IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 3, 2021, accepted February 15, 2021, date of publication February 24, 2021, date of current version March 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062231

Multi-Stream Semantics-Guided Dynamic Aggregation Graph Convolution Networks to Extract Overlapping Relations

XIUSHAN LIU, JUN CHENG, AND QIN ZHANG

School of Electronic and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China

Corresponding author: Jun Cheng (mrchengjunjun@foxmail.com)

This work was supported by the Science and Technology Program of Guangzhou under Grant 201803010081.

ABSTRACT The existing relation extraction approaches select the relevant partial dependency structures and exhibit the limitation associated with long-distance dependencies. Moreover, due to the repeated use of the irrelevant redundant information and the lack of consideration of the key semantic details, the extraction of relations is relatively complex when the entities overlap. To address this limitation and effectively exploit the relevant information while ignoring irrelevant information, this paper proposes a simple but effective multistream semantics-guided dynamic aggregation graph convolution network (SG-DAGCN) to realize the extraction of overlapping relations. The proposed model constructs the entity relation graphs by enumerating the possible candidates and external auxiliary information and adaptively manages the relevant substructure. Subsequently, this framework models the relational graphs between the entities through a dynamic aggregation graph convolution module and gradually produces the discriminative embedded features and a refined graph through the dynamic aggregation of nodes. The proposed approach can effectively leverage the rich multiscale structural information and capture the long-distance dependencies between overlapping entities in long sentences. The results of the experiments conducted on two typical benchmark datasets show that the proposed model can achieve a high level of performance and outperform other state-of-the-art methods in both qualitative and quantitative aspects.

INDEX TERMS Overlapping relation extraction, multiscale structural information, dynamic aggregation, long distance dependencies, refined graph, relevant substructure.

I. INTRODUCTION

Entity and relation extraction are basic tasks of information extraction in natural language processing (NLP) [1]. Relation extraction is aimed at identifying the relations among different entities in a piece of text [2], and this task is of significance in a variety of natural language processing applications, such as knowledge graph construction [3], question answering [4] and machine translation [5].

In recent years, many methods to accomplish the entity and relation extraction tasks have been developed. Most of the existing models of entity and relation extraction can be divided into two classes: pipeline-based methods and joint-based (including sequence labeling) methods. Pipeline-based methods consider the entity extraction and relationship extraction as two independent subtasks [6]. However, in such techniques, the interaction between the two subtasks and errors in the entity recognition module are ignored, thereby affecting the performance of the following relationship classification. In other words, the result of the relationship extraction task depends largely on the result of the entity extraction, since the relation is highly correlated with the entity information. Finally, the effect of the entity relation extraction is affected.

Compared to pipeline-based entity and relation extraction methods, joint-based extraction methods aggregate the information of the entities and relations, thereby effectively reducing the error propagation and capturing the nonlocal syntactic relations. Furthermore, the interaction between the entities and relation is strengthened. Most of the existing joint extraction methods are based on structured learning of the features. For example, Li and Ji [7] designed an incremental joint extraction approach for the entity mentions and relations. However, this joint method relied heavily on the

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

natural language processing toolkits and manual features. Recently, with the development of the deep learning technology, neural networks were used to realize relation extraction through parameter sharing to accomplish joint modeling. However, in such approaches, separate components for entity recognition and relation classification must be established. For instance, Yuntian *et al.* [8] proposed a joint approach to extract the entities and relations by using reinforcement learning and deep learning. Wei *et al.* [9] proposed a novel hierarchical binary tagging framework to realize the joint extraction of entities and relations. Although these methods could effectively capture the interaction between the entities and relations and enhance the relation extraction, the overlapping relations in the entity could be identified. Recently, to solve the problem of relationship extraction in overlapping entities, many researchers attempted to extract the overlapping entity relations. For instance, Takanobu *et al.* [10] proposed a hierarchical reinforcement learning framework to enhance the interaction among the entities. Li *et al.* [11] verified the tree structures necessary to realize the deep learning of representations and compared separate sequence-based and tree-structured LSTM-RNNs in the relation classification task. Fu *et al.* [12] modeled text as relational graphs and used the graph convolutional network to realize the joint entity and relation extraction. Zeng *et al.* [13] proposed a relation classification approach based on a convolutional deep neural network in which the graph convolutional networks were modeled over a pruned tree. However, this end-to-end joint extraction method could not effectively extract the relation when the overlapping situation was relatively complex and likely eliminated several key details during the extraction.

To effectively extract the relation of overlapping entities, increase the utilization of key details and consider the interaction among the entities and relationships, this paper proposes a novel multistream semantics-guided dynamic aggregation graph convolution network (SG-DAGCN) framework to perform the extraction of entity relations, even under complex overlap scenarios. In practice, a sentence usually contains multiple entities and complex relations, and these entities may be nested within one another. Moreover, the relational triplets may overlap, as shown in FIGURE 1. Meanwhile, we can represent a relation by a triplet (Entity1, Relation type, Entity2), and here, Entity1 is the first argument of the relation and can be obtained from detected entity and candidate entity at a triplet or the sentence. The second argument of entity and relation type can be achieved from other detection entities and sentence. Obviously, the first entity can be used multiple times to from overlapping relations.

Consider the case in which one entity in a triplet overlaps with another entity. Specifically, in FIGURE 1, it can be noted that the entity 'Manhattan' overlaps with 'Manhattan office', and the mention 'New York' overlaps with 'New York attorney'. Second, the dependencies among the entities must be considered to realize the inference of a relational triplet. For instance, the triplet (Eliot Spitzer, contains, Manhattan) can be inferred from (Eliot Spitzer, holder, New York attorney) and (New York, contains, Manhattan), and (Maurice Greenberg, contains, Manhattan) can be inferred from (Maurice Greenberg, birthplace, New York) and (New York, contains, Manhattan). In addition, the triplet (Trump, President, United States) can be inferred from (Trump, Governance, United States), and (Trump, President, United States) can be inferred from (Trump, Live in, Whitehouse) and (Whitehouse, Presidential palace, United States). Therefore, it is necessary for the model to be able to capture the dependencies among the entities. Meanwhile, the dependencies between entities are importance for improve the inference of an overlapping relational triplet. However, the existing models cannot capture the long-distance dependencies among the entities. The framework uses the refined external semantic information to compensate for the semantic deficiency of the entity relation triples and build a new multiscale relation graph. Subsequently, the dynamic aggregation technique is used to extract the overlapping relation.

In summary, the primary contributions of the proposed framework are as follows:

(1) We propose a multistream semantics guided dynamic aggregation graph convolution network (SG-DAGCN) to extract the overlapping relations. The network includes two main modules, namely, the semantics-guided module and dynamic aggregation module. The contribution of the joint semantics is considered to identify the interaction between the entities and the relation.

(2) To use the semantic information for different modal and embedded discriminative features of the overlapping relation in an entity relation triplet sequence, the semantics-guided module includes four information stream branches. Two of the information streams branches examine the global and local semantics of the entity relation through different models. The other two information stream branches determine the position information of the entity relation through different models. Finally, the information is gradually refined through self-guided attention layers.

(3) To further enhance the interaction and dependencies among the entity and relation, we design a multiscale dynamic aggregation technique, implemented through a dynamic aggregation graph convolution (DAGCN) module. This module can directly model the cross-spacetime joint dependencies and substantially facilitate the interaction of the entity and relation. Moreover, this approach can also effectively solve the long-distance dependence problem and refine the graph by removing the redundant dependencies among the neighborhoods. In addition, the rich multiscale semantic information can help enhance the ability to distinguish the overlapping relation.

We conducted overlapping relation extraction experiments on the NYT and CoNLL04 public datasets. The experimental results indicated that the proposed SG-DAGCN framework outperforms the other state of the art methods. An extensive analysis indicated that the proposed SG-DAGCN framework can effectively capture the long-distance dependencies between the entities and enhance the interaction between

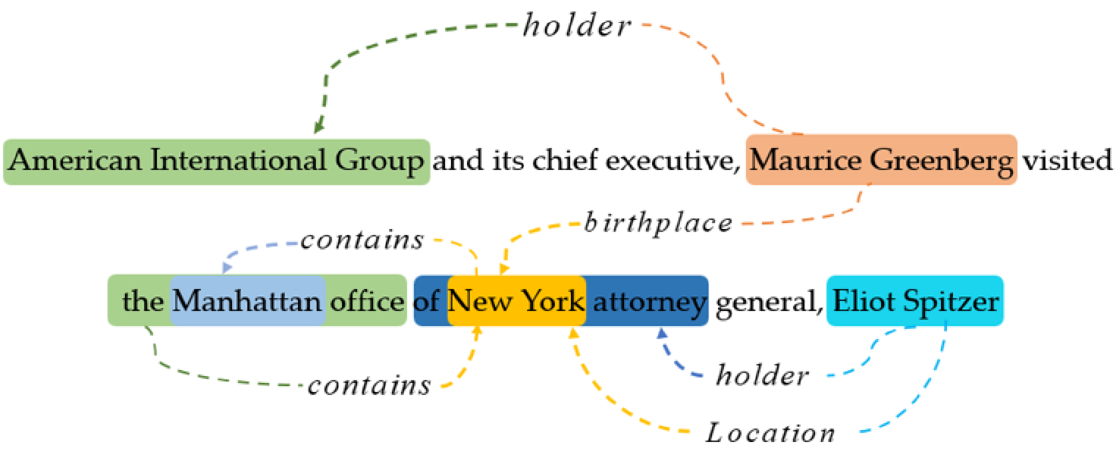


FIGURE 1. Example of the overlapping entities and relations in a sentence.

the entity and relation. The important details (including semantic information) are retained to the greatest extent, while increasing the ability to distinguish among the overlapping entities. The remaining paper is organized as follows: Sect. II provides a review of the related work pertaining to the entity relation extraction. Sect. III describes the multi-stream semantics-guided dynamic aggregation graph convolution network (SG-DAGCN) framework. Sect. IV presents the experiment results for the overlapping relation extraction from three public datasets by using the SG-DAGCN frameworks. The concluding remarks and scope for future research are presented in Sect. V.

II. RELATED WORK

In recent years, relation extraction has been widely applied in the knowledge graph and natural language processing domains. In general, most existing entity and relation extraction methods can be classified as either pipeline-based methods or joint-based methods. Pipeline-based methods treat the entity and relation extraction as two separate subtasks. First, the name entity recognition (NER) approach is used to identify the given entity in a sentence. Second, the entity and relation are distinguished using a classification model. Pipeline-based methods usually exhibit a low performance as the error propagates from the entity recognition to the relation extraction, and such approaches ignore the interaction between the entities and relation. To effectively utilize the interaction information of the entities and relation and reduce the propagation and accumulation of errors, joint-based methods were developed. Compared to pipeline-based methods, joint-based methods achieved more satisfactory results in the entity relation extraction task. For instance, Dai *et al.* [14] proposed a joint extraction approach for the entities and overlapping relations by using position attentive sequence labeling. Zhang *et al.* [15] designed an end-to-end neural relation extraction model with global optimization and utilized the syntactic information of feature learning to enhance the performance. Nguyen and Verspoor [16] proposed an end-to-end neural relation extraction approach using deep biaffine attention. The model extended a BiLSTM-CRF-based entity recognition framework with a deep biaffine attention layer to model the second order interactions between the latent features for relation classification. Hong *et al.* [17] proposed a biomedical entity relation extraction network (BERE), which was a deep learning-based model using the Gumbel tree-GRU to learn sentence structures and joint embedding to incorporate the entity information. Geng *et al.* [18] attempted to enhance the performance of the relation extraction by using the bidirectional sequential LSTM with attention to identify word-based features in combination with structural features to classify the relations. Liu *et al.* [19] introduced the self-attention mechanism in the joint model to learn the word intrasentence dependencies. Although the entity relation joint extraction models exhibit an excellent performance, they do not consider the overlapping in a given sentence; thus, these models are not effective when the entities and relations overlap.

Recently, with the widespread use of the graph neural networks for knowledge graph construction and text classification, among other tasks, an increasing number of graph models have been employed to extract the entity relations. For instance, Peng *et al.* [6] used a relation extraction framework based on the graph long short-term memory networks (graph LSTM) to extend the model to cross-sentence n-ary relation extraction and enhance the dependencies between intra- and intersentence scenarios. Luan *et al.* [20] introduced a general framework for several information extraction tasks, in which the span representations were shared using dynamically constructed span graphs. Zhang *et al.* [21] propose an extension of the graph convolutional networks tailored for relation extraction, in which the information was pooled simultaneously over arbitrary dependency structures in an efficient manner. Sahu *et al.* [22] designed a novel graph convolution network for the intersentence case of relation extraction; specifically, the graph was constructed using the intersentence and intrasentence dependencies to capture the local and nonlocal dependency information. Ning and Qi *et al.* [23] introduced a semisupervised convolution graph kernel model to realize the relation extraction of English text. Zhang *et al.* [24] embedded the relational knowledge in a relation extraction model through an attention mechanism and used the graph convolution networks to learn the explicit relation. Zhang *et al.* [25] used the structural information to perform extraction among the entities in a given text and proposed an attention-guided graph convolutional network for relation extraction. However, these entity relation extraction methods did not exhibit a satisfactory performance when the overlapping relation was complex.

III. PROPOSED METHOD

We consider the interaction between the entities and relation and propose a semantics-guided dynamic aggregation graph convolution network to realize the overlapping relation extraction. The network is composed of a semantic guided module and dynamic aggregation graph convolution module. The objective is to exploit different levels of the semantics information to enhance the interaction among the entities and relation to extract the overlapping relation. The approach removes the irrelevant redundant information between the entities and relation; specifically, the irrelevant information is cropped through self-guided attention layers while the feature mapping information is gradually refined. Moreover, the dependencies among multiple streams of information are established to complement one another. The entity relation graphs are constructed over these multiple streams of refined information at different spatiotemporal scales. Next, the convolution operation is conducted on these multiscale graphs, which simultaneously and adaptively aggregates the multiscale spectral-spatial information and further refines the input entity and relation graph. The nodes of the graph are dynamically updated, owing to which, the powerful multiscale aggregators can effectively capture the node information of entity relation graph. The entity pairs belonging to the same relation class are automatically collected in the embedding space. FIGURE 1 shows the SG-DAGCN overlapping relation framework. Next, we describe all the components used to construct the proposed SG-DAGCN framework, including the input expression (Section III-A), semantics guided module (Section III-B), backbone of the dynamic aggregation graph convolution (DAGCN) module (Section III-C) and exposition of the loss function (Section III-D). As illustrated in FIGURE 2, the framework consists of an input expression layer, a semantics-guided (SG) module and a dynamic aggregation graph convolution network (DAGCNs) module. The input expression layers learn the embedding representation between the entity and relation by fusing the entity relation triples and sentence sequence information. Two types of semantics, namely, the self-guided attention [31], [32] joint type and multiscale dynamic aggregate type, are incorporated into the semantics guided (SG) module and DAGCN module, respectively. To enhance the interaction of the entity and relation in the SG module, and to eliminate the irrelevant redundant information, we use four branch streams, pertaining to the global semantics of the sentence sequences based on the Tree LSTM [33] and Tree CNN [34] (denoted by $f_{sc(R)}$ in FIGURE 2), local semantics of the sentence sequences based on the word embedding MLP and Tree CNN (denoted by $f_{wc(R)}$ in FIGURE 2), semantics information of the head entity of the triples based on the Bi-LSTM and char CNN (denoted by $f_{sc}^{E_h}$ in FIGURE 2), and local semantics information of the tail entity of the triples based on the word embedding MLP and char CNN [35] (denoted by $f_{wc}^{E_t}$ in FIGURE 2). These entities include the position information. To establish the dependencies and interaction among the entity and relation in the DAGCN module, we use three multiscale dynamic graph convolution (MS-DGCNs) blocks, where each MS-DGCN includes three scales, namely, $S = 1, 2, 3$. Next, the semantic information is refined and gathered through the aggregate layers.

A. INPUT REPESENTATION

To effectively accomplish the next series of tasks, we first define the overlapping relation types. Subsequently, we analyze and examine the representation of the entity relation.

1) DEFINITION OF THE OVERLAPPING RELATION

To understand the overlapping relation types, in the section, we discuss and define the overlapping relation types in sentences. [29]. In the first case, no overlapping exists (the entity pair exhibits only one relation). Second, an entity may correspond exhibit different types of relations simultaneously, that is, one entity in a triplet may overlap another entity in a triple. Third, each entity pair in a triple may exhibit different relationships, that is, an entity pair may overlap with other entity and relation triplet.

To sum up, we consider that there are two situations for the relational triplet and entity in a sentence, namely, one of the entities overlaps with other triplet and entity pair overlaps with other triples. We can obvious observations there are have four classes of overlaps relation based on figure 1.

Overlapping relation in simply triplet. Namely, one of the entities is the same as an entity in any other triplet. For instance, the triplet (Eliot Spitzer, contains, Manhattan) shares the entity Eliot Spitzer with (Eliot Spitzer, holder, New York attorney).

An entity overlaps with an entity in any other triplet. For instance, in the triplet (Eliot Spitzer, holder, New York attorney) the entity New York attorney overlaps with the entity New York in (New York, contains, Manhattan). Two entities in one triplet are the same in other triplet(s). For example, the triplet (Trump, President, United States) and (Trump, Governance, United States) have the same entity pair Trump and United States.

Two entities in a triplet overlap with the entities in another triplet, or an entity is the same, and the other is nested. For instance, in (Trump, President, United States) and (Donald

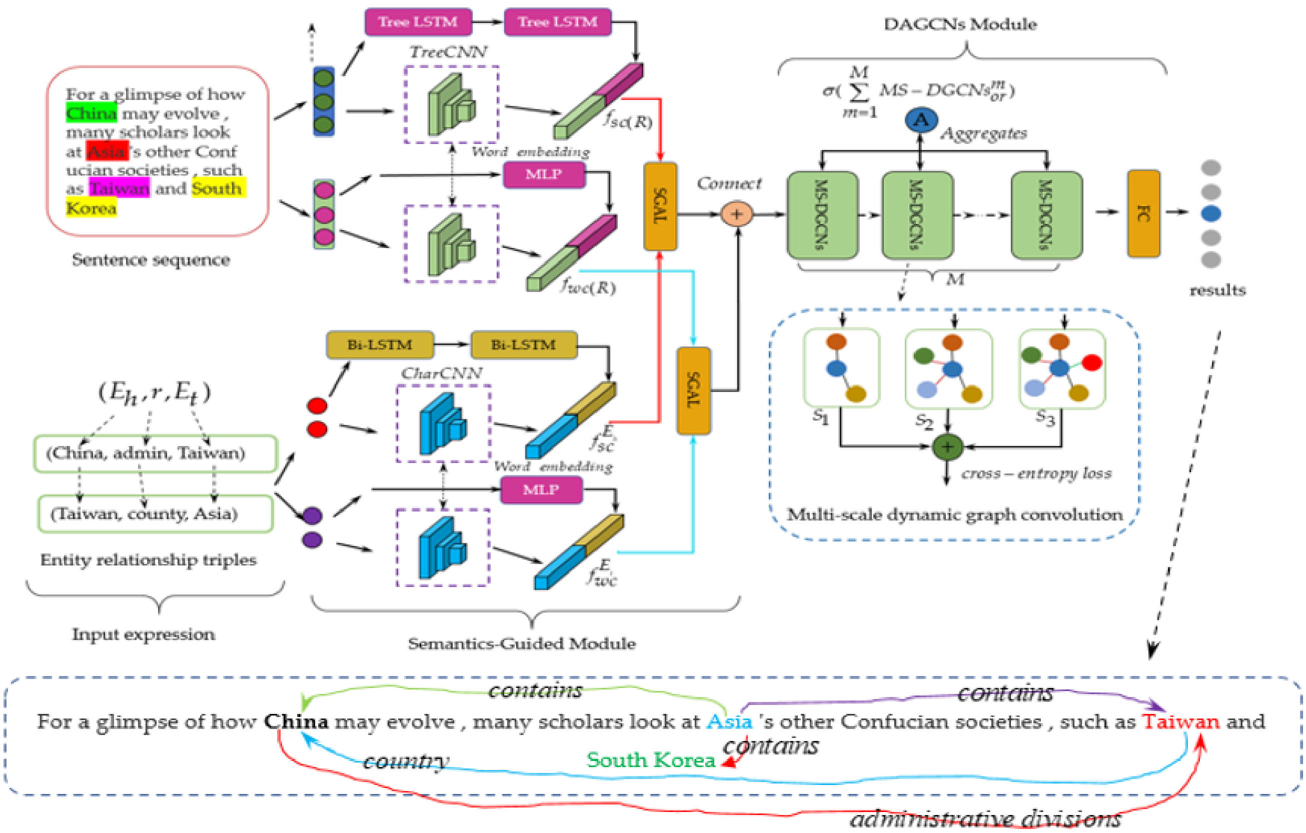


FIGURE 2. Framework of the proposed semantics-guided dynamic aggregation graph convolution network (SG-DAGCN). *M* indicates the number of multiscale dynamic aggregation graph convolution blocks, $MSDGCNs_{D}or^{m}$ indicates the M^{th} overlapping relation of the multiscale dynamic graph convolution blocks. *SGAL* indicates the self-guided attention layers. "results" indicates the extract results of overlapping relation (shown the bottom in FIGURE 2). For instance, (China, administrative divisions, Taiwan).

Trump, Governance, United States), the entity "Trump" nests with "Donald Trump" and the other entity "United States" are identical.

2) RELATION REPRESENTATION

The overlapping relation extraction is considered to be the prediction task of an entity relation triplet [25]. Unlike the existing approaches to extract the overlapping relation that only embed the representation of the entity relation triplets, we embed not only the representation of the entity relation triplet but also the contextual information of the sentence and local information of entity in the triplet to enhance the ability to distinguish the overlapping relation. Specifically, we construct a simple quintuple prediction by considering the local and global context semantics of the head and tail entities in a relation triplet. Suppose a sentence is represented as $Sent = \{w_1, w_2, \ldots, w_i, \ldots, w_n\}$,We can represent this sentence as a collection of entity relation quintuples that contain the local and global context semantics and position information of the head and tail entities. The specific representation is as follows:

$$X = \{(e_h, r, e_t, sc_h, wc_t)|e_h \in E_h, r \in R, e_t \in E_t\}. \quad (1)$$

where sc_h indicates all the words before the head entities in the sentence. wc_t indicates the $1-hop$ neighborhood word of the tail entities in the sentence. R, E_h, E_t indicate the sets of relations, head entities and tail entities, respectively. The quintuples include the position information of the entities.

B. SEMANTICS-GUIDED MODULE

To capture the information for different modalities and learn the discriminative features, the proposed joint semantic extractor with four branches leverages the syntax structure and position information of the entities from both the sentence sequences and entity relation quintuples. Two branches perform the sentence sequence global and local semantics information extraction, and the other two branches extract the position and relation information of the entities of the quintuples. This section describes each component of the semantics-guided module.

1) SENTENCE SEQUENCES

As shown in the top half of the semantics-guided module in FIGURE 2, given a sentence sequence S that contains an entity relation triplet, the corresponding context semantic of the entity relation $f_{sc(R)}$ and $f_{wc(R)}$ including both the global and local information are obtained. $f_{sc(R)}$ is obtained using the Tree LSTM [33] and Tree CNN [34] models, and

$f_{wc(R)}$ is obtained using the Tree CNN and word embedding MLP models. For the global and local semantics branches, the words (or entities) are used as the sentence sequence cure. Different feature extractors are used for the entity (word) in a sentence sequence, as the head and tail entity objects are associated and not associated with the past words, respectively.

Therefore, to extract the feature of the local semantic information $f_{wc(R)}$, we use a two-layer MLP to which the word embedding vector is input. Nevertheless, to balance the contribution of the global and local semantics information, we ensure that the feature vectors have the same dimensionality. Finally, we gradually refine the features by using the self-guided attention layers.

2) ENTITY RELATION QUINTUPLES

As shown in the bottom half of the semantics-guided module in FIGURE 2, the structure of the entity relation quintuple information extractor is similar to the sentence sequence feature extractor, albeit different in three aspects. The first difference pertains to the input. We use the entity pair, position information and character of the common entities as the quintuple cure. Moreover, the embedding representation model is different, that is, we use the character CNN [35] and bidirectional long short term memory networks (Bi-LSTM) as the structure. The features $f_{sc}^{E_h}$ and $fc_{wc}^{E_h}$ are obtained through the self-guided attention layers (SGAL) that concatenate the character information, word embedding information and position information.

3) SELF-GUIDED ATTENTION LAYERS

To enhance the interaction ability of the entities and relation and highlight the differences in the overlapping while reducing the use of irrelevant redundant information, we use the self-guided attention [31][32] layers to gradually refine the feature mapping. Specifically, before feeding the entity relation information into the dynamic aggregation graph convolution module, we fuse the features using the proposed self-guided attention layers. We perform the fusion operation twice in the final entity relation extraction framework. Experiments are conducted to demonstrate the effectiveness of the fusion operation. The fusion feature is represented as.

$$F = SGAL(F_{sc}^{ER}, F_{wc}^{ER}). \tag{2}$$

where F_{sc}^{ER} indicates the fusion of $f_{sc(R)}$ and $f_{sc}^{E_h}$, F_{wc}^{ER} indicates the feature fusion of $f_{wc(R)}$ and $f_{wc}^{E_t}$, and $SGAL(\cdot)$ indicates the fusion operation through the self-guided attention layers (SGAL). F_{sc}^{ER} and F_{wc}^{ER} can be represented as.

$$\begin{cases} F_{sc}^{ER} = SGAL(f_{sc(R), f_{sc}^{E_h}}) \\ F_{wc}^{ER} = SGAL(f_{wc(R), f_{wc}^{E_t}}) \end{cases} \tag{3}$$

Then, these multi-stream information by self-guided attention layers context connect and as input of DAGCNs module. The connect process as.

$$F = F_{sc}^{ER} \oplus F_{wc}^{ER}. \tag{4}$$

where, $\oplus$ indicates connect.

In summary, the self-guided attention technique can selectively fuse the detailed information and highlight the interaction between the entity and relation, thereby enhancing the ability to distinguish the overlapping properties.

C. MS-DAGCNs MODULE

This section describes the graph convolution networks and multiscale dynamic aggregation graph convolution networks, which are the two components of the DAGCN module. Although a multilayer graph convolution network [21], [36] can effectively aggregate and transfer the node information of neighbors, the graph is fixed throughout the operation process, which leads to a deteriorated performance if the input graph structure is unreasonable. In this case, the long-distance dependence and multiscale information of the entity and relation cannot be obtained. Thus, we propose a multiscale dynamic aggregate graph convolution network (MS-DAGCN), in which the graph can be gradually refined during the process. The ability of the graph structure to capture multiscale features is improved by fusing the currently embedded feature and the graph information used in the previous layer. The operation process is as follows.

1) TRADITIONAL GRAPH CONVOLUTION

Graph convolution networks (GCNs) are novel neural networks that operate directly on a graph structure and conduct node embedding through the aggregation and information transfer in the neighborhood. Unlike the existing approaches pertaining to traditional neural networks, the GCNs can operate on data with a discretionary structure. Specifically, given a graph with n nodes, we can define the graph $\zeta = (v, \varepsilon)$, where $v = \{v_1, v_2, \ldots, v_n\}$ and $\varepsilon = \{e_1, e_2, \ldots, e_n\}$ indicate the set of nodes and edges, respectively, and the adjacency matrix indicates the connection among the nodes.

$$A_{i,j} = \begin{cases} 1, if\ v_1 \in \varphi(v_j) and v_j \in \varphi(v_i)\ or\ v_i = v_j \\ 0, otherwise \end{cases} \tag{5}$$

where $\varphi(v_j)$ indicates the neighborhood set of the node v_j, and $A_{i,j} = 1$ indicates the presence of edges between the nodes. The update process of the multilayer graph convolution networks can be expressed as follows.

$$X^{(l)} = \delta(\widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X^{(l-1)} W^{(l-1)}). \tag{6}$$

where $X^{(l)}$ indicates the status information of the l^{th} layer, $\widetilde{A}$ indicates the normalized self-loop adjacency matrix, $\delta(\cdot)$ is the activation function of the graph convolution layers, and W is the weight matrix.

2) MULTI-SCALE GRAPH CONSTRUCTION

The structural information of the multiscale graph is useful for the entity and relation extraction. The main reason is that the contextual semantic information corresponding to the different scales can help exploit the enhanced local performance of the overlapping relation from various levels and effectively compensate for the shortage of single scale information. In the proposed framework, the multiscale interaction information is captured through the entity and relation by building the graph structure at the neighborhood nodes at different hops. Specifically, in an $s - hops$ neighborhood node, every quintuple sequence node of the entity relation X is connected with $s - hops$, as shown in the lower half of the DAGCN module in FIGURE 2. In our paper, we define S as 3, with $s \in S$. The entity relation joint nodes on $s - hops$ can be expressed as.

$$\nu_s(\nu_i) = \nu_{s-1}(\nu_i) \cup \nu_1(\nu_{s-1}(\nu_i)). \tag{7}$$

where ν is a set of the $s - hops$ neighborhood nodes. s is defined as 1, 2, 3.

3) EDGE DYNAMIC REGRESSION

Traditional edge regression methods [36] (as shown in equation (1)) cannot effectively capture the correlation among the entity relationship nodes and reduce the interaction of the entities and relations due to the extensive use of the redundant information of the nodes. To solve these problems, we employ a two-layer MLP that takes the difference of the two-node feature information points as the input and outputs a scalar value between 0 to 1 as the edge. The adjacency matrix can be expressed as follows.

$$A_{i,j} = \begin{cases} Sigmoid(\delta_1(ReLu(\delta_2(v_i - v_j)))) \\ 0 \end{cases} \tag{8}$$

where δ_1 and δ_2 indicate two different linear layers.

Next, we define the adjacency matrix $A^{(l)}$ at the l^{th} layer and use the embedding kernel Γ^Q to encode the similarity of both the nodes. However, the introduction of the embedded information can help explore more accurate graph structures [37], and the detailed information contained in the graph structure is fully utilized. The adjacency matrix $A^{(l+1)}$ from the $(l+1)^{th}$ layer can be dynamically updated as.

$$A^{(l+1)} \leftarrow A_1(A^{(l)} + \tau\Gamma^Q)A^T + \theta^{(l)}I. \tag{9}$$

where $A^{(l+1)}$ is the $(l+1)^{th}$ layer of the adjacency matrix, and A_1 indicates the initial adjacency matrix. τ is the weight of the kernel Γ^Q, defined as.

$$\Gamma^Q = X^{(l)}(X^{(l)})^T. \tag{10}$$

where $X^{(l)}$ indicates the output from the l^{th} layer. According to equation (9), $A_{ij}^{(l+1)}$ can be calculated as.

$$\begin{cases} A^{(l+1)} = \sum_{v_p \in \varphi(v_i)} \sum_{v_\eta(v_j)} A_{i,p}A_{j,q}\phi_{p,q} + \theta^{(l)}I_{i,j} \\ \phi_{p,q} = A_{p,q}^{(l)} + \varpi < X_{p,:}^{(l)}, X_{q,:}^{(l)} > \end{cases} \tag{11}$$

where $< \cdot >$ is the inner product vector $\theta^{(l)}$. $I_{i,j}$ denote the relative importance and identity matrix from the entity relation nodes of v_i and v_j, respectively.

According to equation (6), the dynamic aggregation of the multiscale graph convolution layers can be indicated as.

$$X^{(l)} = delta(\widetilde{D}_s^{-\frac{1}{2}}\widetilde{A}_s\widetilde{D}_s^{-\frac{1}{2}}X^{(l-1)}W_s^{(l-1)}). \tag{12}$$

where $\widetilde{D}_{i,j}^s = \sum_j \widetilde{A}_{i,j}^s$. S indicates the different scale of graph.

4) MULTI-SCALE AGGREGATES

To achieve more effective multiscale spatiotemporal interaction semantic information of the entity relation, the dependence between the different scales of the semantic information and ability to represent the relevant information must be strengthened. This aspect can reduce the use of redundant information and improve the extraction of the overlapping relation. We conducted reaggregation for multiple multiscale dynamic graph convolution modules as follows.

$$X = \delta(\sum_{m=1}^{M} MS - DGCNs_{or}^m). \tag{13}$$

where $MS - DGCNs$ indicate the blocks of multiscale dynamic graph convolution. In summary, the association and interaction among the entity relation nodes can be captured by the multiscale dynamic aggregation graph convolution (DAGCN) module, and the context multiscale structure semantics of different levels can be described.

D. LOSS FUNCTION

The proposed framework employs multiple losses in all the layers during training, namely, the cross-entropy error and affinity loss. The cross-entropy error is adopted to penalize the difference between the network output and labels of the original labeled examples, as follows.

$$\begin{cases} \xi_{bce} = \sum_{g \in \gamma_g} \sum_{f=0}^{C} \Upsilon_{gf} lnO_{gf} \\ \xi_{sf} = -\frac{1}{N} \sum_{i}^{N} log(\frac{exp(A_{ij})}{\sum_i^N exp(A_{ij})}) \end{cases} \tag{14}$$

where C indicates the class of the overlapping relation. O indicates the output of the SG-DAGCN, with $O = X$ (as shown in equation (12)). Υ indicates the matrix of the labels.

We can summarize the loss ξ_{total} as follows.

$$\xi_{total} = \xi_{bce} + \xi_{sf}. \tag{15}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In the section, we conduct a series of experiments to demonstrate the effectiveness of the proposed SG-DAGCN framework for overlapping relation extraction, and provide the detailed analysis of the experiment.

A. DATASET PREPARATION

NYT datasets [27], [28]: this dataset is the annotated New York Times corpus and contains 24 relation types between the entities. This dataset contains 1.18 million articles published between 1987 and 2007. For the smooth implementation of the experiment, we selected 84,832 sentences

TABLE 1. Dataset statistics.

Dataset	Entity	Sentence	Overlapping	Relation
CoNLL04	5347	1441	2382	5
NYT	491,540	84,832	108,493	24

and 491,540 entities (including overlapping entities and relation) by filtering extremely long or short sentences. Next, we randomly divided 60% and 20% of the data into the training and testing sets, respectively, and the remaining data were the validation set.

CoNLL04 datasets [29], [30]: This dataset with 1441 examples consists of news articles, with four entity types and five relation types. We randomly divided 60% of the data into the training sets. Half of the remaining data pertain to the test set, with 288 test and validation instances, respectively.

We did not remove the words after cleaning the CoNLL04 and NYT public datasets. The statistics of the preprocessed datasets are summarized in TABLE 1.

In TABLE 1, "Overlapping" indicates the overlapping relation triples. In the sentence and triple of the CoNLL04 datasets, the overlapping items of the ratio are 99% and 50%, respectively. In the NYT datasets, the overlapping items have a ratio of 94% and 75%.

B. TRAINING AND IMPLEMENTATION DETAILS

1) FRAMEWORK SETUP

To gather a considerable amount of the multiscale structure information and context semantics information of the entities and relation, strengthen the interaction between the entities and relation, and enhance the ability to distinguish the overlapping relation, in the dynamic aggregation graph convolution (DAGC) module, the number and scale of the DAGC layers are set as 2 and 3 ($S = 3$), respectively, and the number of DAGC blocks is also set as 3 ($M = 3$). The semantics-guided (SA) module consists of four different components, with the number of convolution kernels and filters in the char CNN being set as 3 and 64, respectively. The hidden size of the Bi-LSTM is set as 256, and the number of layers of the MLP is 2. During training, we use the Adam optimizer and stop the training when the validation loss does not decrease for 10 consecutive epochs. The initial learning rate and dropout for the framework are set as 0.005 and 0.25, respectively. Subsequently, 300-dimensional glove word embeddings are performed. Finally, to enhance the extraction accuracy, we perform label smoothing with the smoothing factor set as 0.1. The proposed SG-DAGCN framework employs two losses in all the network layers during training.

2) TRAINING ENVIRONMENT

All the experiments are performed on Pytorch version 1.3.0 and Keras 2.1.5 with two Nvidia Tesla P100 GPUs. Python version 3.6 is employed.

C. EVALUTION METRICS AND BASELINE METHODS

1) EVALUATION METRICS

To verify the feasibility and correctness of the proposed SG-DAGCNs framework, we adopt the precision (P), recall (R) and F1 score as the evaluation metrics to evaluate the performance of different entity relation extraction models. Specifically, the relation is considered to be accurate when the entities and relation are both correctly extracted. In the experiments, the entity types and order between the entity pairs are considered in the CoNLL04 public dataset. However, in the NYT public datasets, we only consider the order between the entity pairs. The evaluation metric of F1 can be expressed as.

$$F1 = \frac{2 \times P \times R}{P + R}. \tag{16}$$

where P is the precision, and R denotes the recall.

2) BASELINE METHODS

We compare the proposed SG-DAGCN to multiple state of the art entity relation extraction approaches:

(1) LSTM-SDP:This method involves an entity relation joint extraction model, using the sequential long short term memory networks (LSTM) to decode the entities, the relation is decoded using the Tree LSTM shortest dependency path.

(2) LSTM-ED: The bidirectional long short term memory networks (Bi-LSTM) are used to encode and decode the entities. Moreover, convolution neural networks (CNNs) are used to distinguish the classes of the relation.

(3) Attention-DA: Attention is used to determine the contextual semantics information and perform the encoding process.

(4) LSTM-CRF: This approach represents a typical sequential label entity relation joint extraction model, using the LSTM networks to encode and the LSTM and CRF to decode.

(5) LM-HA: This method is an enhanced joint entity and relation extraction model with an auxiliary training objective for the language modeling. The vital semantic information is captured through the hierarchical multihead attention.

(6) CMAN: This approach is a deep cross-modal attention network to realize the joint entity and relation extraction. The framework is constructed by stacking multiple attention units in depth, with fully dense interactions over token label spaces.

(7) DAG: A graph for the entity relation is constructed based on the transition parsing model.

(8) AGGCN [25]: This method represents a novel attention-guided graph convolution model, directly taking the full dependency trees as the inputs.

(9) DYGIE: The method involves a span graph framework to realize the joint overlapping relation extraction.

(10) BAER [29]: An end-to-end neural network is used for the entity relation extraction, and rearrangement of the boundaries and edges is performed.

TABLE 2. Experiment results on the CONLL04 and NYT datasets.

Dataset-model		Entity(CoNLL04)	Relation(CoNLL04)	Entity(NYT)	Relation(NYT)
Nongraph	LSTM-SDP	73.5	63.4	59.89	47.56
	LSTM-ED	70.43	61.39	62.49	49.38
	Attention-DA	69.98	60.37	60.77	47.58
	LSTM-CRF	70.18	63.25	61.97	48.63
	LM-HA	72.96	63.09	64.09	50.72
	CMAN	74.12	64.38	64.52	50.39
Graph	DAG	85.60	67.80	66.97	52.97
	AGGCNs [25]	—	69.0	—	54.88
	DYGIE	85.91	66.40	66.17	53.84
	BAER [29]	87.48	69.64	68.30	56.47
	SG-DAGCNs	**89.04**	**71.78**	**70.15**	**58.13**

D. COMPARISON WITH THE STATE-OF-THE-ART APPROCHES

To further verify that the proposed SG-DAGCN framework can more effectively extract the overlapping relation among entities compared with the state-of-the-art entity and relation extraction models, we utilized the CoNLL04 and NYT public benchmark datasets. TABLE 2 presents the experiments results of the different entity relation extraction methods on the CoNLL04 and NYT public datasets. The state-of-the-art methods contain the graph (most graph methods extract the relation between the entities) and an end-to-end model without the graph structure.

In TABLE 2, "–" indicates that the joint entity relation extraction models are not experimentally tested on the relevant datasets. The bold value represents the best experiment results on the public benchmark datasets. "Nongraph" indicates that the entity relation extraction frameworks do not contain the graph structure of the end-to-end models. TABLE 2 lists the F1 scores.

The following conclusions can be attained according to the experiment results presented in TABLE 2:

(1) For the CoNLL04 and NYT benchmark datasets, the proposed SG-DAGCN overlapping relation framework achieves the optimal experiment results. For instance, the relation extraction F1 score of the SG-DAGCN framework is higher than that of the BAER by 2.14% and 1.66%, respectively, on the CoNLL04 and NYT datasets. The experiment results indicate that the proposed SG-DAGCN can realize better embedding and enhance the interaction of the relation between the entities. Moreover, the high order semantics information of the entity and relation data is distinctly formulated and combined with the low order important detail information by using the multiple stream semantics-guided framework, which enhances the entity relation extraction performance. The topology of the graph corresponds to adaptive learning for different graph convolution layers, and the multiscale and hierarchical structure information are aggregated through the dynamically updated neighbor nodes, which can facilitate the entity recognition and overlapping relation extraction.

(2) For all the experiment results on the CoNLL04 and NYT public benchmark datasets, the graph-based end-to-end framework outperforms the nongraph structure models in terms of the F1 score on the entity recognition and overlapping relation extraction. For example, the entity recognition F1 score of the LSTM-SDP models is smaller to that of the DAG by 12.10% on the CoNLL04 dataset. The main reason is that the graph-based end-to-end framework can effectively capture the high order semantics information between the entities by transmitting the neighbor node features. Moreover, the interaction between the entities and relation is strengthened. This finding demonstrates the superiority of the graph-based end-to-end framework.

E. ABLATION EXPERIMENTS OF THE COMPONENTS

To further verify the effect of the propose SG-DAGCN overlapping relation extraction framework of the semantics-guided component, different numbers of multiscale dynamic aggregation graph convolution blocks and different scales of the graph convolution layers, we analyze the experiment results for each component on the NYT public benchmark dataset in the overlapping relation extraction task. The experiment results are presented in TABLE 3, in which the evaluation indexes are the precision (P), recall (R) and F1 score (%). In TABLE 3, "S = 1,2,3,4,5" indicate the scale of the MS-GCN layers. "M = 1,2,3,4,5" indicate the number of multiscale graph convolution blocks. "SG" indicates the multistream semantics-guided module. "DAGCN" indicates the dynamic aggregation multiscale graph convolution module. "SC" is the connection feature of the sentence and character. "WC" indicates the connection feature of the word and character. "R" and "E" denote the relation and entities, respectively. "S/DAGCNs (Nonguided)" indicates the entity relation extraction framework that does not contain the self-guided attention layers. "SG/MS-CGNs (No-DA)" indicates that dynamic aggregation is not used in the propose SG-DAGCN model.

The following conclusions can be obtained considering the experiment results in TABLE 3. (1) In terms of the multi-stream semantics-guided module, the "SC(R_E)/DAGCNs" framework achieves better results than those of the other models in terms of the relation extraction, with 54.86% (P), 53.33% (R) and 54.08% (F1) values on the NYT public benchmark datasets. Moreover, the "SC(R_E)/DAGCNs" framework outperforms the "WC(R_E)/DAGCNs" by 2.66% (F1). in addition, the relation extraction performance

TABLE 3. Experiment results of different components on the NYT dataset.

Dataset-model		P (NYT-Relation extraction)	R (NYT-Relation extraction)	F1 (NYT-Relation extraction)
SG Module	SC(R)/DAGCNs	50.29	48.85	49.55
	WC(R)/DAGCNs	49.72	48.27	48.98
	SC(E)/DAGCNs	52.61	51.21	51.90
	WC(E)/DAGCNs	51.94	50.54	51.23
	SC(R_E)/DAGCNs	57.86	54.33	56.04
	WC(R_E)/DAGCNs	54.11	52.67	53.38
DAGCN Module	SG/DAGCN $(S=1, M=1)$	52.61	51.24	51.92
	SG/DAGCN $(S=2, M=2)$	55.93	54.48	55.20
	SG/DAGCN $(S=4, M=4)$	55.03	53.62	54.31
	SG/DAGCN $(S=5, M=5)$	54.27	52.91	53.58
Joint Component	S/DAGCNs $(Nonguided)$	57.49	56.18	56.82
	SG/MS-CGNs $(No-DA)$	56.88	55.74	54.56
	SG-DAGCNs $(S=3, M=3)$	**58.86**	**57.43**	**58.13**

of the "SC" model is higher than that for the "WC" model. Compared to the "WC(R)/DAGCNs" framework, the relation extraction results of the "SC(R)/DAGCNs" model is improved by 0.57%, 0.58% and 0.57%. The extraction performance indicates that the syntax representation from the Bi-LSTM tree is more important than the word embedding MLP representation. The main reason is that the Bi-LSTM Tree model can reduce the use of the irrelevant redundant information.

(2) For the DAGCN module, with an increase in the scale $(S = 1, 2, 3, 4, 5)$ and number of MS-GCN blocks $(M = 1, 2, 3, 4, 5)$, the extraction performance first increases and later decreases. For instance, the "$SG/DAGCN(S = 4, M = 4)$ " framework outperforms the "$SG/DAGCN(S = 5, M = 5)$ " by 0.76% (P), 0.71% (R) and 0.73% (F1), because the large-scale semantic information contains more irrelevant redundant information. Although different scales can capture more information levels for the entities and relation, when the scale and number are excessively large, the important detailed information may be lost. Therefore, the scale size depends on the actual situation.

However, the "$SG/DAGCN(S = 2, M = 2)$ " model outperforms "$SG/DAGCN(S = 1, M = 1)$" by 3.32% (P), 3.24% (R) and 3.28% (F1). The main reason is that more scale information can be captured, and different scales exhibit complementary structural information. This aspect can alleviate the deficiency of the single scale structure information representational capacity of the entities and relation.

(3) In terms of the joint component configuration, the relation extraction methods exhibit different performances when we eliminate one of the components such as the self-guided attention layers or the dynamic aggregation module. Compared to the model without the dynamic aggregation technique (SG/MS-CGNs), the framework without the self-guided attention layer (S/DAGCNs) achieves a higher performance; however, this overlapping relation extraction framework still exhibits a competitive performance compared to that of the other state of the art methods. The main reason is that the multiscale dynamic aggregation graph convolution approach can effectively obtain the deep semantic information of the entities and relation by aggregating and transferring the neighbor node information. Moreover, the interaction between the adjacent nodes is used to strengthen the dependence between the entities and relation. In addition, this information flow is gradually refined through the self-guided attention layers, thereby reducing the use of the redundant information.

F. ANALYSIS AND DISCUSSION

1) ABLATION STUDY

To further analyze the contributions of the two main components, that is, the semantics-guided layers and multiscale dynamic aggregation graph convolution module, the S/DAGCNs, SG/MS-GCNs, proposed SG-DAGCNs (highest performing) and other state of the art methods such as the BAER are evaluated on the CoNLL04 public benchmark dataset. The performance of the four overlapping relation extraction models under different settings for the training with different numbers of training samples is shown in FIGURE 3. In the experiment, we consider four training settings, that is, 1%, 10%, 20%, 30% of the number of training samples.

According to the experiment results shown in FIGURE 3, adding the self-attention module or multiscale dynamic aggregation graph convolution module enhances the performance of the overlapping extraction framework. This finding suggests that both the modules can improve the performance of the entity relation extraction, where the dynamic aggregation technique plays a more important role. Specifically, this technique can assist the multiscale graph convolution networks (MS-GCNs) to aggregate the neighborhood node information and produce the optimal representation for the entity relation graph.

In contrast, as the number of training samples increases, the performance gap gradually increases. The highest performance corresponds to the SG-DAGCN when using 1% of the training samples, and the F1 score is 34.75%, corresponding to an improvement of 7.27% and 5.32% for the BAER and S/DAGCNs methods, respectively. These experiment results

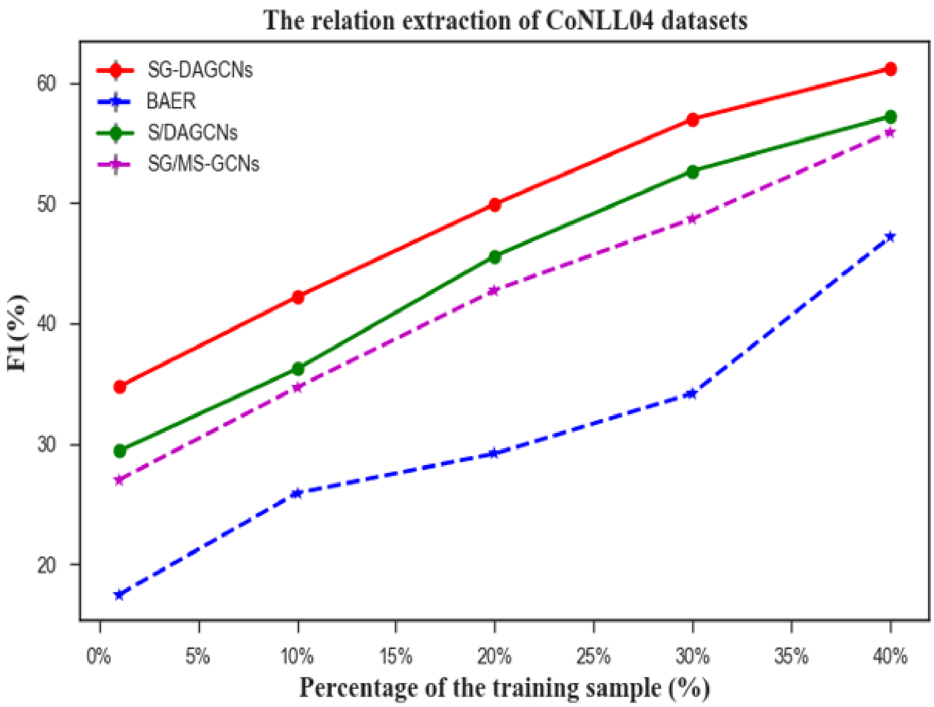


FIGURE 3. Comparison of the S/DAGCNs, SG/MS-GCNs, SG-DAGCNs, and BAER models under different training sample sizes.

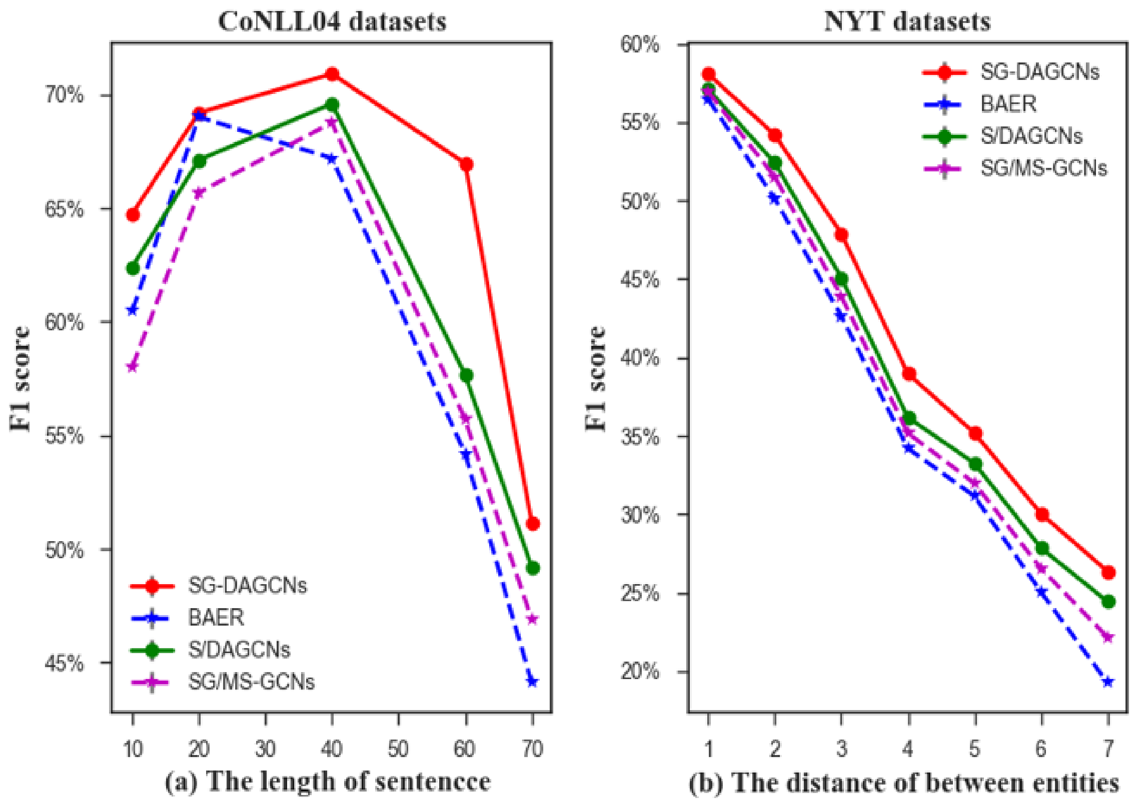


FIGURE 4. Performance pertaining to different sentence lengths and entity distance for the CoNLL04 dataset and NYT dataset, respectively.

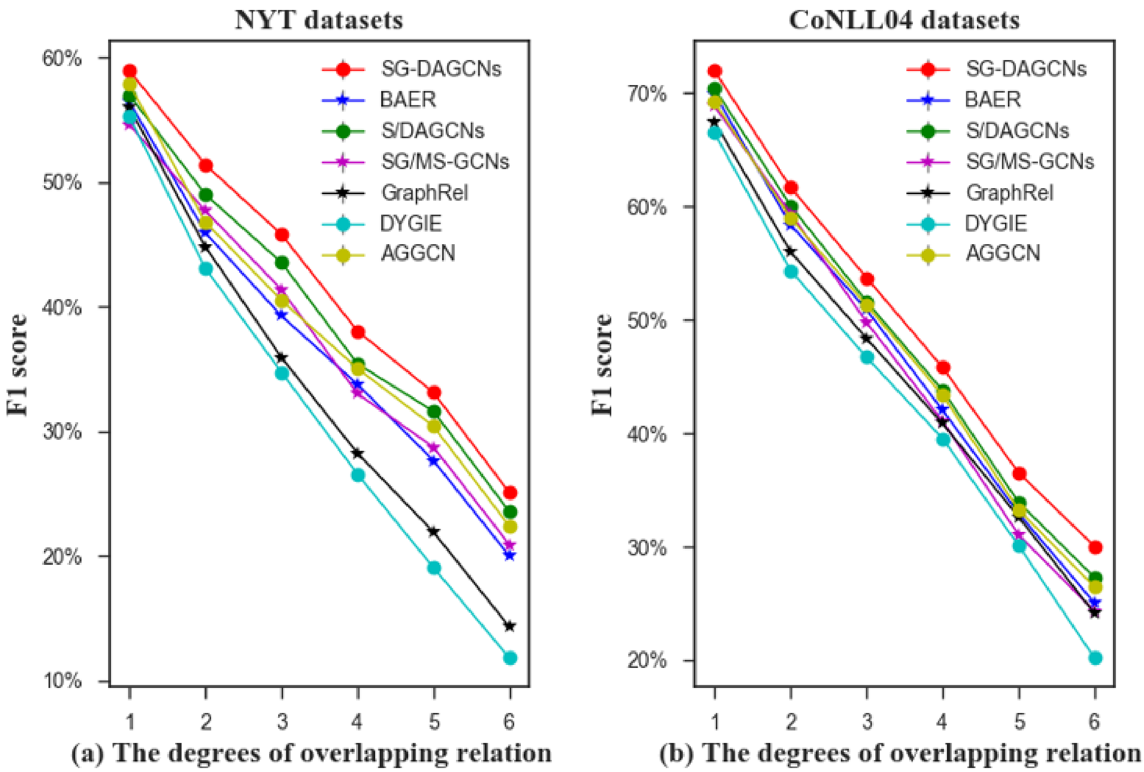


FIGURE 5. Performance for the different degrees of overlapping relation on the CoNLL04 and NYT public benchmark datasets.

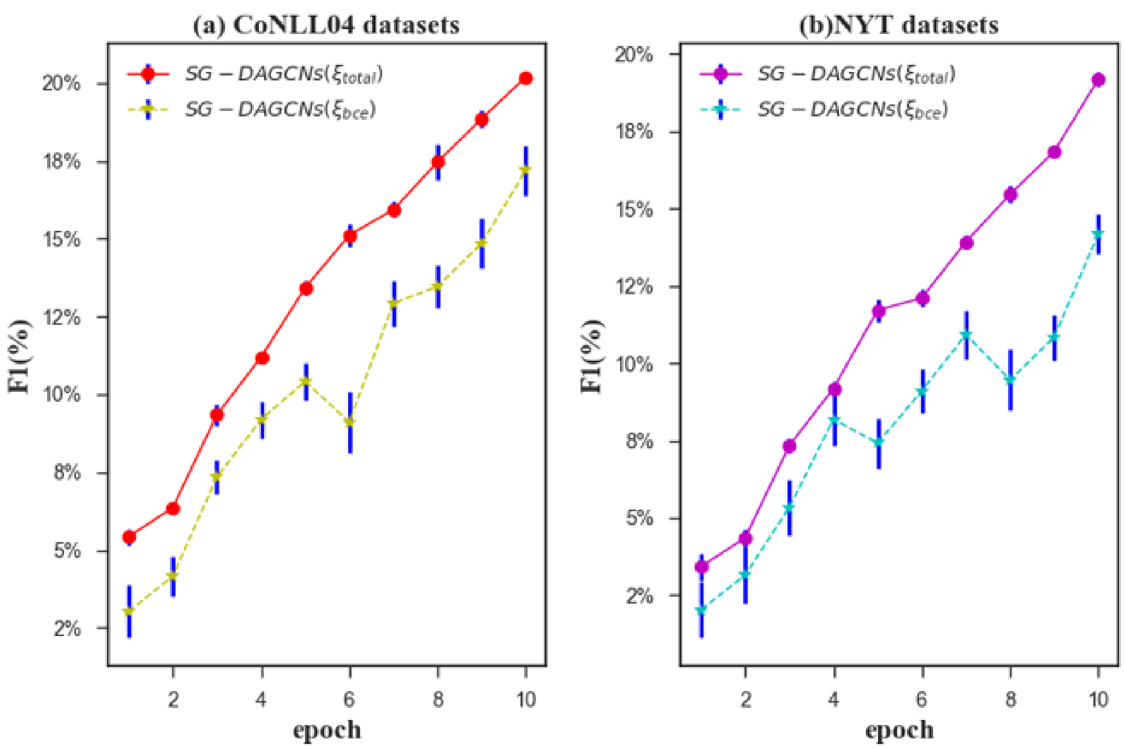


FIGURE 6. Performance of different loss functions. ξ_{total} and ξ_{bce} indicate the total loss function and binary cross-entropy loss function, respectively. The blue vertical line indicates the error of the loss $\xi_{total} = \xi_{bce} + \xi_{sf}$.

demonstrate that the proposed SG-DAGCN overlapping relation extraction framework is more effective when using small scale samples and further indicates the superiority of the proposed method over other methods.

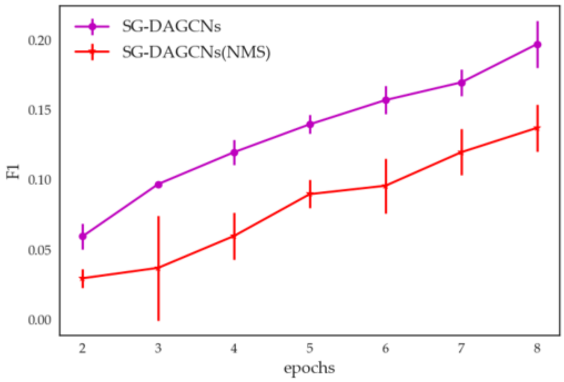


FIGURE 7. Performance of multi-scale information. SG-DAGCNs indicate non multi-scale information.

2) SENTENCE LENGTH AND ENTITY DISTANCES

This section describes the performance of the different methods for different sentence lengths and entity distances. Experimental verification is performed on the NYT public benchmark dataset. The experiment results are shown in FIGURE 4, with the evaluation metric of the F1 score. To examine the accuracy, we divide the sentence length and entity distance into multiple subsequences, with the subsequence of the sentence set as $SL \in \{10, 20, 40, 60, 70\}$. indicates the length of the sentence. The distance between the entities is set as $ED \in \{1, 2, 3, 4, 5, 6, 7\}$, where ED indicates the distance between the entities

According to the performance results in FIGURE 4 (a), the proposed three overlapping extraction frameworks exhibit a higher performance, and the SG-DAGCNs with the semantics-guided module and dynamic aggregation approach outperform the other two models under all the sentence lengths. Moreover, the S/DAGCNs with dynamic aggregation outperform the BAER model in most situations. In addition, the performance first increases and later decreases for the four extraction frameworks as the sentence length increases. The dynamic aggregation technique provides more important detail information of the high order case and the underlying multiscale graph structures. Intuitively, as the sentence length increases, the dynamic aggregation multiscale graph can capture more information of the neighborhood nodes.

According to the results of different distances between the entities, as shown in FIGURE 4 (b), the proposed frameworks outperform the other models. As the distance between the entities increases, the extraction performance degrades considerably for all the methods, although the proposed SG-DAGCN framework is competitive against the state-of-the-art methods. Even when the distance between the entities is 7, the proposed three overlapping relation extraction frameworks can realize a more accurate extraction of the relations among the entities. The F1 score is 26.35%, 24.42% and 22.15%. This finding further suggests that the proposed SG-DAGCNs can effectively solve the problem of the long-distance dependence.

G. EFFECTIVENESS OF OVERLAP EXTRACTION

We further discuss and analyze the extraction performance for difference degrees of overlapping relations for multiple methods and conduct experimental verification on the CoNLL04 and NYT public benchmark datasets. First, we partition the overlapping relation into six subsequences $OLR \in \{1, 2, 3, 4, 5, 6\}$, where OLR indicates the overlapping relation. FIGURE 5 shows the F1 score of the six methods under different degrees of the overlapping relation.

According to the experiment results in FIGURE 5, the proposed SG-DAGCN model achieves the highest performance on the NYT and CoNLL04 benchmark datasets, even as the overlapping degree increases. When the overlapping degrees is 5 or 6, the proposed SG-DAGCN model outperformance the other state of the art methods. These experiment results further suggest the superiority of the semantics-guided layers and dynamic aggregation technique in capturing and aggregating the important detail semantic information of the nodes.

As illustrated in FIGURE 5, Compared with the BEAR and AGGCN, the S/DAGCNs and SG/MS-GCNs models exhibit a higher performance owing to their enhanced ability in capturing the multiscale structure information and reinforcing the interaction of the entities and relation.

H. EFFECTIVENESS OF THE LOSS FUNCTION

To verify the effectiveness of the designed loss function, we use the CoNLL04 and NYT benchmark datasets. FIGURE 6 shows the experiment results.

According to FIGURE 6, the performance of the model using the total loss function is higher than that of the model using the binary cross-entropy loss function. Specifically, $SG - DAGCNs(\xi_{total})$ outperforms $SG - DAGCNs(\xi_{bce})$ on the CoNLL04 and NYT datasets. Moreover, the error of the loss is large when using the binary cross-entropy loss training network, that is, the performance of the method is limited in the overlapping extraction task, and the interaction ability between the entities and relation is weakened. This finding demonstrates the effectiveness of the designed loss function.

I. EFFECTIVENESS OF MULTI-SCALE INFORMATION

To verify the effectiveness of the multi-scale information for improve the information interaction between entities and relationships, we use the CoNLL04 benchmark datasets. FIGURE 7 shows the experiment results.

According to FIGURE 7, the performance of the model using the multi-scale information is higher than that of the model using the non multi-scale information. Specifically, "SG-DAGCNs" outperforms "SG-DAGCNs (NMs)" on the CoNLL04 datasets. Moreover, the error of the loss is large and performance is poor when non-using the multi-scale information training network, that is, the performance of the method is limited in the overlapping extraction task, and the interaction ability between the entities and relation is enhanced. This finding demonstrates the effectiveness of the designed multi-scale information.

V. CONCLUSION AND FUTURE WORK

We propose a semantics-guided dynamic aggregation graph convolution network to realize the overlapping relation extraction. The framework can effectively capture the possible overlapping relation between entities through the semantics-guided and dynamic aggregation techniques. The experiment results show that the proposed SG-DAGCN method outperforms the other methods on the CoNLL04 and NYT benchmark datasets. Moreover, the effectiveness of the proposed SG-DAGCN framework is demonstrated for the overlapping relation extraction tasks. Future research will be focused on the network structure and external feature information. Specifically, we aim to design a simple and efficient semantic framework to achieve accurate extraction of the overlapping relationships. Network structure: Design a more robust graph structure or use graph attention networks to reduce the loss of information flow passing between the nodes and layers. External feature information: Design an effective feature extraction network, such as by improving the Tree LSTM structure, and retain the most discriminative detail information.

AUTHOR CONTRIBUTIONS

Xiushan Liu designed the experiment to evaluate the performance and wrote the article. Jun Chen and Qin Zhang supervised the study and reviewed the article. All the authors have read and agreed to the published version of the article.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments and helpful suggestions.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

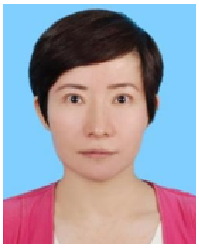
REFERENCES

[1] F. Li, M. Zhang, G. Fu, and D. Ji, "A neural joint model for entity and relation extraction from biomedical text," *BMC Bioinf.*, vol. 18, no. 1, p. 198, Dec. 2017.
[2] Y. Luan, L. He, M. Ostendorf, and H. Hajishirz, "Multi-task identification of entities," in *Proc. Conf. Sci. Knowl. Graph Construct.*, 2018.
[3] M. Zhao, Y. Zhao, and B. Xu, "Knowledge graph completion via complete attention between knowledge graph and entity descriptions," in *Proc. 3rd Int. Conf. Comput. Sci. Appl. Eng.*, 2019, pp. 1–6.
[4] C. Fu, "Tracking user-role evolution via topic modeling in community question answering," *Inf. Process. Manage.*, vol. 56, no. 6, Nov. 2019, Art. no. 102075.
[5] F. Aqlan, X. Fan, A. Alqwbani, and A. Al-Mansoub, "Improved Arabic–Chinese machine translation with linguistic input features," *Future Internet*, vol. 11, no. 1, p. 22, Jan. 2019.
[6] N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-T. Yih, "Cross-sentence N-ary relation extraction with graph LSTMs," *Trans. Assoc. Comput. Linguistics*, vol. 5, pp. 101–115, Dec. 2017.
[7] Q. Li and H. Ji, "Incremental joint extraction of entity mentions and relations," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 402–412.
[8] F. Yuntian, W. Hao, and G. Chen, "Joint extraction of entities and relations using reinforcement learning and deep learning," *Comput. Intell. Neuroence*, vol. 2017, Aug. 2017, Art. no. 7643065.
[9] Z. Wei, J. Su, Y. Wang, Y. Tian, and Y. Chang, "A novel cascade binary tagging framework for relational triple extraction," 2019, *arXiv:1909.03227*. [Online]. Available: https://arxiv.org/abs/1909.03227
[10] R. Takanobu, T. Zhang, J. Liu, and M. Huang, "A hierarchical framework for relation extraction with reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 7072–7079.
[11] J. Li, M.-T. Luong, D. Jurafsky, and E. Hovy, "When are tree structures necessary for deep learning of representations?" 2015, *arXiv:1503.00185*. [Online]. Available: http://arxiv.org/abs/1503.00185
[12] T.-J. Fu, P.-H. Li, and W.-Y. Ma, "GraphRel: Modeling text as relational graphs for joint entity and relation extraction," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1409–1418.
[13] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proc. 25th Int. Conf. Comput. Linguistics*, 2014, pp. 2335–2344.
[14] D. Dai, X. Xiao, Y. Lyu, S. Dou, Q. She, and H. Wang, "Joint extraction of entities and overlapping relations using position-attentive sequence labeling," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 6300–6308.
[15] M. Zhang, Y. Zhang, and G. Fu, "End-to-end neural relation extraction with global optimization," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 1730–1740.
[16] D. Q. Nguyen and K. Verspoor, "End-to-end neural relation extraction using deep biaffine attention," in *Proc. Eur. Conf. Inf. Retr.* Cham, Switzerland: Springer, 2019, pp. 729–738.
[17] L. Hong, J. Lin, J. Tao, and J. Zeng, "BERE: An accurate distantly supervised biomedical entity relation extraction network," 2019, *arXiv:1906.06916*. [Online]. Available: http://arxiv.org/abs/1906.06916
[18] Z. Geng, G. Chen, Y. Han, G. Lu, and F. Li, "Semantic relation extraction using sequential and tree-structured LSTM with attention," *Inf. Sci.*, vol. 509, pp. 183–192, Jan. 2020.
[19] M. Liu, Y. Zhang, W. Li, and D. Ji, "Joint model of entity recognition and relation extraction with self-attention mechanism," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 19, no. 4, pp. 1–19, Jul. 2020.
[20] Y. Luan, D. Wadden, L. He, A. Shah, M. Ostendorf, and H. Hajishirzi, "A general framework for information extraction using dynamic span graphs," 2019, *arXiv:1904.03296*. [Online]. Available: http://arxiv.org/abs/1904.03296
[21] Y. Zhang, P. Qi, and C. D. Manning, "Graph convolution over pruned dependency trees improves relation extraction," 2018, *arXiv:1809.10185*. [Online]. Available: http://arxiv.org/abs/1809.10185
[22] S. Kumar Sahu, F. Christopoulou, M. Miwa, and S. Ananiadou, "Inter-sentence relation extraction with document-level graph convolutional neural network," 2019, *arXiv:1906.04684*. [Online]. Available: http://arxiv.org/abs/1906.04684
[23] X. Ning and Y. Qi, "Semi-supervised convolution graph kernels for relation extraction," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2011, pp. 510–521.
[24] N. Zhang, S. Deng, Z. Sun, G. Wang, X. Chen, W. Zhang, and H. Chen, "Long-tail relation extraction via knowledge graph embeddings and graph convolution networks," 2019, *arXiv:1903.01306*. [Online]. Available: http://arxiv.org/abs/1903.01306
[25] Z. Guo, Y. Zhang, and W. Lu, "Attention guided graph convolutional networks for relation extraction," 2019, *arXiv:1906.07510*. [Online]. Available: http://arxiv.org/abs/1906.07510
[26] J. Dunietz and D. Gillick, "A new entity salience task with millions of training examples," in *Proc. 14th Conf. Eur. Chapter Assoc. Comput. Linguistics*, 2014, pp. 205–209.
[27] X. Ren, Z. Wu, W. He, M. Qu, C. R. Voss, H. Ji, T. F. Abdelzaher, and J. Han, "CoType: Joint extraction of typed entities and relations with knowledge bases," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1015–1024.
[28] T. Tran and R. Kavuluru, "Neural metric learning for fast end-to-end relation extraction," 2019, *arXiv:1905.07458*. [Online]. Available: http://arxiv.org/abs/1905.07458
[29] H. Fei, Y. Ren, and D. Ji, "Boundaries and edges rethinking: An end-to-end neural model for overlapping entity relation extraction," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102311.
[30] R. Chi, B. Wu, L. Hu, and Y. Zhang, "Enhancing joint entity and relation extraction with language modeling and hierarchical attention," in *Proc. Asia–Pacific Web (APWeb) Web-Age Inf. Manage.*, 2019, pp. 314–328.
[31] H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10076–10085.
[32] Y. Sun, Y. Wang, Z. Liu, J. E. Siegel, and S. E. Sarma, "PointGrow: Autoregressively learned point cloud generation with self-attention," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 61–70.

[33] X. Yu, G. Li, C. Chai, and N. Tang, "Reinforcement learning with tree-LSTM for join order selection," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 1297–1308.
[34] D. Roy, P. Panda, and K. Roy, "Tree-CNN: A hierarchical deep convolutional neural network for incremental learning," *Neural Netw.*, vol. 121, pp. 148–160, Jan. 2020.
[35] L. Zhu and H. Zheng, "Biomedical event extraction with a novel combination strategy based on hybrid deep neural networks," *BMC Bioinf.*, vol. 21, no. 1, p. 47, Dec. 2020.
[36] M. Li, S. Chen, Y. Zhao, Y. Zhang, Y. Wang, and Q. Tian, "Dynamic multiscale graph neural networks for 3D skeleton based human motion prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 214–223.
[37] S. Abu-El-Haija, A. Kapoor, B. Perozzi, and J. Lee, "N-GCN: Multi-scale graph convolution for semi-supervised node classification," in *Proc. Uncertainty Artif. Intell. Conf.*, 2020, pp. 841–851.

XIUSHAN LIU was born in Hunan, China. He received the M.S. degree in computer technology from the South China University of Technology, Guangzhou, China. He is currently an Associate Professor with the School of Electronic and Information, Guangdong Polytechnic Normal University. His current research interests include artificial intelligence and pattern recognition.

JUN CHENG was born in Hubei, China. He received the B.S. and M.S. degrees in electronic and information engineering from the South China University of Technology, Guangzhou, China, in 2002 and 2005, respectively. He is currently an Associate Professor with the School of Electronic and Information, Guangdong Polytechnic Normal University. His current research interests include artificial intelligence, machine learning, and deep learning.



QIN ZHANG was born in Hubei, China. She received the M.S. degrees in measuring and testing technology from Wuhan University, Wuhan, China, in 2001. She is currently an Associate Professor with the School of Electronic and Information, Guangdong Polytechnic Normal University. Her current research interests include pattern recognition and deep learning.