Mining Significant Microblogs for Misinformation Identification: An Attention-Based Approach

QIANG LIU, FENG YU, SHU WU, and LIANG WANG, Chinese Academy of Sciences & University of Chinese Academy of Sciences

With the rapid growth of social media, massive misinformation is also spreading widely on social media, e.g., Weibo and Twitter, and brings negative effects to human life. Today, automatic misinformation identification has drawn attention from academic and industrial communities. Whereas an event on social media usually consists of multiple microblogs, current methods are mainly constructed based on global statistical features. However, information on social media is full of noise, which should be alleviated. Moreover, most of the microblogs about an event have little contribution to the identification of misinformation, where useful information can be easily overwhelmed by useless information. Thus, it is important to mine significant microblogs for constructing a reliable misinformation identification method. In this article, we propose an attention-based approach for identification of misinformation (AIM). Based on the attention mechanism, AIM can select microblogs with the largest attention values for misinformation identification. The attention mechanism in AIM contains two parts: content attention and dynamic attention. Content attention is the calculated-based textual features of each microblog. Dynamic attention is related to the time interval between the posting time of a microblog and the beginning of the event. To evaluate AIM, we conduct a series of experiments on the Weibo and Twitter datasets, and the experimental results show that the proposed AIM model outperforms the state-of-the-art methods.

CCS Concepts: • **Information systems** → **Data mining**; **Collaborative and social computing systems and tools**; *Web mining*; *Social networks*;

Additional Key Words and Phrases: Misinformation identification, attention model, social media, significant microblogs

ACM Reference format:
Qiang Liu, Feng Yu, Shu Wu, and Liang Wang. 2018. Mining Significant Microblogs for Misinformation Identification: An Attention-Based Approach. *ACM Trans. Intell. Syst. Technol.* 9, 5, Article 50 (April 2018), 20 pages.
https://doi.org/10.1145/3173458

1 INTRODUCTION

With the rapid growth of social media, e.g., Facebook, Twitter, and Weibo, people are sharing information and expressing their attitudes publicly. Social media brings great convenience to users, and information can be spread rapidly and widely today. However, misinformation can also be

This work was jointly supported by the National Key Research and Development Program (2016YFB1001000) and the National Natural Science Foundation of China (61403390, 61772528, U1435221).
Authors' addresses: Q. Liu, F. Yu, S. Wu (corresponding author), and L. Wang, Institute of Automation, Chinese Academy of Sciences (CASIA), National Laboratory of Pattern Recognition (NLPR), Center for Research on Intelligent Perception and Computing (CRIPAC), University of Chinese Academy of Sciences (UCAS), Beijing, 100080, China; emails: {quiang.liu, feng.yu, shu.wu, wangliang}@nlpr.ia.ac.cn.

https://doi.org/10.1145/3173458

spread on the Internet more easily. Misinformation brings significant harm to daily life, social harmony, and even public security. With the growth of the Internet and social media, such harm will also increase. For instance, as the loss of MH370 has drawn worldwide attention, a great amount of misinformation has spread on social media, e.g., MH370 has landed in China,[1] the loss of MH370 is caused by terrorists,[2] and Russian jets are related to the loss of MH370.[3] This misinformation about MH370 misleads public attitudes toward a wrong direction and delays the search of MH370. Up to March 15, 2017, on the biggest Chinese microblog Web site Sina Weibo,[4] 32,076 pieces of misinformation have been reported and collected in its misinformation management center.[5] Accordingly, it is crucial to evaluate information credibility and detect misinformation on social media.

In this work, we need to identify whether an event on social media is misinformation or true information. An event refers to a piece of news spreading on social media, e.g., "MH370 has been found in the international waters near Perth." Usually, an event, which may be misinformation or true information, contains a group of microblogs, which may include postings, repostings, and comments.

Today, to automatically identify misinformation on social media, some methods have been proposed. Most existing methods identify misinformation at the microblog level [3, 11, 30] or the event level [18, 25, 44]. Some studies investigate the aggregation of credibility from the microblog level to the event level [14]. However, considering dynamic information, some work designs temporal features based on the prorogation properties over time [18] or trains a model with features generated from different time periods [25]. Recently, recurrent neural networks (RNNs) [26] have been incorporated for misinformation identification, and the gated recurrent unit (GRU) structure [7] has proved to have satisfactory performance [24]. Moreover, some methods take usage of users' feedback (comments and attitudes) to evaluate the credibility [9, 32, 44]. For instance, signal tweets, which indicates users' skepticism about factual claims, have been taken out for detecting misinformation [44].

Although the preceding methods succeed in misinformation identification, they have severe drawbacks. Among these methods, some identify misinformation according to global statistical features of an event or a time window, and some calculate the credibility of each microblog and then aggregate them to the credibility of the whole event. However, information on social media is full of noise, which should be alleviated. Moreover, most of microblogs about an event have little contribution to the identification of misinformation. As shown in the example of misinformation on Sina Weibo in Table 1, most users simply repost the fake news or credence to the misinformation. Only a few users express their questions about the misinformation. Thus, it is important to select those significant microblogs and obtain reliable features to construct a misinformation identification method.

Fortunately, the attention mechanism [13] is suitable for selecting the most significant components of information. Via the attention mechanism, components that contribute more to a specific task have large weights for satisfying the objective as much as possible. The attention mechanism has succeeded in multiple tasks, e.g., visual object detection [1], image caption generation [39], machine translation [2], text summarization [33], and text classification [35]. Accordingly, with the

[1] http://www.fireandreamitchell.com/2014/03/07/rumor-malaysia-airlines-mh370-landed-china/.
[2] http://www.csmonitor.com/World/Asia-Pacific/2014/0310/Malaysia-Airlines-flight-MH370-China-plays-down-terrorism-theories-video.
[3] http://www.inquisitr.com/1689765/malaysia-airlines-flight-mh370-russian-jets-in-baltic-may-hold-clue-to-how-flight-370-vanished/.
[4] http://weibo.com.
[5] http://service.account.weibo.com/?type=5&status=4.

Table 1. Example of Misinformation on Sina Weibo

Posting Time	Content
2014/03/20 23:55	Hearing from an Australian friend: The plane has been found in the international waters near Perth. It is proven to be MH370 according to a major component of the plane.
2014/03/01 23:56	May God bless them!
2014/03/20 23:57	Reposting
2014/03/20 23:58	It is serious to spread rumors!
2014/03/20 23:59	Reposting
2014/03/21 00:00	Hopefully it's not true.
2014/03/21 00:02	Reposting
2014/03/21 00:03	Really???
2014/03/21 00:04	Waiting for official confirmation tomorrow.
2014/03/21 00:06	Reposting
2014/03/21 00:07	Reposting
2014/03/21 00:17	Let's watch the exact news tomorrow morning. Anyway, may God bless them!
2014/03/21 00:18	Reposting
2014/03/21 00:21	What a bad news!
2014/03/21 00:22	Reposting
2014/03/21 00:32	Reposting
2014/03/21 00:35	Reposting unreliable information, what an expert!
2014/03/21 00:46	Reposting
2014/03/21 00:51	No! No! No!
2014/03/21 01:06	Dare to post misinformation!
2014/03/21 01:09	Reposting
2014/03/21 01:25	Is it reliable?
2014/03/21 01:32	Reposting

attention mechanism, we are able to mine significant microblogs for identifying misinformation and design a reliable automatic detection method.

Moreover, early detection of misinformation is another important and practical task, in which we need to detect misinformation as early as possible [24, 44]. Thus, we can take immediate action at the beginning stage of the spreading of misinformation and minimize the baneful influence. For early detection, we need to identify misinformation with the first several microblogs. Thus, existing methods are not qualified for practical early detection due to the conflict between the models and the task. Meanwhile, with the attention mechanism, we can identify misinformation with several significant microblogs. This minimizes the conflict between the models and the task. Accordingly, the attention mechanism is naturally suitable for early detection of misinformation.

In this work, we propose AIM. First, for each microblog belonging to an event, we calculate a corresponding attention value based on its textual features. This attention value is named *content attention*. Second, considering that microblogs posted at different times have distinct significance for the event, we calculate dynamic attention for each microblog. Dynamic attention can be determined related to the time interval between the posting time of a microblog and the beginning of the event. Then we aggregate the content attention and the dynamic attention, and we obtain the final attention weights for each microblog belonging to an event. The weighted sum of these microblogs can be performed to generate the representation of the whole event. Finally, true information can be separated from misinformation based on the event representation.

In summary, the main contributions of this work are listed as follows:

- We incorporate the attention mechanism for misinformation identification on social media, which mines the most significant microblogs.
- We design both content attention and dynamic attention for capturing different aspects of significance of microblogs for misinformation identification.
- Experiments conducted on two real-world datasets (i.e., the Weibo and Twitter datasets) show that AIM is effective and outperforms state-of-the-art methods significantly.
- Visualization of the leaned attention mechanism in AIM demonstrates the rationality of our proposed method.

The rest of this article is organized as follows. In Section 2, we review some related work on misinformation identification and the attention mechanism. Then we detail the proposed AIM model in Section 3. In Section 4, we conduct and analyze experiments on two real-world datasets and compare them to several state-of-the-art methods. In Section 5, we illustrate some visualization examples of the leaned attention mechanism. Section 6 concludes this work and discusses future research directions.

2 RELATED WORK

In this section, we briefly review some related works on misinformation identification and attention mechanism.

2.1 Misinformation Identification

Recently, many methods have been put forward for misinformation automatic identification. The work of Kumar et al. [17] analyzes impact and characteristics of hoax articles in Wikipedia and proposes an efficient method to identify these Wikipedia hoaxes. On social media, some researchers identify misinformation at the post level [3, 30], i.e., classifying a single microblog post as being credible or not based on tweet-based features. Some perform a characterization analysis for the spread of fake images of microblog posts during crisis events [11]. Some identify whether an event belongs to misinformation or truth information and extract handcrafted features from the event level [18, 25, 36, 44]. Another work obtains credibility of a microblog post and then aggregates credibility to the event level [14]. Moreover, some other works extract more effective handcrafted features. For instance, the work of Jin et al. [15] takes advantage of "wisdom of crowds" to identify fake news, i.e., mining opposing voices from conflicting viewpoints. Based on the time series of misinformation lifecycle, the temporal characteristics of social context information are captured in Kwon et al. [18] and Ma et al. [25]. The work of Del Giudice [9] and Rieh et al. [32] investigates the Web page credibility through users' feedback. Signals tweets are identified from trending misinformation via finding signature text phrases expressing skepticism about factual claims [44]. Recently, an RNN-based model attempted to capture the dynamic temporal signals in the misinformation diffusion process and incrementally learn both the temporal and textual representations of an event [24].

2.2 Attention Mechanism

The attention mechanism is first applied to a visual attention system for rapid scene analysis [13]. The visual attention system selects attended locations in order of decreasing saliency so that a complex scene can be understood by rapidly selecting saliency locations in a computationally efficient method. In recent years, the deep neural network (DNN) has become increasingly popular. The attention mechanism is once again taken out to be integrated into DNNs. The attention mechanism is incorporated into the RNN in Mnih et al. [29] to attend to different locations within the images one at a time and process them sequentially. The attention mechanism can help control expensive computation independent of the input image size and learn tracking without explicit training signals.

Furthermore, the work of Ba et al. [1] extends the attention RNN model to a multiple-objects detection task that is learning to both localize and recognize multiple objects despite being given only class labels. For an image caption task, an attention-based model is able to automatically fix its attention on salient objects of an input image while generating the corresponding words of the output sentence [39]. Some employ the attention mechanism in a visual question-answering task, e.g., generating question-guided attention to image feature maps for each question [5], a question-guided spatial attention to images for questions of spatial inference [38], and querying an image and inferring the answer multiple times to narrow down the attention to images progressively via stacked attention networks [41]. For fine-grain image classification, the attention-based convolutional neural networks (CNN) model has been found to perform better by identifying which one is to be attended to and what can be extracted without any expensive annotations, e.g., bounding box or part information [37].

In the field of natural language processing, researchers first introduce the attention mechanism to neural machine translation. Based on a primitive encoder-decoder architecture, the work of Bahdanau et al. [2] searches a source sentence to attend to the most relevant words to predict a target word. Some extend the attention mechanism to global and local ones and compare different methods of obtaining attention scores [23]. Moreover, a hierarchical attention mechanism guides layers in a CNN model to model text in Yin et al. [43]. The work of Dhingra et al. [8] integrates a multihop architecture with a gated attention layer based on multiple interactions between the query embedding and the intermediate states of a recurrent document reader. In addition, the attention mechanism is introduced into other research issues, e.g., abstractive text summarization [33], the text comprehension task [8, 16, 43], relation classification [34, 45] and text classification [42]. In Chorowski et al. [6], a novel model for speech recognition is proposed, which incorporates both content-based attention [2, 39] and location-based attention [10].

3 PROPOSED METHOD

In this section, we first define and formulate the problem. We then detail the proposed AIM model. Finally, we present the parameter learning procedure for the AIM model.

3.1 Problem Definition

The problem studied in this article can be defined and formulated as follows. An event refers to a piece of news spreading on social media, e.g., "MH370 has been found in the international waters near Perth." The event can be true or false. Each event is associated with multiple microblogs, containing postings, repostings, and comments. These microblogs, which may have different degrees of importance and credibility, describe various aspects of the event. In this work, based on all of these microblogs, our task is to identify whether an event on social media is misinformation or not.

Suppose that a set of events is denoted as $E = \{e_1, e_2, \ldots, e_n\}$. l_{e_i} is the label of the corresponding event e_i, where $l_{e_i} = 1$ means that event e_i is misinformation and $l_{e_i} = 0$ otherwise. The microblogs of the event e_i can be denoted as $M^{e_i} = [m_1^{e_i}, m_2^{e_i}, \ldots, m_{n_{e_i}}^{e_i}]$, where n_{e_i} is the total number of microblogs of this event. All microblog sets can be written as $M = \{M^{e_1}, M^{e_2}, \ldots, M^{e_n}\}$. Each microblog $m_j^{e_i}$ consists of its content feature vector $\mathbf{f}_j^{e_i}$ and posting time $t_j^{e_i}$. Then content feature vectors and the posting time of event e_i can be denoted as $\mathbf{F}^{e_i} = [\mathbf{f}_1^{e_i}, \mathbf{f}_2^{e_i}, \ldots, \mathbf{f}_{n_{e_i}}^{e_i}]$ and $T^{e_i} = [t_1^{e_i}, t_2^{e_i}, \ldots, t_{n_{e_i}}^{e_i}]$, respectively.

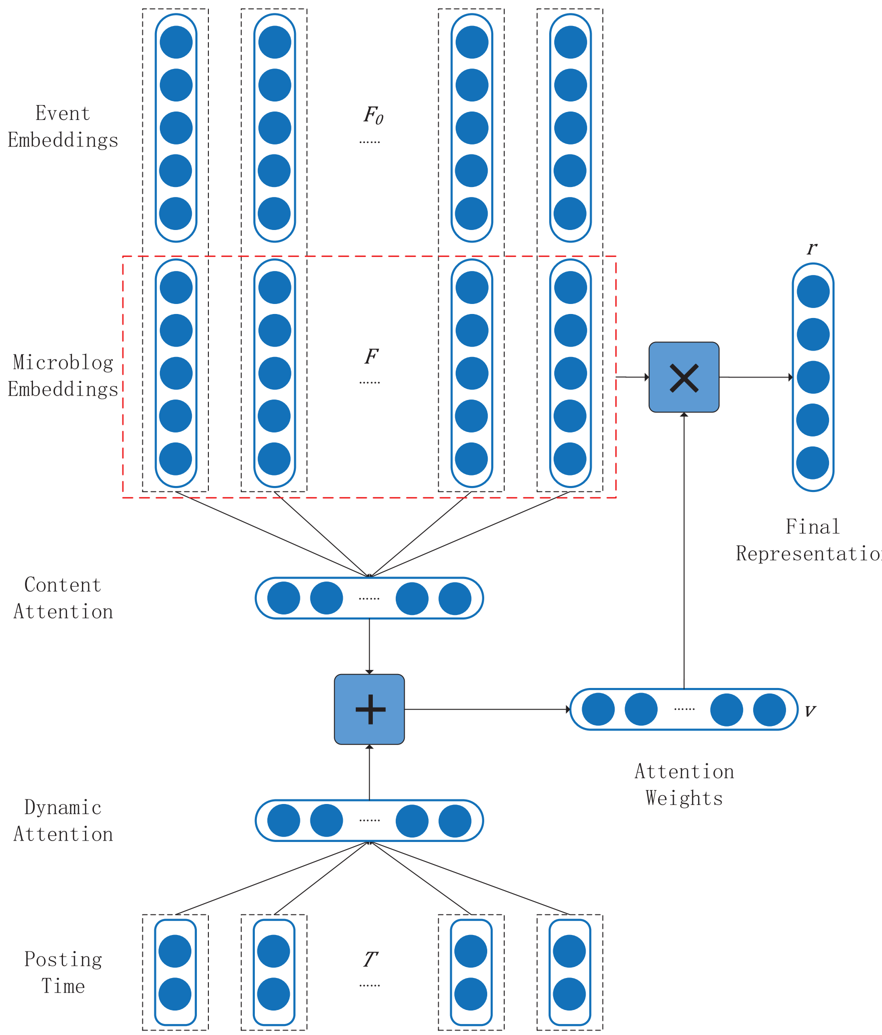


Fig. 1. Illustration of the proposed AIM model. The attention mechanism in AIM contains two parts: content attention and dynamic attention. Content attention is calculated based on the textual features of each microblog. Dynamic attention is related to the time interval between the posting time of a microblog and the beginning of the event.

3.2 The Attention Mechanism for Misinformation Identification

As we know, information on social media is usually full of noise, which should be alleviated. Moreover, most microblogs about an event have little contribution to the identification of misinformation, and useful information can be easily overwhelmed by useless information. Thus, a reliable misinformation identification method can benefit from mining several most significant microblogs. Meanwhile, the attention mechanism [13] is suitable for and widely applied in selecting the most significant components of information [1, 2, 33, 35, 39]. Via the attention mechanism, components that contribute more to a specific task have large weights for satisfying the objective as much as possible. Thus, we propose a reliable automatic detection method to mine signification microblogs for identifying misinformation based on the attention mechanism.

For an event e_i that contains microblogs $M^{e_i} = [m_1^{e_i}, m_2^{e_i}, \ldots, m_{n_{e_i}}^{e_i}]$, we have a function $\mathrm{A}(e_i)$ to calculate attention values for each microblog. To capture content features and dynamic properties simultaneously, we have content attention and dynamic attention that can be calculated via $\mathrm{A_c}(e_i)$ and $\mathrm{A_t}(e_i)$, respectively. We will discuss them further in Sections 3.3 and 3.4. This process can be formulated as follows:

$$\mathrm{A}(e_i) = \mathrm{A_c}(e_i) + \mathrm{A_t}(e_i), \tag{1}$$

which outputs a n_{e_i}-dimensional attention value vector. This vector can be further normalized, and thus we can obtain an attention weight vector:

$$\mathbf{v}^{e_i} = \mathrm{softmax}(\mathrm{A}(e_i)), \tag{2}$$

where $\mathbf{v}^{e_i} \in \mathbb{R}^{n_{e_i}}$, denoting weights for each microblog belonging to event e_i. A large attention weight indicates that the corresponding microblog has a significant effect on misinformation identification.

Then a weighted sum of all microblogs can be performed to generate the representation of the whole event:

$$\mathbf{r}^{e_i} = \mathbf{F}^{e_i}\mathbf{v}^{e_i}, \tag{3}$$

where $\mathbf{F}^{e_i} \in \mathbb{R}^{d \times n_{e_i}}$ and $\mathbf{r}^{e_i} \in \mathbb{R}^d$. $\mathbf{r}^{e_i}$ denotes the final representation of event e_i, and $\mathbf{F}^{e_i} = [\mathbf{f}_1^{e_i}, \mathbf{f}_2^{e_i}, \ldots, \mathbf{f}_{n_{e_i}}^{e_i}]$ denotes content features of microblogs M^{e_i}. Here we use textual embeddings of microblogs as content features. In this work, we apply para2vec [19] for extracting embeddings of microblogs. Para2vec is an extended version of word2vec [28] and is a state-of-the-art method for extracting sentence embeddings. In this work, we empirically set the dimensionality of embeddings as $d = 50$.

Finally, the prediction on e_i can be made with a logistic regression:

$$\hat{l}_{e_i} = \mathrm{sigmoid}\left(\mathbf{W}^T\mathbf{r}^{e_i} + \mathbf{b}\right), \tag{4}$$

where $\mathbf{W} \in \mathbb{R}^d$ and $b \in \mathbb{R}$. $\hat{l}_{e_i} = 1$ means that event e_i is predicted to be misinformation and $\hat{l}_{e_i} = 0$ otherwise. The larger the predicted value, the lower the credibility of event e_i.

3.3 Content Attention

Content features of microblogs are the most important factor for describing an event. These features can tell us what has happened and how people react. Thus, it is vital to generate attention values based on textual embeddings $\mathbf{F}^{e_i} = [\mathbf{f}_1^{e_i}, \mathbf{f}_2^{e_i}, \ldots, \mathbf{f}_{n_{e_i}}^{e_i}]$ of microblogs. Moreover, when people make comments about the information of an event, the original content, i.e., the first microblog that describes what has happened, is usually not included. To model a comment's correlation with the original content for deciding its significance, we concatenate the textual embedding of the very first microblog and obtain the embeddings $\mathbf{F}_0^{e_i} = [\mathbf{f}_1^{e_i}, \mathbf{f}_1^{e_i}, \ldots, \mathbf{f}_1^{e_i}]$ of the event. Based on the microblog embeddings and the event embeddings, we can calculate content attention values for microblogs belonging to the same event.

First, we transfer above embeddings to a hidden space:

$$\mathbf{H}^{e_i} = \tanh\left(\mathbf{W}_h \begin{bmatrix} \mathbf{F}_0^{e_i} \\ \mathbf{F}^{e_i} \end{bmatrix}\right), \tag{5}$$

where $\mathbf{W}^h \in \mathbb{R}^{d_h \times 2d}$ and $\mathbf{H}^{e_i} \in \mathbb{R}^{d_h \times n_{e_i}}$. $\mathbf{H}^{e_i}$ denotes the hidden representations of microblogs M^{e_i}, and d_h denotes the dimensionality of the hidden space. This hidden space allows the interaction between microblog embeddings and event embeddings, and brings a new space for calculating content attention values.

Then we can calculate content attention values as:

$$\mathrm{A_c}\,(e_i) = \mathbf{W}_a^T \mathbf{H}^{e_i}, \tag{6}$$

where $\mathbf{W}^a \in \mathbb{R}^{d_h}$. These generated content attention values can be further managed as shown in Equation (1).

3.4 Dynamic Attention

In addition to content features, the posting time of a microblog is also vital for deciding its significance. At the beginning of an event, microblogs can tell what is going on about the event, which is important for knowing the whole picture of the event. With the spread of information, the reaction of common people is usually to repost, repeat, and echo the message, which has little value for identification. Slowly, some people tend to think about the event, or some people knowing the fact appear. Thus, people start to express their own attitudes, which may include suspicion, affirmation, and denial. This information can help identify misinformation.

Accordingly, we incorporate dynamic attention, which can be determined related to the time interval between the posting time of a microblog and the beginning of the event. The beginning of an event is the posting time of the very first microblog of the corresponding event. For example, the beginning time of event e_i is $t_1^{e_i}$. Accordingly, dynamic attention values can be calculated as follows:

$$\mathrm{A_t}\,(e_i) = \left[\mathbf{c}_{t_1^{e_i}-t_1^{e_i}}, \mathbf{c}_{t_2^{e_i}-t_1^{e_i}}, \ldots, \mathbf{c}_{t_{n_{e_i}}^{e_i}-t_1^{e_i}}\right], \tag{7}$$

where $\mathbf{c}_{t_j^{e_i}-t_1^{e_i}}$ denotes the dynamic attention value of the corresponding time interval $t_j^{e_i} - t_1^{e_i}$. These generated dynamic attention values can be further managed as shown in Equation (1).

Furthermore, if we learn a distinct attention value for every possible continuous time interval value, we have to estimate a great number of parameters and the model tends to overfit. Here, following several methods [20–22], we equally partition the range of all possible time interval values into discrete bins. Specifically, in this work, the range of all possible time interval values is partitioned into 1-hour bins. Only the attention values of the upper and lower bounds of time bins are needed to be estimated in our model. For time interval values in a time bin, their attention values can be calculated via a linear interpolation.

Suppose that we have an arbitrary time interval value $t_d = t_j^{e_i} - t_1^{e_i}$. Mathematically, the corresponding dynamic attention value $\mathbf{c}_{t_d}$ can be calculated as follows:

$$\mathbf{c}_{t_d} = \frac{\mathbf{c}_{L(t_d)}(U(t_d) - t_d) + \mathbf{c}_{U(t_d)}(t_d - L(t_d))}{U(t_d) - L(t_d)}, \tag{8}$$

where $U(t_d)$ and $L(t_d)$ denote the upper bound and lower bound of time interval value t_d, and $\mathbf{c}_{U(t_d)}$ and $\mathbf{c}_{L(t_d)}$ denote the dynamic attention values for $U(t_d)$ and $L(t_d)$, respectively. Such a linear interpolation method can solve the problem of learning attention values for continuous time intervals. Note that although the change of dynamic attention values in each discrete time bin is linear, the global change in the entire range of all possible time interval values is nonlinear.

To make the calculation of dynamic attention clear, we can explain with an example. If we want to calculate the attention value for time interval $t_d = 1.6h$, then the upper bound and lower bound of $1.6h$ will be $U(t_d) = 2h$ and $L(t_d) = 1h$, respectively. The corresponding attention value $\mathbf{c}_{1.6h}$ can be calculated as follows:

$$\mathbf{c}_{1.6h} = \frac{\mathbf{c}_{1h}(2h - 1.6h) + \mathbf{c}_{2h}(1.6h - 1h)}{2h - 1h} = 0.4\mathbf{c}_{1h} + 0.6\mathbf{c}_{2h}. \tag{9}$$

Table 2. Detailed Statistics of the Weibo and Twitter Datasets

Statistics (#)	Weibo	Twitter
Users	2,746,818	491,229
Microblogs	3,805,656	1,101,985
Events	4,664	992
Misinformation	2,313	498
True information	2,351	494

3.5 Parameter Learning

The proposed AIM model can be trained in an end-to-end way by backpropagation. The goal of training is to minimize the following error between l_{e_i} and $\hat{l}_{e_i}$ for each event e_i:

$$J = -\sum_{i=1}^{n} l_{e_i} \ln \hat{l}_{e_i} - \sum_{i=1}^{n} (1 - l_{e_i}) \ln \left(1 - \hat{l}_{e_i}\right) + \frac{\lambda}{2} \|\theta\|, \tag{10}$$

where λ is the L2-regularization term and $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{W}_h, \mathbf{W}_a, \mathbf{c}\}$, denoting all parameters needed to be learned in AIM. Then the derivations of J with respect to all parameters can be calculated, and we can employ stochastic gradient descent (SGD) to estimate the model parameters. The training procedure consists of two parts: the training of the attention mechanism that learns $\mathbf{W}_h$, $\mathbf{W}_a$, and $\mathbf{c}$, and the training of the logistic regression that learns $\mathbf{W}$ and $\mathbf{b}$. These two parts of training are done alternately. This process is repeated iteratively until the convergence is achieved.

4 EXPERIMENTS

In this section, we first present our experimental settings. We then report experiment results of AIM on misinformation identification and compare to several state-of-the-art methods. We also investigate the impact of hyperparameters in AIM. Moreover, we study the performance comparison on early detection of misinformation.

4.1 Experimental Settings

To evaluate the performance of AIM, following some representative previous works [24, 25], we conduct experiments on the Weibo and Twitter datasets. Detailed statistics of the two datasets can be found in Table 2. Misinformation in the Weibo dataset is collected from the Weibo misinformation management center,[6] which reports various misinformation. A similar number of true information is collected by crawling microblogs of general threads that are not reported as misinformation. Misinformation and true information are confirmed on Snopes,[7] which is an online misinformation debunking service.

In our experiments, we randomly choose 10% of events in each dataset for model tuning, and the rest 90% are randomly assigned in a 3:1 ratio for training and testing. Empirically, the L2-regularization term is set to be $\lambda = 0.001$, the learning rate of SGD is set to be 0.01, and the dimensionality of microblog embeddings is set to be $d = 50$. The dimensionality d_h of hidden space for generating content attention is turned in our experiments.

Moreover, we adopt several evaluation metrics for evaluating the performance of AIM and other compared methods: accuracy, precision, recall, and f1-score. Accuracy is a standard metric for

[6]http://service.account.weibo.com/?type=5&status=4.
[7]http://www.snopes.com.

classification tasks, which is evaluated by the percentage of correctly predicted misinformation and true information. Precision, recall, and f1-score are widely used metrics for classification tasks, which are computed according to where correctly predicted misinformation or true information appear in the predicted list. The larger the values of the preceding evaluation metrics, the better the performance.

To demonstrate the effectiveness of AIM, several state-of-the-art methods are compared in our experiments:

- *GRU* has been incorporated for misinformation identification. A model based on two GRU hidden layers [7] and textual features in dynamic time windows achieves satisfactory performance in both misinformation identification and misinformation early detection [24].
- *SVM-TS* is a linear support vector machine (SVM) [4] classifier that uses time-series structures to model the variation of social context features [25]. Handcrafted features based on content, users, and propagation patterns are replicated.
- *DT-Rank* is a ranking model based on a decision tree [31] to identify trending misinformation [44]. DT-Rank searches for enquiry phrases and cluster disputed factual claims, and ranks the clustered results based on statistical features.
- *DTC* is a decision tree classifier [3]. It uses handcrafted features based on the overall statistics of the posts rather than temporal information.
- *SVM-RBF* is an SVM-based model with the radical basis function (RBF) kernel [40]. It also uses handcrafted features based on the overall statistics of the posts.
- *DFC* is a random forest classifier [12] using three parameters to fit the temporal posting volume curve [18]. Same handcrafted features are used as in SVM-TS.

For our proposed AIM model, we implement it with Python[8] and Theano.[9] Moreover, versions of AIM without event embeddings, content attention, and dynamic attention are also implemented and compared for evaluating the impact of different components of AIM. To show the significant improvements of our proposed methods, experiments are done 20 times, and the ranges of accuracies of our proposed methods are illustrated in Table 3.

4.2 Performance Comparison on Misinformation Identification

Table 3 illustrates the performance comparison of misinformation among AIM and several state-of-the-art models on the Weibo and Twitter datasets. According to the conclusion in Section 4.4, the dimensionality of the hidden space for generating content attention is $d_h = 40$. We can see that the performance ranking of misinformation identification methods is as follows: AIM, GRU, SVM-TS, RFC, DTC, SVM-RBF, and DT-Rank. Compared to neural network–based methods, i.e., AIM and GRU, the performance of other methods is relatively poor. The methods using handcrafted features or rules may not adapt to shape dynamic and underlying correlations in social media. In contrast, neural network–based methods, AIM and GRU, can learn high-level interactions among deep latent features, which can better model real-world scenarios.

Among those conventional methods, DT-Rank uses a set of regular expressions selected from signal microblog posts containing skeptical enquiries. But not all microblog posts in both the Twitter and Weibo datasets involve these skeptical enquiries. These selected expressions are insufficient to conclude the information credibility. Moreover, SVM-TS and RFC incorporate the dynamic properties into conventional models, which helps outperform other compared methods e.g., SVM-RBF

[8]https://www.python.org/.
[9]http://deeplearning.net/software/theano/.

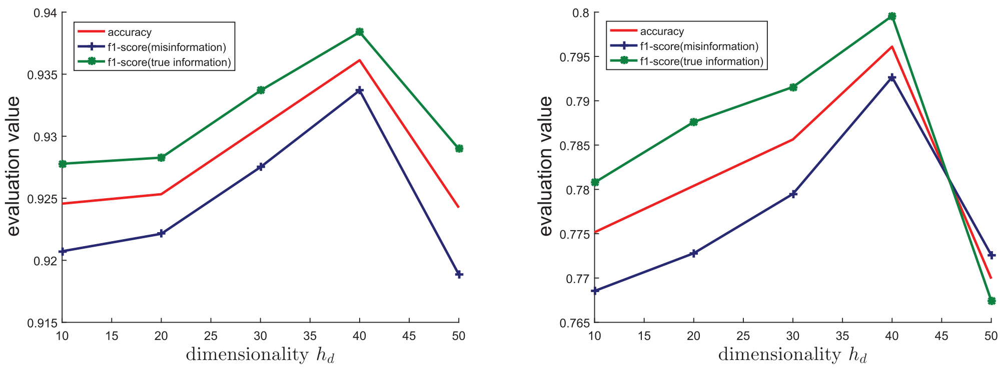
Table 3. Performance Comparison on Misinformation Identification With $d_h = 40$ on the Weibo and Twitter Datasets

		Weibo				Twitter			
Method	Class	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
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	M	0.908	0.874	**0.954**	0.912	0.757	0.732	0.815	0.771
	T		**0.950**	0.862	0.904		0.788	0.698	0.771
AIM	M	0.934	0.920	0.943	0.931	0.791	0.737	0.835	0.783
(-$\mathbf{F}_0$)	T	±0.002	0.947	0.926	0.936	±0.010	0.841	0.732	0.783
AIM	M	0.886	0.871	0.897	0.884	0.752	0.711	0.818	0.761
(-A_c)	T	±0.002	0.895	0.864	0.879	±0.005	0.819	0.707	0.759
AIM	M	0.927	0.915	0.934	0.925	0.770	0.725	0.826	0.772
(-A_t)	T	±0.002	0.939	0.922	0.930	±0.004	0.826	0.725	0.772
AIM	M	**0.936**	**0.922**	0.945	**0.934**	**0.796**	**0.746**	0.846	**0.794**
	T	±0.003	0.949	**0.928**	**0.938**	±0.009	**0.851**	**0.754**	**0.799**

Note: *M* stands for misinformation and *T* stands for true information. Results are evaluated by accuracy, precision, recall, and f1-score. AIM (-$\mathbf{F}_0$), AIM (-A_c), and AIM (-A_t) denote versions of AIM without event embeddings, content attention, and dynamic attention, respectively.

and DTC. Thus, we can conclude that dynamic properties are important features for misinformation identification.

From the experimental results, we can clearly observe that GRU achieves the best performance among all compared methods. It is obvious that AIM obtains significant improvement over GRU. On the Weibo dataset, compared to GRU, AIM improves performance by 2.8%, 2.2%, and 3.4% evaluated by accuracy, f1-score (misinformation), and f1-score (true information), respectively. On the Twitter dataset, the improvements become 3.9%, 2.3%, and 2.8%. Despite the fact that both AIM and GRU learn deep latent features from a sequence of groups of microblog posts, a trained GRU model possesses a constant recurrent transition matrix, which induces unchangeable propagations of sequence signals between every two consecutive time windows. However, in real-world scenarios, social media is so dynamic and complicated that the preceding constant recurrent transition matrix of the GRU model has its limitations in shaping an adequate misinformation identification model. Moreover, the GRU model has a bias toward the latest elements that it takes as input [27], whereas key features of both misinformation and true information do not necessarily appear at the rear part of an input sequence. Meanwhile, all compared methods, including GRU, cannot select significant microblogs for misinformation identification, and thus massive useless information will lower performance. Via overcoming these shortcomings, AIM shows its superiority in misinformation identification, which is proven by the experimental results. Moreover, via experiments



(a) Performances on the Weibo dataset.

(b) Performances on the Twitter dataset.

Fig. 2. Performance of AIM on misinformation identification with varying dimensionality $d_h = [10, 20, 30, 40, 50]$. Results are evaluated by accuracy and f1-score for both misinformation and true information.

performed 20 times, the ranges of accuracies of our proposed methods in Table 3 demonstrate the significant improvements of AIM compared to other methods.

4.3 Impact of Different Components of AIM

To investigate the impact of event embeddings $\mathbf{F}_0$ in content attention on misinformation identification, we implement a version of AIM without event embeddings, denoted as AIM (-$\mathbf{F}_0$). The corresponding performance on misinformation identification is shown in Table 3. Compared to AIM, AIM (-$\mathbf{F}_0$) slightly decreases the performance on both datasets. This indicates that a comment's correlation with the original content indeed has a contribution in deciding its significance. However, such an effect is not very significant.

Furthermore, to investigate the impact of content attention $\mathrm{A_c}$, a version of AIM, i.e., AIM (-$\mathrm{A_c}$), is implemented and compared in Table 3. Compared to AIM, AIM (-$\mathrm{A_c}$) decreases the accuracy by 5.0% and 4.4% on the Weibo dataset and the Twitter dataset, respectively. This shows that the content attention is very important for the attention mechanism. Moreover, from the results in Table 3, it is obvious that content attention has larger significance than dynamic attention for misinformation identification.

Similarly, to investigate the impact of dynamic attention $\mathrm{A_t}$, AIM (-$\mathrm{A_t}$) is compared in Table 3. Compared to AIM, AIM (-$\mathrm{A_t}$) decreases the accuracy by 0.9% and 2.6% on the Weibo dataset and the Twitter dataset, respectively. It is obvious that the decay brought by AIM (-$\mathrm{A_t}$) is relatively large. This indicates that the posting time of a microblog is important for deciding its significance, and dynamic attention is vital for misinformation identification. Moreover, AIM (-$\mathrm{A_t}$) can still outperform GRU, which means that AIM with only content attention can beat the compared methods.

4.4 Impact of Dimensionality

The dimensionality d_h of hidden space for generating content attention is a hyperparameter in AIM. To investigate its impact on the performance of AIM, we illustrate the performance of AIM evaluated by accuracy and f1-score with varying $d_h = [10, 20, 30, 40, 50]$ in Figure 2. The f1-score is evaluated for both misinformation and true information.

From the figure, on both datasets, we can clearly observe that the performance of AIM increases rapidly from $d_h = 10$. It achieves the best performance at $d_h = 40$ and then decreases with the increasing dimensionality. Curves evaluated by accuracy, f1-score (misinformation), and f1-score (true information) share similar trends. From our observation, we select the best dimensionality of AIM as $d_h = 40$ and report the corresponding results in the rest of our experiments. Moreover, these curves show that AIM is not very sensitive to the hidden dimensionality in a large range, and it can still outperform the compared methods even without the best dimensionality.

4.5 Performance Comparison on Early Detection of Misinformation

The early detection of misinformation is an important and practical task. We need to detect misinformation as early as possible. Thus, we can take immediate action at the beginning stage of the spread of misinformation and minimize the baneful influence. To investigate the performance of AIM on early detection of misinformation, we select the most competitive methods, i.e., GRU and SVM-TS, with highest accuracies according to Table 3 and illustrate their performance with varying detection deadlines in Figure 3. The performance is evaluated by accuracy, and $d_h = 40$. For reckoning the average reporting time over misinformation, conventional early detection tasks count on official announcements made by debunking services, e.g., Snopes and Sina Community Management Center. Thus, we take the official report time as a baseline.

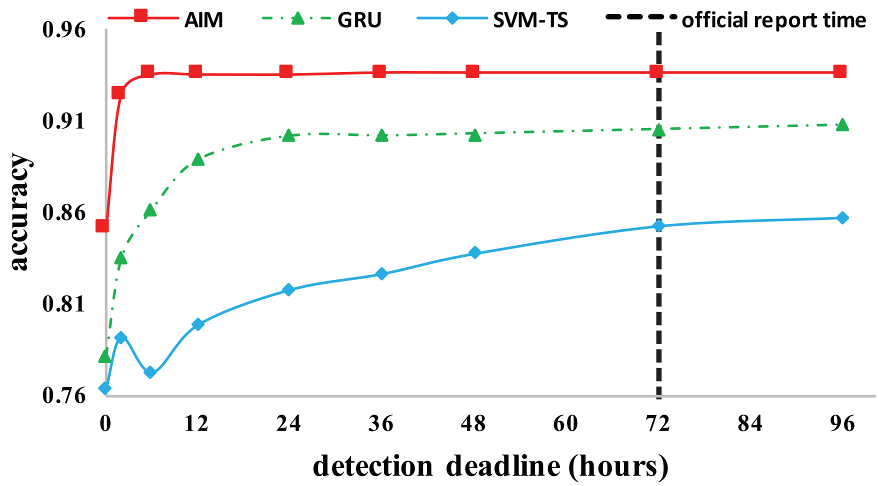
As shown in the figure, the accuracy of most methods will experience a conspicuous climb during the first few hours and then rise with different growth rates, convergence rates, and convergence accuracies. For instance, the accuracy curve of SVM-TS climbs slowly in the early phase and gradually converges to a relatively low accuracy. Moreover, its accuracy curve still fluctuates after the official report time. However, the accuracy curve of GRU climbs rapidly in the early phase and converges to a much higher accuracy on a much earlier deadline than that of SVM-TS.

The proposed AIM models can reach relatively high accuracy at a very early time, whereas other methods will take a longer time to achieve good performance. Furthermore, compared to the performance of GRU and SVM-TS, that of the proposed AIM model takes a relatively large lead at any phase. On the Weibo dataset, the accuracy of AIM can reach more than 90%, which is a very high accuracy and even larger than the performance of GRU on misinformation identification, in just 2 hours. At the official report time on the Weibo and Twitter datasets, the accuracy of AIM reaches about 93% and 77%, respectively. These experimental results show that the proposed AIM model is very practical for early detection of misinformation.

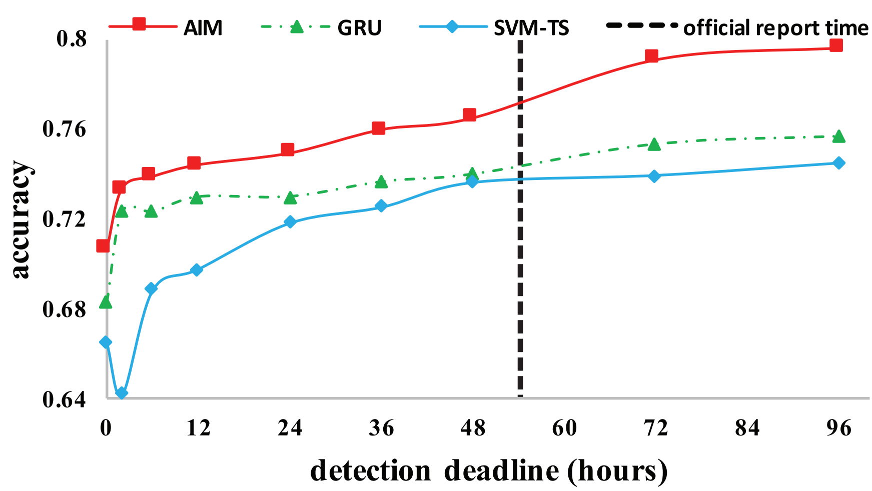
Most state-of-the-art methods for early detection, e.g., GRU and SVM-TS, usually follow the intuitive paradigm to model time-series features in sequences of microblog posts. But these time-series–based models are not qualified for practical early detection due to the conflict between the models and the task. For example, GRU requires, on one hand, that the input sequence of dynamic temporal signals should be long enough to be captured, whereas on the other hand, early detection of misinformation requires that the input sequence must be limited. The limited input sequence may not cover required dynamic temporal signals. Thus, GRU may not be suitable for early detection of misinformation in some cases. With the attention mechanism, AIM identifies misinformation with several significant microblogs. This minimizes the conflict between the models and the task, and makes AIM a naturally suitable model for early detection of misinformation.

5 VISUALIZATION

In this section, we present the visualization of the leaned attention mechanism for demonstrating the rationality of our proposed AIM model. First, we illustrate the curves of the learned dynamic attention values in AIM. Then, we pick several events, including both misinformation and true information, and illustrate microblogs with the largest weights in each event.



(a) Performances on the Weibo dataset.



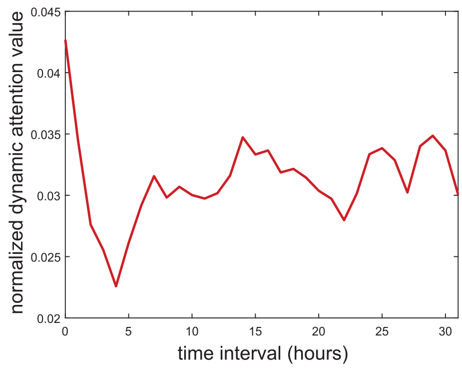
(b) Performances on the Twitter dataset.

Fig. 3. Performance of AIM, GRU, and SVM-TS on early detection of misinformation with $d_h = 40$. Results are evaluated by accuracy. The official report time indicates the average time required for publicly reporting misinformation on the platform.

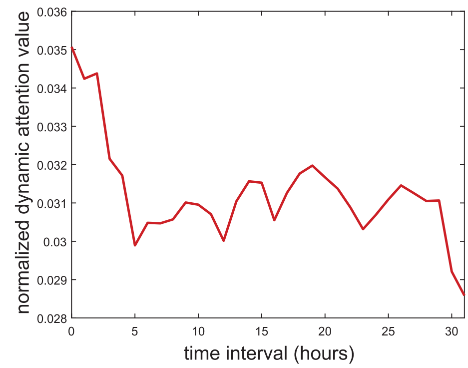
5.1 Visualization of Dynamic Attention

In Figure 4, we illustrate the learned dynamic attention values in AIM on the Weibo and Twitter datasets. The dynamic attention values for different time intervals since the beginning of an event are normalized. A higher dynamic attention value indicates that microblogs posted at the corresponding time interval have larger significance for misinformation identification.

From the curves on the Weibo and Twitter datasets, we can draw similar conclusions. At the very beginning of an event, we have the largest dynamic attention values. This means that the



(a) Normalized dynamic attention values on the Weibo dataset.



(b) Normalized dynamic attention values on the Twitter dataset.

Fig. 4. Illustration of the learned dynamic attention in AIM for different time intervals. The dynamic attention values are normalized. The larger the attention value, the higher the significance.

Table 4. Example 1 of Misinformation in the Weibo Dataset in Which Several Microblogs With the Largest Attention Weights Are Illustrated

Posting Time	Content
2014/03/20 23:55	Hearing from an Australian friend: The plane has been found in the international waters near Perth. It is proven to be MH370 according to a major component of the plane.
2014/03/22 01:25	Is it reliable?
2014/03/22 00:35	Reposting unreliable information, what an expert!
2014/03/22 00:49	Hopefully it's not true.
2014/03/22 09:48	This can't be true.
2014/03/22 00:04	Waiting for official confirmation tomorrow.
2014/03/22 07:33	Really???
2014/03/22 00:17	Let's watch the exact news tomorrow morning. Anyway, may God bless them!
2012/08/22 00:05	Is it true?

very first microblog and some early comments are very important for identifying misinformation. Then the attention values decrease rapidly until about the 5-hour mark. This may indicate that during this time period, people tend to repost and echo the message, and there is little useful information. Then the curve starts to increase and achieves another high level. This shows that at this moment, people start to think and express their own attitudes, which may include suspicion, affirmation, and denial. Finally, the curve tends to shock and then stay stable. Curves in Figure 4 demonstrate the rationality of the learned dynamic attention values in AIM and provide proofs for the contribution of dynamic attention on misinformation identification.

5.2 Microblogs With the Largest Weights

In Tables 4 through 9, we pick several events containing both misinformation and true information, and we illustrate several microblogs with the largest attention weights belonging to each event. Attention weights consist of both content attention and dynamic attention, as in Equations (1) and

Table 5. Example 2 of Misinformation in the Weibo Dataset in Which Several Microblogs With the Largest Attention Weights Are Illustrated

Posting Time	Content
2012/08/29 19:20	This mourning, 30 trucks full of coins arrived at Apple's headquarters. Samsung paid Apple one-billion fine, with 20 billion coins!
2012/08/29 19:22	Is it true?
2012/08/29 19:23	Is it true...
2012/08/29 19:36	What? You must be kidding! I wonder how Steve Jobs feels.
2012/08/29 22:38	How can they get such a great amount of coins?
2012/08/29 21:24	Is this rumor or humor? Laughing...
2012/08/30 01:54	How cheating this is!

Table 6. Example 3 of Misinformation in the Weibo Dataset in Which Several Microblogs With the Largest Attention Weights Are Illustrated

Posting Time	Content
2013/04/05 22:08	China Mobile will charge Weixin and Weibo since September 1st. 10 Yuan per 500 messages.
2013/04/05 22:09	Surprising... Reliable news?
2013/04/05 22:38	Is it true?
2013/04/05 22:12	If this is true, I will not use Weixin and Weibo anymore.
2013/04/06 08:24	Luckily, I'm a user of China Telecommunications.
2013/04/05 22:18	Luckily, I'm not a user of China Mobile.
2013/04/06 09:23	@China Mobile Any confirmation?
2013/04/06 16:33	What kind of logic this is!
2013/04/05 22:31	Surprising!

Table 7. Example 4 of Misinformation in the Weibo Dataset in Which Several Microblogs With the Largest Attention Weights Are Illustrated

Posting Time	Content
2012/08/11 14:23	At 8:20 this morning, a vicious explosion happened in Public Security Bureau of Jieshou City Anhui Province. Seven policemen died on the spot.
2012/08/11 17:09	Is it true?
2012/08/11 20:25	Any confirmation?
2012/08/11 15:30	Official confirmation is need.
2012/08/11 14:46	Any official confirmation?
2012/08/11 14:47	No picture, no truth!
2012/08/11 14:48	Friends in Anhui, explain the real situation.
2012/08/11 18:42	Really???

(2). From these examples, we can observe what types of microblogs can contribute to misinformation identification the most.

Tables 4 through 7 illustrate four examples of misinformation. Example 1 is a piece of misinformation about MH370, saying that it has crashed near Australia. Example 2 is an absurd joke about Samsung and Apple. Example 3 is a fake new policy of China Mobile. Example 4 talks about a fabricated explosion accidence. From these examples, it is clear that the original microblog of an

Table 8. Example 1 of True Information in the Weibo Dataset in Which Several Microblogs With the Largest Attention Weights Are Illustrated

Posting Time	Content
2015/11/28 16:27	Panda Panpan's "100"-year-old birthday! Would you like to send her a heartfelt blessing?
2015/11/28 16:30	Long live!
2015/11/28 16:49	Surprising! Happy birthday, Panpan.
2015/11/28 21:45	Happy birthday. Smiling...
2015/11/28 18:48	She's pretty!
2015/11/28 16:29	Happy birthday, Panpan.
2015/11/28 23:37	When I was little, I have worn a skirt with Panpan on it.
2015/11/28 17:50	Memory in childhood. Happy birthday.

Table 9. Example 2 of True Information in the Weibo Dataset in Which Several Microblogs With the Largest Attention Weights Are Illustrated

Posting Time	Content
2015/11/18 08:50	A thief was caught in a net bar, playing an online game. He asked the policemen to wait for a second, and said he couldn't implicate his teammates.
2015/11/18 09:28	Great teammate! HaHaHa...
2015/11/18 11:42	Great teammate!
2015/11/18 08:56	Laughing...
2015/11/18 19:30	The facial expression is so funny.
2015/11/18 09:18	If I were his teammate, I would be touched.
2015/11/18 09:19	Your teammate is so lucky!
2015/11/18 16:35	Policeman: I'll play for you. You can follow my colleagues.

event has very large attention weight. This is consistent with our intuition, because the original microblog details what has happened, which is important for understanding the event. Among other microblogs with large attention weights, most are about denying the event, questioning the real situation, suspecting the event, or sarcasm. Obviously, the attention mechanism in AIM can mine microblogs that are the most significant for misinformation. This makes AIM a reliable misinformation detection approach, which can alleviated noisy and useless information.

Tables 8 and 9 illustrate two examples of true information. Example 1 is about the birthday of a panda called *Panpan*. Example 2 is an incredible and funny story about a stupid thief. Although the second example is a little absurd and hard to believe, it is a piece of true information and has been correctly classified by AIM. As in Tables 4 through 7, the original microblog of true information has very large attention weight. However, other significant microblogs are different from those in Tables 4 through 7. Comments in Tables 8 and 9 mostly reflect the event or the news itself. Based on significant microblogs, the difference between misinformation and true information can easily be distinguished by AIM.

6 CONCLUSIONS AND FUTURE WORK

In this article, we propose a novel attention-based approach for misinformation identification on social media. The attention mechanism in our proposed AIM model consists of two parts: content attention and dynamic attention. Content attention is calculated based textual features of each microblog. Dynamic attention is related to the time interval between the posting time of a microblog and the beginning of the event. Via aggregation of the content attention and the dynamic attention, we can obtain the final attention weights for each microblog belonging to an event. A weighted sum of these microblogs can be performed to generate the final representation of the whole event. The experimental results on two real datasets, i.e., the Weibo and Twitter datasets, show that AIM can outperform the state-of-the-art methods. In addition, the visualization of the leaned attention mechanism in AIM illustrates the rationality of our proposed model.

In the future, we plan to explore the following directions. In AIM, multiple features, e.g., posted images and propagation structure, are not incorporated. Thus, we can incorporate more features in our model. Moreover, AIM is a static model, i.e., parameters in AIM do not change among different time periods. However, the trend of topics on social media is dynamic over time, and we usually have different hot topics in different time periods. Thus, it is necessary to incorporate time-aware or topic-aware mechanism in AIM.

REFERENCES

[1] Jimmy Ba, Volodymyr Mnih, and Koray Kavukcuoglu. 2015. Multiple object recognition with visual attention. In *Proceedings of the International Conference on Learning Representations*.

[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the International Conference on Learning Representations*.

[3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, New York, NY, 675–684.

[4] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2, 3, 27.

[5] Kan Chen, Jiang Wang, Liang-Chieh Chen, Haoyuan Gao, Wei Xu, and Ram Nevatia. 2015. ABC-CNN: An attention based convolutional neural network for visual question answering. arXiv:1511.05960.

[6] Jan K. Chorowski, Dzmitry Bahdanau, Dmitriy Serdyuk, Kyunghyun Cho, and Yoshua Bengio. 2015. Attention-based models for speech recognition. In *Proceedings of the Conference on Neural Information Processing Systems*. 577–585.

[7] Junyoung Chung, Caglar Gülçehre, Kyunghyun Cho, and Yoshua Bengio. 2015. Gated feedback recurrent neural networks. In *Proceedings of the International Conference on Machine Learning*. 2067–2075.

[8] Bhuwan Dhingra, Hanxiao Liu, William W. Cohen, and Ruslan Salakhutdinov. 2017. Gated-attention readers for text comprehension. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.

[9] Katherine Del Giudice. 2010. 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.

[10] Alex Graves. 2013. Generating sequences with recurrent neural networks. arXiv:1308.0850.

[11] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy. In *Proceedings of the 22nd International Conference on World Wide Web Companion*. 729–736.

[12] Tin Kam Ho. 1995. Random decision forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, Vol. 1. IEEE, Los Alamitos, CA, 278–282.

[13] Laurent Itti, Christof Koch, and Ernst Niebur. 1998. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 11, 1254–1259.

[14] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. 2014. News credibility evaluation on microblog with a hierarchical propagation model. In *Proceedings of the IEEE 14th International Conference on Data Mining*. IEEE, Los Alamitos, CA, 230–239.

[15] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. 2972–2978.

[16] Rudolf Kadlec, Martin Schmid, Ondrej Bajgar, and Jan Kleindienst. 2016. Text understanding with the attention sum reader network. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics Workshop*.

[17] Srijan Kumar, Robert West, and Jure Leskovec. 2016. Disinformation on the Web: Impact, characteristics, and detection of Wikipedia hoaxes. In *Proceedings of the 25th International World Wide Web Conference*. 591–602.

[18] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *Proceedings of the IEEE 13th International Conference on Data Mining*. IEEE, Los Alamitos, CA, 1103–1108.

[19] Quoc V. Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the International Conference on Machine Learning*, Vol. 14. 1188–1196.
[20] Qiang Liu, Shu Wu, Diyi Wang, Zhaokang Li, and Liang Wang. 2016. Context-aware sequential recommendation. In *Proceedings of the 2016 IEEE International Conference on Data Mining (ICDM'16)*. 1053–1058.
[21] Qiang Liu, Shu Wu, and Liang Wang. 2017. Multi-behavioral sequential prediction with recurrent log-bilinear model. *IEEE Transactions on Knowledge and Data Engineering* 29, 6, 1254–1267.
[22] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16)*. 194–200.
[23] Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing*. 1412–1421.
[24] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the International Joint Conference on Artificial Intelligence*. 3818–3824.
[25] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. 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*. ACM, New York, NY, 1751–1754.
[26] Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur. 2010. Recurrent neural network based language model. In *Proceedings of the 2010 INTERSPEECH Conference (INTERSPEECH'10)*. 1045–1048.
[27] Tomáš Mikolov, Stefan Kombrink, Lukáš Burget, Jan Černockỳ, and Sanjeev Khudanpur. 2011. Extensions of recurrent neural network language model. In *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'11)*. 5528–5531.
[28] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS'13)*. 3111–3119.
[29] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. 2014. Recurrent models of visual attention. In *Proceedings of the 27th Conference on Neural Information Processing Systems (NIPS'14)*. 2204–2212.
[30] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing*. 1589–1599.
[31] J. Ross Quinlan. 1987. Simplifying decision trees. *International Journal of Man-Machine Studies* 27, 3, 221–234.
[32] Soo Young Rieh, Grace YoungJoo Jeon, Ji Yeon Yang, and Christopher Lampe. 2014. Audience-aware credibility: From understanding audience to establishing credible blogs. In *Proceedings of the 8th International Conference on Weblogs and Social Media*.
[33] Alexander M. Rush, Sumit Chopra, and Jason Weston. 2015. A neural attention model for abstractive sentence summarization. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing*.
[34] Linlin Wang, Zhu Cao, Gerard de Melo, and Zhiyuan Liu. 2016. Relation classification via multi-level attention CNNs. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.
[35] Yequan Wang, Minlie Huang, Li Zhao, and Xiaoyan Zhu. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'16)*.
[36] Shu Wu, Qiang Liu, Yong Liu, Liang Wang, and Tieniu Tan. 2016. Information credibility evaluation on social media. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence*. 4403–4404.
[37] Tianjun Xiao, Yichong Xu, Kuiyuan Yang, Jiaxing Zhang, Yuxin Peng, and Zheng Zhang. 2015. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 842–850.
[38] Huijuan Xu and Kate Saenko. 2016. Ask, attend and answer: Exploring question-guided spatial attention for visual question answering. In *Proceedings of the European Conference on Computer Vision*. 451–466.
[39] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. 2015. Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the International Conference on Machine Learning*. 2048–2057.
[40] Fan Yang, Yang Liu, Xiaohui Yu, and Min Yang. 2012. Automatic detection of rumor on Sina Weibo. In *Proceedings of the SIGKDD Workshop on Mining Data Semantics*. 13.
[41] Zichao Yang, Xiaodong He, Jianfeng Gao, Li Deng, and Alex Smola. 2016. Stacked attention networks for image question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 21–29.
[42] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 1480–1489.

[43] Wenpeng Yin, Sebastian Ebert, and Hinrich Schütze. 2016. Attention-based convolutional neural network for machine comprehension. arXiv:1602.04341.

[44] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*. 1395–1405.

[45] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 207.

Received July 2017; revised October 2017; accepted December 2017