Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks

Youze Wang[1], Shengsheng Qian[2,3], Jun Hu[1], Quan Fang[2,3], Changsheng Xu[1,2,3,4]
[1]Hefei University of Technology
[2]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences
[4]Peng Cheng Laboratory, ShenZhen, China
youze.wang@mail.hfut.edu.cn,shengsheng.qian@nlpr.ia.ac.cn
hujunxianligong@gmail.com,{qfang,csxu}@nlpr.ia.ac.cn

ABSTRACT

Nowadays, with the rapid development of social media, there is a great deal of news produced every day. How to detect fake news automatically from a large of multimedia posts has become very important for people, the government and news recommendation sites. However, most of the existing approaches either extract features from the text of the post which is a single modality or simply concatenate the visual features and textual features of a post to get a multimodal feature and detect fake news. Most of them ignore the background knowledge hidden in the text content of the post which facilitates fake news detection. To address these issues, we propose a novel Knowledge-driven Multimodal Graph Convolutional Network (KMGCN) to model the semantic representations by jointly modeling the textual information, knowledge concepts and visual information into a unified framework for fake news detection. Instead of viewing text content as word sequences normally, we convert them into a graph, which can model non-consecutive phrases for better obtaining the composition of semantics. Besides, we not only convert visual information as nodes of graphs but also retrieve external knowledge from real-world knowledge graph as nodes of graphs to provide complementary semantics information to improve fake news detection. We utilize a well-designed graph convolutional network to extract the semantic representation of these graphs. Extensive experiments on two public real-world datasets illustrate the validation of our approach.

CCS CONCEPTS

• **Information systems** → *Multimedia and multimodal retrieval.*

KEYWORDS

fake news detection, social media, graph neural networks

ACM Reference Format:

Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, Changsheng Xu. 2020. Fake News Detection via Knowledge-driven Multimodal Graph Convolutional Networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval (ICMR '20), June 8–11, 2020, Dublin, Ireland.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3372278.3390713

ICMR '20, June 8–11, 2020, Dublin, Ireland

ACM ISBN 978-1-4503-7087-5/20/06…$15.00
https://doi.org/10.1145/3372278.3390713

Figure 1: An example of multimodal tweets. We can obtain a lot of information, such as Michael Bloomberg, Democrat, impeachment, Donald Trump from the textual content, and see a speaker, the Stars and the Stripes as visual information, and think of the speaker is a politician, Donald Trump is American President from the background knowledge.

1 INTRODUCTION

With the development of the Internet and multimedia, the way people acquire information has been significantly changed. Nowadays, there are more and more people consuming news through social media, which can provide all kinds of multimedia information on the events taking place all over the world. Unfortunately, social media websites also have fostered various fake news which usually contain misrepresented or even forged multimedia content, to mislead the readers and get rapid spread. Some evil guys even use rumors to mislead public opinion, which can damage the credibility of the government on purpose. Therefore, it is necessary and urgent to use an automatic detector to prevent fake news from causing serious negative effects and make users receive truthful information.

Thus far, various fake news detection approaches [15, 16, 25, 32], including both traditional learning and deep learning-based models, have been used to debunk fake news and minimize their harmful effects. Earlier studies on fake news detection constructed features through heuristic rules or statistical information. For example, SVM-TS [16] utilizes heuristic rules and a linear SVM [27] to classify fake news on Twitter and uses a time-series structure to model the social feature variations. With the great success of the neural network, existing deep learning models have achieved performance improvement over traditional ones due to their superior ability of feature extraction. The convolutional neural networks (CNNs) [32] is introduced to obtain the high-level representations from text content of the post to identify rumors. Recurrent neural networks (RNNs) [15] is used to learn the hidden representation and sequential features from the propagation of fake news.

However, most of the existing deep learning methods only capture local semantic features in small sliding windows (short messages or word-level syntactics) for fake news detection, and ignore the structural information of posts which is a very important aspect for fake news detection. For example, some posts may have many words and understanding the semantics of them needs to model non-consecutive phrases and long-range word dependency [21]. For example, the Figure 1 is the relatively long post text, in which the key words are "Michael Bloomberg", "impeachment", and "Donald Trump" and so on. The combination of these key words indicates that the semantics may be related to "Michael Bloomberg supported the impeachment to Donald Trump". However, these key words are not grouped together and distributed throughout the whole post text. It is hard to capture the dependency of semantic and structure information in a small sliding window. Therefore, how to effectively capture the non-consecutive and long-range semantic relations among words in feature representation becomes more and more important for fake news detection.

Furthermore, fake news detection is quite different from other classification tasks such as text classification, because in most situations, it needs to detect fake news from various fields which the model maybe never seen. However, existing deep learning methods typically focus on inferring clues from the post text content, and think little of the visual information and background knowledge of posts which humans also use in judging the credibility of an event. For example, to judge the credibility of the post in Figure 1, the first thing people usually do is to observe the picture and then read the text content, and realize that Michael Bloomberg is an American politician whose Democratic Party is one of the two major political parties in modern America and he supports the impeachment against Donald Trump, and finally give judgment. This indicates that social media posts carry a great deal of latent knowledge-level connections and multimodal property, which can help us to judge whether the post is fake news. Hence, how to acquire the background knowledge of the post text content, and fuse the textual information, knowledge concepts and visual information of the post in a principled way is the key for fake news detection.

In order to address the aforementioned challenges, we propose a Knowledge-driven Multimodal Graph Convolutional Network (KMGCN) for fake news detection by modeling posts as graphs data structure, and combining the textual information, knowledge concepts and visual information into a unified deep model. (1) To capture long-range semantic features for better content representations, we model each post content as a graph rather than word sequences. (2) To make full use of the background knowledge and multimodel information, we utilize object detection techniques to extract objects in visual content as visual words and obtain knowledge concepts through knowledge distillation, both of which are modeled into the graphs to provide complementary semantic information. Therefore, KMGCN can effectively exploit the latent semantic-level features of posts to learn a robust model for fake news detection. We conduct experiments on the public Twitter and Weibo benchmark datasets, demonstrating that our method outperforms several state-of-the-art methods. In conclusion, the contributions of our work are as follows:

- We propose an end-to-end Knowledge-driven Multimodal Graph Convolutional Network to model the semantic-level representations for fake news detection by jointly modeling the textual information, knowledge concepts and visual information into a unified deep model.
- We model multimodal posts as graphs in classification tasks and propose a multimodal graph convolutional network, which can capture non-consecutive and long-range semantic relations. Moreover, knowledge distillation is used to provide supplementary knowledge concept which can generalize well for the newly emerged posts.
- We evaluate our method on two real-world datasets, and experimental results demonstrate our KMGCN approach outperforms the baseline methods.

2 RELATED WORK

In this section, we briefly review the work related to the proposed model. We mainly focus on the following topics: fake news detection and graph neural network.

2.1 Fake News Detection

There are many tasks related to fake news detection, such as rumor detection and spam detection. Following the previous work, we specify the definition of fake news as news that is intentionally fabricated and can be verified as fake. In the fake news detection task, the main challenge is how to distinguish news according to features. The features can be extracted from posts, attached images and related knowledge concepts. Based on the difference of modality, we roughly summarize existing models into two categories: single modality based and multimodal fake news detection.

2.1.1 Single Modality based Fake News Detection. Existing models mainly extract the textual features or semantic features from the text content of posts, which have been explored in many works of literature of fake news detection [2, 5, 12, 25]. 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. [15] introduce recurrent neural networks (RNNs) to learn the hidden representations from the text content of relevant posts. Yu et al. [32] use convolutional neural networks (CNNs) to obtain key features and their high-level interactions from the text content of the relevant posts. Unfortunately, linguistic patterns are not yet well understood, which are not enough to identify fake news. However, social media platforms have rich multimodal information, such as texts, images and videos, which can complement each other for social media analysis [22, 23].

2.1.2 Multimodal Fake News Detection. In recent years, there are more and more models with deep neural networks to learn feature representations from multiple aspects. In [7], the authors propose a deep learning-based fake news detection model, which extracts the multimodal and social context features and fuses them by attention mechanism. In [28], Wang et al. identify a single post by leveraging both the textual and visual information of each post, and utilizing an adversarial method to remove event-specific features from post representation. In [8], Khattar et al. propose a novel multimodal variational autoencoder, which uses a bimodal variational autoencoder coupled with a binary classifier for the task of fake news detection. In [33], Zhang et al. utilize the multimodal knowledge-aware representation and the event-invariant features to form the event representation, which is fed into a deep neural network for fake news detection.

However, most of the existing methods only capture local semantic features in small sliding windows for fake news detection, while the long-range and non-consecutive semantic relations among words in feature representation are always be neglected.

2.2 Graph Neural Networks

Recently, graph neural networks [1, 29] have been increasingly applied to NLP tasks to learn text representations. Bastings et al. [1] employ graph convolutional networks (GCNs) [10] to incorporate syntactic structure into neural attention-based encoder-decoder models for machine translation. Marcheggiani et al. [17] incorporate information about the predicate-argument structure of source sentences, and use GCNs to inject a semantic bias into sentence encoders. TextGCN [31] builds a single text graph for a corpus-based on word co-occurrence and document-word relations, then learns a Text Graph Convolutional Network for the corpus to learn word and document embeddings. Peng et al. [21] propose a GCN-based deep learning model to first convert texts into graph-of-words, and then use graph convolution operations to convolve the word graph for text classification. SE-GCN [14] is a long document matching approach that builds concept graphs for documents and employs a siamese encoded graph convolutional network to generate the matching results. However, the above methods [21, 31] mainly focus on how to apply the model to textual corpora and may not be appropriate for multimodal posts in social media.

To take advantage of the multimodal content information and additional background knowledge for fake news detection, we present a knowledge-driven multimodal graph convolutional network to extract features of posts to identify fake news by jointly modeling the textual information, knowledge concepts and visual information into a unified framework.

3 METHOD

In this section, we mainly introduce the proposed Knowledge-driven Multimodal Graph Convolutional Network (KMGCN) in detail. We first describe the problem definition, and then, we introduce the overall framework of KMGCN. The details of the proposed model are shown in the following sections.

3.1 Problem Definition

Fake news detection task can be defined as a binary classification problem, which aims to classify a claim in social media as fake news or not. There are two types of claims in social media fake news detection community: the post-level work identifies whether a single post of social media is fake news or not, while the event-level work conducts fake news detection on a group of posts which constitute an event.

The goal of our model is to identify whether a post is fake or not at the post-level. Given a set of multimedia posts from social media $\mathrm{D} = \left\{p_1, \ldots, p_\mathrm{N}\right\}$ where p_i is a post which consists of a set of words and corresponding visual information, N represents the number of posts. We need to learn a model $f : \mathcal{D} \rightarrow \mathcal{Y}$, to classify each post p_i into the predefined categories $\mathcal{Y} = \{0, 1\}$, which is the ground-truth label of the claim (0 denotes Real news, and 1 denotes Fake news).

3.2 Overall Framework

Our purpose is to identify whether a claim is fake news or not. To this end, we present a *Knowledge-driven Multimodal Graph Convolutional Networks* (KMGCN) to model the semantic-level representations in a unified framework. Figure 2 shows the framework of KMGCN, which mainly consists of the following components:

- **Knowledge Distillation**: The background knowledge distilled from a real-word knowledge graph can complement the semantic representation of short texts of posts. Furthermore, the conceptual information extracted from entities can provide additional evidence to enhance fake news detection. Specifically, the goal of this module is to retrieve relevant knowledge from Knowledge Graphs.
- **Graph Construction of Multimodal Content**: We model the multimodal content of the input post as graphs based on the point-wise mutual information (PMI) score of word pairs. In particular, we constructed a graph for each multimedia post. Specifically, we utilize object detection techniques to extract objects in visual content as visual words and obtain knowledge concept through knowledge distillation to model all of them into the graphs.
- **Knowledge-driven Multimodal Graph Convolutional Network (KMGCN)**: Based on the constructed graphs, a multimodal graph convolutional network is used to obtain the semantic-level features for each post. We employ two GCN layers and a global mean pooling to aggregate the nodes of each graph and obtain the representation vector of each post. In the end, the multimodal representation of each post is fed into a binary classifier to calculate the predicted probability to determine whether the news is fake or not.

3.3 Knowledge Distillation

Given a post text, we hope to find a concept set C relevant to it. The process of knowledge distillation consists of four steps. Given the short text content of posts, many entities linking methods, such as Rel-Norm [13], Link Detector [19], EDEL [11] can be utilized to link the ambiguous entity mentions M in a text to the correct entities T in the knowledge graph they refer to. Then, for each identified entity $\mathrm{t} \in \mathrm{T}$, we acquire its conceptual information from an

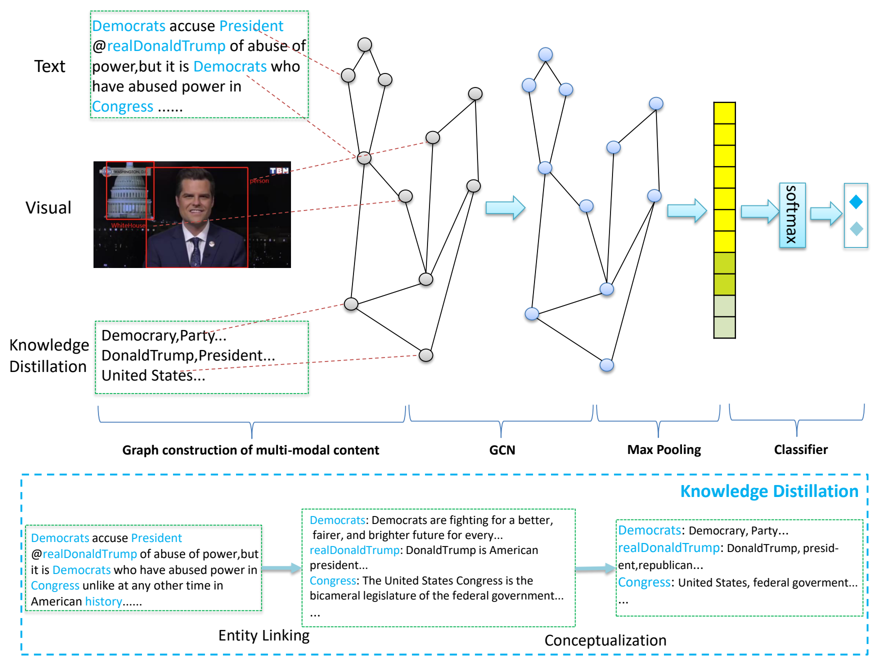

Figure 2: The overall framework of KMGCN. The input consists of the multimodal content of a post and corresponding knowledge concepts. (1) Firstly, our model obtains knowledge concepts according to the text of the post, which is illustrated by the process of Knowledge Distillation in the bottom half of Figure 2. (2) Then, our model converts the multimodal content and knowledge concepts into graphs. (3) Finally, each graph is aggregated to obtain the representation of each post through two GCN layers and the pooling layer. The feature representation of each post is fed into the full-connected layer and softmax layer to identify whether the news is fake or not.

existing knowledge graph, such as YAGO [26] and Probase [30] by conceptualization. This paper takes isA relation as an example. For instance, given a short text "Michael Bloomberg, who ranked fourth in support from Democrat-leaning voters in the latest Reuters/lpsos poll, backed the impeachment vote against President Donald Trump, saying the lawmakers did their "constitutional duty", we obtain the entity set T = $\{$Michael Bloomberg, Democrat, DonaldTrump$\}$ by entity linking. Then, we conceptualize the entities in T and acquire its concept set $C_{\text{Michael Bloomberg}}$ = (*American politician, businessman, author, Bloomberg L.P., New York City*), C_{Democrat} = (*democracy, Party*), $C_{\text{Donald Trump}}$ = (*American president, republican, American politician*) from external knowledge graphs. Given a post p_i, we can conduct the knowledge distillation from knowledge graph and get a set of concepts for every entity contained in p_i. For every concept set $C_t = (c_1, \ldots, c_i, \ldots, c_m)$ where t is the original entity and c_i is the i-th concept about entity t, m is the set size, we aim at producing the concept knowledge for each post.

3.4 Graph Construction of Multimodal Content

An undirected graph is created in our model for each post to model its multimodal content information. Given a post, the words are taken as the graph nodes and the relationship between words is taken as edges. We employ point-wise mutual information (PMI) [4, 6] to calculate the weights of edges, which can preserve the global word co-occurrence information. In detail, we employ a fixed-size window on all posts content for gathering word co-occurrence statistics. Then, we calculate the PMI of word pairs as follows:

$$p(w_i) = \frac{W(w_i)}{|W|} \tag{1}$$

$$p(w_i, w_j) = \frac{W(w_i, w_j)}{|W|} \tag{2}$$

$$PMI(w_i, w_j) = \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \tag{3}$$

where $|W(w_i)|$ is the number of sliding windows that contain the word w_i, $|W(w_i, w_j)|$ is the number of sliding windows that contain both the word w_i and w_j, and $|W|$ is the total number of sliding windows. Note that the statistics are based on the global corpus rather than a specific post content. PMI scores can reflect the correlation between words, and positive PMI scores implies high semantic correlations. Therefore, we only preserve edges with positive PMI scores and discard those with non-positive PMI scores:

$$A_{ij} = \begin{cases} PMI(w_i, w_j) & PMI(w_i, w_j) > 0 \\ 0 & PMI(w_i, w_j) \leq 0 \end{cases} \tag{4}$$

In addition to the text content, we utilize the visual content to facilitate the fake news detection model by extracting semantic objects in images as visual words. We employ the YOLOv3 detector [24] pre-trained on the dataset that we have labeled to search objects in each image. For each post, we may obtain several images and the YOLOv3 detector may detect several semantic objects in each image. The labels of detected objects, such as "person" and "gun", are treated as words that occurred in a post and are added into the text content of each post.

Besides textual content and visual information, we also obtain a set of knowledge concepts from knowledge graphs for every entity in a post to complement the semantic representation of the post. Like visual words, we add knowledge concepts into the text content of each post. In the end, we build a graph which contains textual words, visual words and knowledge concepts by PMI for each post. After this process, we obtain an adjacency matrix A for each post.

3.5 Knowledge-driven Multimodal Graph Convolutional Network (KMGCN)

A GCN is a multi-layer neural network that operates directly on a graph, it takes an undirected graph as input and outputs embedding vectors for vertices based on properties of their neighborhoods. Formally, we consider an undirected graph $G = (V, E)$, where V are sets of nodes and E are sets of edges. Every node is assumed to be connected to itself, i.e., $(v, v) \in \mathbb{E}$ for any v. We use n $(n = |V|)$ to denote the number of nodes. Each node is associated with a d-dimensional feature vector and we use a feature matrix $X \in \mathbb{R}^{n \times l^{(0)}}$ to represent the features of all vertices, where the i_{th} row corresponds to the feature vector of the i_{th} node. We introduce an adjacency matrix $A \in \mathbb{R}^{n \times n}$ to indicate the edge set E, where A_{ij} is the weight of the edge between the i_{th} node and the j_{th} node. The degree matrix D is a diagonal matrix and $D_{ij} = \sum_j A_{ij}$. Based on the adjacency matrix A and the degree matrix D, each GCN layer input feature matrix $Z^{(j)} \in \mathbb{R}^{n \times l^{(j)}}$ (the input feature matrix of first layer is $X \in \mathbb{R}^{n \times l^{(0)}}$) and output a higher-order feature matrix $Z^{(j+1)} \in \mathbb{R}^{n \times l^{(j+1)}}$ for vertices as follows:

$$Z^{(0)} = X \tag{5}$$

$$Z^{(j+1)} = \sigma(D^{-\frac{1}{2}}(I + A)D^{-\frac{1}{2}}Z^{(j)}W) \tag{6}$$

where $W \in \mathbb{R}^{l^{(j)} \times l^{(j+1)}}$ is a weight matrix that can be learned during training, I is the identify matrix and σ is an activation function, e.g. a ReLU $\sigma(x) = max(0, x)$. Note that the adjacency matrix A has been obtained in graph construction of multimodal content. For feature matrix X for each post, we employ the distributed Word2Vec representation for words [18].

After two layers of GCN, we choose a global mean pooling to aggregate the vertices of each graph and get the representation vector O of posts. Finally, we feed each post p_{i}'s representation vector O_{i} into a binary classifier $\mathcal{F}$ and get the prediction r_{i}:

$$r_{\mathrm{i}} = \mathcal{F}(O_{\mathrm{i}}) \tag{7}$$

In order to calculate the classification loss, we employ cross-entropy as follow:

$$\mathcal{L} = \sum_{\mathrm{i}=1}^{N} -[y_{\mathrm{i}} * log(r_{\mathrm{i}}) + (1 - y_{\mathrm{i}}) * log(1 - r_{\mathrm{i}})] \tag{8}$$

where N is the number of posts, y_i is the ground-truth label of i-th post.

4 EXPERIMENT

In this section, we conduct experiments to evaluate our model against state-of-the-art models on two real-world datasets. Furthermore, we give detailed experimental analysis to show more insights on our model.

4.1 Experimental Setup

4.1.1 Datasets. We used two public real-world social media datasets, PHEME [34] and WEIBO [15], to validate the effectiveness of the proposed KMGCN. The PHEME dataset is collected based on 5 breaking news, and news contain a set of claims. The WEIBO dataset is collected based on the claims reported on www.weibo.com, where each claim contains text, image URL, response and so on. Each dataset has a large number of articles and images with labels. Table 1 shows the statistics of the two datasets. .

4.1.2 Implementation Details. For the data, we split the data into a training set and a testing set on a scale of 7:3 without overlapping. In the knowledge distillation process, we acquire an entity set T of a textual post by leveraging the existing entity linking solutions [3]. For conceptualization, we retrieve the entities in Probase [30] and YAGO [26] and only consider the isA relation. In terms of parameter setting, we set the size of the sliding window to 20, and the dimension of the word to 100, We apply two layers of GCN. The output dimension of convolution network of the first layer and second layer are 64 and 32 respectively, the learning rate is 0.001. Our algorithms are implemented on Pytorch deep learning framework [20] and are trained with Adaptive Moment Estimation(Adam) [9] optimizer. Specifically, in order to better extract objects in visual content, we labeled about 80 different kinds of objects on the images in two real-world social media datasets by ourselves. We employ accuracy, precision, recall, and F-measure [28] as evaluation metrics of the following baselines approaches.

Table 1: The Statistics of the Real-World Datasets.

News	PHEME	WEIBO
Fake News	1972	2313
Real News	3830	2351
Images	3670	3989

4.2 Baselines

The experiments on the two datasets use the baselines listed as follows:

- *SVM-TS* [16]: SVM-TS proposes the linear SVM classifier that uses time-series structures to model the variation of social context features.
- *GRU* [15]: GRU uses a multilayer generic GRU network to model the microblog as a variable-length time series, which is effective for the early detection of rumors.
- *CNN* [32]: CNN uses a convolution network to learn rumor representations by framing the relevant posts as fixed-length sequence.
- *TextGCN* [31]: Text Graph Convolution Network (TextGCN) is an algorithm that uses a graph convolution network to

Table 2: Results of comparison among different models on PHEME and WEIBO Datasets.

Dataset	Methods	Accuracy	Precision	Recall	F1
WEIBO	SVM-TS	0.6312	0.6329	0.6301	0.6309
	CNN	0.7112	0.713	0.7112	0.711
	EANN	0.7212	0.7353	0.7228	0.7160
	GRU	0.7927	0.8139	0.7927	0.7891
	TextGCN	0.8571	0.8634	0.8576	0.8565
	KMGCN	**0.8863**	**0.9100**	**0.9645**	**0.8834**
PHEME	SVM-TS	0.6399	0.6391	0.6211	0.6395
	CNN	0.7007	0.7413	0.7074	0.6896
	EANN	0.7177	0.7382	0.7179	0.7104
	TextGCN	0.8282	0.8274	0.8283	0.8277
	GRU	0.8374	0.8382	0.8374	0.8312
	KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

Table 3: Results of comparison among different variants in our model on WEIBO Dataset.

Methods	Accuracy	Precision	Recall	F1
textGCN	0.8571	0.8634	0.8576	0.8565
KMGCN-NoKDVisual	0.8713	0.8718	0.974	0.8692
KMGCN-NoVisual	0.8792	0.8748	0.9712	0.8733
KMGCN-NoKD	0.8799	0.8799	0.9327	0.8761
KMGCN	**0.8863**	**0.91**	**0.9645**	**0.8834**

Table 4: Results of comparison among different variants in our model on PHEME Dataset.

Methods	Accuracy	Precision	Recall	F1
GRU	0.8374	0.8382	0.8374	0.8312
KMGCN-NoKDVisual	0.8553	0.8552	0.8089	0.8323
KMGCN-NoVisual	0.8621	0.852	0.8166	0.8339
KMGCN-NoKD	0.8690	0.8675	0.8690	0.8677
KMGCN	**0.8756**	**0.8762**	**0.8765**	**0.8764**

classify documents. The whole corpus is modeled as a heterogeneous graph. It combines graph neural network to learn words and document embedding.

- *EANN* [28]: this method is a post-level fake news detection model which aims to determine whether a post is fake or not by concatenating textual and visual feature of a post, and utilizing an adversarial method to remove event-specific features from post representation.

In addition to the above baselines, we design several variants to demonstrate the effectiveness of each component in our model. We will introduce these variants in the performance analysis section.

4.3 Performance Analysis

4.3.1 Quantitative Evaluation. Table 2 shows the experimental results of our proposed KMGCN and all baselines approaches. From the Table 2, we can draw the following observations:

(1) SVM-TS model performs worst among all methods, which is possible that the hand-crafted features are weak and not enough to identify fake news.
(2) CNN is a supervised algorithm and captures local feature between different words, training by neural networks. Its performance is only better than SVM-TS on two datasets, which is probably because the CNN ignores the long-range semantic relations among words and local feature is not enough to make judgment for a post.
(3) EANN performs better than CNN. That is because EANN employs TextCNN to extract textual feature and VGG-19 to extract visual feature, which provides complementary information to improve fake news detection.
(3) TextGCN is also a supervised algorithm, and performs best in all baselines in WEIBO dataset and performs better in PHEME dataset, which shows that the graph structure can effectively capture word co-occurences and document-word relations by the flexible graph convolutional network.
(4) GRU performs significantly better than CNN on two datasets , especially the PHEME dataset, because the recurrent neural networks can inherently deal with variable-length sequence of posts, while CNN needs more data to make judgment.
(6) The proposed KMGCN has achieved the best performance compared with all the baselines. The superiority of KMGCN can be attributed to two properties: 1) KMGCN uses visual information and knowledge concepts to enhance the semantic

information of post text. 2) Multimodal graph convolutional networks can better capture non-consecutive phrases and word dependency to obtain more semantic representations.

4.3.2 Comparison among KMGCN variants. In this section, we compare among the variants of KMGCN in PHEME and WEIBO datasets as shown in Table 3 and Table 4. The following variants of our model are designed for comparison.

- KMGCN-NoKD: the variant of KMGCN, which removes the knowledge distillation when modeling the post.
- KMGCN-NoVisual: the variant of KMGCN, which removes visual information when modeling the post.
- KMGCN-NoKDVisual: the variant of KMGCN, which removes knowledge distillation and visual information when modeling the post.

From the Table 3 and Table 4, we can have the following observations:

- Comparing the performance of KMGCN-NoKDVisual with that of TextGCN on WEIBO dataset and that of GRU on PHEME dataset, we can find that graph construction with GCN on pure post text (KMGCN-NoKDVisual) is obviously better than the best of baselines except for Recall on PHEME dataset, which shows graph construction as input can effectively capture the long-range semantic relations among words in feature representation for fake news detection.
- Comparing the performance of KMGCN with that of KMGCN-NoVisual, we can find that the removal of visual information leads to the model's declines in two datasets. Although the introduction of visual content information does not bring huge performance improvement, it can consistently provide complementary information to benefit our model.
- Comparing the performance of KMGCN with that of KMGCN-NoKD, we can find the model equipped with knowledge information achieves consistently better results than the Non-knowledge model in both datasets, which indicates that the knowledge information is an important kind of complementary information for fake news detection.

4.3.3 Qualitative Evaluation. In order to illustrate the effectiveness of the proposed model for fake news detection, we give two multimodal posts that KMGCN can recognize accurately.

(a) **The pilot actually took a selfie in the air.**

(b) **Judge Calls For US Marshals and FBI To Arrest Congress and Obama.**

Figure 3: Illustration of some fake news detected by KMGCN.

By providing two multimodal posts with representative textual and visual information, it is very intuitive for us to realize the importance of multimodal information and background knowledge for fake news detection. As can be seen, in Figure 3(a), the textual content is normal and cannot show evidence to identify that the news is fake, while the attached image looks quite suspicious and the visual detector can detect several semantic objects, such as "people", "airplane" and "clouds", which can be added into textual content of the post to provides complementary information for fake news detection. In Figure 3(b), the attached image looks quite normal, but the corresponding textual description is surprising and seems to be impossible, and our model can utilize knowledge distillation to obtain knowledge concepts about "judge", "FBI", "Congress", and "Obama" as background knowledge, which are added into the textual content of the post and are helpful to detect whether the news is fake. Based on the results, we can confirm that our proposed model can effectively use multimodal information of the post and knowledge concepts for fake news detection.

5 CONCLUSION

In this paper, we propose a novel *Knowledge-driven Multimodal Graph Convolutional Networks* (KMGCN), which exploits the multimodal content and the external knowledge-level connections for fake news detection. In order to utilize multimodal information, we model the multimodal content as graphs including post text, visual information to capture non-consecutive semantic features and make use of external additional knowledge information. Specifically, we conduct the knowledge distillation from a knowledge graph and get a set of concepts to complement the semantic representation of short texts of posts. We experimentally demonstrate that our model is more robust and effective than state-of-the-art baselines based on two public benchmark datasets for fake news detection. In the future, we plan to use a more efficient way to extract the visual information to help our model recognize fake news.

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 61720106006, 61802405, 61702509, 61832002, 61936005, 61572503, 61872424, and U1705262, the Key Research Program of Frontier Sciences, CAS, Grant NO. QYZDJ-SSW-JSC039, and the K.C.Wong Education Foundation.

REFERENCES

[1] Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an. 2017. Graph Convolutional Encoders for Syntax-aware Neural Machine Translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017*. 1957–1967.

[2] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*. ACM, 675–684.

[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*. ACM, 457–466.

[4] Kenneth Ward Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16, 1 (1990), 22–29.

[5] Aditi Gupta, Ponnurangam Kumaraguru, Carlos Castillo, and Patrick Meier. 2014. Tweetcred: Real-time credibility assessment of content on twitter. In *International Conference on Social Informatics*. Springer, 228–243.
[6] Jun Hu, Shengsheng Qian, Quan Fang, and Changsheng Xu. 2019. Hierarchical Graph Semantic Pooling Network for Multi-modal Community Question Answer Matching. In *Proceedings of the 27th ACM International Conference on Multimedia*. ACM, 1157–1165.
[7] 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 25th ACM international conference on Multimedia*. ACM, 795–816.
[8] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *The World Wide Web Conference*. 2915–2921.
[9] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
[10] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.
[11] Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. End-to-End Neural Entity Linking. *CoNLL 2018* (2018), 519.
[12] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*. IEEE, 1103–1108.
[13] 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 (Volume 1: Long Papers)*. 1595–1604.
[14] Bang Liu, Ting Zhang, Di Niu, Jinghong Lin, Kunfeng Lai, and Yu Xu. 2018. Matching Long Text Documents via Graph Convolutional Networks. *CoRR* abs/1802.07459 (2018). arXiv:1802.07459
[15] 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 *Ijcai*. 3818–3824.
[16] 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, 1751–1754.
[17] Diego Marcheggiani, Joost Bastings, and Ivan Titov. 2018. Exploiting Semantics in Neural Machine Translation with Graph Convolutional Networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 2 (Short Papers)*. 486–492.
[18] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[19] David Milne and Ian H Witten. 2008. Learning to link with wikipedia. In *Proceedings of the 17th ACM conference on Information and knowledge management*. ACM, 509–518.
[20] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. (2017).
[21] Hao Peng, Jianxin Li, Yu He, Yaopeng Liu, Mengjiao Bao, Lihong Wang, Yangqiu Song, and Qiang Yang. 2018. Large-Scale Hierarchical Text Classification with Recursively Regularized Deep Graph-CNN. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web, WWW 2018, Lyon, France, April 23-27, 2018*. 1063–1072.
[22] Shengsheng Qian, Tianzhu Zhang, Changsheng Xu, and M. Shamim Hossain. 2014. Social Event Classification via Boosted Multimodal Supervised Latent Dirichlet Allocation. *TOMM* 11, 2 (2014), 27:1–27:22.
[23] Shengsheng Qian, Tianzhu Zhang, Changsheng Xu, and Jie Shao. 2015. Multi-modal event topic model for social event analysis. *IEEE transactions on multimedia* 18, 2 (2015), 233–246.
[24] Joseph Redmon and Ali Farhadi. 2018. YOLOv3: An Incremental Improvement. *CoRR* abs/1804.02767 (2018). arXiv:1804.02767
[25] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* 19, 1 (2017), 22–36.
[26] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2008. Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web* 6, 3 (2008), 203–217.
[27] Simon Tong and Daphne Koller. 2001. Support vector machine active learning with applications to text classification. *Journal of machine learning research* 2, Nov (2001), 45–66.
[28] 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*. ACM, 849–857.
[29] Man Wu, Shirui Pan, Xingquan Zhu, Chuan Zhou, and Lei Pan. 2019. Domain-Adversarial Graph Neural Networks for Text Classification. In *2019 IEEE International Conference on Data Mining, ICDM 2019, Beijing, China, November 8-11, 2019*. IEEE, 648–657.
[30] Wentao Wu, Hongsong Li, Haixun Wang, and Kenny Q Zhu. 2012. Probase: A probabilistic taxonomy for text understanding. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*. ACM, 481–492.
[31] Liang Yao, Chengsheng Mao, and Yuan Luo. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7370–7377.
[32] Feng Yu, Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2017. A convolutional approach for misinformation identification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. AAAI Press, 3901–3907.
[33] Huaiwen Zhang, Quan Fang, Shengsheng Qian, and 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*. 1942–1951.
[34] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting context for rumour detection in social media. In *International Conference on Social Informatics*. Springer, 109–123.