Multi-modal Knowledge-aware Event Memory Network for Social Media Rumor Detection

Huaiwen Zhang[1,2], Quan Fang[1,2], Shengsheng Qian[1,2], Changsheng Xu[1,2,3]
[1]National Lab of Pattern Recognition, Institute of Automation, CAS, Beijing 100190, China
[2]University of Chinese Academy of Sciences
[3]Peng Cheng Laboratory, ShenZhen, China
{huaiwen.zhang,qfang,shengsheng.qian,csxu}@nlpr.ia.ac.cn

ABSTRACT

The wide dissemination and misleading effects of online rumors on social media have become a critical issue concerning the public and government. Detecting and regulating social media rumors is important for ensuring users receive truthful information and maintaining social harmony. Most of the existing rumor detection methods focus on inferring clues from media content and social context, which largely ignores the rich knowledge information behind the highly condensed text which is useful for rumor verification. Furthermore, existing rumor detection models underperform on unseen events because they tend to capture lots of event-specific features in seen data which cannot be transferred to newly emerged events. In order to address these issues, we propose a novel Multi-modal Knowledge-aware Event Memory Network (MKEMN) which utilizes the Multi-modal Knowledge-aware Network (MKN) and Event Memory Network (EMN) as building blocks for social media rumor detection. Specifically, the MKN learns the multi-modal representation of the post on social media and retrieves external knowledge from real-world knowledge graph to complement the semantic representation of short texts of posts and takes conceptual knowledge as additional evidence to improve rumor detection. The EMN extracts event-invariant features of events and stores them into global memory. Given an event representation, the EMN takes it as a query to retrieve the memory network and output the corresponding features shared among events. With the additional information provided by EMN, our model can learn robust representations of events and consistently perform well on the newly emerged events. Extensive experiments on two Twitter benchmark datasets demonstrate that our rumor detection method achieves much better results than state-of-the-art methods.

CCS CONCEPTS

• **Information systems → Data cleaning**;

KEYWORDS

Social Media, Rumor Detection, Multi-Modal, Knowledge Graph, Memory Network

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10...$15.00
https://doi.org/10.1145/3343031.3350850

ACM Reference Format:
Huaiwen Zhang, Quan Fang, Shengsheng Qian, Changsheng Xu. 2019. Multi-modal Knowledge-aware Event Memory Network for Social Media Rumor Detection. In *Proceedings of the 27th ACM International Conference on Multimedia (MM'19), October 21–25, 2019, Nice, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3343031.3350850

Figure 1: An example of multi-modal tweets.

1 INTRODUCTION

Social media has revolutionized the way people acquire information. Unfortunately, it also has fostered various false rumors which usually contain misrepresented or even forged multimedia content, to mislead the readers and get rapid dissemination. The widespread social media rumors may cause negative effects. Some offenders even use rumors to guide public opinion, damage the credibility of the government and even interfere with the general election [1]. Therefore, it is urgent to detect and regulate social media rumors for ensuring users receive truthful information and maintaining social harmony.

To debunk rumors and minimize their harmful effects, many efforts have been made. Fact-checking websites such as snopes.com and politifact.com rely on manual efforts to track and debunk rumors, which has obvious limitations on efficiency and coverage. To automatically detect rumors, early approaches train supervised classifiers learned on various hand-crafted features generated from the media content of posts, and the social context of users [2, 17, 18, 22, 24]. Inspired by the great success of the neural network, various deep learning models are proposed recently and demonstrate state-of-the-art detection performances. For example, Ma et al. [23] introduced recurrent neural networks (RNNs) to learn the hidden representations from text content of relevant posts regarding given claims for detecting rumors. Yu et al. [42] used convolutional neural networks (CNNs) for obtaining key features and their high-level interactions from text content of the relevant posts. Chen et al. [4] imported the attention mechanism (Attentions) to learn selectively temporal hidden representations of sequential posts for identifying rumors.

However, existing deep learning models typically focus on inferring clues from short text content of relevant posts, and largely ignore the visual data and background knowledge of posts which humans also use in judging the credibility of an event [29]. For example, to evaluate the credibility of the post in Figure 1, people usually look at the picture first and then read the text content and realize that the *R.E.M.* is an organization, and *Donald Trump, Ted Cruz* are politicians, and give judgment at the end. This indicates that social media posts share a great deal of latent knowledge-level connections, which benefits the rumor identification. How to acquire the background knowledge of the short text, and fuse the text, knowledge and image data of the post reasonably is the key to the rumor event representation. Furthermore, unlike other classification tasks such as text classification [10], the rumor detection task has a unique challenge that is detecting rumors on newly emerged events which the model may never see [32]. In the real scenario, it is crucial to detect rumors as early as possible so that interventions can be conducted in time. Hence, a rumor detection model should have the ability of the detection of unseen event rumors. Generally, social media rumors could exhibit certain high-level shared patterns that are irrelevant to event-specific semantics, which facilitates the detection of rumor from unverified posts [36]. Therefore, how to learn transferable feature representations among social media rumors is essential to newly emerged rumor recognition. To the best of our knowledge, the idea to unify multi-modal features and latent knowledge connections of posts as well as event-invariant feature modeling in a principled way for rumor identification is unexplored and challenging.

In this paper, we aim to tackle the above issues by introducing a novel end-to-end framework referred to as Multi-modal Knowledge-aware Event Memory Network (MKEMN) for social media rumor detection. The main advantages of MKEMN are: 1) To capture the full semantic meaning of the post, we propose the Multi-modal Knowledge-aware Network (MKN) which generates the post representation by treating word embedding, visual embedding, and knowledge embedding of the post as multiple stacked channels just like colored images while explicitly keeping their alignment relationships. 2) To extract the event-invariant features and enhance the ability of rumor detection model, we present the Event Memory Network (EMN) which builds an external memory shared during the whole training precess to capture the event-independent latent topic information of events. MKEMN processes event posts to obtain multi-modal knowledge-aware representation and the event-invariant features to form the event representation, which is finally fed into a deep neural network (DNN) for rumor detection. Extensive experiments on two Twitter benchmark datasets [23, 45] demonstrate that our rumor detection method outperforms several state-of-the-art methods.

The main contributions of our paper are three-fold:

- We propose a multi-modal knowledge-aware network, which exploits the multi-modal content and the external knowledge-level connections for accurate rumor identification.
- The proposed event memory network uses memory network to measure the dissimilarities among different events, and further learns the event invariant features which can generalize well for the newly emerged events.
- We experimentally demonstrate that our model is more robust and effective than state-of-the-art baselines based on two public benchmark datasets for the tasks of rumor detection on Twitter.

2 RELATED WORK

In this section, we briefly review the work related to the proposed model. We mainly focus on the following three topics: rumor detection, entity linking and memory networks.

2.1 Rumor Detection

Social psychology literature defines a rumor as a story or a statement whose truth value is unverified or deliberately false [6]. Most existing studies solve the automatic rumor detection problem in feature-based approaches. Based on the source of features, we roughly summarize existing models into two categories: the media content features, social context features. Existing models [2, 4, 23, 42] mainly extract the textual features from text content of posts. For example, Castillo et al. [2] extract message-based features and topic-based features from text content and classify the post by decision tree. Ma et al. [23] introduce recurrent neural networks (RNNs) to learn the hidden representations from text content of relevant posts. Yu et al. [42] use convolutional neural networks (CNNs) to obtain key features and their high-level interactions from text content of the relevant posts. Few approaches [14, 36] are conducted on rumor detection based on the multimedia content. Wang et al. [36] simply concatenate the visual features and textual features of post to get a multi-modal features and identify rumors. Jin et al. [14] propose a recurrent neural network with an attention mechanism to fuse image and text features of the post for rumor detection. Social context features are extracted from the social network of users which represent the engagements of post on social media. Ma et al. [24] utilize the times series of social content information in Weibo to verify the credibility of the post. Wu et al. [38] aim to capture propagation patterns such as graph structure of the message propagation in social media to detect rumors.

Different from all the aforementioned work, in this paper, we extend the media content based methods with a novel features: the background knowledge behind the highly condensed text, and consider multiple types of features simultaneously when identifying rumors on social media.

2.2 Entity Linking

Entity linking is the task of assigning entity mentions in a text to corresponding entries in the knowledge graph. However, an entity mention could possibly denote different named entities, and a named entity may have multiple surface forms, such as its aliases, abbreviations, and alternate spellings. An entity linking system has to disambiguate the entity mention in the textual context and identify the mapping entity for each entity mention. Recently, the entity linking research has largely focused on two types of contextual information for disambiguation: local information [19, 40] based on words that occur in a context window around an entity mention, and global information [7–9, 12] exploiting document-level coherence of the referenced entities. Many state-of-the-art methods aim to combine the benefits of both [3, 30].

2.3 Memory Network

Memory network [26, 35, 37] is a recurrent attention model that utilizes an external memory module for question answering and language modeling. Generally, the memory network consists of two components: an external memory and a controller which perform operations on the memory including reading and writing. The memory component increases model capacity and provides an internal representation of knowledge The iterative reading operations on the external memory enable memory networks to extract features shared among document. Researchers have also proposed to use memory networks in other tasks such as sentiment classification [21] and recommendation [13]. Note that these approaches usually focus on entry-level or sentence-level memories, while our work addresses the global event-level memory which shares across the training data. In addition, unlike traditional memory networks, we treat the input memory and output memory as the same one.

3 PROBLEM STATEMENT

Rumor detection task can be defined as a binary classification problem, which aims to classify a claim in social media as a rumor or non-rumor. There are two types of claims in social media rumor detection community: the post-level work [14, 15, 38] identifies whether a single post of social media is a rumor or not, while the event-level work [4, 11, 23, 25, 31, 42] conducts rumor detection on a group of posts which constitute an event.

The task of this paper is to identify whether a claim is a rumor or not at the event-level, where each claim comprises a sequence of correlative posts (reposts and comments) and each post is associated with a timestamp. Specifically, we represent an event-level claim as $x = (y, [p_1, p_2, \cdots, p_T])$ where $y \in \{0, 1\}$ is the ground-truth label of the claim (i.e., Non-rumor or Rumor) and $[p_1, p_2, \cdots, p_T]$ is a sequence of relevant posts, where each post p_t consists of text s_t and image v_t. Formally, rumor detection on event-level aims to learn a projection $F(X) \rightarrow \{0, 1\}$.

4 MULTI-MODAL KNOWLEDGE-AWARE EVENT MEMORY NETWORK

In this section, we present the proposed Multi-modal Knowledge-aware Event Memory Network (MKEMN) in detail. We first introduce the overall framework of MKEMN which contains Multi-modal Knowledge-aware Network (MKN) and Event Memory Network (EMN) to conduct rumor detection on social media. Then, we detail the architecture of MKN which can flexibly fuse the multi-modal content and the background knowledge into sentence representation learning in Section 4.2, and the design of the event memory network which can capture the features shared among events in Section 4.3.

4.1 MKEMN framework

The framework of MKEMN is illustrated in Figure 2. We describe the architecture of MKEMN from the left to right. As shown in Figure 2, MKEMN takes an event (a set of posts) as input.

- For each post, MKN processes its text content, image content and background knowledge, and aligns them as different input channels of CNN to incorporate them into post representation learning. Details can be found in Section 4.2.
- To get the robust representations of the events, we use an event memory network to extract and store the event-invariant features. The EMN takes the event vector as a query to retrieve memory network and output the corresponding features shared among events. The details of EMN are presented in Section 4.3
- The final multi-modal knowledge-aware representation and event-invariant feature of event are fed into a deep neural network (DNN) to calculate the predicted probability to determine whether the event is a rumor or not.

4.2 Multi-modal Knowledge-aware Network

To learn the representation of a post, most of the existing models treat it as the text content representation, and largely ignore the visual content and knowledge information which are useful for rumor detection. In this section, we propose the multi-modal knowledge-aware network, which takes the text content, visual content and knowledge information of post as input, and incorporates them into the post representation efficiently. As shown in Figure 2, the MKN consists of four parts: the text encoder, the knowledge encoder, the visual encoder, and the final multi-modal knowledge-aware CNN.

4.2.1 Text Encoder. Text encoder module aims to produce the short text representation for a given short text s. After word embedding, the sentence s will be projected to a sequence of word vectors: $s = [w_1, w_2, \cdots, w_n]$, where $w_i \in \mathbb{R}^{d_w}$ is the vector for the i-th word in the sentence, n is the word length and d_w is the dimension. Then, we feed them into a Bidirectional GRU [5] to capture the contextual information of the sequence:

$$\begin{aligned} \overrightarrow{h_t} &= GRU(w_t, \overrightarrow{h_{t-1}}) \\ \overleftarrow{h_t} &= GRU(w_t, \overleftarrow{h_{t+1}}) \end{aligned} \tag{1}$$

We concatenate each $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ to obtain a hidden state h_t. Let the hidden unit number for each unidirectional GRU be $\frac{d_h}{2}$. For simplicity, we denote all the h_t as $H \in \mathbb{R}^{n \times d_h}$:

$$H = (h_1, h_2, \ldots h_n) \tag{2}$$

where the d_h is the dimension of hidden state.

4.2.2 Knowledge Encoder. Knowledge encoder module aims to distill background knowledge from real-world knowledge graph to complement the semantic representation of short texts of posts.

As illustrated in Figure 3, the process of knowledge distillation consists of three steps. First, to distinguish knowledge entities in the text content of post $s = [w_1, w_2, \cdots, w_n]$, we utilize the technique of entity linking [16, 20, 28] to disambiguate mentions in texts by associating them with predefined entities in a knowledge graph.

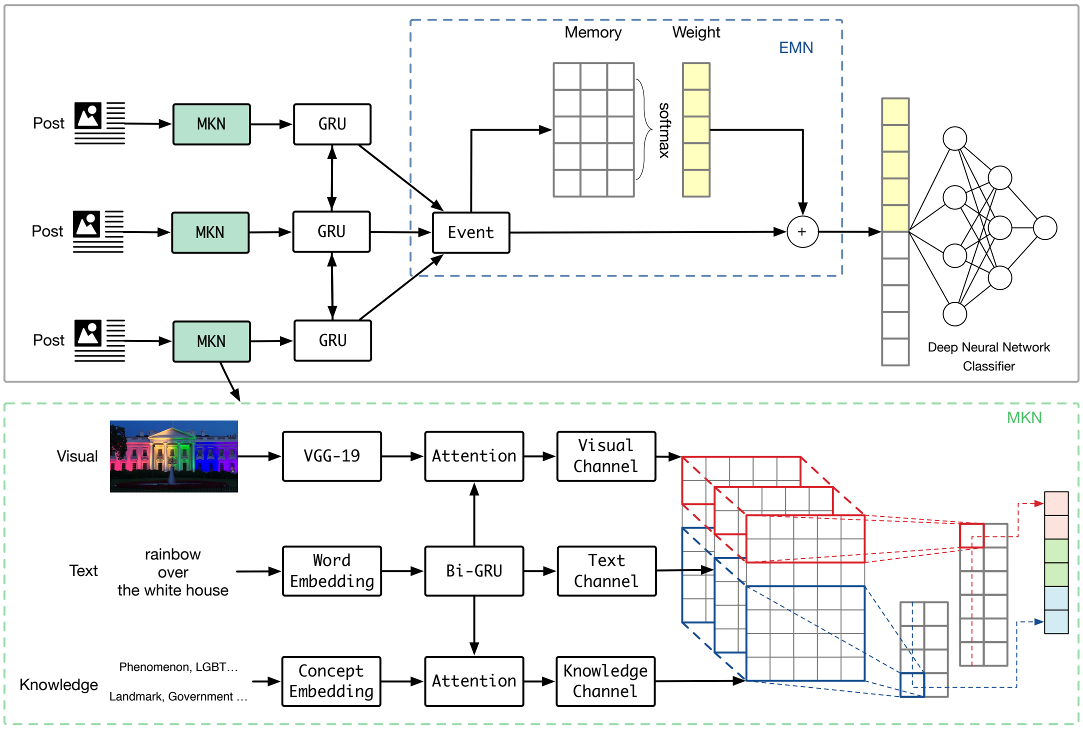

Figure 2: Illustration of the Multi-modal Knowledge-aware Event Memory Network (MKEMN).

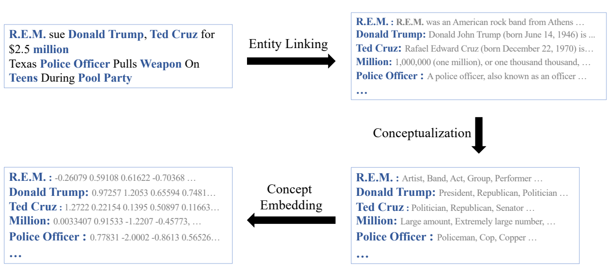

Figure 3: Illustration of knowledge distillation process.

Then, for each identified entity $e \in E_s$, we acquire its conceptual information $C_e = (c_e^1, c_e^2, \ldots, c_e^m)$ from existing knowledge graphs [34, 39] and taxonomies [27, 43], where c_e^i is the i-th related concept of entity e and m is the concept set size of entity e. This paper takes *isA* relation as an example, and other semantic relations such as *isPropertyOf* can also be applied in a similar way.

For instance, given a text “R.E.M. sue Donald Trump, Ted Cruz for \$2.5 million”, we obtain the entity set E={R.E.M., Donald Trump, Ted Cruz} by entity linking. Then, we conceptualize the entities in E and acquire its concept set $C_{\text{R.E.M.}}$ = {artist, band, act, group, performer}, $C_{\text{Donald Trump}}$= {president, republican, politician}, $C_{\text{Ted Cruz}}$ = {republican, politician} from external knowledge graphs.

Given the conceptual information C_e for entity e, we aims to producing the concept knowledge vector $k_e \in \mathbb{R}^{d_k}$ for it, where d_k is the dimension of the concept embedding. To reduce the negative effects of some inappropriate concepts introduced due to the fuzziness of entities or the noise in knowledge graphs, we propose concept attention to measure the semantic similarity between the i-th concept c_e^i and the word representation h_t. We use the following formula to calculate the attention:

$$\alpha_t^i = \text{softmax}(\mathcal{F}(W_c[c_e^i \oplus h_t] + b_c)) \tag{3}$$

The attention network $\mathcal{F}$ receives embeddings of current word and the corresponding concept as input and outputs the impact weight, where $W_c \in \mathbb{R}^{d_k + d_h}$ is a learnable weight matrix, $\oplus$ is the concatenation operator and b_c is the offset. The α_t^i denotes the weight of attention from the i-th concept towards the word representation h_t. A larger α_t^i means that the word w_t is more semantically similar to the i-th concept. In the end, the final attention weights are employed to calculate a weighted sum of the concept vectors, resulting in a semantic vector that represents the concepts k_t[1]:

$$k_t = \sum_{i=1}^{m} \alpha_t^i c_e^i \tag{4}$$

4.2.3 Visual Encoder. Visual Encoder take the visual content of the post as input and produces the weighted visual features of the post. Given a picture, we first resize the images to 224×224 pixels, then feed it to the 19-layer VGGNet [33] to extract the image features. We chose the features from the last pooling layer which has a dimension of $512 \times 7 \times 7$ as in most time a word is related to a small region of the input image.

[1]For the word w_t which not identified as an entity, we set the k_t to 0.

Therefore, an image could be represented as :

$$\tilde{v} = \{\tilde{v}_i | \tilde{v}_i \in \mathbb{R}^{d_v}., i = 1, 2, \cdots, N\} \tag{5}$$

where $\tilde{v}_i$ is the feature vector of image region i, and $d_v = 512$ is the dimension of $\tilde{v}_i$ and $N = 7 \times 7$ is the number of image regions.

To filter out the noise and pinpoint the regions $\tilde{v}_i \in \mathbb{R}^{d_v}$ that are highly relevant to the current word, we apply a word-guided visual attention module. Given a word w_t, the feature h_t can be obtained by Equation 2 and the image feature matrix $\tilde{v}$ is obtained by Equation 5, we feed them through a single layer neural network followed by a softmax function to generate the attention distribution over the N regions of the image:

$$\beta_t^i = \text{softmax}(\mathcal{G}(W_{\tilde{v}[\tilde{v}_i \oplus h_t] + b_{\tilde{v}}})) \tag{6}$$

where $\mathcal{G}$ is the attention network, $W_{\tilde{v}} \in \mathbb{R}^{d_h + d_v}$ is the trainable parameters, and β_t^i corresponds to the attention probability of image region $\tilde{v}_i$ given the word feature h_t. Based on the attention distribution β_t , which is the weight corresponding to each image region, the new image vector related to word h_t can be obtained by:

$$v_t = \sum_i^N \beta_t^i \tilde{v}_i \tag{7}$$

4.2.4 Multi-modal knowledge-aware CNN. After the text encoding, knowledge encoding and image encoding, each word w_t will be associated with an word representation $h_t \in \mathbb{R}^{d_h}$, an knowledge embedding $k_t \in \mathbb{R}^{d_k}$ and the corresponding visual content embedding $v_t \in \mathbb{R}^{d_v}$, where d_h, d_k, d_v are the dimensions of word embedding, knowledge embedding and visual embedding, respectively.

Given the input above, a straightforward way to combine words and associated knowledge and visual data is to concatenate them to the word sequence, i.e.,

$$W = [h_1, h_2, \cdots, h_n, k_1, k_2, \cdots, k_n, v_1, v_2, \cdots, v_n] \tag{8}$$

However, this simple concatenating strategy has the following limitations: 1) The concatenating strategy breaks up the connection among words and associated knowledge and visual data, and is unaware of their alignment. 2) Word embedding, knowledge embedding and visual embedding are learned by different methods, meaning it is not suitable to process them together in a single vector space. 3) The concatenating strategy implicitly forces word, knowledge and visual embeddings to have the same dimension, which may not be optimal in practical settings since the optimal dimensions for representation dimensions may differ from each other.

Being aware of the above limitations, we propose a multi-channel and word-knowledge-visual-aligned CNN for combining all the information of post p_t. The architecture of CNN is illustrated in the left lower part of Figure 2. For each text sentence s, in addition to using its word embeddings $h_{1:n} = [h_1, h_2, \ldots, h_n]$ as input, we also introduce the transformed knowledge embeddings: $\mathcal{H}_k(k_{1:n}) = [\mathcal{H}_k(k_1), \mathcal{H}_k(k_2), \ldots, \mathcal{H}_k(k_n)]$ and transformed visual embeddings $\mathcal{H}_v(v_{1:n}) = [\mathcal{H}_v(v_1), \mathcal{H}_v(v_2), \ldots, \mathcal{H}_v(v_n)]$ as source of input , where $\mathcal{H}_k$ and $\mathcal{H}_v$ are the continuous transformation function which it can map the knowledge embeddings and visual embeddings from their original space to the word space while preserving their original spatial relationship.

Note that the word embeddings $h_{1:n}$, transformed knowledge embeddings $\mathcal{H}_k(k_{1:n})$ and transformed context embeddings $\mathcal{H}_v(v_{1:n})$ are the same size, and act as multiple channels similar to color images.

$$G = \begin{bmatrix} h_1, & h2, & \cdots & h_n \\ \mathcal{H}_k(k_1) & \mathcal{H}_k(k_2) & \cdots & \mathcal{H}_k(k_n) \\ \mathcal{H}_v(v_1) & \mathcal{H}_v(v_2) & \cdots & \mathcal{H}_v(v_n) \end{bmatrix} \in \mathbb{R}^{3 \times n \times d_h} \tag{9}$$

After getting the multi-channel input G, we apply multiple filters $f \in \mathbb{R}^{3 \times l \times d_h}$ with varying window sizes l to extract specific local patterns in the sentence. The local activation of sub-matrix $G_{i:i+l-1}$ with respect to f can be written as

$$r_i^f = \mathcal{I}(f * G_{i:i+l-1} + b) \tag{10}$$

where $\mathcal{I}$ is the activation function, and we use a max-over-time pooling operation on the output feature map to choose the largest feature:

$$\tilde{r}^f = \max\{r_1^f, r_2^f, \cdots, r_{n-l+1}^f\} \tag{11}$$

All features $\tilde{r}^{f_i}$ are concatenated together and taken as the final representation p_t of the post, i.e.,

$$p_t = [\tilde{r}^{f_1}, \tilde{r}^{f_2}, \cdots, \tilde{r}^{f_j}] \tag{12}$$

where j is the number of filters.

4.3 Event Memory Network

Given an event x whose posts representation list are $[p_1, p_2, \cdots, p_T]$, existing rumor detection methods [23] learn the representation of event by feeding this sequence of posts in to a post-level GRU, and take the hidden output of the last step as the event representation x:

$$x = GRU(p_T, h_{T-1}^p) \tag{13}$$

where h_{t-1}^p is the hidden state of post-level GRU at $T-1$ time step. However, these models underperform on unseen events because they tend to capture lots of event-specific features in seen data which cannot be transferred to newly emerged events.

For this reason, we present an event memory network to extract event-invariant features of events and stores them into global memory. As shown in Figure 2, the architecture of EMN mainly consists of two parts: 1) the event representation x and 2) the memory $M \in \mathbb{R}^{d_m \times K}$, an external memory shared during the whole training process which can capture the internal latent topic information of events, the d_m is vector dimension of each latent topic, and K is number of latent topic cluster. To extract the features shared among event, given the event representation x as query, we construct the memory query process as follows:

$$\begin{aligned} q_k &= \text{softmax}(x^T M_k) \\ Q_K &= \textstyle\sum_{k=1}^K q_k M_k \\ X &= x \oplus Q_K \end{aligned} \tag{14}$$

First, the similarity between the query x and each latent topic memory is calculated by dot-product. Then the resulting K values are normalized by the softmax function to produce a similarity probability q_k . After calculating the latent topic probability q_k , the event-invariant features $Q_K \in \mathbb{R}^{d_m}$ can be retrieved from summing

Table 1: Results of comparison among different models on TWITTER and PHEME Dataset.

Dataset	Methods	Accuracy	Rumor			Non-Rumor		
			Precision	Recall	F_1	Precision	Recall	F_1
TWITTER	DTC	0.687	0.667	0.702	0.684	0.665	0.700	0.682
	SVM-TS	0.731	0.735	0.730	0.733	0.744	0.720	0.732
	EANN	0.777	0.762	0.783	0.772	0.772	0.803	0.787
	CNN	0.810	0.807	0.820	0.813	0.806	0.813	0.809
	GRU	0.815	0.826	0.812	0.819	0.814	0.826	0.820
	CallAtRumors	0.824	0.815	**0.862**	0.838	0.823	0.863	0.841
	MKEMN	**0.866**	**0.893**	0.847	**0.870**	**0.866**	**0.867**	**0.866**
PHEME	DTC	0.548	0.581	0.536	0.558	0.577	0.523	0.548
	SVM-TS	0.634	0.667	0.611	0.638	0.633	0.638	0.635
	EANN	0.620	0.614	0.644	0.629	0.613	0.636	0.624
	CNN	0.651	0.655	0.631	0.643	0.663	0.644	0.653
	GRU	0.736	0.729	0.739	0.734	0.745	0.714	0.729
	CallAtRumors	0.773	0.776	0.771	0.773	0.751	0.776	0.763
	MKEMN	**0.816**	**0.809**	**0.819**	**0.814**	**0.816**	**0.801**	**0.809**

over $M \in \mathbb{R}^{d_m \times K}$ weighted by the q_k . Then we concatenate this feature vector Q_K with the original encoding vector x to generate the new event representation X. Note that, if the dimension size of x is different from that of memory vector, additional output projection layer should be placed after x before applying dot-product to the memory.

4.4 Deep Neural Network Classifier

Finally, given the memory-enhanced event representation X,the probability of this event belonging to rumor or not can be predicted by the DNN $\mathcal{D}$:

$$z = \mathcal{D}(X) \tag{15}$$

For a batch of event $\mathcal{X} = [X_1, X_2, \cdots, X_J]$ and their class label $y = [y_1, y_2, \cdots, y_J]$, the loss function can be written as follow:

$$\mathcal{L} = \sum_{j=1}^{J} -[y_j \times \log(z_j) + (1 - y_j) \times \log(1 - z_j)] \tag{16}$$

5 EXPERIMENTS AND RESULTS

5.1 Experimental Setup

5.1.1 Datasets. Two public dataset TWITTER [23] and PHEME [45] are used to validate the effectiveness of the proposed MKEMN on detecting social media rumors. Both datasets are constituted by tweets of Twitter. The TWITTER dataset are collected based on the claims reported on snopes.com, where each claim contains a sequence of tweets as the Section 3 described. The PHEME dataset are collected based on 5 breaking news, and each news contains a set of claims. Following the previous work, we filter out claims with less than 10 tweets and balance the number of instances of the two classes.

5.1.2 Implementation Details. In the knowledge distillation process, we acquire an entity set E of a short text s by leveraging the existing entity linking solutions [3]. For conceptualization, we retrieve concepts in Probase and YAGO and only consider the isA relation. The pre-trained Glove word vector on twitter data is used to initialize the word embedding, and the embedding size is set to 25. The hidden size of GRU in text encoder is 64, hence the d_h is 128. The concept embedding c_i is initialized by the average of the word embeddings of the word it contains, which means d_e is 25. We implement the transform function $\mathcal{H}_k$ and $\mathcal{H}_v$ by the non-linear function $Tanh(Wx + b)$. The kernel size of multi-modal knowledge-aware CNN is $[1, 2, 3, 4]$ and the kernel number is 25. For the topic memory, we set the cluster number to 2 and 3 for TWITTER and PHEME datasets, respectively. The memory size is set to 128.

We hold out 10% of the claims in each dataset for tuning the hyper parameters, and for the rest of the claims, we conduct 5-fold cross-validation and use accuracy, precision, recall and F-measure as evaluation metrics.

5.1.3 Comparison Models. We make comparisons among the following models:

DTC [2]: a decision tree classifier for modeling Twitter information credibility using various hand-crafted features.

SVM-TS [24]: utilizes a linear SVM to classify rumors on twitter and uses time-series structure to model the social feature variations

GRU [23]: uses a multilayer generic GRU network to model the microblog event as a variable-length time series, which is effective for early detection of rumors.

CNN [41]: uses a convolution neural network to learn rumor representations by framing the relevant posts as fixed-length sequence.

EANN [36]: a post-level rumor detection model which aims to classify a single post as rumor or not by leveraging both the textual and visual information of post, and utilizing an adversarial method to remove event-specific features from post representation. To apply EANN to the event-level situation, we run the EANN for every post of an event and identify the event as rumor or not by voting.

CallAtRumors [4]: presents an LSTM model to automatically identify rumors. By using the standard attention mechanism at word level, this method could detect rumors effectively.

MKEMN: the proposed method in this paper, which considers the multi-modal and knowledge data to learn the post representation and uses EMN to extract the event-invariant features for rumor detection.

5.2 Result and Analysis

5.2.1 Comparison with Existing Methods. Table 1 shows the performance of all the compared models based on the two datasets. From Table 1, we can draw the following observations:

- The DTC and SVM-TS models perform the worst among all methods. This is because they are built with the hand-crafted features, which are weak and have limited rumor detection utility.
- EANN performs better than hand-crafted models in TWITTER dataset, but achieves slightly lower results in PHEME dataset. This is probably because the event-level rumor detection task has more challenges than the post-level task, as all the events of both datasets come from the real-world news, and the posts within an event have the similar topics and complex interaction.
- GRU and CNN have comparable performance in TWITTER Dataset due to the superior ability of feature extraction of deep learning models. Besides, in PHEME dataset GRU achieves a better result. This suggests that the recurrent networks can inherently deal with variable-length sequence of posts, while CNN may need more data to make judgement.
- CallAtRumors achieves better results in all baselines. This is because CallAtRumors employs the attention mechanism and can better extract the specific local patterns in posts.
- Compared with all the baselines, the MKEMN has achieved the best performance and outperformed other rumor detection methods in most cases. We attribute the superiority of MKEMN to its two properties: 1) MKEMN uses the multi-modal knowledge-aware network for post representation, which could better capture the multi-channel semantic information. 2) MKEMN employs the event memory network which provides the event-invariant features to help the model consistently perform well on the newly emerged events.

5.2.2 Comparison among MKEMN variants. In this section, we compare among the variants of MKEMN with respect to the following two aspects to demonstrate the effectiveness of the design of the MKEMN framework: the usage of event memory network, the usage of multi-modal and knowledge. The following variants of our model are designed for comparison.

MKN: the variant of MKEMN, which removes the EMN module and learns the events representations by GRU.

Non-Knowledge MKN: the variant of MKN, which removes the knowledge information when model the post.

Non-Visual MKN: the variant of MKN, which removes the visual information when model the post.

The results are shown in Table 2, and Table 3 from which we can conclude that:

- MKN with EMN (a.k.a MKEMN) is better than MKN. The results prove that the event-invariant features extracted by EMN can help the model to consistently perform well on the newly emerged events.
- Compared with the best baseline, the usage of visual embedding (Non-Knowledge MKN) and knowledge embedding (Non-Visual MKN) can improve the accuracy by 0.5% and 1.2%, in TWITTER, respectively, and we can achieve even better performance by combining them together (MKE). This finding confirms the benefit brought by using visual and knowledge in our model.
- The Non-Knowledge MKE performs poorly than CallAtRumors in PHEME, which may be because the visual data in 5 breaking news of PHEME have highly similarities with each other.
- The models equipped with knowledge information achieve consistently better results than Non-Knowledge models in both datasets, which indicates that the knowledge information is an important kind of complementary information for rumor detection.

Table 2: Comparison among MKEMN variants on Twitter dataset, where R is rumor and N is Non-rumor.

Methods	Class	Accuracy	Precision	Recall	F_1
MKEMN	R	0.866	0.893	0.847	0.870
	N		0.866	0.867	0.866
MKN	R	0.841	0.85	0.833	0.841
	N		0.818	0.854	0.836
Non-Knowledge MKN	R	0.829	0.85	0.822	0.835
	N		0.813	0.834	0.823
Non-Visual MKN	R	0.836	0.834	0.846	0.841
	N		0.838	0.826	0.832
CallAtRumors	R	0.824	0.815	0.862	0.838
	N		0.823	0.863	0.841

Table 3: Comparison among MKEMN variants on PHEME dataset, where R is rumor and N is Non-rumor.

Methods	Class	Accuracy	Precision	Recall	F_1
MKEMN	R	0.816	0.809	0.819	0.814
	N		0.816	0.801	0.809
MKN	R	0.808	0.813	0.802	0.808
	N		0.803	0.810	0.806
Non-Knowledge MKN	R	0.767	0.766	0.769	0.767
	N		0.752	0.797	0.774
Non-Visual MKN	R	0.790	0.78	0.81	0.795
	N		0.794	0.781	0.788
CallAtRumors	R	0.773	0.776	0.771	0.773
	N		0.751	0.776	0.763

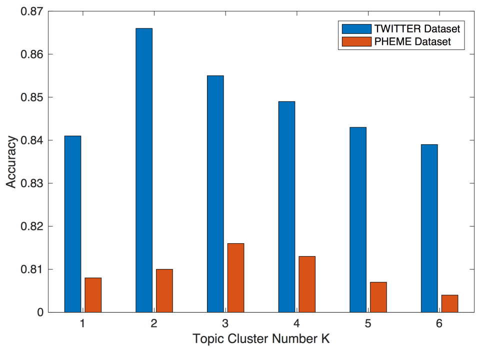

Figure 4: Results of different memory cluster numbers

5.3 Effects of the Memory Cluster Numbers

The effectiveness of event memory network has been proved in Section 5.2.2. In this section, we analyze the event memory network with different numbers of latent clusters. Specifically, we set the range of latent clusters number K as $[1, 2, 3, 4, 5, 6]$ in both datasets. Figure 4 shows the performance of MKEMN with different values of K. From Figure 4, we can observe that the model performance increases as the number of latent clusters grows until 2 for the TWITTER and 3 for the PHEME. This is probably because there are different number of subjects in each dataset. Specifically, the PHEME dataset is constructed by collecting thousands of events which are associated with the five breaking news (charliehebdo, ferguson, germanwings-crash, ottawashooting,and sydneysiege) [44]. There are naturally internal categories in PHEME dataset, which could be captured by event memory with 3 topic clusters. The TWITTER, however, has diverse contents, which means the features shared among event is sparse, thus the EMN module has fewer clusters compared with the PHEME.

5.4 Early Rumor Detection

Identifying rumours early can prevent the further propagation and minimize their damage. We evaluate the early detection performance of models by rough incrementally adding test data in the chronological order. The results are shown in Figure 5, from which we can conclude that:

- At the early stage with 10% to 60% test data, the accuracies of all methods increase.
- The EANN model have a stable performance, this is probably because it conducts rumor detection at post-level and can not model the interactions between posts.
- Our model outperforms six comparative methods by a significant margin. Particularly, MKEMN using 50% data, has already outperformed all the baselines using 80% data.This is due to the event memory network that takes the advantage of capturing the event-invariant features and saving them in external memory, which can boost the performance by querying the existing information when the model faces newly emerged events.

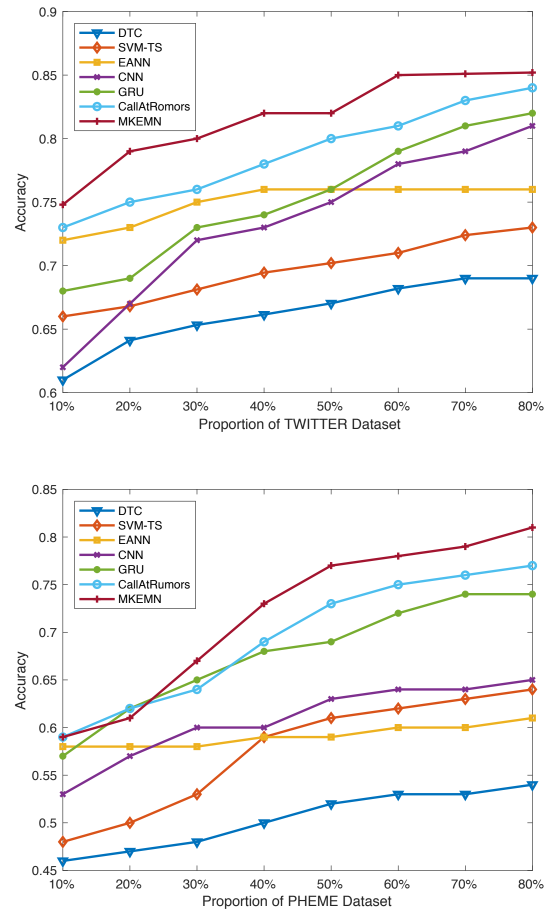

Figure 5: Results of rumor early detection

6 CONCLUSION

We propose a multi-modal knowledge-aware network, which exploits the multi-modal content and the external knowledge-level connections for accurate rumor identification. Besides, the proposed event memory network uses memory network to measure the dissimilarities among different events, and further learns the event invariant features which can generalize well for the newly emerged events. We experimentally demonstrate that our model is more robust and effective than state-of-the-art baselines based on two public benchmark datasets for rumor detection on Twitter. In our future work, we plan to use memory network to capture the rumor propagation to boost rumor detection performance.

ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China (No. 2017YFB1002804), the National Natural Science Foundation of China under Grants 61432019, 61702509, 61802405, 61832002, 61572503, 61872424 and 61720106006, the Key Research Program of Frontier Sciences, CAS, Grant NO. QYZDJ-SSW-JSC039, and the K.C.Wong Education Foundation.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[2] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011*, Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar (Eds.). ACM, 675–684. https://doi.org/10.1145/1963405.1963500

[3] Lihan Chen, Jiaqing Liang, Chenhao Xie, and Yanghua Xiao. 2018. Short Text Entity Linking with Fine-grained Topics. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*. ACM, New York, NY, USA, 457–466. https://doi.org/10.1145/3269206.3271809

[4] Tong Chen, Xue Li, Hongzhi Yin, and Jun Zhang. 2018. Call Attention to Rumors: Deep Attention Based Recurrent Neural Networks for Early Rumor Detection. In *Trends and Applications in Knowledge Discovery and Data Mining - PAKDD 2018 Workshops, BDASC, BDM, ML4Cyber, PAISI, DaMEMO, Melbourne, VIC, Australia, June 3, 2018, Revised Selected Papers (Lecture Notes in Computer Science)*, Mohadeseh Ganji, Lida Rashidi, Benjamin C. M. Fung, and Can Wang (Eds.), Vol. 11154. Springer, 40–52. https://doi.org/10.1007/978-3-030-04503-6_4

[5] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *CoRR* abs/1406.1078 (2014). arXiv:1406.1078 http://arxiv.org/abs/1406.1078

[6] Nicholas DiFonzo and Prashant Bordia. 2007. *Rumor psychology: Social and organizational approaches.* American Psychological Association.

[7] Octavian-Eugen Ganea, Marina Ganea, Aurélien Lucchi, Carsten Eickhoff, and Thomas Hofmann. 2016. Probabilistic Bag-Of-Hyperlinks Model for Entity Linking. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016.* 927–938. https://doi.org/10.1145/2872427.2882988

[8] Octavian-Eugen Ganea and Thomas Hofmann. 2017. Deep Joint Entity Disambiguation with Local Neural Attention. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017.* 2619–2629. https://aclanthology.info/papers/D17-1277/d17-1277

[9] Amir Globerson, Nevena Lazic, Soumen Chakrabarti, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2016. Collective Entity Resolution with Multi-Focal Attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers.* http://aclweb.org/anthology/P/P16/P16-1059.pdf

[10] Edouard Grave, Tomas Mikolov, Armand Joulin, and Piotr Bojanowski. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 2: Short Papers*, Mirella Lapata, Phil Blunsom, and Alexander Koller (Eds.). Association for Computational Linguistics, 427–431. https://aclanthology.info/papers/E17-2068/e17-2068

[11] Han Guo, Juan Cao, Yazi Zhang, Junbo Guo, and Jintao Li. 2018. Rumor Detection with Hierarchical Social Attention Network. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, Alfredo Cuzzocrea, James Allan, Norman W. Paton, Divesh Srivastava, Rakesh Agrawal, Andrei Z. Broder, Mohammed J. Zaki, K. Selçuk Candan, Alexandros Labrinidis, Assaf Schuster, and Haixun Wang (Eds.). ACM, 943–951. https://doi.org/10.1145/3269206.3271709

[12] Zhaochen Guo and Denilson Barbosa. 2018. Robust named entity disambiguation with random walks. *Semantic Web* 9, 4 (2018), 459–479. https://doi.org/10.3233/SW-170273

[13] Haoran Huang, Qi Zhang, and Xuanjing Huang. 2017. Mention Recommendation for Twitter with End-to-end Memory Network. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, Carles Sierra (Ed.). ijcai.org, 1872–1878. https://doi.org/10.24963/ijcai.2017/260

[14] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. 2017. Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs. In *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017*, Qiong Liu, Rainer Lienhart, Haohong Wang, Sheng-Wei "Kuan-Ta" Chen, Susanne Boll, Yi-Ping Phoebe Chen, Gerald Friedland, Jia Li, and Shuicheng Yan (Eds.). ACM, 795–816. https://doi.org/10.1145/3123266.3123454

[15] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian. 2017. Novel Visual and Statistical Image Features for Microblogs News Verification. *IEEE Transactions on Multimedia* 19, 3 (March 2017), 598–608. https://doi.org/10.1109/TMM.2016.2617078

[16] Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-End Neural Entity Linking. In *Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL 2018, Brussels, Belgium, October 31 - November 1, 2018*, Anna Korhonen and Ivan Titov (Eds.). Association for Computational Linguistics, 519–529. https://aclanthology.info/papers/K18-1050/k18-1050

[17] S Kwon, M Cha, and K Jung. 2017. Rumor Detection over Varying Time Windows. *Plos One* 12, 1 (2017), e0168344–.

[18] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang. 2013. Prominent Features of Rumor Propagation in Online Social Media. In *2013 IEEE 13th International Conference on Data Mining.* 1103–1108. https://doi.org/10.1109/ICDM.2013.61

[19] Nevena Lazic, Amarnag Subramanya, Michael Ringgaard, and Fernando Pereira. 2015. Plato: A Selective Context Model for Entity Resolution. *TACL* 3 (2015), 503–515. https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/637

[20] Phong Le and Ivan Titov. 2018. Improving Entity Linking by Modeling Latent Relations between Mentions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, 1595–1604. https://aclanthology.info/papers/P18-1148/p18-1148

[21] Zheng Li, Yu Zhang, Ying Wei, Yuxiang Wu, and Qiang Yang. 2017. End-to-End Adversarial Memory Network for Cross-domain Sentiment Classification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, Carles Sierra (Ed.). ijcai.org, 2237–2243. https://doi.org/10.24963/ijcai.2017/311

[22] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-time Rumor Debunking on Twitter. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM '15)*. ACM, New York, NY, USA, 1867–1870. https://doi.org/10.1145/2806416.2806651

[23] 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 Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, Subbarao Kambhampati (Ed.). IJCAI/AAAI Press, 3818–3824. http://www.ijcai.org/Abstract/16/537

[24] 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 Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, James Bailey, Alistair Moffat, Charu C. Aggarwal, Maarten de Rijke, Ravi Kumar, Vanessa Murdock, Timos K. Sellis, and Jeffrey Xu Yu (Eds.). ACM, 1751–1754. https://doi.org/10.1145/2806416.2806607

[25] Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, 1980–1989. https://aclanthology.info/papers/P18-1184/p18-1184

[26] Alexander H. Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-Value Memory Networks for Directly Reading Documents. *CoRR* abs/1606.03126 (2016). arXiv:1606.03126 http://arxiv.org/abs/1606.03126

[27] George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39–41. https://doi.org/10.1145/219717.219748

[28] David N. Milne and Ian H. Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008*, James G. Shanahan, Sihem Amer-Yahia, Ioana Manolescu, Yi Zhang, David A. Evans, Aleksander Kolcz, Key-Sun Choi, and Abdur Chowdhury (Eds.). ACM, 509–518. https://doi.org/10.1145/1458082.1458150

[29] Shengsheng Qian, Tianzhu Zhang, Changsheng Xu, and Jie Shao. 2016. Multi-Modal Event Topic Model for Social Event Analysis. *IEEE Trans. Multimedia* 18, 2 (2016), 233–246. https://doi.org/10.1109/TMM.2015.2510329

[30] Jonathan Raiman and Olivier Raiman. 2018. DeepType: Multilingual Entity Linking by Neural Type System Evolution. *CoRR* abs/1802.01021 (2018). arXiv:1802.01021 http://arxiv.org/abs/1802.01021

[31] Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A Hybrid Deep Model for Fake News Detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, Ee-Peng Lim, Marianne Winslett, Mark Sanderson, Ada Wai-Chee Fu, Jimeng Sun, J. Shane Culpepper, Eric Lo, Joyce C. Ho, Debora Donato, Rakesh Agrawal, Yu Zheng, Carlos Castillo, Aixin Sun, Vincent S. Tseng, and Chenliang Li (Eds.). ACM, 797–806. https://doi.org/10.1145/3132847.3132877

[32] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explorations* 19, 1 (2017), 22–36. https://doi.org/10.1145/3137597.3137600

[33] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1409.1556

[34] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. YAGO: A Large Ontology from Wikipedia and WordNet. *J. Web Semant.* 6, 3 (2008), 203–217. https://doi.org/10.1016/j.websem.2008.06.001

[35] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-To-End Memory Networks. In *Advances in Neural Information Processing Systems*

28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada, Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett (Eds.). 2440–2448. http://papers.nips.cc/paper/5846-end-to-end-memory-networks

[36] Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. ACM, New York, NY, USA, 849–857. https://doi.org/10.1145/3219819.3219903

[37] Jason Weston, Sumit Chopra, and Antoine Bordes. 2015. Memory Networks. *CoRR* abs/1410.3916 (2015).

[38] Ke Wu, Song Yang, and Kenny Q. Zhu. 2015. False rumors detection on Sina Weibo by propagation structures. In *31st IEEE International Conference on Data Engineering, ICDE 2015, Seoul, South Korea, April 13-17, 2015*, Johannes Gehrke, Wolfgang Lehner, Kyuseok Shim, Sang Kyun Cha, and Guy M. Lohman (Eds.). IEEE Computer Society, 651–662. https://doi.org/10.1109/ICDE.2015.7113322

[39] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Qili Zhu. 2012. Probase: a probabilistic taxonomy for text understanding. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2012, Scottsdale, AZ, USA, May 20-24, 2012*, K. Selçuk Candan, Yi Chen, Richard T. Snodgrass, Luis Gravano, and Ariel Fuxman (Eds.). ACM, 481–492. https://doi.org/10.1145/2213836.2213891

[40] Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, and Yoshiyasu Takefuji. 2017. Learning Distributed Representations of Texts and Entities from Knowledge Base. *TACL* 5 (2017), 397–411. https://transacl.org/ojs/index.php/tacl/article/view/1065

[41] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A Convolutional Approach for Misinformation Identification. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, Carles Sierra (Ed.). ijcai.org, 3901–3907. https://doi.org/10.24963/ijcai.2017/545

[42] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, Tieniu Tan, et al. 2017. A Convolutional Approach for Misinformation Identification.. In *IJCAI*. 3901–3907.

[43] Huaiwen Zhang, Quan Fang, Shengsheng Qian, and Changsheng Xu. 2018. Learning Multimodal Taxonomy via Variational Deep Graph Embedding and Clustering. In *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018, Seoul, Republic of Korea, October 22-26, 2018*. 681–689. https://doi.org/10.1145/3240508.3240586

[44] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2016. Learning Reporting Dynamics during Breaking News for Rumour Detection in Social Media. *CoRR* abs/1610.07363 (2016). arXiv:1610.07363 http://arxiv.org/abs/1610.07363

[45] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting Context for Rumour Detection in Social Media. In *Social Informatics - 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I (Lecture Notes in Computer Science)*, Giovanni Luca Ciampaglia, Afra J. Mashhadi, and Taha Yasseri (Eds.), Vol. 10539. Springer, 109–123. https://doi.org/10.1007/978-3-319-67217-5_8