

Received August 4, 2020, accepted August 20, 2020, date of publication August 26, 2020, date of current version September 9, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019586

Robust Reasoning Over Heterogeneous Textual Information for Fact Verification

YONGYUE WANG¹, CHUNHE XIA^{1,2}, (Member, IEEE), CHENGXIANG SI³,
BEITONG YAO¹, AND TIANBO WANG⁴, (Member, IEEE)

¹Beijing Key Laboratory of Network Technology, School of Computer Science and Engineering, Beihang University, Beijing 100191, China

²Guangxi Key Laboratory of Multi-Source Information Mining and Security, School of Computer Science and Information Technology, Guangxi Normal University, Guilin 541004, China

³National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China

⁴School of Cyber Science and Technology, Beihang University, Beijing 100191, China

Corresponding author: Tianbo Wang (wangtb@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1636208, Grant 61862008, and Grant 61902013; and in part by the Beihang Youth Top Talent Support Program under Grant YWF-20-BJ-J-1038.

ABSTRACT Automatic fact verification (FV) based on artificial intelligence is considered as a promising approach which can be used to identify misinformation distributed on the web. Even though previous FV using deep learning have made great achievements in single dataset (e.g., FEVER), the trained systems are unlikely to be capable of extracting evidence from heterogeneous web-sources and validating claims in accordance with evidence found on the Internet. Nevertheless, the heterogeneity covers abundant semantic information, which will help FV system identify misinformation in a more accurate way. The current work is the first attempt to make the combination of knowledge graph (KG) and graph neural network (GNN) to enhance the robustness of FV systems for heterogeneous information. As a result, it can be generalized to multi-domain datasets after training on a sufficient single one. To make information update and aggregate well on the collaborative graph, the present study proposes a double graph attention network (DGAT) framework which recursively propagates the embeddings from a node's neighbors to refine the node's embedding as well as applies an attention mechanism to classify the importance of the neighbors. We train and evaluate our system on FEVER, a single and benchmark dataset for FV, and then re-evaluate our system on UKP Snopes Corpus, a new richly annotated corpus for FV tasks on the basis of heterogeneous web sources. According to experimental results, although DGAT has no excellent advantages in a single dataset, it shows outstanding performance in more realistic and multi-domain datasets. Moreover, the current study also provides a feasible method for deep learning to have the ability to infer heterogeneous information robustly.

INDEX TERMS Fact verification, heterogeneous information, graph neural network, robust reasoning.

I. INTRODUCTION

With the explosive growth of Internet, fake news has already posed serious threats to the public's factual judgment and the credibility of the governments. Rumor or misinformation makers usually use creative language to embellish their content and attract netizens, which will hide this important features and thus traditional FV methods fail to recognize the misinformation [1].

Recent achievements in scattered information fusion technology and multi-hop reasoning method (such as natural

language inference) explains this problem, which improve the performance of fact verification (FV) by integrating the entire network information. That is to say, natural language inference (NLI) models can use scattered web information comprehensively, which will lead them to obtain sufficient web sources. For example, as shown in Figure 1, a FV system first searches related evidence sentences from one dataset, then conducts joint reasoning over these evidences, and finally aggregates the information to confirm the claim integrity.

NLI models either concatenate all relevant information (often referred to as evidence) into a single string, which is often used in FV systems in the FEVER challenge [2], [3],

The associate editor coordinating the review of this manuscript and approving it for publication was Arianna Dulizia.

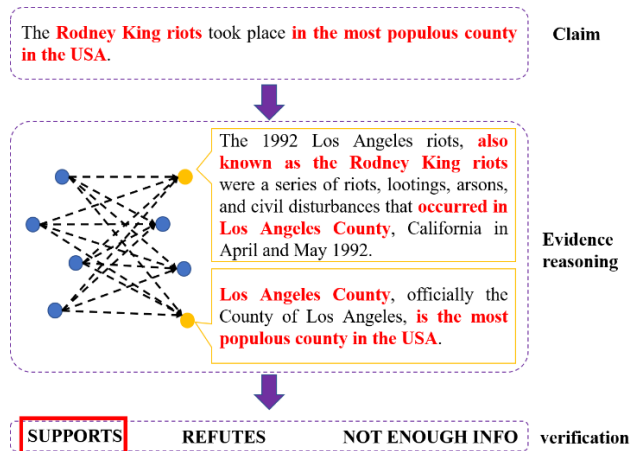


FIGURE 1. An example of fact verification system.

or adopt fusion technology to aggregate the features of isolated relevant information [4], [5]. However, the proposed traditional NLI methods fail to grasp the correlation between evidences, causing that these FV systems cannot make full use of the valuable evidences.

To make full use of the structural information of the provided context, some recent work has integrated the graph neural network into FV systems (see Section II-A for detail), and has also made a significant improvement. Both train and test of the recent methods are on a single dataset (e.g., Wikipedia) with the same structural information. This single dataset bases on synthetic information derived from Wikipedia rather than natural information that is originated from heterogeneous web sources. Web sources with heterogeneous datasets may contain more valuable information which can help better verify claims. Since the single datasets only cover Wikipedia, the trained systems are unlikely to be capable of extracting evidence from heterogeneous web-sources and validating claims on the basis of evidence found on the Internet. Though effective in single datasets, current methods are challenging to model the explicit relationship between heterogeneous information. The main challenge is how to effectively integrate large-scale textual information with different structure from various sources to recognize the embellished misinformation with hidden characteristics. Recently, although people have proposed numerous heterogeneous information fusion methods [6]–[12], the necessary condition of these methods is that the structures of the heterogeneous information are specific and known before training and testing, which is impracticable to apply in the unpredictable web source. Fortunately, the emerging success on knowledge-graph-based inference may explain this problem. Inspired by [13], [14] and based on state-of-the-art model [15], in the present study, we first proposed a novel double-graph-based reasoning framework over knowledge graphs for FV. The main building block is a collaborative knowledge graph (CKG) construction by combining the information of evidences and world knowledge as a unified graph. This new CKG can help NLI model

complete self-taught learning on heterogeneous datasets. Furthermore, we also adopt knowledge-aware attention [14] to propagate and aggregate information over the collaborative graph structure. In this way, the reasoning process simultaneously employs semantic and the relevant word knowledge of entity in this graph. We conduct experiments both on the most influential benchmark datasets FEVER [2] and richly annotated corpus based on heterogeneous web source UKP Snopes Corpus [16] for fact verification. Ablation study demonstrates that the integration of double-graph-driven representation learning mechanisms enhances the performance of reasoning on heterogeneous information. Our contributions are briefly summarized as follows:

- We first combine the intrinsic information of evidences and world knowledge with evidences as a unified relation graph and propose a double-graph-based graph reasoning approach to enhance the robustness of the FV system when inferring heterogeneous information.
- Results demonstrate that although our double-graph-based mechanisms have no excellent advantages in the single source dataset (FEVER), it achieves state-of-the-art performance on the UKP Snopes Corpus. Our study illustrates that the state-of-the-art methods on one dataset with the same structural information may not perform well on realistic scenario with heterogeneous data. In addition, we also provide a feasible method for deep learning to infer this heterogeneous information robustly.

II. RELATED WORK

A. ADVANCED AUTOMATIC FACT VERIFICATION

Reference [17] uses the attention model (DAM) [18] to produce NLI predictions for each claim(sentence to be verified)-evidence pair individually and aggregate all these predictions for final checking. Then, [19]–[21] employ a more effective NLI model enhanced by sequential inference model [22] to predict the relevance between individual evidence and claim. As pre-trained language models can fully grasp the semantic information of context, [23] adopts the generative pretraining transformer (GPT) [24] to enhance the understanding of claim-evidence logical relationships. Reference [25] improves the pre-trained process by updating Transformers, which can natively represent structure textual information. Using neural matching kernel, [26] proposes more fine-grained evidence selection and reasoning method KGAT for FV. To capture rich semantic-level structures among multiple evidences, [15] presents a graph-based approach for fact checking based on semantic role labeling (AllenNLP [27]) and achieves state-of-the-art performance.

B. FEVER: FACT EXTRACTION AND VERIFICATION

FEVER Shared task [2] aims to evaluate a FV system's ability to verify the claims based on evidences from Wikipedia. The dataset [3] for the current task consists of 185,445 claims. Given a claim, FV systems first extract relevant evidence

(the sentences supporting or refuting the claim). Then, NLI methods process the claim and evidence, and finally label the claim as Supported, Refuted or NotEnoughInfo. The formal expression of these verification process is presented as follows:

Suppose $\mathcal{D} = \{D_1, D_2 \dots\}$ denotes a set of documents, and one certain document D_i is an array of sentences, namely, $D_i = \{S_i^1, S_i^2 \dots\}$. The inputs of a FV system are a textual claim C_i and the relevant documents $\cup D_i$. The output should be a tuple (\hat{E}_i, \hat{y}_i) where $\hat{E}_i = \{s_i^{e_0}, s_i^{e_1}, \dots\} \subset \cup D_i$, denoting the set of evidence sentences specific to the claim, and $\hat{y}_i \in \{SUPPORTS, REFUTES, NOTENOUGHINFO\}$, the prediction to the given claim. Suppose the standard evidence and ground truth label of the claim are E_i and y_i . Regarding a successful verification of a claim c_i , the tuple (\hat{E}_i, \hat{y}_i) predicted by the FV system should satisfy the following conditions: $E_i \in \hat{E}_i$ and $y_i = \hat{y}_i$.

C. NATURAL LANGUAGE INFERENCE

Natural Language Inference (NLI) is a task related to the judgement of semantic relationship between a premise and a hypothesis [28]. Recently, the availability of much larger annotated language corpus, e.g., Multi-NLI [29] and SNLI [30], have made it practicable to train more complicated deep learning models which have achieved state-of-the-art performance [31]–[33]. NLI which is regarded as a typical semantic matching problem almost exists in every NLP task. NLI is targeted at comprehensively analyzing scattered sentences, and then can infer the desired information. Coincidentally, the FV requires the comprehensive use of web information resources as the evidence to label the given claim. It is intuitive to transfer NLI models into the FV task and several methods based on NLI model, which has also achieved significant improvement in FEVER shared task. The current study enriches neural-network-based NLI models with external knowledge. Different from early study on NLI that analyzes internal knowledge of relatively small NLI datasets, the present work aims to merge the advantage of powerful modeling ability of neural networks with extra external robust knowledge so as to finish reasoning task on heterogeneous information. It can be proved that the proposed model enhances the state-of-the-art NLI models' adaptability in heterogeneous information.

D. GNN FOR NLI

To obtain both the global and local information representation of graph network, Graph Neural Networks (GNNs) learn nodes' effective feature vectors through a recursive neighborhood aggregation [34] on the information graph. Reference [35] proposes Graph Convolutional Network (GCN). Recently, self-attention mechanism has been introduced into GNNs (Graph Neural Network [36]) to encourage FV models to focus on the most significant parts of data, aiming to address the shortcomings of prior approaches based on GCNs. Therefore, in order to learn more information of the relationships between the given claim and evidence

sentences, [36] first proposes Graph Attention Network (GAT), introducing the attention mechanism into graphs learning. GAT is a new type of GCN, which uses attention mechanism for homogeneous graphs containing single type of nodes and edges. Although GAT has made great process in FV tasks, it has not been effectively applied to heterogeneous information [12]. Based on these two models, there exist a lot of improved methods such as [4], [5], [12], [15], [26], achieving good performance in NLI task. However, the above GNNs can only provide NLI with relation representation of one graph and cannot deal with various types of nodes and edges in one more different graph. The incorporation of external and inference-related knowledge graph in the process of natural language inference is studied in the current work. For example, intuitive knowledge concerning synonymy, antonymy, hypernymy and hyponymy at semantic level may help NLI models achieve soft-alignment between premises and hypotheses. Knowledge about hypernymy and hyponymy may help NLI models to capture entailment relation. Knowledge about antonymy and co-hyponyms (words sharing the same hypernym) may be beneficial for the modeling of contradiction.

III. METHODOLOGY

We adopt a three-step pipeline with components for document retrieval, sentence selection and claim verification to complete the FV task.

A. PIPELINE

In the stages of document retrieval and sentence selection, we simply employ the method from [5], since their method has achieved the best performance. Finally, the claim verification model outputs the veracity of the given claim after the FV system obtains the claim and evidential sentences. The overview pipeline of the proposed method is illustrated in Figure 2. In Figure 2, ev_i denotes the i -th evidential sentence, h_{cei} denotes the representation of i -th claim-evidence pair, MPL denotes Multilayer Perceptron, softmax denotes softmax function, KG denotes the Knowledge Graph, CKG denotes the Collaborative Knowledge Graph, h_{evi} denotes the representation of ev_i , $\sum \alpha h$ denotes the weighted sum of multiple representations an align denotes the align function. Based on majority of previous models, we first introduce a double graph attention network (DGAT) framework in the final claim verification stage, which is explained detailly in Section III-B. The main contribution of the current work is the double-graph-based reasoning approach for claim verification.

B. CLAIM VERIFICATION WITH GRAPH-BASED REASONING

The present section describes our DGAT framework for claim verification. According to Figure 2, given a claim and the retrieved evidence, we first employ a sentence encoder to achieve representations for the claim and the evidence. Subsequently, we build a fully-connected evidence graph which is

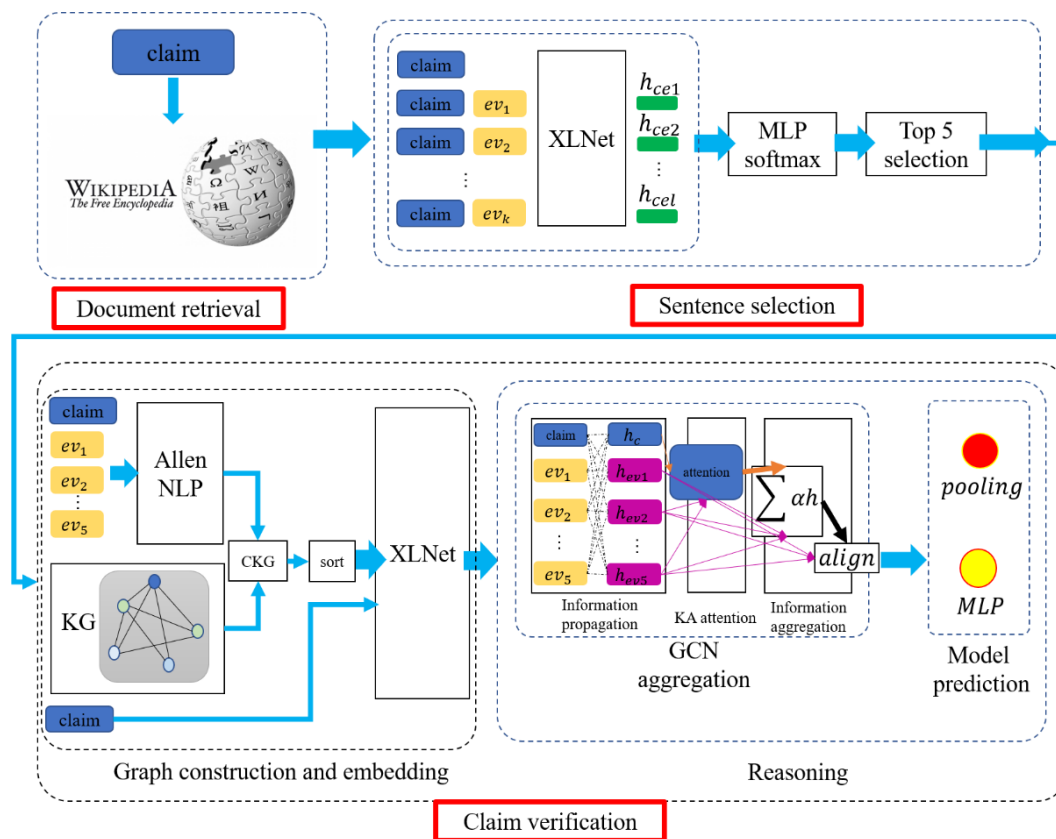


FIGURE 2. The pipeline of our method.

integrated with knowledge graph and put forward an evidence reasoning network to propagate information among evidence and reason over this collaborative graph. Eventually, we utilize an evidence aggregator to infer the final claim labels.

1) CONTEXTUAL SEMANTIC GRAPH CONSTRUCTION

After feeding the evidential sentences, to represent the structural information of these sentences, we construct a contextual semantic graph. During this process, we use a practical and ready-made labeling tool based on semantic role [37] to construct this semantic relation graph. Specifically, with the injected evidential sentences, this construction process will proceed in the following three steps.

1. We first use Semantic Role Labeling (SRL) toolkit [27], named AllenNLP, to parse each provided sentences (including the *claim* and evidence ev_i) into tuples. The accuracy of this module can reach 96.88%, trained on the CoNLL 2009 dataset.
2. Following the semantic-level graph construction process [15], we then use the specific type elements in the tuple described in the first step as nodes of the semantic graph. We attempt to set these types as subject, predicate, object, verb, tense, voice, position, and time. The study connects any two nodes as their edges within a tuple.

3. To learn the relationship between evidential sentences, we create edges across different tuples. [15] provides a very good idea: creating edges by finding similarity relationships across two different tuples. Assuming tuple A as the subject(s) and B as the object(o), this similarity relationship includes the following three aspects: (1) tuple A equals tuple B; (2) tuple A contains tuple B and (3) tuple A and tuple B have the number of overlapped tokens. However, this coarse evidential relationship across different sentences may lose a lot of important structural information. In order to refine the external relationship between evidences, we employ WordNet [38] to further find these relationships including synonymy, antonymy, hypernymy and meronymy.

This developed contextual semantic graph construction method provides an operational and more fine-grained strategy for representing structural information of multiple evidential sentences.

We denote the contextual semantic graph as G_1 . Next, we integrate world knowledge into the constructed semantic graph G_1 to improve the robustness of FV system's inference on heterogeneous information.

2) CONTEXTUAL KNOWLEDGE GRAPH CONSTRUCTION

In addition to the intrinsic information of evidences, we have to integrate world knowledge within evidences into the

cognitive graph. Typically, such auxiliary data consists of real-world entities and relationships among them to profile the evidence sentences at semantic level. We organize supplementary information in the form of knowledge graph G_2 , which is a directed graph composed of subject-property-object triple facts [39]. More formally, it is presented as $\{(h, r, t) | h, t \in E, r \in R\}$, where each triplet describes that there exists a relationship r from head entity h to tail entity t . For example, *(Beijing, CapitalOf, China)* points out the fact that Beijing is the capital of china. Note that R contains relations in both canonical direction (e.g., *CapitalOf*) and inverse direction (e.g., *MotherlandOf*). We train and evaluate this module that generate ConceptNet [40] relations, and this module achieved 92.7% accurate.

3) CONTEXTUAL COLLABORATIVE KNOWLEDGE GRAPH CONSTRUCTION

The present section defines an enhanced graph, named Collaborative Knowledge Graph (CKG), which encodes intrinsic information of evidence sentences and world knowledge related to evidences as a unified graph. We first represent each entity's entailment behavior as a triplet, $(s, predicate, o)$, where $y_{so} = 1$ denotes as an additional relation interact between subject s and object o . The intrinsic information of these evidence graph can be seamlessly integrated with KG as a unified graph

$$G = \{(h, r, t) | h, t \in E', r \in R'\}, \quad (1)$$

where $E' = E \cap S$ and $R' = r \cap \{predicate\}$.

- Input: a CKG (denoted G) that contains the intrinsic information of evidence graph G_1 and knowledge graph G_2 .
- Output: the probability of the labels for the given claim.

C. CONTEXTUAL WORD REPRESENTATIONS ENHANCED BY COLLABORATIVE KNOWLEDGE GRAPH

In the current section, we proposed to use CKG structural features to enhance contextual word representations. Specifically, in order to make better use of the CKG structural information, our idea is that the stronger semantic relationship in the CKG, the closer the relative position of the nodes. Therefore, we reordered the evidence sentences by employing a *sort* module based on [15].

Fortunately, the pre-trained language model XLNet [41] naturally includes the concept of relative position. Thus, to enhance the representation ability of the pre-trained model, we injected the relative position information of CKG nodes to the input of XLNet. TransR [42] learns the CKG representation of each entity and relation by optimizing the translation principle:

$$e_h^r + e_r \approx e_t^r. \quad (2)$$

If a triplet (h, r, t) exists in the CKG, where $e_h, e_t \in \mathbb{R}^d$ and $e_r \in \mathbb{R}^k$ are the embedding vector for h, t and r , respectively; e_h^r and e_t^r are the projected representations of e_h and e_t in the

relation r 's space. Therefore, for a given triplet (h, r, t) , its plausibility score is formulated as following:

$$g(h, r, t) = \|W_r e_h + e_r - W_r e_t\|_2^2, \quad (3)$$

where $W_r \in \mathbb{R}^{k \times d}$ denotes the transformation matrix of relation r , which projects entities from the d -dimension entity space into the k -dimension relation space. A lower score of $g(h, r, t)$ suggests that there exists strong relationship between the two entities. As a result, we can define the energy score as the relative position distance between the two entities. We train our representation module on FB15K [43], and the accuracy of triple classification can reach 91.2%

After reordering the sequences, we inject them into XLNet model to obtain the contextual representations. Finally, we obtain the contextual representations h ($[CLS]$). The h ($[CLS]$) captures the semantic interaction between the claim and the evidence at word level.

D. ATTENTIVE INFORMATION PROPAGATION LAYERS

After obtaining the contextual word representations, to further exploit graph information at the semantic level, we take advantage of the GCN [35] architecture to recursively propagate information along high-order connectivity. We closely follow the experimental setting in [35], achieving 80.4% classification accuracy on Cora. Moreover, in order to reveal the importance of different node connections, we employ graph attention network [36] to create attentive weights of cascaded propagations. We follow the setting of GAT [36], and the classification accuracy can reach 83.5%. We first describe a single layer architecture, which consists of three parts including information propagation, knowledge-aware attention, and information aggregation. Then, we will succinctly study how to extend it to multiple layers.

1) INFORMATION PROPAGATION

An entity in multiple triplets may act as the bridge, connecting two different triplets and propagating information. Our intuitive idea is to operate information propagation between one entity and its neighbors in the CKG.

Formally, we denote G as the collaborative graph and $H \in \mathbb{R}^{N^v \times d}$ as a matrix containing representation of all nodes, where N^v and d signify the number of nodes and the dimension of node representations, respectively. Each row $H_i \in \mathbb{R}^d$ is the representation of node i . We denote the adjacency of graph G as A and its degree matrix as D . Considering an entity h , we use

$$N_h = \{(h, r, t) | (h, r, t) \in G\} \quad (4)$$

to denote the set of triplets as termed ego-network [44], where h and t represent the head and tail entity respectively. In order to describe the first-order connectivity structure of entity h in a formal way, based on [44], we calculate the linear combination value of h 's ego-network:

$$e_{N_h} = \sum_{(h,r,t) \in N_h} \pi(h, r, t) e_t, \quad (5)$$

where $\pi(h, r, t)$ can be learned through datasets to control the decay factor on each edge (h, r, t) propagation, demonstrating how much information being propagated from t to h conditioned on r .

More generally, we organize the information propagation among the nodes into matrix form to facilitate GPU calculation. GCNs will aggregate information from the multi-hop neighboring nodes as following:

$$H_i^{(j+1)} = \rho\left(\tilde{A}H_i^{(j)}W_j\right), \quad (6)$$

where j denotes the layer number, $H_i^{(j+1)} \in R^d$ is the updated d -dimension representation of node i , \tilde{A} is the normalized symmetric adjacency matrix, W is the weight matrix of j -th layer, and ρ signifies an activation function. The propagation method learns the claim-based and evidence-based graph separately. Meanwhile, we obtain the representations of all nodes in claim-based graph H_c and evidence-based H_e graphs respectively.

2) KNOWLEDGE-AWARE (KA) ATTENTION

For the sake of simplicity, we use e_{ij} for $\pi(h, r, t)$. We conduct $\pi(h, r, t)$ based on attention mechanism, which is formulated as following:

$$e_{ij} = \left(W_r h_c^i\right)^T \tanh\left(W_r h_e^j + h_r\right), \quad (7)$$

where the function \tanh [36] is used as the nonlinear activation function. The \tanh nonlinear activation function is talented at making the attention degree dependent on the distance between H_e and H_c in the relation r 's space, which will propagate more instructive information for closer entities. Note that, we only use dot product on these representations progress for simplicity.

Then, we normalize the coefficients based on the soft-max function across all triplets connected with h :

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N_e^v} \exp(e_{ik})} \quad (8)$$

As a result, the final attention score has the ability to indicate which neighbor nodes should be paid more attention to capturing more useful CKG information. When operating propagation forward, the attention flow indicates the partial data given close attention, which can also be considered as explanations for the fact verification.

Different from the information propagation in GCN [35] and GraphSage [45], our model not only employs the proximity structure of CKG, but also specifically calculates the variable importance of neighbors. Additionally, different from graph attention network [15], [36] which only uses each node representation as the model's inputs, we calculate the relation h_r between h_c and h_e , which will obtain more information during the propagation progress.

3) INFORMATION AGGREGATION

After attention, we obtain claim-centric evidence information representation:

$$x_i = \sum_{j \in N_e^v} \alpha_{ij} h_e^j. \quad (9)$$

Besides, we use the matrix \mathbf{X} to denote the claim-centric evidence representation of all nodes in CKG.

To obtain the node-to-node alignment, based on fine-grained word alignment function [46], we employ the claim-centric evidence information representation and claim node representation, as shown below:

$$a_i = \text{align}\left(h_c^i, x_i\right). \quad (10)$$

Consequently, the aligned vector is $A = [a_1, \dots, a_{N_e^v}]$. We use the max pooling over A to obtain the final output q . Finally, q is concatenated and injected with final hidden vector h (CLS) to an MLP layer for the label prediction p .

IV. EXPERIMENTAL METHODOLOGY

In the current section, we describe the dataset, evaluation metrics, baselines, and implementation details in our experiments.

A. DATASETS

The success of FV system-centered experiments depends on the compliance of the corpus. First of all, considerable tagged cases should be introduced into discrepant FV sub-task. Secondly, given the possible sources of misconceptions varying from official announcements to blogs or Twitter posts., the training data should not be chosen from the same web source alone. Specifically, references related to FV can be drawn from the datasets (PolitiFact14 [45], Emergent16 [46] and RumourEval17 [47]). However, it is impractical to classify the deep learning model into the category of unobserved claims according to the training on these datasets which just incorporate hundreds of verified claims. In spite that the Snopes17 [48] covers more verified claims, most of the documents involved have nothing to do with the claim. In another word, it is devoid of useful information for verification. The FEVER [2], the largest dataset useful in the development of FV system, is composed of 185,455 tagged claims and a total of 5,416,537 Wikipedia documents. Users may also make adjustments on available sentences to propose subjective claims and mark other sentences. FEVER supports the training of deep learning systems with full access to evidence on Wikipedia. Whereas, as FEVER is solely based on Wikipedia and overall claims are raised collectively, it is unlikely to retrieve evidence from diverse web-sources or verify the reliability of claims with such evidence gathered by training systems. UKP Snopes Corpus [27] comes up with a novel tagged corpus fit for FV tasks on the basis of heterogeneous web sources, in which the original corpus1 is licensed to be free.

To train the model proposed in the research, experiments have been conducted on the paramount dataset FEVER [2]. As mentioned above, DGAT has been endowed with the general applicability in heterogeneous drawings. To examine the universality of the model, UKP Snopes Corpus is then adopted to generalize multi-domain FV tasks.

B. BASELINES

In our experiments, we use the typical FV methods as baselines, including three previous top models on FEVER (Athene [19], UCL MRG [20], UNC NLP [5]) and BERT based models (GEAR [4], and a concurrent work KGAT by [26]).

C. IMPLEMENTATION DETAILS

During training, we adopt the batch size of 32, the learning rate of $2e - 5$. All models are evaluated with LA on the test set and trained for two epochs. All claims are given five pieces of evidence. In our experiments, the max length is set to 150. All models are implemented with PyTorch.

D. EVALUATION METRICS

In the FEVER shared task, we follow the official evaluation, metrics for claim verification, including Label Accuracy (LA) and FEVER score. We use recall and F1 on UKP Snopes Corpus. It needs to be pointed out that although F1 contains recall and accuracy components, recall still remains the most important factor because FV system bases on the retrieval predictions, and low recall makes many given claims have no probable evidence.

1) LA

LA calculates classification accuracy rate of claims without the consideration of retrieved evidence.

2) FEVER SCORE

The FEVER score is the LA conditioned on providing at least one complete evidence set, which better reflects the inference model ability.

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we first compare our DGAT framework with baseline models on FEVER test sets to evaluate whether our model can function well on homogeneous dataset. Then, we compare the performance of the proposed method with others in UKP Snopes Corpus, a rich heterogeneous dataset. Next, we provide an ablation study to obtain the effect of individual module. Finally, the current work presents a case study to demonstrate the effectiveness of our framework.

A. OVERALL PERFORMANCE

Table 1 shows the FV performances. Several testing scenarios are conducted to compare DGAT effectiveness with transformer-based baselines. In comparison with baseline models, although our DGAT is not the best on all scenarios

TABLE 1. Results on both homogenous and heterogeneous datasets.

Model	FEVER test set		UKP Snopes Corpus	
	LA (%)	FEVER score (%)	Recall	F1
Athene[19]	65.46	61.58	0.585	0.596
UCL MRG[20]	67.62	62.52	0.397	0.402
UNC NLP[5]	68.21	64.21	0.483	0.496
GEAR[4]	71.60	67.10	0.471	0.480
KGAT[26]	74.07	70.38	0.532	0.549
Transformer-XH[25]	72.39	69.07	0.462	0.473
DREAM[15]	76.85	70.60	0.625	0.638
Our method	71.79	66.91	0.762	0.783

in homogeneous dataset, it can work well on the benchmark dataset for fact checking. Most notably, our model has achieved the results in heterogeneous datasets, demonstrating that the reasoning model with knowledge graph can improve the robustness of the FV system on heterogeneous information.

B. MODEL COMPARISON

Table 1 reports the performance of our model and baselines on the FEVER test set (The public leaderboard for perpetual evaluation of FEVER is <https://competitions.codalab.org/competitions/18814#results>. BUAA_KLBNT is our Team Name), showing that regarding label accuracy and FEVER score, the performance of our model is inferior to KGAT, Transformer-XH and DEARM on the test set, within 5.5% gaps. It is worth noting that, the proposed approach, which exploits improving the robustness of FV systems in heterogeneous datasets through combining the intrinsic information of evidences and world knowledge within evidences, outperforms other baselines. In addition, we also find that the performance of baseline models on heterogeneous datasets is not in consistence with that on homogeneous datasets, for instance, Transformer-XH performs better than GEAR in FEVER test set, conversely in UKP Snopes Corpus. The advantage of our DGAT is that it can fill the knowledge gap between contexts. However, our model has no advantage in single-domain dataset FEVER, since there is almost no or very little knowledge gap in FEVER, and the complex structure is easy to make the model overfit. Heterogeneous data and evidential sentences from complex textual sources, as found in UKP Snopes Corpus and in the real web, make it difficult to correctly label the claims with previous methods.

C. ABLATION STUDY

The section reports extensive ablation studies for each module in our DGAT model, and also makes comparison with our full-model results. Table 2 presents the recall rate and F1 on the UKP Snopes Corpus after removing different components (including the knowledge graph fusion mechanism (III-B-3 and III-C) and graph convolutional network and graph attention network (III-D) respectively in our DGAT model. The last row in Table 2 is equivalent to simply

TABLE 2. Ablation study on UKP Snopes Corpus.

Model	Recall rate	F1
DGAT	0.762	0.783
-external knowledge collaboration	0.618	0.625
-graph based reasoning and attention	0.633	0.646
-both above modules	0.486	0.497

connecting all the evidence sentences into a string for reasoning, without any graph structure information for FV task. As shown in Table 2, compared with not using graph structure information, the use of graph network structure can make full use of semantic information and bring 27.6% improvement on recall rate. Eliminating the external knowledge collaboration module decreases 14.4% and 15.8% in terms of recall rate and F1 respectively. Eliminating the graph-based reasoning and attention module reduces 12.9% (Recall rate) and 13.7% (F1) because graph-based reasoning module learns more graph network information.

As shown in Table 2, after removing the two main modules, the performance of DGAT is quite poor compared to the other approaches (Athene, KGAT and DREAM) in Table 1. We integrate our DGAT approach to these three strong baselines, and report the results in Table 3.

TABLE 3. Integrating the DGAT approach to strong baselines on UKP Snopes Corpus.

Model	Recall rate	F1
DGAT	0.762	0.783
Athene	0.732(↑0.158)	0.743(↑0.147)
KGAT	0.712(↑0.180)	0.727(↑0.178)
DREAM	0.758(↑0.133)	0.785(↑0.147)

From Table 3, the performance of baselines incorporated with the DGAT approach can be improved effectively on the Snopes Corpus, showing that incorporating with external knowledge provides a feasible method for FV based on deep learning to have the ability to infer multi-domain information robustly.

D. CASE STUDY

Figure 3 shows two examples in our experiments which needs scattered evidence from textual sources with multi-structures to make the right inference. Additionally, these two evidences also rank at the first and second in our retrieved evidence set. To verify whether Claim 1 is “refuted”, “supported” or “not enough information”, our model needs the evidence “1-1” and the evidence “1-2”. The two evidential sentences are from different web structural textual information and our collaborative graph bridges the gap of different textual sources. According to the top two evidential sentences for Claim 1, they use different tokens (*Hilter* and *Nazi* respectively) to

Claim 1: Upon assuming power in Germany, <i>Hitler</i> ordered the defunding and elimination of police departments.
Evidence text fragment: (1-1) <i>The United States Holocaust Memorial Museum</i> observed, neither “a wholesale purge nor a wholesale resignation of policemen” occurred in Germany during the <i>Nazi</i> era. (1-2) <i>The United States Holocaust Memorial Museum</i> noted that “ <i>Hitler</i> posed as a champion of law and order” and maintained the trust of the police because the “ <i>Nazi</i> state in fact alleviated many of the frustrations the police [had previously] experienced”.
Label 1: refuted
claim 2: The <i>Fox News</i> will be shutting down for routine maintenance on 21 Jan. 2013.
Evidence text fragment: (2-1) <i>Fox News Channel</i> announced today that it would <i>shutdown</i> for what it called “routine maintenance”. (2-2) <i>President Roger Ailes</i> explained the timing of the <i>shutdown</i> : “We wanted to pick a time when nothing would be happening that our viewers want to see.” <i>The time is on 21 January 2013.</i>
Label 2: supported

FIGURE 3. A case study of our approach. Fact shared across the claim and the evidential sentences are highlighted with different colors. The red font represents the involvement of knowledge graph helps NLI model reasoning in heterogeneous textual information.

convey relevant semantics and the KG can establish the connection between these two sentences from different web sources. In addition, it can also be found all evidence nodes in the collaborative graphs tend to attend the top 2 evidence nodes, providing the most beneficial information for the inference process. The attention weights in other nodes are extremely low, indicating that our model has the ability to choose useful information from scattered evidence even in various textual sources with their respective structures.

E. ERROR ANALYSIS

In the current section, we present the error analysis for the incorrectly predicted instances and summarize the 4 primary types of errors as follows, which can also provide experience for later model improvement of FV systems.

Table 4 presents the confusion matrix for the FGE sets predictions. Obviously, the system finds it easiest to classify instances labelled as REFUTES. However, using the NOT ENOUGH INFO label correctly is the most difficult. Then,

TABLE 4. Confusion matrix on the heterogeneous datasets.

	Gold	supports	refutes	Not enough info	total
predict					
supports		5,268	158	1246	6,672
Refuses		654	1,858	1,756	4,268
Not		812	250	4,506	5,568
enough info					
Total		6,734	2,266	7,508	16,508

we describe some frequent failure cases of our model in the description below.

1) CONFUSING SEMANTICS AT SENTENCE LEVEL

In the semantic level of the whole sentence, our DGAT does not have good alignment ability to predict the relationship of two complex sentences. For example, “*Andrea Pirlo is an American professional footballer*” vs “*Andrea Pirlo is an Italian professional footballer who plays for an American club.*” This defect results in that FV system fail to correctly label the “not enough information” instances. For example, the given claim “*Terry Crews played on the Los Angeles Chargers.*” (annotated as NotEnoughInfo) is labeled as refuted, with the retrieved top-1 sentence “*In football, Crews played...for the Los Angeles Rams, San Diego Chargers and Washington Redskins,...*”. Although this evidential sentence is highly associated with the given claim, the evidence cannot conclude that the claim is wrong, making this kind of cases even more difficult.

2) MULTIPLE SEMANTIC COMPLEXITY

In some complex situations, just using alignment methods in the NLI system is not enough to make the prediction of correct relationships. In these cases, the model needs to obtain the relationship between scattered words. For example, “*China keeps all mobile phone chips manufactured within the Huawei for use in Chinese electronics.*” vs. “*Huawei’s mobile phone chips became the Chinese leading electronic by monetary value.*”

3) NUMBERS COMPLEXITY AT SENTENCE LEVEL

Due to the method’s disability to interpret numerical values, a great many claims with numerical component factors are labeled incorrectly. For instance, the claim “*To maintain one’s current weight, one must eat at least 30 calories per kilogram of body weight per day.*” is not classified as refuted based on the evidence sentence “*To maintain their current weight, humans need to consume nearly 10 calories per kilogram of body weight per day.*”. The number is the most important factor in the determination of what the claim is labeled as. Nevertheless, the information representation module cannot embed numbers distinctly enough.

VI. CONCLUSION

To conclude, the current work presents a DGAT approach with semantic relations in CKG for knowledge-aware FV. When assessing the veracity of a claim providing multiple evidential sentences, our approach is first built upon an automatically constructed semantic graph and a knowledge graph. Compared with the previous research, our DGAT has the greatest advantage of completing the NLI task on heterogeneous web textual information with the help of knowledge graph. In addition, the core of reasoning modules is the attentive embedding propagation layer, which adaptively propagates the embeddings from a node’s neighbors to update the node’s representation. Experiments demon-

strate that DGAT can function well on the FEVER shared test datasets, which are the common homogeneity datasets. However, more importantly, extensive experiments on heterogeneous datasets prove the rationality and effectiveness of DGAT. The present study first analyzes the potential of joint knowledge graph and GNN, and stands for an initial attempt to exploit structural knowledge based on information propagation mechanism in FV task. Besides knowledge graph, a lot of other structural textual information with higher intelligence and wisdom exists in real-world scenarios, such as social networks. For instance, through the integration of social network with our CKG, we can study how social influence influences the fact checking. Another exciting direction refers to the integration of information propagation and decision process, also opening up research possibilities of explainable FV.

4) FUTURE WORK

There is an important point about training methods on multi-domain datasets. In some cases, the interpretation of facts may differ between multiple sources. Therefore, it may be better to use multiple models trained on multiple single datasets for fact validation. To test this, we have trained our DGAT on multiple single datasets (Partial FEVER dev set, PoliFact and GossipCop). Unfortunately, under such experimental conditions, DGAT will suffer from learning degradation (The models only perform well on the latter single-domain dataset on which they are trained, while achieve degraded results on the former single-domain dataset), which makes FV performance worse. We make the following analysis on this interesting experimental result. Compared with the current advanced models only based on GNNs, the advantage of our DGAT is that it can fill knowledge gaps by training on multi-domain datasets. However, there is less knowledge gap in one single-domain dataset, which will lead to the lack of the ability to fill the knowledge gaps between different domains for the trained model. In fact, our training method has disadvantages in some cases. In other words, how does a GNNs trained on a given dataset perform well on a new and potentially significantly different dataset? This issue has not been addressed in previous work that applies GNNs for FV and I think this problem may need more advanced training mechanism to solve. We further try to improve this deficiency by studying new training methods, which may require us to devote more time to study and is beyond the scope of this research.

ACKNOWLEDGMENT

The authors feel appreciated for all anonymous reviewers for their hard work!

REFERENCES

- [1] S. Jiang, S. Baumgartner, A. Ittycheriah, and C. Yu, “Factoring fact-checks: Structured information extraction from fact-checking articles,” in *Proc. Web Conf.*, Apr. 2020, pp. 1592–1603.
- [2] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, “FEVER: A large-scale dataset for fact extraction and verification,” 2018, *arXiv:1803.05355*. [Online]. Available: <http://arxiv.org/abs/1803.05355>

- [3] J. Thorne, A. Vlachos, O. Cocarascu, C. Christodoulopoulos, and A. Mittal, "The fact extraction and verification (FEVER) shared task," 2018, *arXiv:1811.10971*. [Online]. Available: <http://arxiv.org/abs/1811.10971>
- [4] J. Zhou, X. Han, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "GEAR: Graph-based evidence aggregating and reasoning for fact verification," 2019, *arXiv:1908.01843*. [Online]. Available: <http://arxiv.org/abs/1908.01843>
- [5] Y. Nie, H. Chen, and M. Bansal, "Combining fact extraction and verification with neural semantic matching networks," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 6859–6866.
- [6] C.-T. Fan, Y.-K. Wang, and C.-R. Huang, "Heterogeneous information fusion and visualization for a large-scale intelligent video surveillance system," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 4, pp. 593–604, Apr. 2017.
- [7] X. Zhang, Y. Zhang, S. Wang, Y. Yao, B. Fang, and P. S. Yu, "Improving stock market prediction via heterogeneous information fusion," *Knowl.-Based Syst.*, vol. 143, pp. 236–247, Mar. 2018.
- [8] F. Wang, L. Hu, J. Zhou, J. Hu, and K. Zhao, "A semantics-based approach to multi-source heterogeneous information fusion in the Internet of Things," *Soft Comput.*, vol. 21, no. 8, pp. 2005–2013, Apr. 2017.
- [9] G. Li, G. Kou, and Y. Peng, "A group decision making model for integrating heterogeneous information," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 6, pp. 982–992, Jun. 2018.
- [10] R. G. Qiu, "Towards ontology-driven knowledge synthesis for heterogeneous information systems," *J. Intell. Manuf.*, vol. 17, no. 1, pp. 99–109, 2006.
- [11] B. Hu, C. Shi, W. X. Zhao, and T. Yang, "Local and global information fusion for top-N recommendation in heterogeneous information network," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 1683–1686.
- [12] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 2022–2032.
- [13] Q. Ai, V. Azizi, X. Chen, and Y. Zhang, "Learning heterogeneous knowledge base embeddings for explainable recommendation," *Algorithms*, vol. 11, no. 9, p. 137, Sep. 2018.
- [14] X. Wang, X. He, Y. Cao, M. Liu, and T.-S. Chua, "KGAT: Knowledge graph attention network for recommendation," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 950–958.
- [15] W. Zhong, J. Xu, D. Tang, Z. Xu, N. Duan, M. Zhou, J. Wang, and J. Yin, "Reasoning over semantic-level graph for fact checking," 2019, *arXiv:1909.03745*. [Online]. Available: <http://arxiv.org/abs/1909.03745>
- [16] A. Hanselowski, C. Stab, C. Schulz, Z. Li, and I. Gurevych, "A richly annotated corpus for different tasks in automated fact-checking," 2019, *arXiv:1911.01214*. [Online]. Available: <http://arxiv.org/abs/1911.01214>
- [17] J. Luken, N. Jiang, and M.-C. de Marneffe, "QED: A fact verification system for the FEVER shared task," in *Proc. 1st Workshop Fact Extraction Verification (FEVER)*, 2018, pp. 156–160.
- [18] A. P. Parikh, O. Täckström, D. Das, and J. Uszkoreit, "A decomposable attention model for natural language inference," 2016, *arXiv:1606.01933*. [Online]. Available: <http://arxiv.org/abs/1606.01933>
- [19] A. Hanselowski, H. Zhang, Z. Li, D. Sorokin, B. Schiller, C. Schulz, and I. Gurevych, "UKP-athene: Multi-sentence textual entailment for claim verification," 2018, *arXiv:1809.01479*. [Online]. Available: <http://arxiv.org/abs/1809.01479>
- [20] T. Yoneda, J. Mitchell, J. Welbl, P. Stenetorp, and S. Riedel, "UCL machine reading group: Four factor framework for fact finding (HexaF)," in *Proc. 1st Workshop Fact Extraction Verification (FEVER)*, 2018, pp. 97–102.
- [21] C. Hidey and M. Diab, "Team SWEEPer: Joint sentence extraction and fact checking with pointer networks," in *Proc. 1st Workshop Fact Extraction Verification (FEVER)*, 2018, pp. 150–155.
- [22] Q. Chen, X. Zhu, Z. Ling, S. Wei, H. Jiang, and D. Inkpen, "Enhanced LSTM for natural language inference," 2016, *arXiv:1609.06038*. [Online]. Available: <http://arxiv.org/abs/1609.06038>
- [23] C. Malon, "Team papelo: Transformer networks at FEVER," 2019, *arXiv:1901.02534*. [Online]. Available: <http://arxiv.org/abs/1901.02534>
- [24] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. (2018). *Improving Language Understanding by Generative Pre-Training*. [Online]. Available: <https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/languageunderstandingpaper.pdf>
- [25] C. Zhao, C. Xiong, C. Rosset, X. Song, P. Bennett, and S. Tiwary, "Transformer-XH: Multi-evidence reasoning with extra hop attention," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–16.
- [26] Z. Liu, C. Xiong, M. Sun, and Z. Liu, "Fine-grained fact verification with kernel graph attention network," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 7342–7351.
- [27] P. Shi and J. Lin, "Simple BERT models for relation extraction and semantic role labeling," 2019, *arXiv:1904.05255*. [Online]. Available: <http://arxiv.org/abs/1904.05255>
- [28] I. Dagan, O. Glickman, and B. Magnini, "The PASCAL recognising textual entailment challenge," in *Proc. Mach. Learn. Challenges Workshop*. Berlin, Germany: Springer, 2005, pp. 177–190.
- [29] A. Williams, N. Nangia, and S. R. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," 2017, *arXiv:1704.05426*. [Online]. Available: <http://arxiv.org/abs/1704.05426>
- [30] S. R. Bowman, G. Angeli, C. Potts, and C. D. Manning, "A large annotated corpus for learning natural language inference," 2015, *arXiv:1508.05326*. [Online]. Available: <http://arxiv.org/abs/1508.05326>
- [31] L. Sha, B. Chang, Z. Sui, and S. Li, "Reading and thinking: Re-read LSTM unit for textual entailment recognition," in *Proc. COLING, 26th Int. Conf. Comput. Linguistics, Tech. Papers*, 2016, pp. 2870–2879.
- [32] A. Conneau, D. Kiela, H. Schwenk, L. Barrault, and A. Bordes, "Supervised learning of universal sentence representations from natural language inference data," 2017, *arXiv:1705.02364*. [Online]. Available: <http://arxiv.org/abs/1705.02364>
- [33] R. Ghaeini, S. A. Hasan, V. Datla, J. Liu, K. Lee, A. Qadir, Y. Ling, A. Prakash, X. Z. Fern, and O. Farri, "DR-BiLSTM: Dependent reading bidirectional LSTM for natural language inference," 2018, *arXiv:1802.05577*. [Online]. Available: <http://arxiv.org/abs/1802.05577>
- [34] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" 2018, *arXiv:1810.00826*. [Online]. Available: <http://arxiv.org/abs/1810.00826>
- [35] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: <http://arxiv.org/abs/1609.02907>
- [36] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*. [Online]. Available: <http://arxiv.org/abs/1710.10903>
- [37] X. Carreras and L. Márquez, "Introduction to the CoNLL-2005 shared task: Semantic role labeling," in *Proc. 9th Conf. Comput. Natural Lang. Learn. (CONLL)*, 2005, pp. 152–164.
- [38] C. Fellbaum, "WordNet," in *The Encyclopedia of Applied Linguistics*. Oxford, U.K.: Blackwell, 2012.
- [39] Y. Cao, X. Wang, X. He, Z. Hu, and T.-S. Chua, "Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 151–161.
- [40] R. Speer, J. Chin, and C. Havasi, "ConceptNet 5.5: An open multilingual graph of general knowledge," in *Proc. Nat. Conf. Artif. Intell.*, 2017, pp. 4444–4451.
- [41] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 5753–5763.
- [42] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2181–2187.
- [43] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, "Freebase: A collaboratively created graph database for structuring human knowledge," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2008, pp. 1247–1250.
- [44] J. Qiu, J. Tang, H. Ma, Y. Dong, K. Wang, and J. Tang, "DeepInf: Social influence prediction with deep learning," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, London, U.K., 2018, pp. 2110–2119.
- [45] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 1024–1034.
- [46] D. Shen, X. Zhang, R. Henao, and L. Carin, "Improved semantic-aware network embedding with fine-grained word alignment," 2018, *arXiv:1808.09633*. [Online]. Available: <http://arxiv.org/abs/1808.09633>



YONGYUE WANG received the B.S. degree in information science and technology from the China University of Mining and Technology, Xuzhou, China, in 2015, and the M.S. degree in aerospace science and technology from the National University of Defense Science and Technology, Changsha, China, in 2017. He is currently pursuing the Ph.D. degree in computer science with the School of Computer Science and Engineering, Beihang University, Beijing, China. His

main research interests include natural language processing, deep learning, and knowledge graph.



BEITONG YAO received the B.S. degree in applied chemistry from Beihang University, Beijing, China, in 2018, where he is currently pursuing the M.S. degree in computer science with the School of Computer Science and Engineering. His main research interests include natural language processing, data mining, and machine learning.

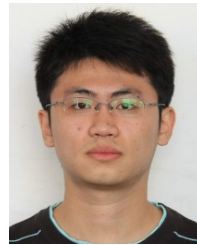


CHUNHE XIA (Member, IEEE) received the Ph.D. degree in computer application from Beihang University, Beijing, China, in 2003. He is currently a Supervisor and a Professor with Beihang University, where he is also the Director of the Beijing Key Laboratory of Network Technology. He has participated in different national major research projects and published more than 70 research papers in important international conferences and journals. His current research interests

include network and information security, data mining, information countermeasure, cloud security, and deep neural networks.



CHENGXIANG SI received the Ph.D. degree from Beihang University, Beijing, China, in 2018. His main research interest includes network and information security.



TIANBO WANG (Member, IEEE) received the Ph.D. degree in computer application from Beihang University, Beijing, China, in 2018. He is currently a Lecturer with Beihang University. He has participated in several national natural science foundations and other research projects. His research interests include network and information security, intrusion detection technology, and information countermeasure.

...