory of Pattern Recognition
[2]Center for Excellence in Brain Science and Intelligence Technology
Institute of Automation, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences
{feng.yu, qiang.liu, shu.wu, wangliang, tnt}@nlpr.ia.ac.cn

Abstract

The fast expanding of social media fuels the spreading of misinformation which disrupts people's normal lives. It is urgent to achieve goals of misinformation identification and early detection in social media. In dynamic and complicated social media scenarios, some conventional methods mainly concentrate on feature engineering which fail to cover potential features in new scenarios and have difficulty in shaping elaborate high-level interactions among significant features. Moreover, a recent Recurrent Neural Network (RNN) based method suffers from deficiencies that it is not qualified for practical early detection of misinformation and poses a bias to the latest input. In this paper, we propose a novel method, Convolutional Approach for Misinformation Identification (CAMI) based on Convolutional Neural Network (CNN). CAMI can flexibly extract key features scattered among an input sequence and shape high-level interactions among significant features, which help effectively identify misinformation and achieve practical early detection. Experiment results on two large-scale datasets validate the effectiveness of CAMI model on both misinformation identification and early detection tasks.

1 Introduction

Nowadays, the increasingly easy access and extensive application of social media provide fertile breeding ground for misinformation dissemination, which will mislead public opinion, impact political election[1] and further pose huge threat to public security and social stability. Moreover, a feasible solution to prevent the spread of misinformation is early detection of misinformation and launch of directed and effective counter campaigns [Kumar and Geethakumari, 2014]. Therefore, it is more and more urgent to identify misinformation from a massive of social media information and detect misinformation as early as possible.

Conventional misinformation identification models leverage handcrafted features from user credibility and microblog post[2] content at post level [Castillo *et al.*, 2011; Qazvinian *et al.*, 2011; Gupta *et al.*, 2013], at event[3] level [Kwon *et al.*, 2013; Ma *et al.*, 2015; Zhao *et al.*, 2015] or aggregating from post level to event level [Jin *et al.*, 2014]. Some other works adopt more effective handcrafted features, such as conflict viewpoints [Jin *et al.*, 2016], temporal properties [Kwon *et al.*, 2013; Ma *et al.*, 2015], users' feedback [Giudice, 2010; Rieh *et al.*, 2014] and signals tweets containing skepticism [Zhao *et al.*, 2015]. However, handcrafted features may not cover potential features in dynamic and complicated social media scenarios. What's more, a rough mergence of different handcrafted features cannot shape high-level interactions among significant features. Last, these feature engineering methods is labor-intensive for so much designs.

In order to mine key features in dynamic and complicated social media sceneries, deep neural network is a good choice. Without any handcrafted features, a 2-layer Gated Recurrent Unit model (termed GRU-2) is adopted for misinformation identification [Ma *et al.*, 2016]. GRU-2 treats text content of microblog posts in an event as a variable-length time series, which can capture the dynamic temporal signals characteristic during the diffusion process. But GRU-2 has the following limitations. *First*, GRU-2 is not qualified for practical early detection tasks with limited input sequence of misinformation. The limited input sequence may not be long enough to embody the dynamic temporal sequential signals, so GRU-2 will not capture the dynamic temporal signals characteristic in some cases. *Second*, a trained RNN model possesses a constant recurrent transition matrix and induces unchangeable propagations of sequence signals between every two consecutive inputs, which is inadequate for dynamic and complicated scenarios. *Third*, the above GRU-2 model has a bias towards the latest elements of input sequence [Mikolov *et al.*, 2011]. But key features do not necessarily appear at the rear part of an input sequence.

As mentioned above, feature engineering based methods fail to shape elaborate high-level interactions among significant features to model real-world social media scenarios, while CNN can not only automatically extract local-global

[1]http://www.npr.org/2016/11/08/500686320/did-social-media-ruin-election-2016

[2]A post refers to a tweet or a posting on microblog websites.

[3]An event includes many microblog posts relevant to the event.

significant features from an input instance but reveal those high-level interactions. Moreover, RNN based methods are not qualified for the task of early detection. Besides these methods involve a bias towards the latest input elements and attempt to obtain unchangeably propagating sequential characteristics, while the convolutional architecture and k-max pooling operation in CNN can flexibly extract key features scattered among one input sequence.

On the other hand, CNN based approaches to speech recognition [Abdel-Hamid *et al.*, 2012], semantic analysis [Kalchbrenner *et al.*, 2014], click-through rate prediction [Liu *et al.*, 2015], semantic segmentation [Zhao *et al.*, 2017] and reinforcement learning tasks [Tamar *et al.*, 2016] have achieved much improvement in respective fields.

We propose a CAMI model for misinformation identification and early detection tasks. First, we investigated the data distribution in adopted datasets (detailed in Section 3) and observe the *long-tailed* distribution of misinformation and truth information. Then, we put forward a proper method to split every event into several phases based on the above observation. Subsequently, all events are split into several groups of microblog posts. And representation of each group is learnt though paragraph vector [Le and Mikolov, 2014]. So an input sequence of CAMI is composed by groups of an event. CAMI cannot only automatically extract local-global significant features from an input instance, reveal those high-level interactions but flexibly extract key features scattered among one input sequence. Finally, we obtain some observations from visualization experiments of the CAMI model, which contribute to better understand human behaviors in cyberspace and more exactly shape real-world social media scenarios.

The main contributions of this work are as follows:

- We use an unsupervised method, paragraph vector, to learn representation of input microblog posts and a supervised method, CNN, to automatically obtain key features of both misinformation and truth information.
- We visualize what the proposed model has captured, which will help us comprehend the inherent properties possessed by information on social media.
- Experiments conducted on two real-world datasets show that CAMI is much more effective and clearly outperforms the state-of-the-art methods in both misinformation identification and early detection tasks.

2 Related Work

In this section, we review some related works on misinformation identification, early detection and convolutional neural network.

2.1 Misinformation Identification and Early Detection

Recently, many methods have been put forward for misinformation automatic identification on wiki websites [Kumar *et al.*, 2016] and social media. Some works treat a microblog post [Castillo *et al.*, 2011; Qazvinian *et al.*, 2011] or an image [Gupta *et al.*, 2013] as object to be identified. Some identify whether an event belongs to misinformation or truth information and extract handcrafted features from the event level [Kwon *et al.*, 2013; Ma *et al.*, 2015; Zhao *et al.*, 2015]. Another work obtains credibility of a microblog post and then aggregate credibility to the event level [Jin *et al.*, 2014]. Moreover, some other works extract more effective handcrafted features, including conflict viewpoints [Jin *et al.*, 2016], temporal properties [Kwon *et al.*, 2013; Ma *et al.*, 2015], users' feedback [Giudice, 2010; Rieh *et al.*, 2014] and signals tweets containing skepticism [Zhao *et al.*, 2015]. All above feature engineering based methods fail to cover potential features in dynamic and complicated social media scenarios and shape elaborate high-level interactions among significant features. For the sake of overcome these deficiencies, a RNN based model attempts to capture the dynamic temporal signals in the misinformation diffusion process and incrementally learn both the temporal and textual representations of an event not relying on any handcrafted features [Ma *et al.*, 2016].

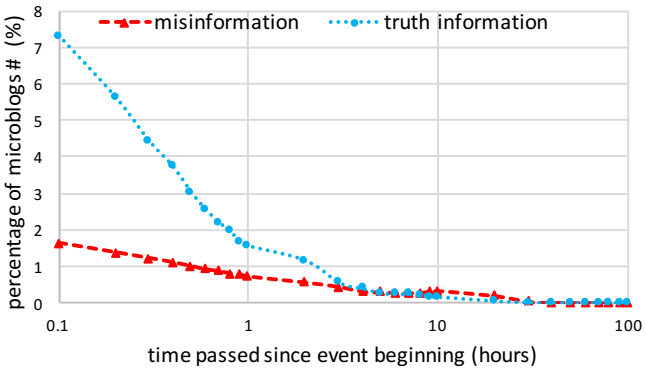


Figure 1: The **long-tailed** distribution of both misinformation and truth information in the Weibo dataset in a semi logarithmic coordinate.

2.2 Convolutional Neural Network

CNN is made up of stacked convolutional and pooling layers, the architectures of which help model significant semantic features and achieve much improvement in respective fields. For instance, CNN has been successfully applied in speech recognition [Abdel-Hamid *et al.*, 2012], sentence semantic analysis [Kalchbrenner *et al.*, 2014], click-through rate prediction [Liu *et al.*, 2015], image semantic segmentation [Zhao *et al.*, 2017] and reinforcement learning tasks [Tamar *et al.*, 2016]. CNN is usually trained through stochastic gradient descent (SGD), with backpropagation to compute gradients.

3 Dataset Analysis

We evaluate models on two large microblog datasets: Weibo and Twitter dataset, which is developed and used by [Castillo *et al.*, 2011; Kwon *et al.*, 2013; Ma *et al.*, 2016]. The numbers of events respectively belonging to misinformation and truth information are 498 and 494 in the Twitter dataset and 2,313 and 2,351 in the Weibo dataset.

We investigated the data distribution of misinformation and truth information in datasets. Take Weibo dataset as an example, the data distribution is illustrated in Figure 1. Each point represents percentage of microblog posts during a time

window of 0.1 hours at the corresponding time point. The *long-tailed* distribution of both misinformation and truth information can be shown even in the semi logarithmic coordinate (otherwise the curves almost coincide with the general coordinates).

4 Proposed CAMI Model

In this section, we introduce the proposed CAMI model. We first present the problem definition. Then we detail the proposed model.

4.1 Problem Definition

Given a set of events, each event comprises a sequence of correlative microblog posts and each microblog post is associated with a timestamp. The task here is to identify whether an event is misinformation or not at the event-level, namely, detect whether an event is misinformation or not by analyzing a sequence of correlative microblog posts of the event.

4.2 Proposed Model

As illustrated in Figure 2, we will introduce the framework of proposed CAMI model. From the bottom up, there are roughly three mini modules as follows.

Misinformation may also be described in a truth-telling way, so it is difficult to identify misinformation merely from one specific microblog post. Relatively it is reasonable to detect misinformation from a sequence of correlative microblog posts of an event. Inherent properties of misinformation and truth information play a pivotal role in misinformation identification. To model these properties for an event, we need to handle all microblog posts of the event as a whole.

Splitting all correlative microblog posts of an event into several groups. We intend to group all correlative microblog posts of an event into a sequence of time windows and extract overall features through modeling microblog post groups.

Why split into several groups? *First*, an event generally consists of thousands of correlative microblog posts on average and there is huge difference in quantity of events. *Moreover*, microblog posts during some specific time windows are so relevant that we can treat these neighbor microblog posts as a group which represents a specific event phase.

How to split? There are two things should be taken into consideration. *First*, all events need to be split in a unified way so that extracted distinguishing features make sense. For instance, truth information tends to be posted or reposted at the beginning and vanish very fast, while misinformation usually draws comparatively sustained attention at the middle phase. So quantities of microblog posts of different information at the same time window may be different. We should compare the number of microblog posts during the same time window and the obtained diversity can make sense. *Second*, we make sure to keep a phase of an event unbroken as possible, i.e., those most relevant microblog posts are within one group representing the phase of an event.

Considering the *long-tailed* distribution of the adopted datasets, equal time intervals adopted by [Ma *et al.*, 2016] may result in groups with unbalanced number of microblog posts, which is not a good strategy to learn appropriate representations for phases of an event. We'd better slice chronological microblog posts of all events into groups with equal amount. To be specific, we collect timestamps of all correlative microblog posts and subtract the starting timestamp of the corresponding event from all timestamps for every event. Then these timestamps are normalized to 0-1 scale. Finally the whole set of timestamps is equally split into 20 shares in chronological order and each time window is formulated as

$$T_i = [t_{i-1}, t_i), i = 1, 2, \cdots, 20, \quad (1)$$

where t_i is a end point of the i-th share. Note that there may be groups of some time windows in some events without any microblog posts.

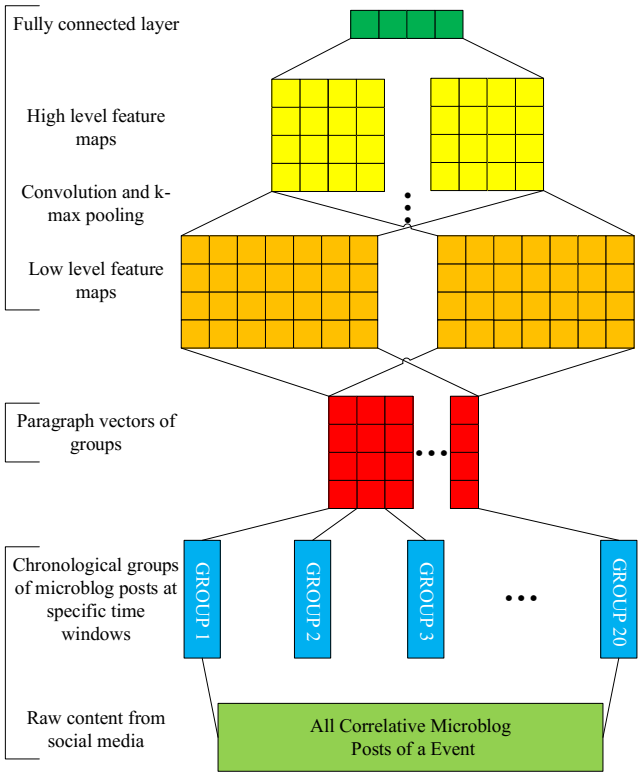


Figure 2: The framework of CAMI. From the bottom up: split raw content into chronological groups based on the distribution; learn paragraph vectors for all groups; extract features from low level to high level with CNN.

Learning representation for each group via paragraph vector. We treat microblog posts of one time window as an *event phase* and model overall features of the event with a sequence of phases. For convenience, paragraph vector [Le and Mikolov, 2014] is employed here. And an event phase of a group of microblog posts within a time window can be seen as a paragraph to learn the paragraph representation $\mathbf{g}_j$,

$$\arg\max_{\mathbf{D},\mathbf{W}} \frac{1}{N} \sum_{n=k}^{N-k} \log p\left(\mathbf{w}_n | \mathbf{w}_{n-k}, \cdots \mathbf{w}_{n+k}\right), \quad (2)$$

The prediction is made via softmax,

$$p\left(\mathbf{w}_n|\mathbf{w}_{n-k},\cdots\mathbf{w}_{n+k}\right)=\frac{\exp(\boldsymbol{\theta}^T\mathbf{x}_n)}{\Sigma_i\exp(\boldsymbol{\theta}^T\mathbf{x}_i)}\ , \tag{3}$$

$$\mathbf{x}_n=h\left(\mathbf{g}_j,\mathbf{w}_{n-k},\cdots,\mathbf{w}_{n+k};\mathbf{D},\mathbf{W}\right)\ , \tag{4}$$

Given a paragraph of N words, a word is represented by a column vector $\mathbf{w}_n$ in $\mathbf{W}$ and the paragraph is represented by a column vector $\mathbf{g}_j$ in $\mathbf{D}$. Moreover $\boldsymbol{\theta}$ is the softmax parameter and h is a concatenation or average operation. Context words and paragraph memory are leveraged to predict the current word.

In addition those groups without any microblog posts are represented by zero vectors. It is worthwhile to point out that input of the proposed model has a fix size of 20 and paragraph vectors of the input layer of CAMI will not be updated in following training process.

Modeling high-level interactions by CNN. A commonly used architecture of CNN comprises convolutional layers, k-max pooling layers and a fully connected layer.

For an input event instance e_i with n phases, each phase is embedded as $\mathbf{g}_i\in\mathbb{R}^d$ and we can get the instance matrix $\mathbf{G}\in\mathbb{R}^{d\times n}$. In the convolutional network, a convolutional layer is obtained by convolution operations of a weight matrix $\mathbf{C}\in\mathbb{R}^{d\times\omega}$ on the activation matrix at the layer below in a row-wise way. Followed by a nonlinearity function applied to the convolution result, an element of a feature map can be obtained as:

$$\mathbf{f}[i]=\tanh\left(\langle\mathbf{G}[:,\ i:i+\omega-1],\mathbf{C}\rangle_F\right)\ , \tag{5}$$

where $\mathbf{G}[:,\ i:i+\omega-1]$ is the i to $(i+\omega-1)$-th columns of $\mathbf{G}$ and the subscript F is the Frobenius inner product, i.e., the summation of products of corresponding elements of both matrices. At last, we take k-max pooling over the feature map $\mathbf{f}$ to capture the most significant features $\mathbf{f}^k_{max}$, i.e., k largest values of the feature map in response to the specific kernel $\mathbf{f}$ and the order of the values in $\mathbf{f}^k_{max}$ stays the same as their original order in $\mathbf{f}$.

Moreover, the above convolutional and pooling operations can be repeated to yield deeper layers. Finally, there is a fully connected layer and the ultimate output p_{e_i} is obtained via softmax. Where p_{e_i} is the probability predicting whether the event e_i belongs to misinformation.

5 Experiments

In this section, we first present compared methods and settings used in our proposed method. Then we report experiment results of misinformation identification and early detection on the datasets comparing above methods. To further demonstrate effectiveness of our model, we conduct some visualization experiments which help apparently illustrate what the proposed model has learnt.

5.1 Experiment Settings

To empirically evaluate the performance of our method on misinformation identification, we perform experiments on two large microblog datasets. Several methods are used for empirical comparison with ours:

(1) **GRU-2** is equipped with two GRU hidden layers and an embedding layer following the input layer. The enhanced GRU hidden layer conduce to obtain high-level interactions of features [Ma *et al.*, 2016].

(2) **SVM-TS** is a linear SVM classifier that uses time-series structures to model the variation of social context features and these handcrafted features are extracted based on contents, users and propagation patterns [Ma *et al.*, 2015].

(3) **DT-Rank** is a decision-tree-based ranking model to identify trending rumors through ranking the clustered disputed factual claims based on statistical features [Zhao *et al.*, 2015]. **DTC** is a Decision Tree Classifier modeling information credibility [Castillo *et al.*, 2011].

(4) **SVM-RBF** is a SVM-based model with the RBF kernel [Yang *et al.*, 2012].

(5) **RFC** is a Random Forest Classifier with three parameters to fit the temporal tweets volume curve [Kwon *et al.*, 2013].

In all experiments, we randomly choose 10% of dataset for model tuning and the rest 90% are randomly assigned in a 3:1 ratio for training and test. Similar to [Ma *et al.*, 2016], we adopt accuracy, precision, recall and F-measure as the evaluation metrics to measure the performance of misinformation identification. For the proposed CAMI, we apply a CNN architecture with two layers in this work, which is implemented with Theano[4]. The parameters of CAMI are set as $d=72, m=[6,4], w=[7,5]$ for the Weibo dataset, and $d=56, m=[6,4], w=[7,5]$ for Twitter dataset (m, w are the numbers of feature maps and filter width of two layers).

5.2 Results of Misinformation Identification

Performance results of all methods are illustrated in Table 1, from which we can see that the performance ranking of misinformation identification methods is as follows, CAMI, GRU-2, SVM-TS, RFC, DTC, SVM-RBF and DT-Rank. Compared with deep neuron network (DNN) based methods, the performance of other methods is relatively poor. These methods using handcrafted features or rules may not adapt to shape dynamic and complicated scenarios in social media. In contrast, DNN based methods, either CAMI or GRU-2, can learn high-level interactions among deep latent features, which make the models closer to real-world scenarios.

Examining those conventional methods. DT-Rank uses a set of regular expressions selected from signal microblog posts containing skeptical enquiries. But not many microblog posts in both Twitter and Weibo dataset involve these skeptical enquiries and limited selected expressions are not enough to conclude the information credibility. Moreover, SVM-TS and RFC incorporate the temporal structure into conventional models, which helps outperform other compared methods like SVM-RBF and DTC. So we can see that modeling these temporal features is workable and effective.

With regard to DNN based methods, the proposed CAMI model obtain significant improvement comparing GRU-2. Despite the fact that both models learn deep latent features from a sequence of groups of microblog posts, a trained GRU model possesses a constant recurrent transition matrix, which

[4]http://deeplearning.net/software/theano/

Table 1: Misinformation identification (**M**: Misinformation; **T**: Truth Information)

Method	Class	Weibo Accuracy	Weibo Precision	Weibo Recall	Weibo F_1	Twitter Accuracy	Twitter Precision	Twitter Recall	Twitter F_1
DT-Rank	M	0.732	0.738	0.715	0.726	0.681	0.711	0.698	0.704
	T		0.726	0.749	0.737		0.647	0.662	0.655
SVM-RBF	M	0.818	0.822	0.812	0.817	0.715	0.698	0.809	0.749
	T		0.815	0.824	0.819		0.741	0.610	0.669
DTC	M	0.831	0.847	0.815	0.831	0.718	0.721	0.711	0.716
	T		0.815	0.847	0.830		0.715	**0.725**	0.720
RFC	M	0.849	0.786	0.959	0.864	0.728	0.742	0.737	0.740
	T		0.947	0.739	0.830		0.713	0.718	0.716
SVM-TS	M	0.857	0.839	0.885	0.861	0.745	0.707	**0.864**	0.778
	T		0.878	0.830	0.857		0.809	0.618	0.701
GRU-2	M	0.910	0.876	**0.956**	0.914	0.757	0.732	0.815	0.771
	T		**0.952**	0.864	0.906		0.788	0.698	**0.771**
CAMI	M	**0.933**	**0.921**	0.945	**0.933**	**0.777**	**0.744**	0.848	**0.793**
	T		0.945	**0.921**	**0.932**		**0.820**	0.705	0.758

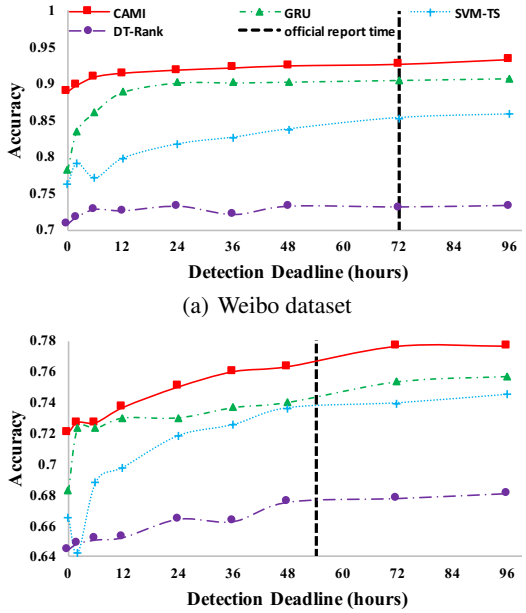
induces unchangeable propagations of sequence signals between every two consecutive time windows. However, in real-world scenarios, social media is so much dynamic and complicated that the above constant recurrent transition matrix of GRU-2 model has its limitation to shape an adequate misinformation identification model. Furthermore, like the conventional RNN, the above GRU-2 model has a bias towards the latest elements that it takes as input [Mikolov *et al.*, 2011]. While key features of both misinformation and truth information do not necessarily appear at rear part of one input sequence, which can be further demonstrated in the following visualization experiment (detailed in Section 5.4). The convolutional architecture and k-max pooling operation in the proposed CAMI model, by contrast, can flexibly extract key features scattered among one input sequence, which will also be demonstrated by the following visualization experiment.

5.3 Early Detection of Misinformation

In order to evaluate performance of early detection of compared methods, we set a series of detection deadlines and only use microblog posts from the initial broadcast to corresponding deadlines during the test process.

Several methods are selected for comparison: GRU-2 and SVM-TS are state-of-the-art methods and DT-Rank is specially designed for early detection of misinformation. Moreover, conventional early detection tasks count on official announcements. So we take as a reference the average reporting time over misinformation, which is announced by the debunking services like Snopes and Sina community management center.

Performance of the proposed CAMI model versus above methods with various deadlines are illustrated in Figure 3. The proposed CAMI model can reach relatively high accuracy at a very early time while other methods will take longer time to perform well enough. Furthermore, accuracy of the proposed CAMI model takes a commanding leading at any phase. Only in this way can the proposed CAMI model shot misinformation at first appearance and achieve more practical early detection.



(a) Weibo dataset

(b) Twitter dataset

Figure 3: Early detection of Misinformation

Accuracy of most methods will experience a conspicuous climbing during the first few hours and then rise with different growth rate, convergence rate and convergence accuracy. For instance, accuracy curves of DT-Rank and SVM-TS both climb slowly at early phase and gradually converge to relatively low accuracy. Moreover, their accuracy curves still fluctuate after the official report time. While accuracy curve of GRU-2 climb rapidly at early phase and converge to much higher accuracy on a much earlier deadline than those of DT-Rank and SVM-TS as well as the mean official report time.

Most state-of-the-art methods for early detection, such as GRU-2 and SVM-TS, usually follow the intuitive paradigm to model time series features in sequences of microblog posts. But these time series based models are not qualified for prac-

tical early detection task due to the *conflict* between the models and the task. Take GRU-2 as an example. On the one hand, the input sequence should be long enough to embody these possibly existing dynamic temporal signals to be captured by GRU-2 [Ma *et al.*, 2016]. On the other hand, the practical early detection means limited input sequence can be used. The limited input sequence may not cover required dynamic temporal signals. So GRU-2 may not work for early detection of misinformation in some cases. Nonetheless convolutional and max pooling operations of proposed CAMI model can flexibly extract key features even from a limited input sequence, which make the proposed CAMI model more effectively applied to early detection of misinformation.

5.4 Visualizing the CAMI Model

The visualization experiments of the CAMI model demonstrate the following two things. *First*, key features scatter among one input sequence but not focus on a fix part. *Second*, the CAMI can flexibly extract these scattered key features.

Visualizing convolutional kernels. We obtain all convolutional kernels from the first convolutional layer of a learnt CAMI model. With regard to a kernel matrix $\mathbf{W} \in \mathbb{R}^{d\times\omega}$ corresponding to a specific feature map, we sum all the rows into a row vector $\mathbf{v}_i \in \mathbb{R}^{\omega}$. Suppose there are m feature maps, we can stack these row vectors, $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m$, into a visualization matrix $\mathbf{V} \in \mathbb{R}^{m\times\omega}$ and then plot it in a checkerboard which is illustrated in Figure 4. Taking the adopted one-dimension convolution into consideration, each row in the visualization figure illustrate general response of a corresponding kernel with respect to the input sequence.

From Figure 4, we can see that the forepart of input usually obtains relatively stronger response than the rear part. After all, main description of misinformation and most relative replies may locate at the forepart. So GRU-2 model may not make the best of key features with a bias towards the latest elements of input. In addition, some kernels respond strongly at middle and rear part, such as ones in the third and fourth rows, which shows that the proposed CAMI model can flexibly extract key features scattered among one input sequence.

Visualizing saliency maps. Inspired by visualizing work in computer vision [Simonyan *et al.*, 2013; Vondrick *et al.*, 2013], we visualize key features grabbed by the CAMI model. In a feedback pass during test process, we compute the gradient of one class label value with respect to the input embedding matrix. More concretely, for a test instance, we perform a feedforward pass and obtain the output value and corresponding class label. Then we treat the class label value as loss and implement back propagation algorithm to acquire the gradient matrix of the class label value with respect to the input embedding matrix. Finally we can get the most salient part of the input instance from the gradient matrix.

Table 2 demonstrates extracted salient parts of an identified misinformation about "Donald Trump Said Republicans Are the Dumbest Group of Voters", in which many questioning and denial signals can be observed in corresponding groups of microblog posts. Such groups with indicating signals could be flexibly grabbed by the proposed CAMI.

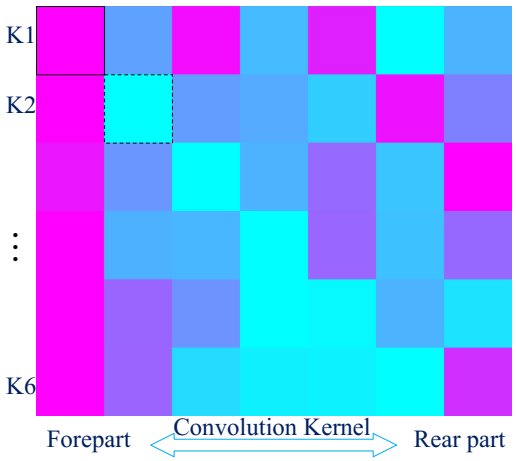


Figure 4: Visualization of convolutional kernels from the first convolutional layer (better viewed in color and rows). Each row represents a convolution kernel of size 7 and there are kernels (termed K1, K2, $\cdots$, K6) from 6 feature maps. Colors varying from bright blue (dashed line box) to bright red (black box) map values from low to high, representing response intensity of kernels with respect to input.

Table 2: Extracted salient microblog posts

time window #1	what???? IS IT TRUE? probably faked I doubt the Trump2016 folks do
time window #2	untrue... False, darn it. Didn't think so... it pays to fact check
time window #6	this is false Fake. False. Deceitful. but no proof exists that he said this... Just another graphic created by a pundit

6 Conclusion

In this paper, we have proposed a novel CAMI model for both misinformation identification and early detection tasks. Extensive experiments on two large social media datasets have demonstrated the effectiveness of the proposed CAMI model than both conventional feature engineering based methods and a RNN based method. We also illustrate inherent properties of information in social media and visualize what the proposed model can captured, which will help comprehend human behaviors in cyberspace to shape more exact real-world social media scenarios. Then we can better accomplish the task of misinformation identification and early detection.

Acknowledgments

This work is jointly supported by National Key Research and Development Program (2016YFB1001000), National Natural Science Foundation of China (61403390, U1435221), CCF-Tencent Open Fund and CCF-Venustech Hongyan Research Fund.

References

[Abdel-Hamid *et al.*, 2012] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid nn-hmm model for speech recognition. In *ICASSP*, pages 4277–4280, 2012.

[Castillo *et al.*, 2011] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *WWW*, pages 675–684, 2011.

[Giudice, 2010] Katherine Del Giudice. Crowdsourcing credibility: The impact of audience feedback on web page credibility. *Proceedings of the American Society for Information Science and Technology*, 47(1):1–9, 2010.

[Gupta *et al.*, 2013] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *WWW*, pages 729–736, 2013.

[Jin *et al.*, 2014] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. News credibility evaluation on microblog with a hierarchical propagation model. In *ICDM*, pages 230–239, 2014.

[Jin *et al.*, 2016] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. News verification by exploiting conflicting social viewpoints in microblogs. In *AAAI*, pages 2972–2978, 2016.

[Kalchbrenner *et al.*, 2014] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. In *ACL*, pages 655–665, 2014.

[Kumar and Geethakumari, 2014] KP Krishna Kumar and G Geethakumari. Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, 4(1):1, 2014.

[Kumar *et al.*, 2016] Srijan Kumar, Robert West, and Jure Leskovec. Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In *WWW*, pages 591–602, 2016.

[Kwon *et al.*, 2013] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. Prominent features of rumor propagation in online social media. In *ICDM*, pages 1103–1108, 2013.

[Le and Mikolov, 2014] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. In *ICML*, pages 1188–1196, 2014.

[Liu *et al.*, 2015] Qiang Liu, Feng Yu, Shu Wu, and Liang Wang. A convolutional click prediction model. In *CIKM*, pages 1743–1746, 2015.

[Ma *et al.*, 2015] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. Detect rumors using time series of social context information on microblogging websites. In *CIKM*, pages 1751–1754, 2015.

[Ma *et al.*, 2016] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. In *IJCAI*, pages 3818–3824, 2016.

[Mikolov *et al.*, 2011] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. Extensions of recurrent neural network language model. In *ICASSP*, pages 5528–5531, 2011.

[Qazvinian *et al.*, 2011] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In *EMNLP*, pages 1589–1599, 2011.

[Rieh *et al.*, 2014] Soo Young Rieh, Grace YoungJoo Jeon, Ji Yeon Yang, and Cliff Lampe. Audience-aware credibility: From understanding audience to establishing credible blogs. In *ICWSM*, 2014.

[Simonyan *et al.*, 2013] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013.

[Tamar *et al.*, 2016] Aviv Tamar, Sergey Levine, Pieter Abbeel, Yi Wu, and Garrett Thomas. Value iteration networks. In *NIPS*, pages 2146–2154, 2016.

[Vondrick *et al.*, 2013] Carl Vondrick, Aditya Khosla, Tomasz Malisiewicz, and Antonio Torralba. Hoggles: Visualizing object detection features. In *ICCV*, pages 1–8, 2013.

[Yang *et al.*, 2012] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. Automatic detection of rumor on sina weibo. In *SIGKDD Workshop on Mining Data Semantics*, page 13, 2012.

[Zhao *et al.*, 2015] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. Early detection of rumors in social media from enquiry posts. In *WWW*, pages 1395–1405, 2015.

[Zhao *et al.*, 2017] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *CVPR*, 2017.