

Compare to The Knowledge: Graph Neural Fake News Detection with External Knowledge

Linmei Hu¹, Tianchi Yang¹, Luhao Zhang², Wanjun Zhong³, Duyu Tang⁴,
Chuan Shi^{1*}, Nan Duan⁴, Ming Zhou⁴

¹School of Computer Science, Beijing University of Posts and Telecommunications

²Meituan ³Sun Yat-sen University ⁴Microsoft Research Asia

{hulinmei, yangtianchi, shichuan}@bupt.edu.cn

zhangluhao@meituan.com

zhongwj25@mail2.sysu.edu.cn

{dutang, nanduan, mingzhou}@microsoft.com

Abstract

Nowadays, fake news detection, which aims to verify whether a news document is trusted or fake, has become urgent and important. Most existing methods rely heavily on linguistic and semantic features from the news content, and fail to effectively exploit **external knowledge** which could help determine whether the news document is trusted. In this paper, we propose a novel end-to-end graph neural model called CompareNet, which compares the news to the knowledge base (KB) through entities for fake news detection. Considering that fake news detection is correlated with topics, we also incorporate topics to enrich the news representation. Specifically, we first construct a *directed heterogeneous document graph* for each news incorporating topics and entities. Based on the graph, we develop a *heterogeneous graph attention network* for learning the topic-enriched news representation as well as the contextual entity representations that encode the semantics of the news content. The contextual entity representations are then compared to the corresponding KB-based entity representations through a carefully designed *entity comparison network*, to capture the consistency between the news content and KB. Finally, the topic-enriched news representation combining the entity comparison features are fed into a fake news classifier. Experimental results on two benchmark datasets demonstrate that CompareNet significantly outperforms state-of-the-art methods.

1 Introduction

With the rapid development of the Internet, there are increasingly huge opportunities for fake news

production, dissemination and consumption. Fake news are news documents that are intentionally and verifiably false, and could mislead readers (Allcott and Gentzkow, 2017). Fake news can easily misguide public opinion, cause the crisis of confidence, and disturb the social order (Vosoughi et al., 2018). It is well known that fake news exerted an influence in the past 2016 US presidential elections (Allcott and Gentzkow, 2017). Thus, it is very important to develop effective methods for early fake news detection based on the textual content of the news document.

Some existing fake news detection methods rely heavily on various hand-crafted linguistic and semantic features for differentiating between news documents (Conroy et al., 2015; Rubin et al., 2016; Rashkin et al., 2017; Khurana and Intelligentie, 2017; Shu et al., 2020). To avoid feature engineering, deep neural models such as Bi-LSTM and convolutional neural networks (CNN) have been employed (Oshikawa et al., 2020; Wang, 2017; Rodríguez and Iglesias, 2019). However, they fail to consider the sentence interactions in the document. Vaibhav et al. showed that trusted news and fake news have different patterns of sentence interactions (Vaibhav et al., 2019). They modeled a news document as a fully connected sentence graph and proposed a graph attention model for fake news detection. Although these existing approaches can be effective, they fail to fully exploit external KB which could help determine whether the news is fake or trusted.

External KB such as Wikipedia contains a large amount of high-quality structured subject-predicate-object triplets and unstructured entity descriptions, which could serve as evidence for detecting fake news. As shown in Figure 4, the news

The work was done while visiting Microsoft Research Asia.

document about “mammograms are not effective at detecting breast tumors” is likely to be detected as fake news with the knowledge that “The goal of mammography is the early detection of breast cancer” in the Wikipedia entity description page¹. Pan et al. proposed to construct knowledge graphs from positive and negative news, and apply TransE to learn triplet scores for fake news detection (Pan et al., 2018). Nevertheless, the performance is largely influenced by construction of the knowledge graph. In this paper, to take full advantage of the external knowledge, we propose a novel end-to-end graph neural model **CompareNet** which directly compares the news to the KB through entities for fake news detection. In CompareNet, we also consider using topics to enrich the news document representation for improving fake news detection, since fake news detection and topics are highly correlated (Zhang et al., 2020; Jin et al., 2016). For example, the news documents in the “health” topic are inclined towards false, while the documents belonging to the “economy” topic are biased to be trusted instead.

Particularly, we first construct a *directed heterogeneous document graph* for each news document, containing sentences, topics and entities as nodes. The sentences are fully connected in bi-direction. Each sentence is also connected with its top relevant topics in bi-direction. If a sentence contains an entity, one directed link is built from the sentence to the entity. The reason for building one-way links from sentences to entities is to ensure that we can learn contextual entity representations that encode the semantics of the news, while avoiding the influence of the true entity knowledge to the news representation. Based on the *directed heterogeneous document graph*, we develop a *heterogeneous graph attention network* to learn topic-enriched news representations and contextual entity representations. The learned contextual entity representations are then compared to the corresponding KB-based entity representations with a carefully designed *entity comparison network*, in order to capture the semantic consistency between the news content and external KB. Finally, the topic-enriched news representations and the entity comparison features are combined for fake news classification. To facilitate related researches, we release both our code and dataset to the public².

In summary, our main contributions include:

- 1) In this paper, we propose a novel end-to-end graph neural model CompareNet which compares the news to the external knowledge through entities for fake news detection.
- 2) In CompareNet, we also consider the useful topic information. We construct a *directed heterogeneous document graph* incorporating topics and entities. Then we develop heterogeneous graph attention networks to learn topic-enriched news representations. A novel entity comparison network is designed to compare the news to the KB.
- 3) Extensive experiments on two benchmark datasets demonstrate that our model significantly outperforms state-of-the-art models on fake news detection by effectively incorporating external knowledge and topic information.

2 Related Work

Fake news detection has attracted much attention in recent years (Zhou and Zafarani, 2020; Oshikawa et al., 2020). A lot of works also focus on the related problem, i.e., fact checking, which aims to search evidence from external knowledge to verify the veracity of a claim (e.g., a subject-predicate-object triple) (Thorne et al., 2018; Zhou et al., 2019; Zhong et al., 2020). Generally, fake news detection usually focuses on news events while fact-checking is broader (Oshikawa et al., 2020). The approaches for fake news detection can be divided into two categories: social-based and content-based.

2.1 Social-based Fake News Detection

Social context related to news documents contains rich information such as user profiles and social relationships to help detect fake news. Social based models basically include stance-based and propagation-based. Stance-based models utilize users’ opinions to infer news veracity (Jin et al., 2016; Wu et al., 2019). Tacchini et al. constructed a bipartite network of user and posts with ‘like’ stance information, and proposed a semi-supervised probabilistic model to predict the likelihood of posts being hoaxes (Tacchini et al., 2017). Propagation-based approaches for fake news detection are based on the basic assumption that the credibility of a news event is highly related to the credibilities of relevant social media posts. Both

¹<https://en.wikipedia.org/wiki/Mammography>

²<https://github.com/ytc272098215/FakeNewsDetection>

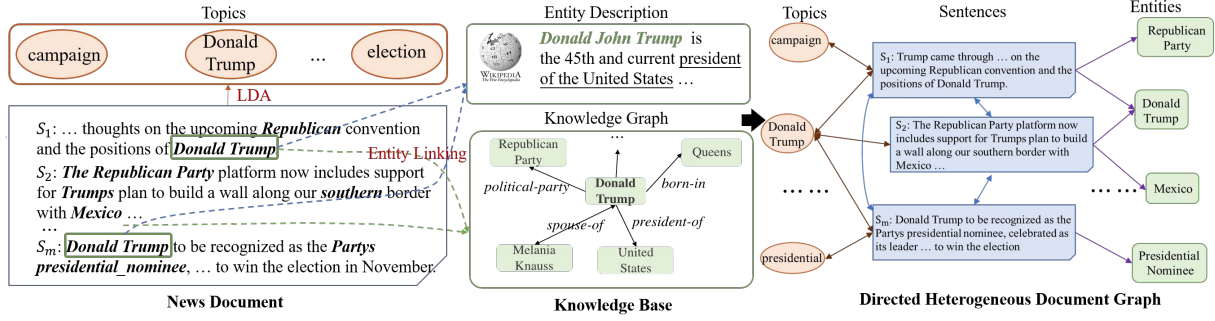


Figure 1: An example of directed heterogeneous document graph incorporating topics and entities.

homogeneous (Jin et al., 2016) and heterogeneous credibility networks (Gupta et al., 2012; Shu et al., 2019; Zhang et al., 2020) have been built to model the propagation process. For instance, (Zhang et al., 2020) constructed a heterogeneous network of news articles, creators and news subjects, and proposed a deep diffusive network model for incorporating the network structure information to simultaneously detect fake news articles, creators and subjects.

2.2 Content-based Fake News Detection

On the other hand, news contents contain the clues to differentiate fake and trusted news. A lot of existing works extract specific writing styles such as lexical and syntactic features (Conroy et al., 2015; Rubin et al., 2016; Khurana and Intelligentie, 2017; Rashkin et al., 2017; Shu et al., 2020; Oshikawa et al., 2020) and sensational headlines (Potthast et al., 2018; Sitaula et al., 2019) for fake news classifier. To avoid hand-crafted feature engineering, neural models have been proposed (Wang, 2017; Rodríguez and Iglesias, 2019). For example, Ibrain et al. applied deep neural networks, such as Bi-LSTM and convolutional neural networks (CNN) for fake news detection (Rodríguez and Iglesias, 2019). However, these works fail to consider different sentence interaction patterns between trusted and fake news documents. Vaibhav et al. proposed to model a document as a sentence graph capturing the sentence interactions and applied graph attention networks for learning document representation (Vaibhav et al., 2019). Pan et al. proposed to construct knowledge graphs from positive and negative news, and apply TransE to learn triplet scores for fake news detection (Pan et al., 2018). Nevertheless, they relied heavily on the quality of the construction of knowledge graphs. In this paper, we propose a novel graph neural model Com-

pareNet which directly compares the news to external knowledge for fake news detection. Considering that the detection of fake news is correlated with topics, we also use topics to enrich the news representation for improving fake news detection.

Some works (Wang, 2017; Khattar et al., 2019; Wang et al., 2020) also consider incorporating multi-modal features such as images for improving fake news detection.

3 Our Proposed CompareNet

In this section, we detail our proposed fake news detection model CompareNet, which directly compares the news to external knowledge for fake news detection. As shown in Figure 2, we also consider topics for enriching news representation since fake news detection is highly correlated with topics (Zhang et al., 2020). Specifically, we first construct a *directed heterogeneous document graph* for each news document incorporating topics and entities as shown in Figure 1. The graph well captures the interactions among sentences, topics and entities. Based on the graph, we develop a *heterogeneous graph attention network* to learn the topic-enriched news representation as well as the contextual entity representations that encode the semantics of the news document. To fully leverage external KB, we take the entities as the bridge between the news document and the KB. We compare the contextual entity representations with the corresponding KB-based entity representations using a carefully designed *entity comparison network*. Finally, the obtained entity comparison features are combined with the topic-enriched news document representation for fake news detection.

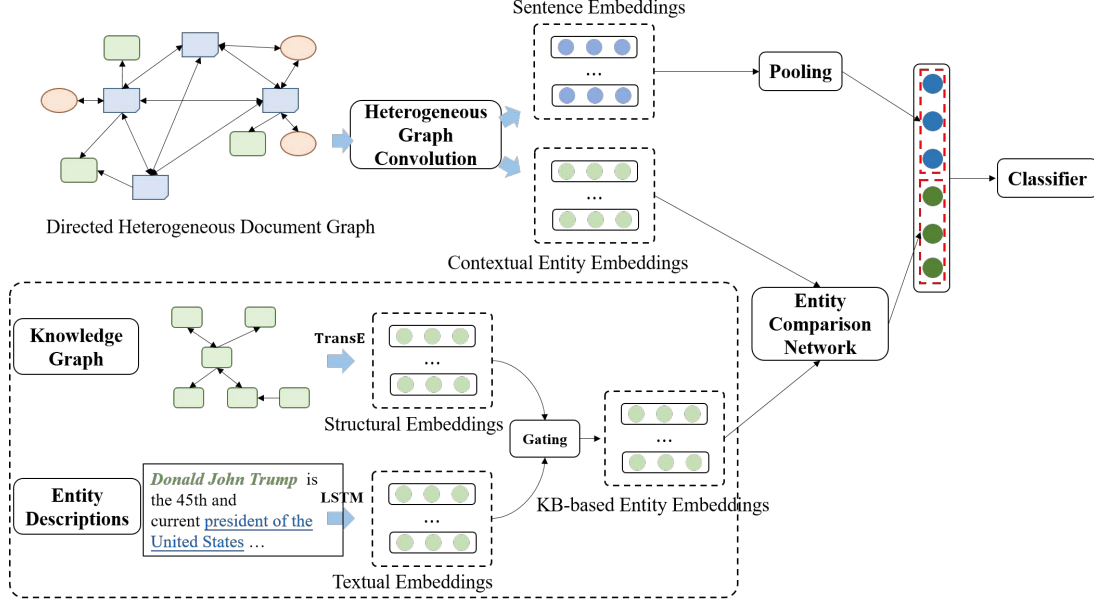


Figure 2: The overview of our proposed model CompareNet.

3.1 Directed Heterogeneous Document Graph

For each news document d , we construct a directed heterogeneous document graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ incorporating topics and entities, as shown in Figure 1. There are three kinds of nodes in the graph: sentences $S = \{s_1, s_2, \dots, s_m\}$, topics $T = \{t_1, t_2, \dots, t_K\}$ and entities $E = \{e_1, e_2, \dots, e_n\}$, i.e., $\mathcal{V} = S \cup T \cup E$. The set of edges \mathcal{E} represent the relations among sentences, topics and entities. The details of constructing the graph are described as follows.

We first split the news document as a set of sentences. Sentences are bidirectionally connected with each other in the graph, capturing the interaction of each sentence with every other sentence. Since topic information is important for fake news detection (Zhang et al., 2020), we apply the unsupervised LDA (Blei et al., 2003) (the total topic number K is set as 100) to mine the latent topics T from all the sentences of all the documents in our dataset. Specifically, each sentence is taken as a pseudo-document and is assigned to the top P relevant topics with the largest probabilities. Thus, each sentence is also connected with its top P assigned topics in bi-direction, allowing the useful topic information to propagate among the sentences. Note that we can also deal with new coming news documents by inferring the topics with trained LDA. We identify the entities E in the document d and map them to Wikipedia using the entity linking

tool TAGME³. If a sentence s contains an entity e , we build a *one-way directed edge* from a sentence to the entity e , in order to allow only information propagation from sentences to entities. In this way, we can avoid integrating *true* entity knowledge directly into news representation, which may mislead the detection of fake news.

3.2 Heterogeneous Graph Convolution

Based on the above directed heterogeneous document graph \mathcal{G} , we develop a heterogeneous graph attention network for learning the news representation as well as the contextual entity representations. It considers not only the weights of different nodes with different types (Hu et al., 2019) but also the edge directions in the heterogeneous graph.

Formally, we have three types $\mathcal{T} = \{\tau_1, \tau_2, \tau_3\}$ of nodes: sentences S , topics T and entities E with different feature spaces. We apply LSTM to encode a sentence $s = \{w_1, \dots, w_m\}$ and get its feature vector $\mathbf{x}_s \in \mathbb{R}^M$. The entity $e \in E$ is initialized with the entity representations $\mathbf{e}_{KB} \in \mathbb{R}^M$ learned from the external KB (see Subsection 3.3.1). The topic $t \in T$ is initialized with one-hot vector $\mathbf{x}_t \in \mathbb{R}^K$.

Next, consider the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where \mathcal{V} and \mathcal{E} represent the set of nodes and edges respectively. Let $X \in \mathbb{R}^{|\mathcal{V}| \times M}$ be a matrix containing the nodes with their features $\mathbf{x}_v \in \mathbb{R}^M$ (each row \mathbf{x}_v is a feature vector for a node v). A and D are

³<https://sobigdata.d4science.org/group/tagme/>

the adjacency matrix and the degree matrix, respectively. The heterogeneous convolution layer updates the $(l + 1)$ -th layer representation of the nodes $\mathbf{H}^{(l+1)}$ by aggregating the features of their neighboring nodes $\mathbf{H}_\tau^{(l)}$ with different types τ . (Initially, $\mathbf{H}^{(0)} = X$):

$$\mathbf{H}^{(l+1)} = \sigma\left(\sum_{\tau \in \mathcal{T}} \mathcal{B}_\tau \cdot \mathbf{H}_\tau^{(l)} \cdot \mathbf{W}_\tau^{(l)}\right), \quad (1)$$

where $\sigma(\cdot)$ denotes the activation function. Nodes with different types τ have different transformation matrix $\mathbf{W}_\tau^{(l)}$. The transformation matrix $\mathbf{W}_\tau^{(l)}$ considers the different feature spaces and projects them into an implicit common space. $\mathcal{B}_\tau \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}_\tau|}$ is the attention matrix, whose rows represent all the nodes and columns represent their neighboring nodes with the type τ . Its element $\beta_{vv'}$ in the v -th row and the v' -th column is computed as follows:

$$\beta_{vv'} = \text{Softmax}_{v'}(\sigma(\nu^T \cdot \alpha_\tau [\mathbf{h}_v, \mathbf{h}_{v'}])), \quad (2)$$

where ν is the attention vector and α_τ is the type-level attention weight. \mathbf{h}_v and $\mathbf{h}_{v'}$ are respectively the representation of the current node v and its neighboring node v' . Softmax function is applied to normalize across the neighboring nodes of node v .

We calculate the type-level attention weights α_τ based on the current node embedding \mathbf{h}_v and the type embedding $\mathbf{h}_\tau = \sum_{v'} \tilde{\mathbf{A}}_{vv'} \mathbf{h}_{v'}$ (the weighted sum of the neighboring node embeddings $\mathbf{h}_{v'}$ with the type τ , where the weight matrix $\tilde{\mathbf{A}} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$ is the normalized adjacency matrix with added self-connections) as follows:

$$\alpha_\tau = \text{Softmax}_\tau(\sigma(\mu_\tau^T \cdot [\mathbf{h}_v, \mathbf{h}_\tau])), \quad (3)$$

where μ_τ is the attention vector for the type τ . Softmax function is applied to normalize across all the types.

After L -layer heterogeneous graph convolution, we can finally get all the node (including sentences and entities) representations aggregating neighborhood semantics. We use max pooling over the representations of the sentence nodes $\mathbf{H}_s \in \mathbb{R}^N$ to obtain the final topic-enriched news document embedding $\mathbf{H}_d \in \mathbb{R}^N$. The learned entity representations that encode the contextual semantics of the document are taken as contextual entity representations $\mathbf{e}_c \in \mathbb{R}^N$.

3.3 Entity Comparison Network

In this subsection, we detail our entity comparison network which compares the learned contextual entity embeddings \mathbf{e}_c to the corresponding KB-based entity embeddings \mathbf{e}_{KB} . We believe entity comparison features could improve fake news detection based on the assumption that \mathbf{e}_c learned from trusted news document can be better aligned with the corresponding \mathbf{e}_{KB} ; while inverse for fake news.

3.3.1 KB-based Entity Representation

We first illustrate how to take full advantage of both *structured* subject-predicate-object triplets and *unstructured* textual entity descriptions in the KB (i.e., Wikipedia) to learn KB-based entity representations \mathbf{e}_{KB} .

Structural Embedding. A wide range of knowledge graph embedding methods can be applied to obtain structured entity embeddings. Due to the simplicity of TransE (Bordes et al., 2013), we adopted TransE to learn entity representations $\mathbf{e}_s \in \mathbb{R}^M$ from the triplets. Formally, given a triplet (h, r, t) , TransE regards a relationship r as a translation vector \mathbf{r} from the head entity h to the tail entity t , namely $\mathbf{h} + \mathbf{r} = \mathbf{t}$.

Textual Embedding. For each entity, we take the first paragraph of the corresponding Wikipedia page as its text description. Then we apply LSTM (Hochreiter and Schmidhuber, 1997) to learn entity representations $\mathbf{e}_d \in \mathbb{R}^M$ that encode the entity descriptions.

Gating Integration. Since both the structural triplets and textual description provide valuable information for an entity, we integrate these information into a joint representation. Particularly, as we have the structural embedding \mathbf{e}_s and textual embedding \mathbf{e}_d , we adopt a learnable gating function to integrate entity embeddings from the two sources. Formally,

$$\mathbf{e}_{\text{KB}} = \mathbf{g}_e \odot \mathbf{e}_s + (\mathbf{1} - \mathbf{g}_e) \odot \mathbf{e}_d, \quad (4)$$

where $\mathbf{g}_e \in \mathbb{R}^M$ is a gating vector (w.r.t. the entity e) to trade-off information from the two sources and its elements are in $[0, 1]$. \odot denotes element-wise multiplication. The gating vector \mathbf{g}_e means that each dimension of \mathbf{e}_s and \mathbf{e}_d are summed by different weights. To constrain the value of each element in $[0, 1]$, we compute the gate \mathbf{g}_e with the Sigmoid function:

$$\mathbf{g}_e = \sigma(\tilde{\mathbf{g}}_e), \quad (5)$$

where $\tilde{\mathbf{g}}_e \in \mathbb{R}^M$ is a real-value vector and is learned in the training process.

After fusing the two types of embeddings with the gating function, we obtain the final KB-based entity embeddings $\mathbf{e}_{KB} \in \mathbb{R}^M$ which encode both structural information from the triplets and textual information from the entity descriptions in the KB.

3.3.2 Entity Comparison

We then perform entity-to-entity comparison between the news document and the KB, to capture the semantic consistency between the news content and the KB. We calculate a comparison vector \mathbf{a}_i between each contextual entity representation $\mathbf{e}_c \in \mathbb{R}^N$ and its corresponding KB-based entity embedding $\mathbf{e}_{KB} \in \mathbb{R}^M$.

$$\mathbf{a}_i = f_{\text{cmp}}(\mathbf{e}_c, \mathbf{W}_e \cdot \mathbf{e}_{KB}), \quad (6)$$

where $f_{\text{cmp}}()$ denotes the comparison function, and $\mathbf{W}_e \in \mathbb{R}^{N \times M}$ is a transformation matrix. To measure the embedding closeness and relevance (Shen et al., 2018), we design our comparison function as:

$$f_{\text{cmp}}(x, y) = W_a[x - y, x \odot y], \quad (7)$$

where $W_a \in \mathbb{R}^{N \times 2N}$ is a transformation matrix and \odot is hadamard product, i.e., element-wise product. The final output comparison feature vector $\mathbf{C} \in \mathbb{R}^N$ is obtained by the max pooling over the alignment vectors $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ of all the entities $E = \{e_1, e_2, \dots, e_n\}$ in the news document.

3.4 Model Training

After obtaining the comparison vector $\mathbf{C} \in \mathbb{R}^N$ and the final news document representation vector $\mathbf{H}_d \in \mathbb{R}^N$, we concatenate and feed them into a Softmax layer for fake news classification. Formally,

$$Z = \text{Softmax}(\mathbf{W}_o[\mathbf{H}_d, \mathbf{C}] + \mathbf{b}_o), \quad (8)$$

where \mathbf{W}_o and \mathbf{b}_o are the parameter matrix and vection of a linear transformation. During model training, we exploit the cross-entropy loss over the training data with the L2-norm of the parameters:

$$\mathcal{L} = - \sum_{i \in D_{\text{train}}} \sum_{j=1} Y_{ij} \cdot \log Z_{ij} + \eta \|\Theta\|_2, \quad (9)$$

where D_{train} is the set of news documents for training, Y is the corresponding label indicator matrix, Θ is the model parameters, and η is regularization factor. For model optimization, we adopt the gradient descent algorithm.

4 Experiments

We conduct extensive experiments across various settings and datasets. Following the previous work (Vaibhav et al., 2019), we use SLN: Satirical and Legitimate News Database (Rubin et al., 2016), and LUN: Labeled Unreliable News Dataset (Rashkin et al., 2017) for our experiments. Table 1 shows the statistics.

Our baseline models include deep neural models: LSTM (Hochreiter and Schmidhuber, 1997), CNN (Kim, 2014), BERT+LSTM (Vaibhav et al., 2019) (BERT for sentence encoder and then LSTM for document encoder) and BERT (Devlin et al., 2019) (directly for document encoder). We also compare our model with graph neural models: GCN and GAT based on an undirected fully-connected sentence graph, which use attention pooling or max pooling for learning news document representation. For fair comparison with the previous work (Vaibhav et al., 2019), we use LSTM to encode sentences with randomly initialized word embeddings, which is the same as all the graph neural baselines. We run our model 5 times and report the micro-averaged (Precision = Recall = F1) and macro-averaged scores (Precision, Recall, F1) in all the settings including 2-way and 4-way classification.

2-way classification: We use the satirical and trusted news articles from LUN-train for training, LUN-test for validation and evaluate our model on the entire SLN dataset. This is done to emulate a real-world scenario where we want to see the performance of our model on an out-of-domain dataset.

4-way classification: We split the LUN-train into a 80:20 split to create our training and validation set. We use the LUN-test as our in-domain test set.

Experimental Setting. In our experiments, we set the number of topics $K = 100$ in LDA. Each sentence is assigned to top $P = 2$ topics with the largest probabilities. The layer number of our heterogeneous graph convolution is set as $L = 1$. These parameters are chosen according to the best experimental results on validation set. The other hyper-parameters are set as the same as the baseline (Vaibhav et al., 2019) for fair comparison. Specifically, all the hidden dimensions used in our model are set as $M = 100$. The node embedding dimension $N = 32$. For GCN, GAT and CompareNet, we set the activation function as LeakyReLU with slope 0.2. For model training, we train the models for a

Dataset	Trusted (#Docs)	Satire (#Docs)	Hoax (#Docs)	Propaganda (#Docs)
LUN-train	GN except 'APW' and 'WPB' (9,995)	The Onion (14,047)	American News (6,942)	Activist Report (17,870)
LUN-test	GN only 'APW' and 'WPB' (750)	The Borowitz Report, Clickhole (750)	DC Gazette (750)	The Natural News (750)
SLN	The Toronto Star, The NY Times (180)	The Onion, The Beaverton (180)	-	-

Table 1: Statistics of the datasets. GN refers to Gigaword News.

maximum of 15 epochs and use Adam optimizer with learning rate 0.001. We set L2 normalization factor η as $1e-6$.

4.1 Overall Results

Table 2 shows the results for the two-way classification between satirical and trusted news articles. We report only micro F1 since micro Precision=Recall=F1. As we can see, our proposed model CompareNet significantly outperforms all the state-of-the-art baselines in terms of all the metrics. Compared to the best baseline model, CompareNet improves both micro F1 and macro F1 by nearly 3%. We can also find that the graph neural network based models GCN and GAT all perform better than the deep neural models including CNN, LSTM and BERT. The reason is that the deep neural models fail to consider the interactions between sentences, which is important for fake news detection since different interaction patterns are observed in trusted and fake news documents (Vaibhav et al., 2019). Our model CompareNet further improves fake news detection by effectively exploiting the topics as well as the external KB. The topics enrich the news representation, and the external KB offers evidences for fake news detection.

We also present the results of four-way classification in Table 3. Consistently, all graph neural models capturing sentence interactions outperform the deep neural models. Our model CompareNet achieves the best performance in terms of all metrics. We believe that our model CompareNet benefits from the topics and external knowledge.

4.2 Ablation Study

In this subsection, we conduct experiments to study the effectiveness of each module in CompareNet and the way we incorporate external knowledge. We study the average performance of 5 runs on the LUN-test set. As shown in Table 4, we test the performance of CompareNet removing structured triplets, removing the entire external knowledge, removing topics, and removing both topics and external knowledge. In the last two rows, we further

Model	Micro	Macro		
	F1	Prec	Recall	F1
CNN	67.50	67.79	67.50	67.37
LSTM	81.11	82.12	81.11	80.96
BERT+LSTM	75.83	76.62	75.83	75.65
BERT	84.16	84.73	84.16	84.10
(Rubin et al., 2016)	-	88.00	82.00	-
GCN + Max	85.83	86.16	85.83	85.80
GCN + Attn	85.27	85.59	85.27	85.24
GAT + Max	86.39	86.44	86.38	86.38
GAT + Attn (2019)	84.72	85.65	84.72	84.62
CompareNet	89.17	89.82	89.17	89.12

Table 2: 2-way classification results on SLN dataset.

Model	Micro	Macro		
	F1	Prec	Recall	F1
CNN	54.03	54.50	54.03	52.60
LSTM	55.06	58.88	55.06	52.50
BERT+LSTM	55.56	57.45	54.86	54.00
BERT	64.66	60.89	64.46	58.80
(Rashkin et al., 2017)	-	-	-	65.00
GCN + Max	65.00	66.75	64.84	63.79
GCN + Attn	67.08	68.60	67.00	66.42
GAT + Max	65.50	69.45	65.33	63.83
GAT + Attn (2019)	66.95	68.05	66.86	66.37
CompareNet	69.05	72.94	69.04	68.26

Table 3: 4-way classification results on LUN dataset.

examine the constructed *directed heterogeneous document graph* and the designed entity comparison function. The variant CompareNet (undirected) does not consider the edge directions of the directed heterogeneous document graph. The variant model CompareNet (concatenation) replaces the entity comparison function as the simple concatenation operation. As we can see from Table 4, removing structural entity knowledge (i.e., w/o Structured Triplets) leads to slight performance drop. If we remove the entire external knowledge (i.e., w/o Entity Cmp), the performance decreases by around 1.3% and 1.8% on micro F1 and macro F1, respectively. Removing topics (i.e., w/o topics) will comparably impair the performance, which shows that the topic

Variants	Micro	Macro		
	F1	Prec	Recall	F1
CompareNet	69.05	72.94	69.04	68.26
- w/o Structured Triplets	68.74	69.34	68.79	68.17
- w/o Entity Cmp	67.46	70.38	67.43	66.35
- w/o Topics	67.40	69.75	67.41	66.73
- w/o Both	65.00	66.75	64.84	63.79
CompareNet (undirected)	66.35	68.11	66.36	65.74
CompareNet (concatenation)	67.40	70.05	67.39	66.25

Table 4: Ablation study of modules.

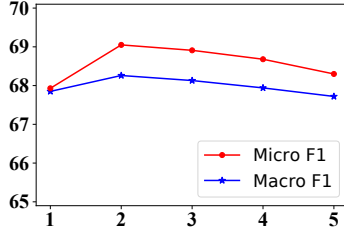


Figure 3: Effect of top assigned topic number P .

information is as important as the external knowledge. Removing both topics and external knowledge (i.e., w/o Both) will lead to substantial performance drop (4.0-5.0%). It demonstrates the importance of both topics and external knowledge. The variant model CompareNet (undirected) although incorporating both topics and external knowledge achieves lower performance than CompareNet w/o Entity Cmp and CompareNet w/o Topics. The reason could be that CompareNet (undirected) directly aggregates the *true* entity knowledge into the news representation in graph convolution without considering the directed edges, which misleads the classifier for differentiating fake news. This verifies the appropriateness of our constructed directed heterogeneous document graph. The last variant CompareNet (concatenation) also performs lower than CompareNet w/o Entity Cmp, further indicating that directly concatenating *true* entity knowledge is not a good way for incorporating entity knowledge. Its performance drops by around 2.0% compared to CompareNet. These demonstrate the effectiveness of the carefully designed entity comparison network in CompareNet.

4.3 Analysis of Top Assigned Topic Number

Figure 3 shows the performance (micro and macro F1) of our model CompareNet on LUN validation set with different number of top assigned topics P to each sentence. As we can see clearly, micro F1 and macro F1 first consistently rises with the increase of P and then drops when P is larger than

News	Entity Description
that may easily be misused by the FDA to target and threaten the natural health community... the FDA could have illegitimately used it to target practically any company it wanted to.	... The FDA is responsible for protecting and promoting public health through the control and supervision of food safety, tobacco products, dietary supplements ...
... women referred to oncologists for treatment after mammograms did not actually have cancer. ... mammograms are not effective at detecting breast tumors ...	Mammography is the process of using low-energy X-rays to examine the human breast for diagnosis and screening. The goal of mammography is the early detection of breast cancer ...

Figure 4: Two news examples from the LUN-test set.

2. This may be because that connecting too many low-probability topics will introduce some noise. Thus, in our experiments, we set $P = 2$.

4.4 Case Study

To further illustrate why our model outperforms state-of-the-art baseline GAT+Attn (Vaibhav et al., 2019), we present two real news examples from the LUN-test set. The baseline model GAT+Attn and the variant model CompareNet w/o Entity Cmp mistakenly predict these two examples as trusted news, while our model CompareNet can successfully predict both of them. As we can see from Figure 4, the content of the news document is in conflict with the entity description from Wikipedia. Specifically, the news about “FDA target and threaten the natural health community” delivers contrary meaning from the entity description that “FDA is responsible for protecting and promoting public health”⁴. Similarly, the news document about “mammograms are not effective at detecting breast tumors” conveys different meaning from the entity description of “mammograms”. We believe that our model CompareNet benefits from the comparison to Wikipedia knowledge by the entity comparison network. We find there are also unsuccessful cases since an entity could be mistakenly linked to a wrong entity in the Wikipedia.

5 Conclusion

In this paper, we propose a novel end-to-end graph neural model CompareNet which compares the news to the external knowledge for fake news detection. Considering that the detection of fake news is correlated with topics, in our model, we also use topics to enrich the news document representation for improving fake news detection. Particularly, we first construct a directed heterogeneous document graph for each news document capturing the interactions among sentences, topics and entities.

⁴https://en.wikipedia.org/wiki/Food_and_Drug_Administration

Based on the graph, we develop a heterogeneous graph attention network for learning topic-enriched news representation as well as contextual entity representations that encode the semantics of the content of the news document. To capture the semantic consistency of the news content and the KB, the learned contextual entity representations are then compared to the KB-based entity representations, with a carefully designed entity comparison network. Finally, the obtained entity comparison features are combined with the news representation for an improved fake news classifier. Experiments on two benchmark datasets have demonstrated the effectiveness of the way we incorporate the external knowledge and topics.

In future work, we will explore a better way to combine multi-modal data (e.g., images) and external knowledge for fake news detection.

Acknowledgments

The work is supported by the National Natural Science Foundation of China (No. 61806020, U1936220, 61972047, 62076245) and the Microsoft Research Asia's Star Track project.

References

- Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–236.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26*, pages 2787–2795.
- Nadia K Conroy, Victoria L Rubin, and Yimin Chen. 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology*, 52(1):1–4.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186.
- Manish Gupta, Peixiang Zhao, and Jiawei Han. 2012. Evaluating event credibility on twitter. In *Proceedings of the Twelfth SIAM International Conference on Data Mining*, pages 153–164.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Linmei Hu, Tianchi Yang, Chuan Shi, Houye Ji, and Xiaoli Li. 2019. Heterogeneous graph attention networks for semi-supervised short text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4821–4830.
- Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 2972–2978.
- 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*, page 2915–2921.
- Urja Khurana and Bachelor Opleiding Kunstmatige Intelligentie. 2017. *The linguistic features of fake news headlines and statements*. Ph.D. thesis, Master's thesis, University of Amsterdam.
- Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751.
- Ray Oshikawa, Jing Qian, and William Yang Wang. 2020. A survey on natural language processing for fake news detection. *ArXiv*, abs/1811.00770.
- Jeff Z. Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. Content based fake news detection using knowledge graphs. In *The Semantic Web - ISWC 2018 - 17th International Semantic Web Conference*, volume 11136, pages 669–683.
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A stylistic inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, pages 231–240.
- Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2931–2937.
- Álvaro Ibraín Rodríguez and Lara Lloret Iglesias. 2019. Fake news detection using deep learning. *CoRR*, abs/1910.03496.
- Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, pages 7–17.

- Dinghan Shen, Xinyuan Zhang, Ricardo Henao, and Lawrence Carin. 2018. Improved semantic-aware network embedding with fine-grained word alignment. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1829–1838.
- Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3):171–188.
- Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 312–320.
- Niraj Sitaula, Chilukuri K. Mohan, Jennifer Grygiel, Xinyi Zhou, and Reza Zafarani. 2019. Credibility-based fake news detection. *CoRR*, abs/1911.00643.
- Eugenio Tacchini, Gabriele Ballarin, Marco L. Della Vedova, Stefano Moret, and Luca de Alfaro. 2017. Some like it hoax: Automated fake news detection in social networks. *CoRR*, abs/1704.07506.
- James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a large-scale dataset for fact extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 809–819.
- Vaibhav Vaibhav, Raghuram Mandyam, and Eduard Hovy. 2019. Do sentence interactions matter? leveraging sentence level representations for fake news classification. In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*, pages 134–139.
- Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.
- William Yang Wang. 2017. “liar, liar pants on fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 422–426.
- Youze Wang, Shengsheng Qian, Jun Hu, Quan Fang, and Changsheng Xu. 2020. Fake news detection via knowledge-driven multimodal graph convolutional networks. In *Proceedings of the 2020 International Conference on Multimedia Retrieval*, pages 540–547.
- Lianwei Wu, Yuan Rao, Haolin Jin, Ambreen Nazir, and Ling Sun. 2019. Different absorption from the same sharing: Sifted multi-task learning for fake news detection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 4643–4652.
- Jiawei Zhang, Bowen Dong, and Philip S. Yu. 2020. Fakedetector: Effective fake news detection with deep diffusive neural network. In *Proceedings of the 36th IEEE International Conference on Data Engineering*, pages 1826–1829.
- Wanjuan Zhong, Jingjing Xu, Duyu Tang, Zenan Xu, Nan Duan, Ming Zhou, Jiahai Wang, and Jian Yin. 2020. Reasoning over semantic-level graph for fact checking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6170–6180.
- Jie Zhou, Xu Han, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. 2019. GEAR: Graph-based evidence aggregating and reasoning for fact verification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 892–901.
- Xinyi Zhou and Reza Zafarani. 2020. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5):1–40.