

Received October 11, 2020, accepted November 5, 2020, date of publication November 11, 2020,  
date of current version December 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3037086

# Relation Extraction for Chinese Clinical Records Using Multi-View Graph Learning

CHUNYANG RUAN<sup>1</sup>, YINGPEI WU<sup>2</sup>, GUANG SHENG LUO<sup>1</sup>, (Member, IEEE),  
YUN YANG<sup>3</sup>, AND PINGCHUAN MA<sup>3</sup>

<sup>1</sup>Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University, Shanghai 200083, China

<sup>2</sup>School of Software Engineering, Fudan University, Shanghai 200433, China

<sup>3</sup>Department of Oncology, Longhua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai 201204, China

Corresponding author: Guang Sheng Luo (luoguangsheng03@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61672161 and in part by the Youth Research Fund of Shanghai Municipal Health and Family Planning Commission under Grant 2015Y0195.

**ABSTRACT** Relation extraction is a necessary step in obtaining information from clinical medical records. In the medical domain, there have been several studies on relation extraction in modern medicine clinical notes written in English. However, very limited relation extraction research has been conducted on clinical notes written in Chinese, especially traditional Chinese medicine (TCM) clinical records (e.g., herb-symptom, herb-disease). Instead of independently extracting each relation from a single sentence or text, we propose to globally and reasonably extract multiple types of relations from the Chinese clinical records with a novel heterogeneous graph representation learning method. Specifically, we first construct multiple view medical entity graphs based on the co-occurring relations, knowledge obtained from the clinic, and domain texts with the corresponding information of two medical entities from the Chinese clinical records, in which each edge is a candidate relation; we then build a Graph Convolutional Network (GCN)-based representation learning with the attention mechanism to simultaneously infer the existence of all the edges via classification. The experimental data were obtained from the Chinese medical records and literature provided by previous work. The main experimental results on Chinese clinical records show that our proposed model's precision, recall, and F1-score reach 10.2%, 13.5%, 12.6%, demonstrating significant improvements over state-of-the-art.

**INDEX TERMS** Relation extraction, graph attention networks, Chinese clinical records, representation learning.

## I. INTRODUCTION

Despite the rise of semi-structured and structured data, the text is still the most widespread content in the real world. To extract meaningful information and knowledge from free text is the subject of considerable research interest in the natural language processing (NLP) fields. Text mining has become one of the most active research sub-fields in data mining. Relation extraction (RE) is an essential sub-task of text mining, which aims to discover relations between entity pairs  $e_1$  and  $e_2$  given unstructured text data.

With the tremendous growth in the adoption of electronic medical records (EMRs) contains a vast of medical information, such as the relation between treatment and disease, is becoming available as a treasure trove for large-scale health

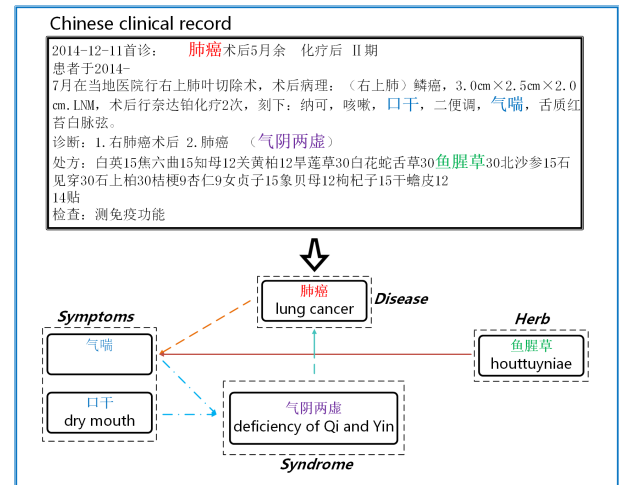
data analysis [1], [2]. Extracting the relevant information existing among medical concepts from EMRs automatically is of great importance to many medical applications, which further support the clinical decision making and government health policy formulation. However, most of the information in current medical records is stored in natural language texts, which makes data mining algorithms unable to process these data directly. To extract the medical information from the EMRs, researchers generally use entity and relation extraction algorithm, which can be processed by conventional data mining algorithms directly. As a kind of EMR, Chinese electronic medical records contain much information about clinical diagnoses and treatment events. Since the publication of the Basic Norms of EMRs in China, a solid body of data has been generated as a result of the unprecedented expansion of EMRs. Although the Chinese Ministry of Health has issued a series of relevant regulations [3], Chinese

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

medical records, especially traditional Chinese medicine (TCM), contain a large amount of clinical information, such as the chief complaint, four diagnoses, and treatment measures, stored as unstructured data in the clinical narrative. Previous research on RE has mainly focused on English text. However, with the widespread use of EMRs in China and continuous development of machine learning techniques, an increasing number of researchers have conducted RE research in the Chinese medical domain. Due to the lack of digitalization and formalization in Chinese clinical records, studies focusing on Chinese clinical records are relatively limited, especially on TCM clinical records.

On the other hand, Chinese clinical medicine is a very complicated medical system in which multiple types of entities are involved, such as “herb,” “formula” (a composition that consists of certain herbs), “symptom,” and “syndrome” (disharmonies at the core of the body, a complex pattern of signs and symptoms, which has no direct parallel in western medicine). Multiple types of intricate relations can exist between these heterogeneous medical entities, such as composition relations between herbs and formulae, treatment relations between formulae and diseases, effectiveness relations between herbs and symptoms, and association relations between syndromes and diseases. Detecting these latent and useful relations is the goal of RE for Chinese clinical records mining. Every researcher contributes their discoveries to the Chinese medical knowledge base to form a large-scale, multi-source, and the unstructured pool of natural language text data [4].

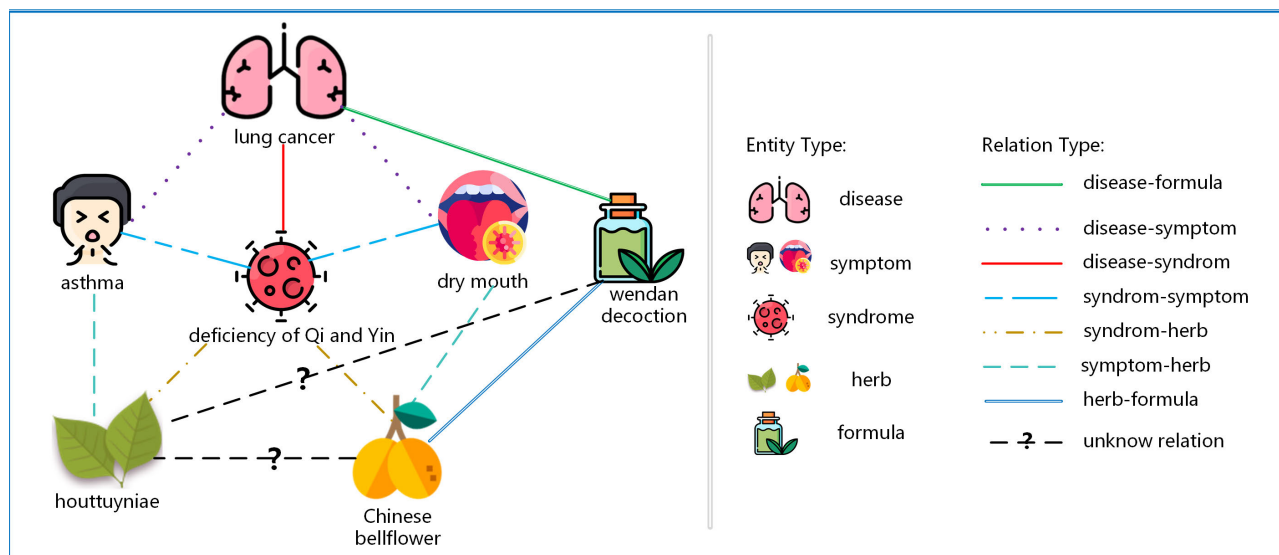
In this study, we aim to solve the issue of extracting relations from this corpus of Chinese text. Specifically, given a set of Chinese clinical records, our task is to discover all relations between the instances of multiple types of entities in corpus. One of the main objectives of relation extraction from Chinese clinical records is to help discover new knowledge about effective personalized treatment. In other words, the most significant relations between medical entities can be used to assist clinical treatment or other clinical research. In addition, the extracted relations may promote the understanding of Chinese medicine in western medicine. Relation extraction from Chinese clinical text is more complicated than relation extraction from biomedical text. The main challenge of relation extraction from Chinese clinical text is the complexity of the structure and system of Chinese clinical records. Multiple types of complex relations exist between a lot of heterogeneous medical entities, which is not advisable to extract a single type of relation independently of the others. Fig. 1 gives an example of extracting multiple types of correlations between heterogeneous Chinese medical entities. In addition, most Chinese clinical records are written in Chinese, a language in which the sentences have no spaces between words and, therefore, word segmentation is needed to automatically divide sentences into words. Errors in word segmentation obstruct feature generation in the relation extraction process. At last, the vast majority of Chinese clinical records are not exactly like natural language sentences. These texts keep their



**FIGURE 1.** An example of extracting multiple types of correlated relations between heterogeneous traditional Chinese medicine (TCM) entities. Words in red font indicate a disease, words in blue font indicate a formula, and words in green font indicate a herb. The relations extracted from different documents are correlated with one another through their common entities.

own way of organizing the medical entities, which are often put in a weakly ordered way. In this article, we propose a new method to complete the task of relation extraction, to globally and reasonably extract relations from the entire corpus of Chinese clinical records from the perspective of graph representation learning. Specifically, we first construct a heterogeneous entity graph from the co-occurring relations, which can be obtained easily from the records. Specifically, we take all types of medical entities that occur in the records as nodes and create an edge for each pair of heterogeneous entities co-occurring in the same text. All the edges are treated as candidate relations to be identified. Then, we use domain knowledge to enhance the semantic information of the above graph. Our hypothesis is that the prior knowledge of associations among medical entities in the domain documents can be used to discover latent relations, thereby improving the quality of relation extraction for clinical records. Fig. 2 gives a simple example of the heterogeneous entity graph. We then propose a novel graphical relation extraction model, called the multi-view graph model to extract relation (MVG2RE), to simultaneously infer the labels of all the candidate relations by employing the node representation learning and classification. To evaluate the performance of our proposed method, we trained and evaluated our *MVG2RE* on a high-quality corpus provided by previous work. We set up extensive experiments, including quantitative analysis, qualitative analysis, and expert evaluation on this dataset. As expected, our approach leads to more accurate matching between entities in records than baseline approaches. The main contributions of this article include the following:

- We propose a focused heterogeneous graph representation learning model to jointly learn the task of RE. The model integrates graph neural network model as a shared parameter layer to achieve better generalization



**FIGURE 2.** An example of a simple heterogeneous traditional Chinese medicine (TCM) network. Different icons and line styles indicate different types of TCM entities and candidate relations, respectively, which are closely correlated with one another. Many triadic closures are formed in the network.

performance. Our proposed model *MVG2RE* is able to learn a better heterogeneous medical graph representation benefit from attention mechanism.

- Co-occurrence, lexical feature representation and semantic information are jointly utilized to extract the of each entity relation. We propose a model that makes full use of rich external knowledge to enhance the RE performance for Chinese clinical records.
- Our model achieves new state-of-the-art results without additional computational overhead when compared with previous works.

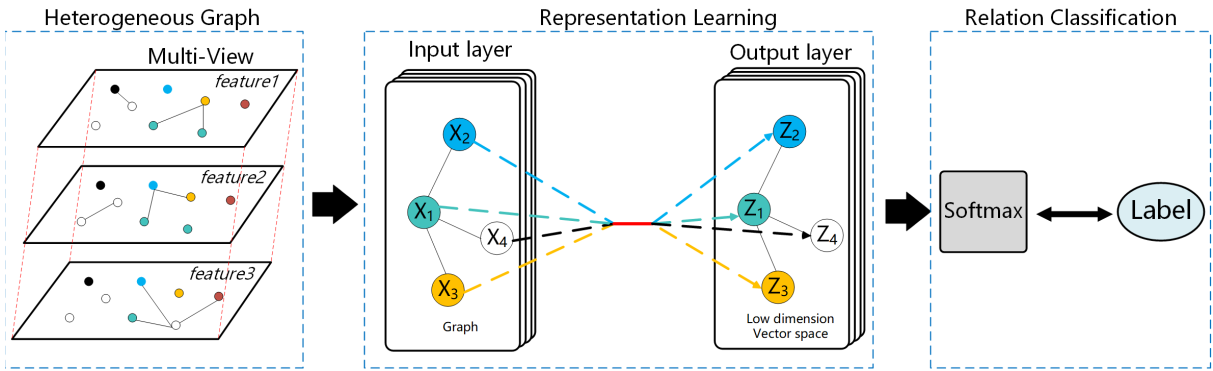
## II. RELATED WORK

The concept of RE was first put forward at the Message Understanding Conference (MUC) and supported by the Defense Advanced Research Projects Agency (DARPA) at the end of the 1980s [5]. RE for medical data refers to the detection and classification of relation mentions among different medical entities within the primary texts. The goal of RE is to detect the occurrences of pre-specified types of relationships between entity pairs. Comparing with the types of medical entities, the types of entity relations are more diverse [6]. There have been many existing studies on medical relation extraction. There has been some work with different methods for medical relation extraction for English EMRs. Dependency trees were used to support the process of querying interactions between genes and proteins [7]. The traditional model, such as support vector machines (SVM), was early employed to identify the relations among diseases, symptoms, drugs, and treatments [8]. A model was proposed in a study [9] to identify the semantic relations among medical concepts from the medical texts and to analyze them to discover medical knowledge. In order to extract more complex relations in medical records, a hybrid

method was proposed based on machine learning, dictionary, and rules [10]. [11] proposed a knowledge-guided SVM approach to discover relations between medical problems, treatments, and tests mentioned in electronic medical records. This model extract feature from records used knowledge datasets that Wikipedia, WordNet and general inquirer [12].

At present, the research studies of relation extraction in Chinese mainly focus on the open domain, and the methods of relation extraction in Chinese EMRs are still in the preliminary stage [5]. Some work focused on using rule, such as co-occurrence, and traditional Chinese medicine integrative database to extract relation among herb, gence, syndrome and disease, and to complete molecular mechanism analysis. On the other hand, a series of NLP methods, such as word segmentation and syntactic parsing, have been proposed to handle Chinese open relation extraction [13]. For example, A semi-supervised model was proposed in work [14] based on bootstrapping to discover the knowledge of gene functional, including extracting the relation among disease, symptom and gene in TCM bibliographic literature and MEDLINE.<sup>1</sup> The topic model is a popular method for relation extraction in TCM literature. These work utilized topic model to construct classifier that can learn the latent topic distribution and discover the intricate relations among herb, symptom and syndrome [15], [16]. In the medical field, the dependency graph was used to automatically learn the syntactic pattern of relation extraction and extracted the relation between disease, drug and symptom by this model [16], [17]. Some work using graph model [4], [18], [19] achieved good performance in the task of RE for Chinese text discovery, which can avoid the pitfalls that the characteristic in Chinese texts.

<sup>1</sup><https://jgc128.github.io/mednli>



**FIGURE 3.** Framework of the RE model. The model consists of five parts: graph input layer based on different knowledge, vector representation layer based on graph convolutional network layer, and output layer.

All the above-mentioned approaches only focus on mining the corpus itself, however, these heterogeneous relations may be dependent on another knowledge, through their common entities, which failed to comprehensively describe how a relation is generated using medical knowledge well. Just as different people view the same clinical record and get different understandings, models need to understand the information in the data from different perspectives according to prior knowledge.

### III. METHODOLOGY

In this section, we first introduce the problem definition of relations extraction in Chinese clinical records. Second, we describe three basic features for relation extraction of Chinese clinical records in more detail. And we use the features as different views to build different heterogeneous entity graphs. Finally, by fusing the attention mechanism, a method of relation extraction based on graph embedding is proposed. The framework of our proposed model multi-view graph attention model termed *MVG2RE* is shown in Fig. 3.

#### A. PROBLEM DEFINITION

In this article, we attempt to extract relations in the context of a heterogeneous medical graph from clinical records. So, we first give the definition of the heterogeneous medical graph, then present the problem formulation.

Let  $V$  and  $E$  be a set of node and a set of edge respectively. We use  $V_\phi$  ( $V_\phi \in V$ ) to denote the set of medical entities with type  $\phi$ , and use  $E_\varphi$  ( $E_\varphi \in E$ ) to denote the set of relation with type  $\varphi$ . Now, we define a heterogeneous medical graph as the  $G = (V, E)$ , where  $V = \cup_{\phi \in \Phi} V_\phi$ ,  $E = \cup_{\varphi \in \Psi} E_\varphi$ .

Given a heterogeneous medical graph  $G = (V, E)$ , the goal of relation extraction is to learn the function  $F$  to predict the label of candidate relations between clinical entities, which can be defined as:

$$F : G = (V, E; \theta) \mapsto l \quad (1)$$

where  $\theta$  is a parameter to be solved,  $l$  is the label. In this article, we aim to identify the reliability of each candidate relation. Therefore, we have  $l = 0, 1$ , where the value of

label is 1 means an edge is reliable; otherwise 0 means it is unreliable.

#### B. OBTAINING VIEW-ORIENTED FEATURES

Aspect-based (also known as aspect-level) sentiment classification aims at identifying the sentiment polarities of aspects explicitly given in sentences [20]. Learning from this idea of mining text from multiple view, the features of graph should be able to reflect prior knowledge of the labels of edges. we aim to obtain view-oriented features by applying multi-view graph neural networks over the context of a clinical record, and imposing an view-specific graph according to different features. Inspired by work [4], we employed three view that categories of features to construct our heterogeneous medical graph:

- **Co-occurring view:** Intuitively, two words exist in the same sentence, and they may have latent relation. This is the simplest feature to built a edge, which represents the instances in which the two entities of a relation co-occur in the same record. We use this feature to built the first-view graph, which is a naive topological structure.
- **Lexical view:** Intuitively, the lexical feature that a context surrounding the two entities is very important for identifying a relation from a record. After removing the infrequent words, we first use k-surrounding words were collected for each instance of an entity. We then calculate the relative distance between each word and entity pair, which can indicate the frequency that each word appeared around the two end-entities of a relation. We use this feature to built the second-view graph, which is a context-level graph.
- **Semantic view:** The latent semantic relevance of two medical entities may also be helpful for identifying a useful relation. For calculating semantic distance between words in a record, we first employ word embedding [21] to represent the semantic meanings of each entity. This distributed vector representations facilitate learning word meanings from large collections of text. According to the similar study [22], we automatically extract entity that we need from three Chinese medical

graph (KG).<sup>234</sup> Specifically, a word  $w_k$  has a corresponding entity  $e_i$  in each KG, we will extract its word embedding  $x_k$ , entity embedding  $e_i$ . In Chinese clinical records, an entity is usually consisting of several words (e.g., “chronic colitis” is an entity that composed of two words: ‘chronic’ and ‘colitis’). Therefore, the knowledge-enhanced word embedding  $X_k$  is calculated by integrating word embedding and entity embedding, which is formulated in the following equation:

$$X_k = e_i \bigoplus \sum_{k=1}^n x_k \quad (2)$$

where  $n$  is the number of words in the entity  $e_i$ , and  $\bigoplus$  is the concatenation operation. If a word has no corresponding entity in the KG, we will learn the word embedding directly. Notably, we employ TransE [23] to achieve the KG representation learning.

### C. HETEROGENEOUS GRAPH EMBEDDING

Before aggregating the information from different heterogeneous graph based on above three view for each entity, we first should get node embedding in each heterogeneous graph according to the study [24]. In each heterogeneous graph, we should notice that the meta-path [25] based neighbors of each node play a different role and show different importance in learning node embedding for the specific task [26]. Graph convolutional network [27] is a multi-layer neural network that operates directly on graph data and induces the embedding vectors of nodes based on the properties of their neighborhoods. Here we mathematically illustrate how multi-layer GCN work on a graph. Given a graph  $G = (V, E)$  where  $V$  and  $E$  denote the set of nodes and edges respectively. We can get the adjacency matrix  $A$  and degree matrix  $D$  based on the topology of the graph  $G$ . The convolution computation for node  $i$  at the  $l$ -th layer of GCN, which takes the input feature representation  $h^{(l-1)}$  as input and outputs the induced representation  $h_i^{(l)}$ , can be defined as:

$$\mathbf{h}_i^{(l)} = \sigma \left( \sum_{j=1}^n \mathbf{A}_{ij} \mathbf{W}^{(l)} \mathbf{h}_j^{(l-1)} + \mathbf{b}^{(l)} \right) \quad (3)$$

where  $\mathbf{W}^{(l)}$  is the weight matrix,  $\mathbf{b}^{(l)}$  is the bias vector, and  $\sigma$  is an activation function (e.g., tanh).  $\mathbf{h}_i^{(0)}$  is the initial input  $x_i$  that the initial representation of node  $i$ , where  $\mathbf{x}_i \in \mathbb{R}^d$  and  $d$  is the input feature dimension.

Generally, a heterogeneous graph contains multi-type node [28] leads to GCN cannot be directly applied to the HIN. Here we draw on the idea of the work [29] that considers the difference of various types of information and projects them into an implicit common space with a dual-level attention mechanism, which had achieve a good performance

for semi-supervised short text classification. We use the heterogeneous graph convolution to learn the representation of heterogeneous entities formulated as follows:

$$H^{(l+1)} = \sigma \left( \sum_{\tau \in \mathcal{T}} \tilde{A}_\tau \cdot H_\tau^{(l)} \cdot W_\tau^{(l)} \right) \quad (4)$$

where  $\tilde{A}_\tau \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}_\tau|}$  is the sub-matrix of  $A$ , whose rows represent all the nodes and columns represent their neighboring nodes with the type  $\mathcal{T}$ . The representation of the nodes  $H^{(l+1)}$  is obtained by aggregating information from the features of their neighboring nodes  $H_\tau^{(l)}$  with different types using different transformation matrix  $W_\tau^{(l)} \in \mathbb{R}^{q^{(l)} \times q^{(l+1)}}$ . The transformation matrix  $W_\tau^{(l)}$  considers the difference of different feature spaces and projects them into an implicit common space  $\mathbb{R}^{q^{(l+1)}}$ . Intuitively, different types of neighboring nodes for a specific node may have different impacts on it. Especially in clinical records, different types of entities may keep a different weight with the specific entity. For example, both drugs are prescribed to treat a disease, but one is an adjunct, which result in a different weights of edges between the two drugs and the disease. To solve this problem, we also dual-level attention to capture both the different importance at both node level and type level mechanism. We can replace Eq.4 with the following layer-wise propagation rule:

$$H^{(l+1)} = \sigma \left( \sum_{\tau \in \mathcal{T}} \mathcal{B}_\tau \cdot H_\tau^{(l)} \cdot W_\tau^{(l)} \right) \quad (5)$$

where,  $\mathcal{B}_\tau$  was calculated by the dual-level attention, which is a attention matrix. Its  $i$ -th row represents the  $i$ -th node  $v_i$ , and its  $j$ -th column represents the  $j$ -th attention weight between node  $v_i$  and  $v_j$ .

### D. MULTI-VIEW GRAPH AGGREGATION

Getting representation of each heterogeneous graph, we here introduce the architecture of graph aggregation. The aim of each graph representation learning is to extract the significant information from a view, and to obtain the low-dimension embedding of each node. Now, we should aggregate the information on each graph to get a multi-view representation for each entity. We employ self-attention [30] graph transformer networks [24] to capture the global features of multi-view graph. Formally, given three heterogeneous entity graph with multiple types of meta-paths, they can be fused and defined by a tensor  $\mathbb{H} \in \mathbf{R}^{N \times N \times C}$ ,  $C$  is the number of view. Furthermore, we average the importance of all the graph representation, which can be explained as the importance of each view. The importance of each view, denoted as  $w_{V_i}$ , is shown as follows:

$$\alpha_P = \frac{1}{|V|} \sum_{i \in V} \mathbf{q}^T \cdot \sigma \left( \mathbf{W} \cdot h_i^P + \mathbf{b} \right) \quad (6)$$

where  $\mathbf{W}$  is the weight matrix,  $\mathbf{b}$  is the bias vector,  $q$  is the view-level attention vector and  $\sigma$  is a activation function.

<sup>2</sup><http://www.tcminformatics.org/wiki/index.php/graph>

<sup>3</sup><http://cmekg.pcl.ac.cn/>

<sup>4</sup>[http://www.tcmkb.cn/knowledge\\_graph/](http://www.tcmkb.cn/knowledge_graph/)

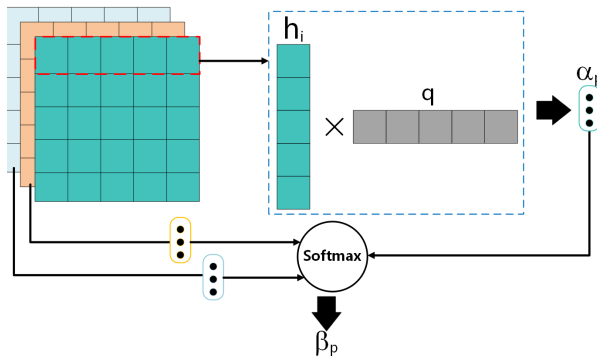


FIGURE 4. Explanation of aggregating process in node representation for multi-view graph.

Note that for the meaningful comparison, all above parameters are shared for all view and semantic-specific embedding. After obtaining the importance of each view, we normalize them via the softmax function. The weight of global graph  $\beta$ , denoted as each  $\alpha$ , can be obtained by normalizing the above importance of all views using softmax function, which is formulated as follows:

$$\beta_P = \frac{\exp(\alpha_P)}{\sum_{P=1}^K \exp(\alpha_P)} \quad (7)$$

where  $K$  is the number of view, and in our study the  $K = 3$ . Eq.7 can be interpreted as the contribution of the view  $P$  for global information. Obviously, the higher  $\beta_P$ , the more important view  $P$  is. With the learned weights as coefficients, we can fuse these semantic-specific representation to obtain the final representation  $Z$  as follows:

$$Z = \sum_{P=1}^3 \beta_P \cdot H_P \quad (8)$$

where  $H_P$  is the heterogeneous graph constructed by view  $P$ , it is a  $N \times N \times 1$  tensor. To better understand the aggregating process of semantic-level, we also give a brief explanation in in Fig. 4.

### E. RELATION CLASSIFICATION

The final embedding is aggregated by all semantic-specific embedding. Then we can apply the final embedding to specific tasks and design different loss function. For supervised node classification, we can minimize the Cross-Entropy over all labeled node between the ground-truth and the prediction. we can obtain the final the  $i$ -th entity vector  $Z_i$  and then feed it into softmax layer for relation classification, the probability distribution of the relation labels are shown in the following equation:

$$O = \text{softmax}(W_o Z_i + b) \quad (9)$$

Then, the conditional probability of the relation label  $L$  of the entity is described as follow:

$$L_Z = \text{argmax}(O) \quad (10)$$

TABLE 1. The Statistics of the Dataset TCMRelExtr.

Relation Type	Number of Unique Candidate Relations		
	Labelled		Unlabelled
	Positive	Negative	
Herb-Syndrome	538	582	10077
Herb-Disease	534	642	10579
Formula-Syndrome	392	574	8693
Formula-Disease	377	411	7094
Syndrome-Disease	431	532	8681

TABLE 2. The Statistics of Parameter Settings.

Parameter	Size
Batch Size	50
Epoch	25
Max sentence length	100
Dropout	0.3
Learnign rate	1e-3
Hidden dimension of model	512
Embedding dimension of graph	150

## IV. EXPERIMENTS AND RESULTS

### A. DATASETS

We evaluate our proposed model *MVG2RE* on real-world datasets TCMRelExtr<sup>5</sup> from hospitals and literature. The datasets contain five kinds of relations: 'herb-syndrome', 'herb-disease', 'formula-disease', 'formula-syndrome', and 'syndrome-disease'. The datasets are annotated by a group of medical doctors. The statistics of our dataset are summarized in Tab. 1. The ratio between the training set and the testing set is 4:1. We use common performance measures such as Accuracy, Precision, Recall and F1-score to evaluate our model [31]. The parameter settings for model training is shown in Tab. 2.

### B. BASELINES

In this study, the task of entity relation extraction can be transformed into a multi-classification problem. In order to verify the effectiveness of our model more comprehensively. We compare our proposed model with traditional classifier, and classifier based on topic model deep learning, GCN. We adopt the following approaches for comparison.

#### 1) TRADITIONAL CLASSIFICATION APPROACH

- **SVM-C** [32] is a Support Vector Machine (SVM) classifier based on feature vectors including the influential entities' own features and entities' context features to extract the five main relationships in medical literature, including 'ISA', 'PART\_OF', 'CAUSES', 'TREATS' and 'DIAGNOSES'.

#### 2) TOPIC MODEL BASED APPROACH

- **PTM** [15]: It proposes a novel topic model integrating knowledge to capture the correlations of symptoms, diseases, and herbs. We set the number of topics is 20 with 5 herbs predicted as best parameters set for comparison.

<sup>5</sup><http://arnetminer.org/TCMRelExtr>

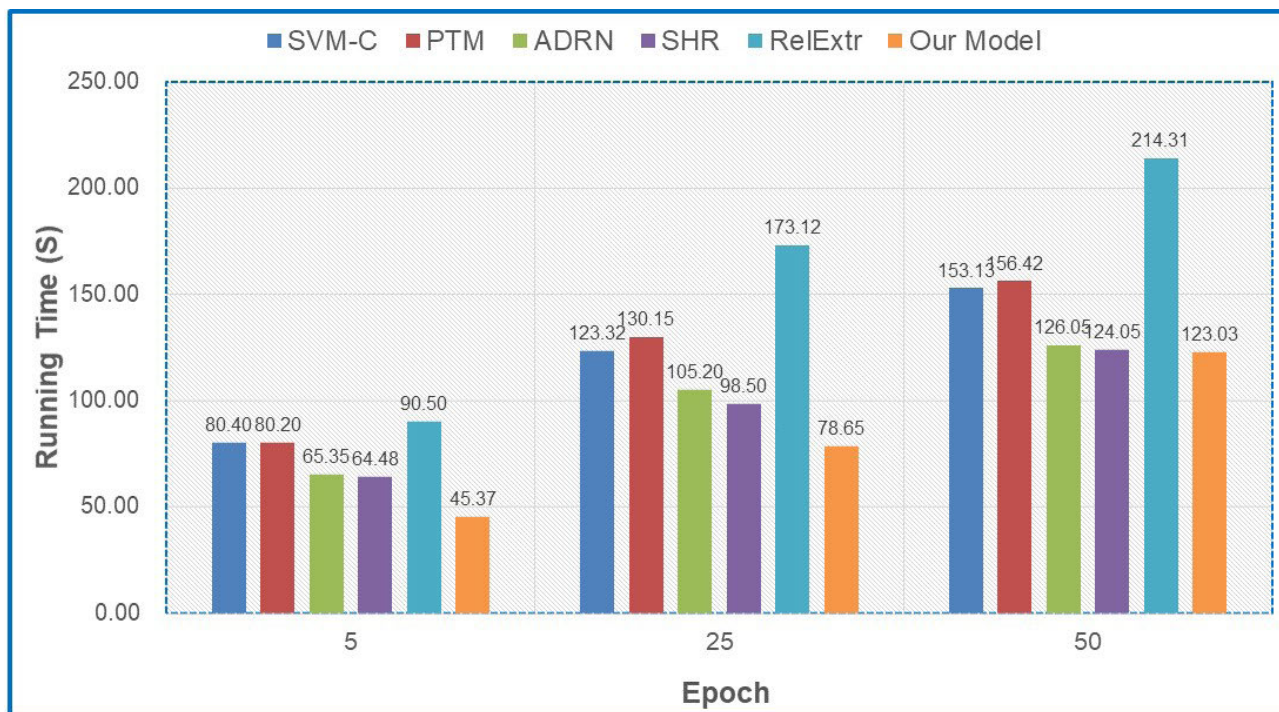


FIGURE 5. Comparison of the running time for different model in three epoch.

TABLE 3. Performance of Relation Extraction by Different Models (%).

Model		Relation Type						Average	Improve Rate(%)
		HS	HD	FS	FD	SD			
SVM-C	Precision	77.65	78.15	79.32	80.13	76.54	78.36	-	
	Recall	71.36	73.50	71.32	72.09	70.21	71.70	-	
	F1	74.37	75.75	75.10	75.90	72.13	74.65	-	
PTM	Precision	78.35	79.12	80.31	81.05	79.23	79.61	+1.000	
	Recall	73.61	74.32	75.71	76.31	74.05	74.80	+4.000	
	F1	75.91	76.64	77.94	78.61	75.84	77.01	+3.100	
ADRN	Precision	76.54	77.28	80.21	81.07	77.04	78.43	-0.008	
	Recall	71.13	72.31	74.63	75.25	74.37	70.54	-3.900	
	F1	74.27	75.24	77.85	78.59	76.20	76.43	-2.400	
SHR	Precision	78.29	79.25	80.18	81.24	77.50	79.29	+1.100	
	Recall	72.56	74.39	72.31	73.12	73.02	73.08	+1.900	
	F1	76.32	76.74	76.04	76.97	75.20	76.23	+2.100	
RelExtr	Precision	83.05	82.91	83.24	85.15	82.00	83.54	+6.600	
	Recall	77.62	79.31	77.65	78.31	78.39	78.36	+9.200	
	F1	80.24	81.07	80.91	81.82	80.07	80.87	+8.300	
MVG2RE	Precision	85.12	86.06	87.23	88.35	85.15	86.38	+10.200	
	Recall	80.23	81.03	82.32	83.21	80.31	81.42	+13.500	
	F1	82.60	83.47	85.94	85.70	82.66	84.07	+12.600	

3) DEEP LEARNING-BASED APPROACH

- **ADRN** [33]: is a deep residual network model based on the attention mechanism to classify the relation of entity pairs in Chinese EMRs. It reduced the influence of data noise on the model training, and enhance entity discrimination feature with position attention mechanism.

4) GRAPH LEARNING-BASED APPROACH

- **SHR** [18] used a series of GCNs to simultaneously learn the symptom embedding and herb embedding from the relationships of symptom-herb, and recommend herbs with syndrome induction from symptoms.
- **RelExtr** [4] is to exploit the power of collective inference in the context of heterogeneous entity networks to

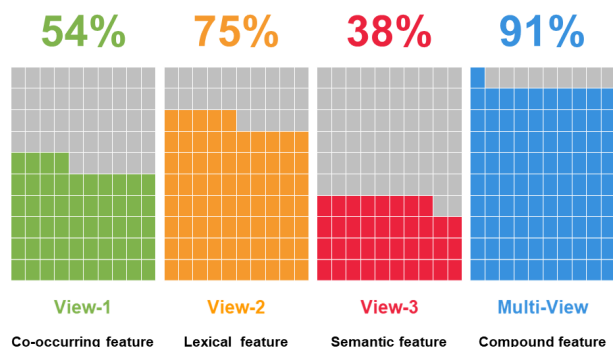
simultaneously and globally extract all types of relations using a semi-supervised learning algorithm to estimate the parameters of the model.

C. QUANTITATIVE EVALUATION

We performed five-fold cross-validation to evaluate the performance of our proposed model *MVG2RE*. Tab. 3 shows the performance of our proposed model *MVG2RE* as compared with that of four other approaches. Our results show that our methods significantly outperform all the baselines by a large margin, which shows the effectiveness of our proposed method on entity relations classification in Chinese clinical records. We set the SNM-C as a benchmark and

**TABLE 4.** Performance of Expert Judgment Evaluation for the Relation Extraction in Clinical Records Manually.

Relation Type	Annotator Agreements		Average Score
	expert-1	expert-2	
Herb-Syndrome	0.723	0.7180	0.721
Herb-Disease	0.814	0.794	0.804
Formula-Syndrome	0.783	0.823	0.803
Formula-Disease	0.794	0.817	0.805
Syndrome-Disease	0.745	0.841	0.793

**FIGURE 6.** Accuracy versus the number of views in relation extraction.

our proposed model *MVG2RE* gets the best improve rate. The traditional method SVMs based on the human-designed features, achieve better performance than the deep models with the random initialization, i.e., ADNR in most cases. We can see that PTM performs better than SVM-C and ADNR, which means considering efficacy knowledge leads to the best predictive power. The performances of graph-based approaches with regard to the extraction of different types of relations and found that the difference between them, which is extremely statistically significant features. Our proposed model consistently outperforms all the state-of-the-art models by a large margin, which not only captures the importance of different neighboring nodes but also the importance of different types of nodes. And, our proposed model gets more context semantics in clinical records enhanced by more knowledge fusion. In addition, our proposed method uses GCN to extract the relations between different entities, which is better at aggregating information from different nodes; at the same time, we use the attention mechanism to distinguish the relations between different nodes. Based on the above analysis, our model *MVG2RE* is capable of achieving good performance in task of relation extraction.

To further determine the effectiveness of our approach, we plotted running time of the all approaches as shown in Fig. 5, where the y-axis represents the value of running time in seconds, and the x-axis represents the three number of epoch, where epoch is set as 5, 25 and 50 respectively. It is clear that our proposed significantly outperforms the baselines on extracting all types of relations. And the traditional machine learning method SVM-C takes the longest time to train.

#### D. EXPERT JUDGMENT EVALUATION

In this part, we conduct a case study to verify the rationality of our proposed relation extraction approach. Tab. 4 shows the statistics to measure inter-annotator agreements that experts

evaluate the results of relation extraction in clinical records manually. The scores among the two annotators are mostly above 0.7, indicating substantial agreement for all relation types. The average inter-annotation agreement was 78.5%. Through the above qualitative analysis, we can find that our proposed model *MVG2RE* has the ability to extract relations in clinical records reasonably.

#### E. EVALUATION OF MULTIPLE VIEWS

In this study, there exist multiple view features in the task. Thus, we intend to measure whether such view would affect the effectiveness of our proposed model. We employ accuracy as metric, and test the difference accuracy of relation extraction of our proposed model under each single view and compound views. It can be seen in Fig. 6 that when the number of views in the task is 3, the accuracy becomes the best, indicating a low robustness in capturing multiple-view correlations and suggesting the need of modelling multi-view dependencies in similar work such as biomedical relations extraction.

#### V. CONCLUSION

We have re-examined the challenges encountering existing models for relation extraction in Chinese clinical records, and pointed out the suitability of the graph model for solving these problem. Accordingly, we have proposed a multi-view graph representation learning that adopts heterogeneous GCN and attention mechanism for complex relation extraction among multi-type entities. Experimental results have indicated that our proposed model *MVG2RE* brings benefit to the overall performance by leveraging both global information and local structure dependencies.

This study may be further improved in the following aspects. First, the edge information of the entity, i.e., the weight of each edge, is not exploited in this work. We plan to design a specific graph neural network that takes into consideration the edge information. Second, more domain knowledge can be incorporated. Last but not least, our model *MVG2RE* may be extended to jointly recognize medical entity diseases and classify the relations between entities.

#### ACKNOWLEDGMENT

We gratefully acknowledge all reviewers for their thoughtful comments and efforts towards improving our manuscript and the clinical data providers, Prof. Tian and Dr. Yang, for their hard work.

#### REFERENCES

- [1] Q. Zhao, J. Li, C. Xu, J. Yang, and L. Zhao, "Knowledge-enhanced relation extraction for chinese EMRs," *IT Prof.*, vol. 22, no. 4, pp. 57–62, Jul. 2020.
- [2] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," *IEEE J. Biomed. Health Informat.*, vol. 22, no. 5, pp. 1589–1604, Sep. 2018.
- [3] *Basic Specification for Electronic Medical Records (Trial)*, M. of Health, 2010.
- [4] H. Wan, M.-F. Moens, W. Luyten, X. Zhou, Q. Mei, L. Liu, and J. Tang, "Extracting relations from traditional chinese medicine literature via heterogeneous entity networks," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 2, pp. 356–365, Mar. 2016.



- [5] P. Zhai, X. Huang, B. Zhang, and Y. Fang, "Relation extraction based on fusion dependency parsing from chinese EMRs," *Sci. Program.*, vol. 2020, pp. 1–9, Jun. 2020.
- [6] S. Zhao, C. Su, Z. Lu, and F. Wang, "Recent advances in biomedical literature mining," *Briefings Bioinf.*, May 2020.
- [7] F. Rinaldi, G. Schneider, K. Kaljurand, M. Hess, C. Andronis, O. Konstandi, and A. Persidis, "Mining of relations between proteins over biomedical scientific literature using a deep-linguistic approach," *Artif. Intell. Med.*, vol. 39, no. 2, pp. 127–136, Feb. 2007.
- [8] O. Uzuner, J. Mailoa, R. Ryan, and T. Sibanda, "Semantic relations for problem-oriented medical records," *Artif. Intell. Med.*, vol. 50, no. 2, pp. 63–73, Oct. 2010.
- [9] X. Zhu, C. Cherry, S. Kiritchenko, J. Martin, and B. de Bruijn, "Detecting concept relations in clinical text: Insights from a state-of-the-art model," *J. Biomed. Informat.*, vol. 46, no. 2, pp. 275–285, Apr. 2013.
- [10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," 2013, *arXiv:1301.3781*. [Online]. Available: <http://arxiv.org/abs/1301.3781>
- [11] B. Rink, S. Harabagiu, and K. Roberts, "Automatic extraction of relations between medical concepts in clinical texts," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 594–600, Sep. 2011.
- [12] G. Psathas, "The general inquirer: Useful or not?" *Comput. Humanities*, vol. 3, no. 3, pp. 163–174, Jan. 1969.
- [13] Y.-H. Tseng, L.-H. Lee, S.-Y. Lin, B.-S. Liao, M.-J. Liu, H.-H. Chen, O. Etzioni, and A. Fader, "Chinese open relation extraction for knowledge acquisition," in *Proc. 14th Conf. Eur. Chapter Assoc. Comput. Linguistics, Short Papers*, vol. 2, 2014, pp. 12–16.
- [14] X. Zhou, B. Liu, Z. Wu, and Y. Feng, "Integrative mining of traditional chinese medicine literature and MEDLINE for functional gene networks," *Artif. Intell. Med.*, vol. 41, no. 2, pp. 87–104, Oct. 2007.
- [15] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, "A topic modeling approach for traditional Chinese medicine prescriptions," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1007–1021, Jun. 2018.
- [16] W. Ji, Y. Zhang, X. Wang, and Y. Zhou, "Latent semantic diagnosis in traditional Chinese medicine," *World Wide Web*, vol. 20, no. 5, pp. 1071–1087, Sep. 2017.
- [17] M. Sun, S. Zhao, C. Gilvary, O. Elemento, J. Zhou, and F. Wang, "Graph convolutional networks for computational drug development and discovery," *Briefings Bioinf.*, vol. 21, no. 3, pp. 919–935, May 2020.
- [18] Y. Jin, W. Zhang, X. He, X. Wang, and X. Wang, "Syndrome-aware herb recommendation with multi-graph convolution network," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 145–156.
- [19] C. Ruan, J. Ma, Y. Wang, Y. Zhang, and Y. Yang, "Discovering regularities from traditional chinese medicine prescriptions via bipartite embedding model," in *Proc. 28th Int. Joint Conf. Artif. Intell.*, S. Kraus, Ed., Macao, China, Aug. 2019, pp. 3346–3352.
- [20] C. Zhang, Q. Li, and D. Song, "Aspect-based sentiment classification with aspect-specific graph convolutional networks," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds. Stroudsburg, PA, USA: Association for Computational Linguistics, Nov. 2019, pp. 4567–4577.
- [21] M. Biesialska, B. Rafieian, and M. R. Costa-jussà, "Enhancing word embeddings with knowledge extracted from lexical resources," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, S. Rijhwani, J. Liu, Y. Wang, and R. Dror, Eds. Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 271–278.
- [22] Q. Zhao, J. Li, C. Xu, J. Yang, and L. Zhao, "Knowledge-enhanced relation extraction for chinese EMRs," *IT Prof.*, vol. 22, no. 4, pp. 57–62, Jul. 2020.
- [23] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 12, pp. 2724–2743, Dec. 2017.
- [24] S. Yun, M. Jeong, R. Kim, J. Kang, and H. J. Kim, "Graph transformer networks," in *Proc. Adv. Neural Inf. Process. Syst. 32, Annu. Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, and R. Garnett, Eds., Dec. 2019, pp. 11960–11970.
- [25] X. Fu, J. Zhang, Z. Meng, and I. King, "MAGNN: Metapath aggregated graph neural network for heterogeneous graph embedding," in *Proc. Web Conf.*, Y. Huang, I. King, T. Liu, and M. van Steen, Eds., Taipei, Taiwan, Apr. 2020, pp. 2331–2341.
- [26] 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.*, L. Liu, R. W. White, A. Mantrach, F. Silvestri, J. J. McAuley, R. Baeza-Yates, and L. Zia, Eds., San Francisco, CA, USA, May 2019, pp. 2022–2032.
- [27] 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>
- [28] C. Shi, Y. Li, J. Zhang, Y. Sun, and P. S. Yu, "A survey of heterogeneous information network analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 1, pp. 17–37, Jan. 2017.
- [29] H. Linmei, T. Yang, C. Shi, H. Ji, and X. Li, "Heterogeneous graph attention networks for semi-supervised short text classification," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 4820–4829.
- [30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. 30, Annu. Conf. Neural Inf. Process. Syst.*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., Long Beach, CA, USA, Dec. 2017, pp. 5998–6008.
- [31] S. Gao, J. Du, and X. Zhang, "Research on relation extraction method of Chinese electronic medical records based on BERT," in *Proc. 6th Int. Conf. Comput. Artif. Intell.*, Tianjin, China, Apr. 2020, pp. 487–490.
- [32] X. Zhao, S. Lin, and Z. Huang, "Extraction of semantic relations from medical literature based on semantic predicates and SVM," in *Proc. Health Inf. Sci. 7th Int. Conf. (HIS)*, Cairns, QLD, Australia, in Lecture Notes in Computer Science, vol. 11148, S. Stuly, I. Lee, Z. Huang, R. Zhou, H. Wang, and W. Xiang, Eds. Cham, Switzerland: Springer, Oct. 2018, pp. 17–24.
- [33] Z. Zhang, T. Zhou, Y. Zhang, and Y. Pang, "Attention-based deep residual learning network for entity relation extraction in chinese EMRs," *BMC Med. Informat. Decis. Making*, vol. 19, no. S2, pp. 171–177, Apr. 2019.



**CHUNYANG RUAN** is currently pursuing the Ph.D. degree with the School of Computer Science, Fudan University, Shanghai, China. He is currently a Lecturer with the Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University, Shanghai. His current research interests include graph embedding, graph neural networks, heterogeneous information networks, data analysis, and knowledge discovery.



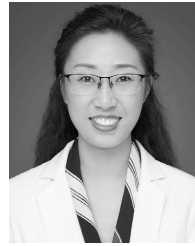
**YINGPEI WU** is currently pursuing the Ph.D. degree with the School of Computer Science, Fudan University, Shanghai, China. His current research interests include heterogeneous information networks, graph neural networks, representation learning, and traditional Chinese medicine informatization.



**GUANG SHENG LUO** (Member, IEEE) received the B.S. degree in computer science from the Huazhong University of Technology, Wuhan, China, in 2005, and the M.S. degree in computer science from Fudan University, Shanghai, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Computer Science. His research interests include deep learning, smart financial, and automatic speech recognition.



**YUN YANG** received the B.S. degree in traditional Chinese medicine and the M.S. degree in internal medicine of traditional Chinese medicine from the Longhua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai, China, in 2008 and 2012, respectively, where she is currently pursuing the Ph.D. degree. Her research interests include the basic and clinical research of traditional Chinese medicine treatment of tumor, the combination of traditional Chinese medicine, and artificial intelligence.



**PINGCHUAN MA** received the master's degree in computer science from East China Normal University. She is currently a Senior Lecturer with the Computer Department, Shanghai Economic Management School. Her current research interest includes programming education.

• • •