

Received 8 October 2022, accepted 1 November 2022, date of publication 4 November 2022, date of current version 16 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3219988

RESEARCH ARTICLE

MetaGNN-Based Medical Records Unstructured Specialized Vocabulary Few-Shot Representation Learning

HONGXING LING¹, GUANGSHENG LUO², AND YU YANG³

¹Business and Information College, Shanghai 200235, China

²Jiangxi University of Finance and Economics, Nanchang 330088, China

³AISHU Information Technology Company Ltd., Shanghai 201900, China

Corresponding author: Guangsheng Luo (luoguangsheng03@126.com)

This work was supported by the Postdoctoral Station of Management Science and Engineering of Jiangxi University of Finance and Economics and the Technology Research Institute of Shanghai AISHU Information Technology Corporation.

ABSTRACT With the continuous breakthroughs in artificial intelligence technology, it has become easier to extract general-purpose knowledge using machine learning, but it is a challenging task to extract and learn small samples of knowledge in medical expertise. On the one hand, it is difficult to represent medical expertise entities, and on the other hand, the training samples of such expertise are small, and deep learning methods often require a large number of samples to complete the learning task. To this end, we propose a graph network learning method for specialized vocabulary representation. Specifically, a contextual knowledge representation model based on graph meta-learning is proposed, which combines text, phrase, vocabulary, and other information to solve the problem of sparse data of medical electronic medical record entities that cannot be extracted and learned. In this method, a text-independent lexical representation learning method, a context-aware graph neural network, and a combined LSTM language model are used to model information from different perspectives as a way to learn semantic representations of professional discourse entities. The experimental results show that the accuracy of the method outperforms other similar methods and proves its effectiveness.

INDEX TERMS Graph network learning, vocabulary representation, knowledge representation, graph meta-learning.

I. INTRODUCTION

In the past two years, significant progress has been made in the field of graph neural network research. Several graph neural network aggregation models have emerged, each suitable for specific scenarios. In the text relationship extraction task, the combination of Graph and LSTM models, the combination of graph network and meta-learning, etc. have achieved good results [2], [3]. According to the research of related scholars, graph neural networks have three general frameworks: the first one, message passing neural network (MPNN) [4], unified graph neural network, and graph

convolutional network (GCN) method [5]; The second non-local neural network (NLNN) [6] combines various Attention mechanisms; the last graph network (GN) unifies the first two MPNN and NLNN, such as GGNN model, relation network (Relation Network [7], Interaction Networks, etc.

However, although these network models perform well in large-sample learning, they still perform poorly in small-sample professional vocabulary representation learning such as electronic medical record representation, Chinese bioinformatics relationship extraction, electronic health record information extraction, medical patient diagnosis classification, etc [8], [9], [10], [11]. The text corpus of medical professional vocabulary is generally small and beyond the

The associate editor coordinating the review of this manuscript and approving it for publication was R. K. Tripathy¹.

expertise of most people, and requires a certain level of medical professional background to understand. The current mainstream approach for extracting medical expertise and textual representation is to extract expertise directly from the text corpus through a data-driven approach. However, this extraction approach expresses a single relationship, which is not sufficient to accurately model the conceptual knowledge of expertise in the schema. In addition, the lexical knowledge of these specializations has a large number of polysemous words and is ambiguous. These ambiguities have a serious impact on the representation learning of the text task, causing semantic bias in the representation vector. From RNN to LSTM, it works well in processing contextual text data [12], but it is difficult to obtain the grammatical structure information of words. Some scholars have studied syntactic information of sentences, such as tree-like LSTM methods [13]. However, a large amount of data is required to obtain syntactic information. If the amount of data is small, it is difficult to obtain the dependency structure information in the vocabulary. Therefore, it is a challenging task to accurately model the dependency structure expressed by the contextual relations of specialized lexical knowledge.

The contributions of this paper are as follows.

- 1) To address the problem of medical professional vocabulary representation learning, we propose a network representation of graph meta-learning classification method. Specifically, we use graph networks to represent the feature information of medical professional vocabularies such as the semantic information of word context and grammatical structure of words and then combine them with meta-learning methods to solve the problem that medical professional samples are difficult to learn.
- 2) For the difficulty of extracting professional medical vocabulary, phrases, and text information, we propose a combination of a graphical neural network and LSTM language model to solve the problem of learning unstructured medical professional vocabulary with multiple meanings.

This study is divided into 5 parts, I. Outlines the research background and contributions of this paper, II. Introduces the work related to graph networks, meta-learning, LSTM fusion, and knowledge representation learning based on graph networks, III. The method section analyzes the methods proposed in this paper, IV. Does comparison experiments on public and real datasets, and does ablation experiments and case analysis, respectively. V. The research work of this paper is summarized.

II. RELATED WORK

A. GRAPH NEURAL NETWORKS AND KNOWLEDGE REPRESENTATION

Graph networks and graph embeddings are currently very mature in the fields of images [14]. Commonly used methods include DeepWalk [15], node2vec [16], SDNE [17] and

graph2vec [18]. These methods have achieved great success in the field of network representation learning. However, in the field of text, because the context relationship is asymmetric, if the vocabulary is polysemous, the learning effect will be negatively affected. In order to reduce these negative effects, more semantic information needs to be learned, so adopting a combined network model is a good choice. In general, GNNs are classified into structured scenarios, unstructured scenarios, and other three categories. The graph neural network is divided into five categories [26], [27], See table 1 below for details.

Previous work generally uses one-hot encoding to represent entity knowledge, but using graph neural networks to encode structural knowledge and attribute entity representation for fine-grained modeling of entities, vocabulary, etc., allows the introduction of multi-angle feature knowledge [19]. Knowledge representation based on graph networks can flexibly use local, spatial, and other structural information, and can effectively address the polysemy of words [20]. Therefore, using graph networks to implement knowledge representation has become a hot research topic in recent years such as Ebisu, Balazevic Zhang, and other related scholars used graphs to accomplish different tasks respectively [21], [22], [23].

B. META-KNOWLEDGE REPRESENTATION LEARNING OF SPECIALIZED VOCABULARY

Meta-learning is based on task learning. The target task is to learn the prior knowledge of sample data, so that the model has the ability to learn, and then can quickly learn new tasks of the same kind. The definition of a meta-learning task is shown below:

$$\mathcal{T} = \{L(x_1, a_1, \dots, x_H, a_H), q(x_1), q(x_{t+1}|x_t, a_t), H\}. \quad (1)$$

The equation (1) L is the loss function of the task, x_1 and x_H are the 1 and H observations, and a_1 and a_H are the 1 and H state output, $q(x_1)$ is the distribution of the initial observation of the first sample; $q(x_{t+1}|x_t)$ is the distribution of the next observation x_{t+1} , and a_t is the output value.

The complementarity of small samples and knowledge graphs has attracted extensive interest in recent years, and Yuan X and other related scholars have completed the missing facts of long-tail relationships by supporting instances of small samples in knowledge graphs [24]. Previous graph models encode small-sample relationships by aggregating information about each entity's neighbors but do not make fine-grained distinctions between neighbor relationships and entities. Inspired by humans and the organization of knowledge represented by hierarchical features, some scholars have explicitly unpacked the meta-learner as a meta-hierarchical graph with different blocks of expertise [25]. When encountering a new task, it constructs meta-knowledge paths by exploiting the most relevant blocks of knowledge or exploring new knowledge.

TABLE 1. Graph neural network classification.

Category	Type	Node Aitributes/Labels	Counterparts in Traditional Domains
Graph Neural Networks	Semi-supervised	Yes	Recursive Neural Networks
Graph Convolutional Networks	Semi-supervised	Yes	Convolutional Neural Networks
Graph Autoencoders	Unsupervised	Partial	Autoencoders/Variational Autoencoders
Graph Recurrent Neural Networks	Various	Partial	Recurrent Neural Networks
Graph Reinforcement Learning	Semi-supervised	Yes	Reinforcement Learning

C. INTEGRATION OF GRAPHS AND LSTM MODELS

A huge quantity of labeling is necessary for relation extraction using supervised learning, and labeling is likely to consume a significant amount of manpower and material expenditures. The graph network's development goal is to reduce expenses by dealing with the problem of a small number of sample annotations or even no annotations. In recent years, using a minimal number of samples to train a graph neural network has become a research hotspot [28]. The goal of few-shot relation extraction is to train a model Graph neural network model using a short quantity of labeled data and then use an LSTM model to learn the sample's context [29]. This paper explains how to accurately classify the category of very few-shot using graph neural network and LSTM model approaches.

There are two recursive propagation units in the Graph-LSTM model: GraphLSTM, a liner LSTM, and a Graph Model. For each connection in the phrase, liner LSTM generates a vector equation(2).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2)$$

The LSTM model is a time-recurrent neural network designed to solve the long-term dependence of RNNs. Compared with RNN, there is one more value representing cell memory. There are three forgetting gates, input gates, and output gates. The forget gate can be decided to discard and retain that information. The input gate can update the unit state. The output gate first passes the previous hidden state and the current input to the Sigmoid function; then the newly obtained unit state is given to the Tanh function; then the Tanh output and The Sigmoid output are multiplied to determine the information that the hidden state should carry, and finally, the hidden state is output as the current unit. The model of LSTM is illustrated below equation (3).

$$\begin{aligned}
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\
 \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \\
 C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \\
 o_t &= \sigma(W_o [h_{t-1}, x_t] + b_o) \\
 h_t &= o_t * \tanh(C_t)
 \end{aligned} \quad (3)$$

The model of Graph LSTM is as follows equation (4).

$$i_t = \sigma \left(W_i x_t + \sum_{j \in P(t)} U_i^{m(t,j)} h_j + b_i \right)$$

$$\begin{aligned}
 o_t &= \sigma \left(W_o x_t + \sum_{j \in P(t)} U_o^{m(t,j)} h_j + b_o \right) \\
 \tilde{c}_t &= \tanh \left(W_c x_t + \sum_{j \in P(t)} U_c^{m(t,j)} h_j + b_c \right) \\
 f_{ij} &= \sigma \left(W_f x_t + U_f^{m(t,j)} h_j + b_f \right) \\
 c_t &= i_t \odot \tilde{c}_t + \sum_{j \in P(t)} f_{ij} \odot c_j \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned} \quad (4)$$

The difference between equation (4) and (3) is that Graph LSTM is flexible and versatile. It can extract the LSTM relationship of the entity binary relationship and another LSTM entity identification while seamlessly combining in together. In addition, Graph LSTM obtains the dependency vector of the sentence and multiplies it by the output state point of the hidden layer unit. The length must be equivalent to the output vector h_t . This problem can be solved by a tensor product. The algorithm is illustrated below equation(5).

$$\begin{aligned}
 i_t &= \sigma \left(W_i x_t + \sum_{j \in P(t)} U_i \times_T (h_j \otimes e_j) + b_i \right) \\
 f_{ij} &= \sigma \left(W_f x_t + U_f \times_T (h_j \otimes e_j) + b_f \right) \\
 o_t &= \sigma \left(W_o x_t + \sum_{j \in P(t)} U_o \times_T (h_j \otimes e_j) + b_o \right) \\
 \tilde{c}_t &= \tanh \left(W_c x_t + \sum_{j \in P(t)} U_c \times_T (h_j \otimes e_j) + b_c \right) \\
 c_t &= i_t \odot \tilde{c}_t + \sum_{j \in P(t)} f_{ij} \odot c_j \\
 h_t &= o_t \odot \tanh(c_t)
 \end{aligned} \quad (5)$$

where, e_j is the vector of dependency relation, and the final result is the matrix form of the tensor product of L dimension vector and D dimension vector, see as follow:

$$T \times_T A = \sum_d (T_{::d} \cdot A_{:,d}) \quad (6)$$

On the other hand, in the research of related scholars in the context of specialized vocabulary, the model-based method in 2001 proved to be useful for meta-learning. A neural network using memory is used for meta-learning, where updating the weights can adjust the output. It is also useful to embed the

training set of the matching network in the method of meta-learning metrics. When gradient optimization is not possible with small samples, the meta-learner can be represented by LSTM. The LSTM model uses the parameter update method of gradient descent to learn the small-sample training set. From the equations of (3) and (4), Graph Model can also be used in few-shot learning. This paper is based on this idea and proposes Graph Model and few-shot meta-learning. And verifies the scenario of dealing with the contextual semantic relation understanding of specialized vocabulary. Meta-learning pseudo-code see algorithm 1.

Algorithm 1 Meta-Learner Training

Require: Meta training set $\mathcal{D}_{\text{meta-train}}$, Learner M with parameter θ , Meta-learner R with parameter θ

Ensure:

- 1: Random initial θ_0
 - 2: **for** $d = 1, n$ **do**
 - 3: $\mathcal{D}_{\text{meta-train}}$ Random training set in meta training set $\mathcal{D}_{\text{train}}$ and Test set $\mathcal{D}_{\text{test}}$
 - 4: $\theta_0 \leftarrow c_0$
 - 5: **for** $t = 1, T$ **do**
 - 6: $\mathcal{D}_{\text{meta-train}}$ Random training set in meta training set $\mathbf{X}_t, \mathbf{Y}_t$
 $\mathcal{L}_t \leftarrow \mathcal{L}(M(\mathbf{X}_t; \theta_{t-1}), \mathbf{Y}_t)$
 - 7: $c_t \leftarrow R((\nabla_{\theta_{t-1}} \mathcal{L}_t, \mathcal{L}_t); \Theta_{d-1})$
 $\theta_t \leftarrow c_t$
 - 8: **end for**
 $\mathbf{X}, \mathbf{Y} \leftarrow D_{\text{test}}$
 - 9: $\mathcal{L}_{\text{test}} \leftarrow \mathcal{L}(M(\mathbf{X}; \theta_T), \mathbf{Y})$
 Update Θ_d using $\nabla_{\Theta_{d-1}} \mathcal{L}_{\text{test}}$
 - 10: **end for**
-

In the natural language processing task, in the simple binary relationship extraction, the main method is the closest dependency path between entities. In multivariate relationship extraction, n-ary relationships can be divided into n-binary relationships, but the inter-sentence dependency relationship cannot be realized, and the Graph and LSTM model can solve this problem well. Graph and LSTM is a kind of multi-task learning based on a sub-relationship. In each sub-task learning, a meta-learner is utilized to accumulate knowledge about the similarities and differences between tasks.

When dealing with the relationship between document sentences, you can use the Graph and LSTM skeleton to process the relationship between sentences through graphs. The nodes of the graph represent the vocabulary, and the edges of the graph represent the relationship between different vocabulary. Are these relations related to whether the utterance is related and whether it exists Syntactic dependence, whether there is a contextual correlation, etc?

D. INTEGRATION OF GRAPHS AND META-LEARNING

Combining Meta-GNN learning and text application, relevant scholars have already conducted research on

primitives [30]. The literature proposes a learning framework called Meta-GNN to solve the problem of node classification in the primitive learning environment. Through learning and training multiple similar few-shot learning tasks, according to the a priori knowledge of the classifier, the new class with few labeled samples is used to classify the node. The cross-entropy loss function for Meta-GNN training is the equation (7).

$$\mathcal{L}_{T_i}(f_\theta) = - \sum_{\mathbf{x}_{ix}, \mathbf{y}_{is}} y_{ix} \log f_\theta(\mathbf{x}_{is}) - (1 - y_{is}) \log (1 - f_\theta(\mathbf{x}_{is})) \quad (7)$$

When updating the parameters, each step or several steps in the task T_i uses a gradient descent algorithm, and the parameters of gradient descent are updated to the equation (8).

$$\theta'_i = \theta - \alpha_1 \frac{\partial \mathcal{L}_T(f_\theta)}{\partial \theta} \quad (8)$$

The α_1 is the learning rate of the learning task and the parameters of the model, and the optimized performance after meta-training $f_{\theta'}$, see equation (9).

$$\theta = \arg \min_{\theta} \sum_{T_i \sim p(T)} \mathcal{L}_{T_i}(f_{\theta'_i}) \quad (9)$$

In essence, the goal of Meta-GNN is to optimize model parameters and achieve the best performance of node classification on new tasks. The new task is updated with a small number of gradient descent algorithms. Perform stochastic gradient descent (SGD) to perform cross-task. Meta-optimization, and the model parameters θ are updated accordingly, see equation (10).

$$\theta \leftarrow \theta - \alpha_2 \frac{\partial \sum_{T_i \sim p(T)} \mathcal{L}_{T_i}(f_{\theta'_i})}{\partial \theta} \quad (10)$$

Summary: in recent years, more significant progress has been made in graph-based representation learning. However, these methods are difficult to apply effectively to representation learning in specialist domains for two reasons: firstly, specialist domain knowledge bases are often sparse and difficult to model accurately; secondly, some particular specialist knowledge bases have a large number of polysemous words, and these polysemous words can have a serious negative impact on the representation learning of their neighbors, causing a semantic shift in the representation vectors of their neighbors.

Different from data-driven acquisition methods, meta-knowledge-based representation learning can collect a large amount of tacit knowledge. In addition, researchers can “query” a large amount of tacit knowledge from it without designing pattern engineering. However, efficient “query” methods are often difficult to design. Secondly, although a large amount of tacit common sense can be obtained based on the acquisition method of meta-knowledge, it is difficult to acquire specialized knowledge other than meta-knowledge due to the reasoning process of learning through learning.

In the current related works, the acquisition method based on text mining has higher acquisition efficiency, but the knowledge acquired by this single method is less accurate. In addition, although this method is a widely used knowledge acquisition method, it is often only able to extract the professional knowledge explicitly mentioned in the text corpus, and it is difficult to obtain the implicit professional knowledge.

III. METHODOLOGY

In this study, we propose a deep representation learning model for primitive learning based on contextual semantic information. The model fuses information about different sources of specialized lexical triples, including vocabulary Phrases, text, and graph information, to solve the problem of low data volume. In this method, a word representation learning method and a graph GNN network are first used to model the information of the sentence, so as to learn the semantic representation of the subject and object. In particular, in the process of textual sentence modeling, a long and short-term memory network is introduced to learn the contextual information of different sentences. The overall architecture description is shown in Figure 1.

A. LEARNING THE REPRESENTATION OF GRAPHS

Given a triple (s, p, o), graph information refers to subgraphs g_1 and g_2 centered on subject s and object o, respectively, both of which are part of a specialized knowledge base G. g_1 and g_2 provide rich structured information about the subject and object, which is useful for determining the relationship between them.

To obtain the representation of subject s and object o from the graph structure information, GNN is used to model the graph information. GNN is a multilayer neural network that operates directly on the graph and learns the semantic representation of nodes through the information transfer from neighbors. It is important to note that the representation learning of relations is not considered in graph information because of: first, the sparsity of the specialized knowledge base; and second, it can improve the computational efficiency. Formally, the expertise base is defined as a graph $G = (V, E)$, where V and E denote the nodes and edges in the graph, respectively. In the modeling process of GNN, for node $v \in V$, there exists an edge association of node v with itself, i.e., $(v, v) \in E$. GNN introduces the adjacency matrix A and degree matrix D of graph G, where $D_{ii} = \sum_j A_{ij}$.

B. META-GNN-BASED CONTEXT CLASSIFICATION

Traditional CNN is difficult to deal with intensive classification tasks that assign unique categories to individual nodes in the graph structure, or processing text sentence classification tasks, or processing word-level sequence labeling. To solve these problems, some scholars have proposed a deep learning model based on graph CNN [31], which first converts text into word graphs, and then uses graph convolution operations to calculate word graphs. The method is to convert documents or Sentences that are regarded as graphs of word nodes,

and a graph is constructed by relying on document reference relations.

The combination of GNN and meta-learning [32] solves the problem of node classification in the graph element learning environment. The method is to train multiple similar few-shot learning tasks through learning and training. According to the prior knowledge of the classifier, use the mark the new class with few samples classifies the nodes. The primitive learner can be divided into a multi-task meta-learner and a meta-tester, where the meta-tester continuously updates the learning parameters of each task through testing. Specifically, the algorithm is shown in algorithm 2.

The pseudo-code of the algorithm for primitive learning is: The Meta-learning loss function is the cross-entropy loss as

Algorithm 2 Graph Meta-Learning Algorithm

Require: distributed meta-learning task $p(\mathcal{T})$, meta-learning task T_{mt} , task learning rate α_1 , meta-learning rate α_2

Ensure: The label of the node in the query set T_{mt}

- 1: random initialization θ
 - 2: **while** *notconverged* **do**
 - 3: batchofmeta – learningtask $\mathcal{T}_i \sim p(\mathcal{T})$
 - 4: **for** Task in \mathcal{T}_i **do**
 - 5: Use \mathcal{S}_i to evaluate $\mathcal{L}_\tau(f_\theta)$ through the equation (11)
 - 6: calculate and update the parameter θ through the equation (9)
 - 7: uses \mathcal{Q}_i to evaluate $\mathcal{L}_{\tau_i}(f_{\theta'_i})$
 - 8: **end for**
 - 9: Update θ by equation (10)
 - 10: **end while**
 - 11: Use the support set T_{mt} to calculate the updated parameters through the equation (8) θ'_{mt}
 - 12: Use model $f_{\theta'_{mt}}$ in query set T_{mt}
-

follows:

$$\mathcal{L}_{T_i}(f_\theta) = - \left(\sum_{\mathbf{x}_{ix}, y_{is}} y_{ix} \log f_\theta(\mathbf{x}_{is}) + (1 - y_{is}) \log(1 - f_\theta(\mathbf{x}_{is})) \right) \quad (11)$$

Deep learning is a mapping from a certain data distribution X to another distribution Y. Meta-learning is a mapping from a set of tasks D to the optimal function $f(x)$ corresponding to the task. Primitive optimization learning is the mapping of each task from Euclidean space to non-Euclidean space, and fast learning on new tasks is achieved by training model parameters, as long as a small number of gradient updates are required. θ is continuously optimized according to the loss gradient of the new task, and finally, the optimal parameters are obtained. The specific definition and algorithm of meta-learning will not be expanded in this section.

Meta-learning uses prior information to find new categories that don't exist. The meta-learning method involves building a brand-new network model based on empirical knowledge, letting the network learn the empirical knowledge, and then completing the learning using the empirical knowledge learned by the network model, or building

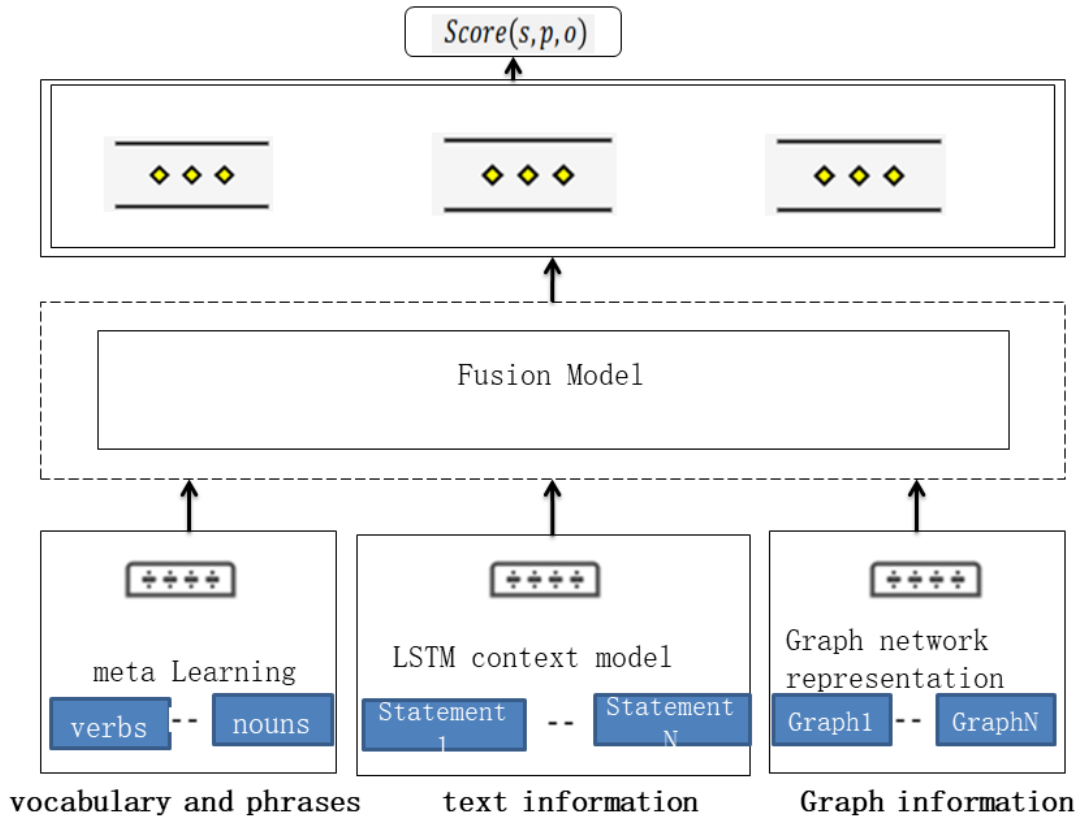


FIGURE 1. General architecture for few-shot representation learning of unstructured professional vocabulary of medical records.

an external memory network to introduce prior knowledge, which can be LSTM, GRU, GNN, and other models. Twin networks, relational networks, and other metric computation methods are used to split meta-learning. In the context of specialized vocabulary, with the prototype network as the benchmark, the data enhancement method process is used, and the cross-domain metric is calculated as in equation (12).

$$\begin{aligned}
 p_c &= \frac{1}{|S_c|} \sum_{(x_i, y_i) \in S_c} \phi(x_i) \\
 \rho(y_i = c | x_i) &= \frac{\exp(-d(\phi(x_i), p_c))}{\sum_{c'} \exp(-d(\phi(x_i), p_{c'}))} \\
 L_p &= - \sum_{(x_i, y_i) \in Q} \log p(y_i = c | x_i). \quad (12)
 \end{aligned}$$

C. EXTRACTION OF SPECIALIZED VOCABULARY

Pixel learning is processed on few-shot learning such as text processing. Unlike Graph-LSTM, the former focuses on text node classification, while the latter focuses on text relationship extraction. The same is that both are currently based on the GNN model. The theoretical basis of GNN is Banach’s Fixed Point Theorem (Banach’s Fixed Point Theorem), see the literature [35]. For a graph, all nodes on the graph finally converge to a certain fixed point after constant

iteration, which is called a fixed point. The points on the graph are mapped by a function, and the two points (x, y) on the plane are compressed into another space to become points $f(x), f(y)$.

In this study, the graph long and short-term memory network meta-learning is applied to the real hospital electronic medical record data for few-shot application detection. The preprocessing of few-shot electronic medical record data extraction in hospital departments is shown in Figure 2. The input is the first layer, followed by a context vector acquired by LSTM for each word and fine-tuned using a meta-learning technique for the context of specialized vocabulary, in which the context vectors of numerous entities are coupled as an input to a relational graph meta-classifier. The average of each word is taken, together with the meta-knowledge of the contextual phrases, and fed into the graph network structure for multi-word entities. The relational classifier and the graph part of the model are unconnected.

D. GraphLSTM-BASED KNOWLEDGE REPRESENTATION

This section first builds a graph network vocabulary and phrase representation learning model to learn contextual semantic representations such as subjects and objects from the perspective of graphs.

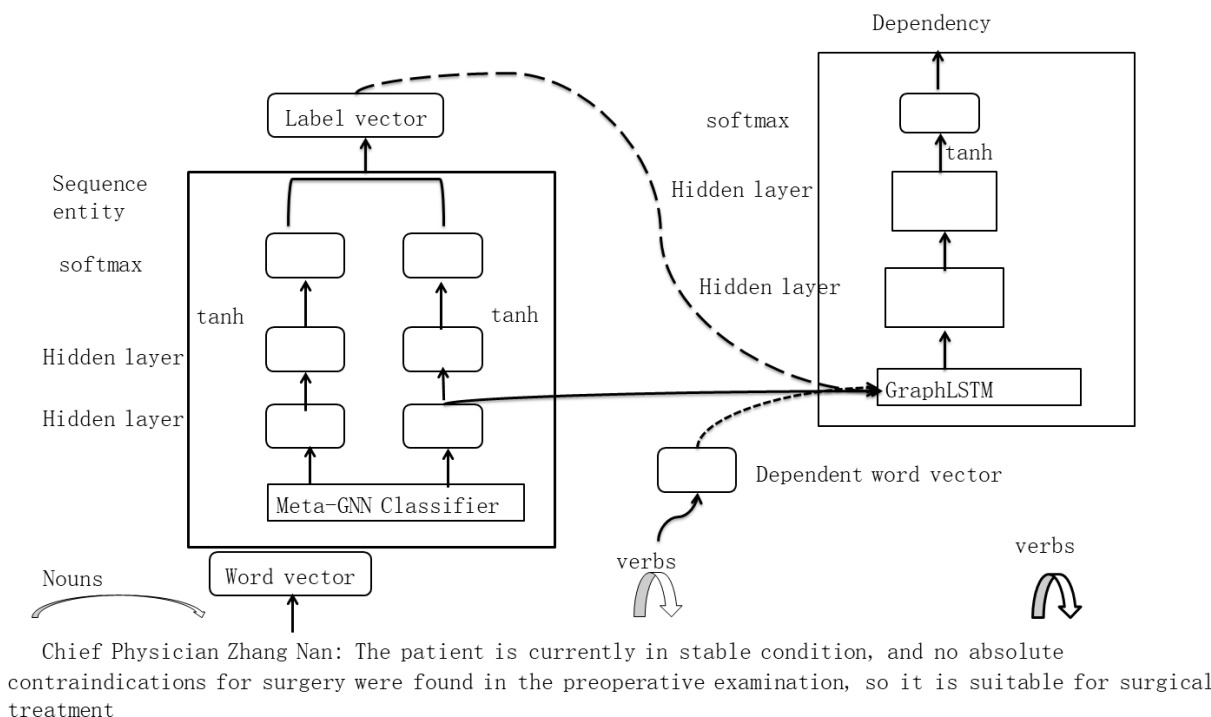


FIGURE 2. Application architecture for text vocabulary extraction.

The relationship between sentences is processed through the graph. The nodes of the graph represent the vocabulary, and the edges of the graph represent the relationship between different vocabularies, such as whether the language is related, whether the syntactic dependency exists, whether there is a contextual connection, etc. The Figure 3 application architecture is also suitable for multi-task learning. For each sub-task, an LSTM can be used as a classifier to learn simple binary relations. Under the framework of meta-learning application, an end-to-end LSTM model of a small number of sample context learning classifiers is constructed. The LSTM model maps the data in each category to space and extracts the mean value of these data features as the prototype. The specific algorithm is in the next Section details.

IV. EVALUATION AND EXPERIMENT

In order to evaluate the effectiveness of the method, this section will conduct a large number of experiments on the part-of-speech and word segmentation knowledge representation tasks of text sentences, and compare the accuracy of the model with the corresponding baseline models on public datasets. In addition, this section will conduct case analysis experiments of the model to further verify the effectiveness of each module method in this study.

A. DATASETS

This section first chooses to validate the Meta-GNN method using the Reddit dataset, which consists of posts from the Reddit forum. If two posts are commented by the same

TABLE 2. Accuracy of “OURS” fusion model public data set.

Method	Accuracy	Recall	F1 Value
Co-CNN [37]	95.23	74.42	76.95
LG-LSTM [38]	96.18	79.43	80.97
OURS	97.60	83.28	83.76

person, then they are considered relevant and the tags are corresponding to each post. Also, the public dataset GDKD is used to validate and evaluate the effectiveness of our method. To validate the effectiveness of our practical application, we used 110 actual hospital electronic medical records to complete the effect validation. In addition, we compare different knowledge representation models on the accuracy of entity extraction from hospital electronic medical records in expertise experiments, and the experimental results show that our method works better.

B. EXPERIMENTAL RESULTS

The results of Accuracy, Recall, and F1 Value of the graph network meta-learner LSTM model on the public data set are as follows table 2.

Compared with deep learning CNN in the public data set ATR, the Graph-LSTM model has an accuracy rate of 97.60%, a recall rate of 83.28%, and an F1 score of 83.76%. The accuracy of the three indicators of CNN, the recall rate, and the F1 score are respectively. Only 95.23%, 74.42%,

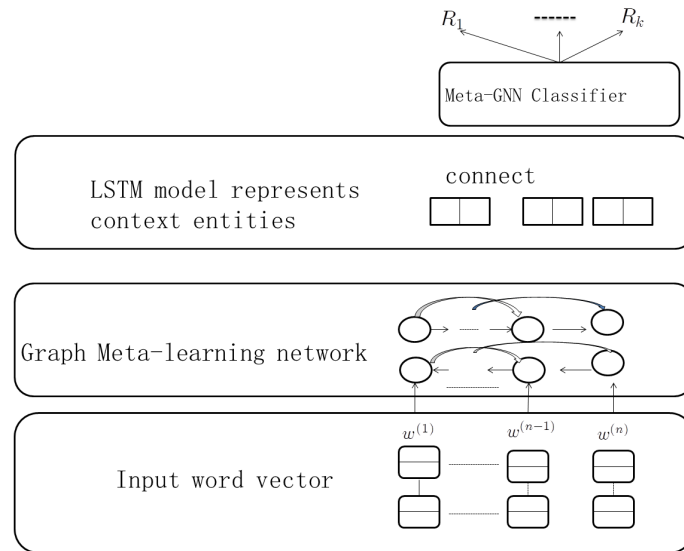


FIGURE 3. The processing architecture of the vocabulary graph.

TABLE 3. Accuracy rate of “OURS” fusion model public data set GDKD.

model	single sentence	cross sentence
CNN [44]	73.0%	74.9%
BiLSTM [44]	73.9%	76.0%
Tree LSTM [44]	74.6%	75.9%
OURS	75.9%	76.7%

76.95%. It can be seen that the combination of the image and the meta-learner LSTM can achieve the SOAT(State-Of-The-Art) effect.

The “OURS” model uses a 3×3 convolution kernel, a meta-GNN classifier, and the Softmax activation function. Aiming at the extraction of sentence relations, the experiment chose the public data set GDKD, which is an open-source drug-gene data set, which is mainly to explore whether there is an association relationship between the data and then perform the treatment. This experiment is not to mine genetic relationships but to verify the effectiveness and accuracy of the model’s application in sentence relationship extraction. Specifically as shown in table 3. Comparing the deep learning CNN model, bidirectional LSTM model, and Tree LSTM model, The OURS model single sentence and cross sentence relationship extraction can reach 75.9% and 76.7%, which is higher than the best Tree LSTM. It can be observed that the “OURS” method has more advantages in the few-shot field. For practical applications, being able to extract more spatial information from the sample set is quite beneficial.

We selected 110 medical record data samples to mark the effect as shown in Figure 4. From Figure 4, we can see that the average accuracy of lexical annotation is 82.35%, which can basically meet the practical application requirements.

TABLE 4. Comparison of model accuracy of different methods for entity extraction.

Models	Accuracy
BERT	84.7%
ALBERT	82.3%
RoBERTa	85.7%
OURS	86.5%

The effect of the entity extraction model for different knowledge representation methods is shown in table 4. Through table 4, it can be seen that our fusion model can reach 85.7% in expertise entity extraction accuracy, which is somewhat higher than all other methods.

C. ABLATION ANALYSIS

To evaluate the effectiveness of each feature and method proposed in this study, we next conduct a series of ablation experiments using 5-way 1-shot and 3-shot on the public dataset Reddit. Comparison with current state-of-the-art related working model results in dataset Reddit. These related works include 2019-2022 related graph representation learning models such as Meta SGC [40], GCN [41], SemiGNN [42], CGNP [43], etc. Based on these models, we have modified the dataset to meet the knowledge representation of meta-learning. The experimental results are shown in table 5, in which our method based on Meta-GNN fusion achieves better 3-shot learning accuracy on small sample datasets than all other models, but there is no CGNP learning for one sample. The effect is good, and the effect will be better as the number of samples increases, indicating that the method based on Meta-GNN fusion is better than other methods. It can be seen from table 5 that the meta-learning classifier based on graph knowledge

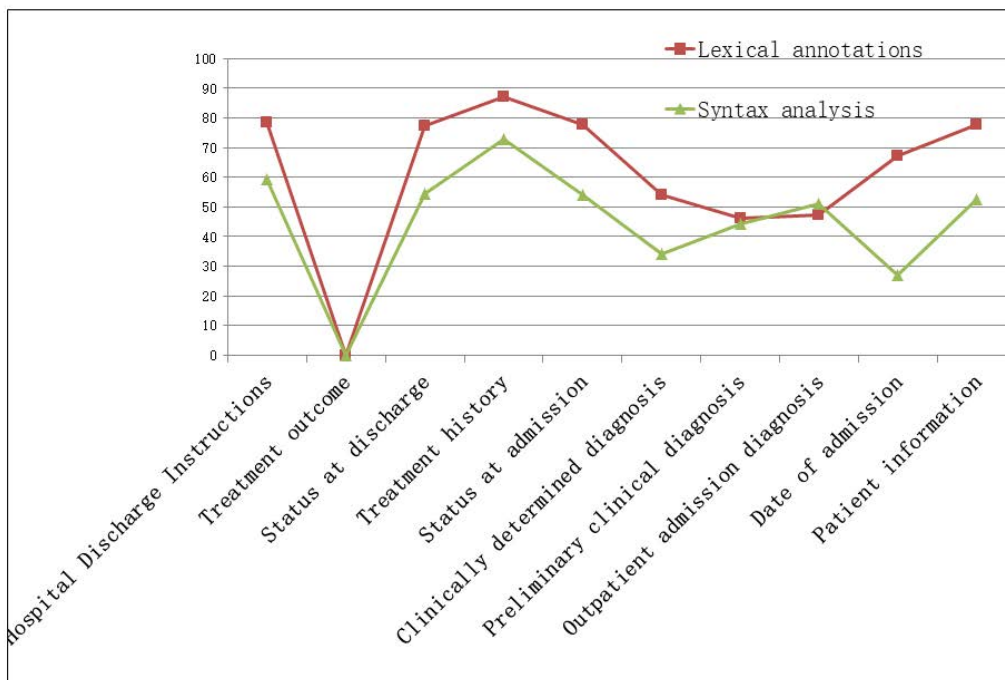


FIGURE 4. Lexical annotation model annotation accuracy results.

```

2018-9-28 17:39 rescue record
Li XX Ling, gender: female, age: 57 years old.
Rescue time: 2018-9-28 16:40
Physicians and nursing staff involved in the rescue at the scene: Deputy Chief Physician Yang Fei, Physician Zhang Fangfang,
Physician Kong Shanshan, Physician Zhang Zhuo, Physician Yu Hailiang, Supervisor Liu Jinghuan, Nurse Liu Nannan, Nurse Liu Yu.
The rescue process is as follows: the patient suddenly experienced loss of consciousness, superficial breathing, wet and cold
around the body, cyanosis of the nail bed, bilateral pupils, etc. at 16:40, a large isometric circle, a diameter of 2mm, the
light reflection disappeared, and ECG monitoring showed a heart rate of 99 beats / min , Finger pulse oxygen 69 %, blood
pressure 136 / 90mmHg, emergency blood gas analysis: PH6.96, oxygen partial pressure 100mmHg, carbon dioxide partial pressure,
remaining alkali, actual bicarbonate can not be determined, Deputy Chief Physician Yang Fei was present to guide the rescue ,
Immediately give a simple breathing apparatus to assist breathing, oxygen flow 1L / min, establish a venous access, emergency
call anesthesiologist, 16:55 Anesthesiologist Huang Lianjun attending the site to perform tracheal intubation for the patient,
the intubation process is smooth, ECG Monitoring showed that the patient's heart rate was 114 beats / min, the pulse oxygen
saturation was 96 %, the blood pressure was 110/70 mmHg, the ventilator assisted breathing, the mode was SIMV, the oxygen
concentration was 100 %, the tidal volume was 460 ml, PEEP4cmH2O, and the respiratory rate was 16 beats / min. The patient's
consciousness gradually recovered at 17:00. The pupils of both sides were of equal circle, 3mm in diameter, and were sensitive
    
```

FIGURE 5. The original text of the rescue record.

```

2018-9-28 17:39 rescue record
Li XX Ling, gender: female, age: 57 years old.
Rescue time: 2018-9-28 16:40
Physicians and nursing staff involved in the rescue at the scene: Deputy Chief Physician Yang Fei, Physician Zhang Fangfang, Physician Kong Shanshan, Physician Zhang Zhuo,
Physician Yu Hailiang, Supervisor Liu Jinghuan, Nurse Liu Nannan, Nurse Liu Yu.
The rescue process is as follows: the patient suddenly experienced loss of consciousness, superficial breathing, wet and cold around the body, cyanosis of the nail bed, bilateral pupils, etc.
at 16:40, a large isometric circle, a diameter of 2mm, the light reflection disappeared, and ECG monitoring showed a heart rate of 99 beats / min , Finger pulse oxygen 69 %, blood
pressure 136 / 90mmHg, emergency blood gas analysis: PH6.96, oxygen partial pressure 100mmHg, carbon dioxide partial pressure, remaining alkali, actual bicarbonate can not be
determined, Deputy Chief Physician Yang Fei was present to guide the rescue , Immediately give a simple breathing apparatus to assist breathing, oxygen flow 1L / min, establish a
venous access, emergency call anesthesiologist, 16:55 Anesthesiologist Huang Lianjun attending the site to perform tracheal intubation for the patient, the intubation process is smooth,
ECG Monitoring showed that the patient's heart rate was 114 beats / min, the pulse oxygen saturation was 96 %, the blood pressure was 110/70 mmHg, the ventilator assisted breathing,
the mode was SIMV, the oxygen concentration was 100 %, the tidal volume was 460 ml, PEEP4cmH2O, and the respiratory rate was 16 beats / min. The patient's consciousness
gradually recovered at 17:00. The pupils of both sides were of equal circle, 3mm in diameter, and were sensitive to light reflection.
Relatives and wishes of the patients present: Explain the condition to the patients' families, and the families express their understanding and have no objection to the rescue process.
    
```

FIGURE 6. Data annotation of rescue records.

representation can achieve 20.71% accuracy on a 5-way 3-shot compared with other classifiers.

Analyzing from the comparative results, graph representation learning models such as GCN, SGC, and CGNP are

TABLE 5. Learning results of fusion meta-learning based on graph knowledge representation in reddit dataset.

Datasets	Reddit	
	1-shot	3-shot
ModelK-shot		
GraphSAGE-Mean [39]	9.47%	15.89%
GraphSAGE-Pool [39]	9.31%	15.36%
SGC [40]	9.80%	16.98%
GCN [41]	9.87%	17.17%
SemiGNN [42]	13.04%	15.07%
CGNP [43]	15.38%	16.04%
OURS	14.22%	20.71%

not very effective with few samples, but when combined with meta-learning, more improvement in accuracy can be obtained, so this fused approach is better adapted to new tasks and more node representations can be learned through meta-learning methods

D. CASE ANALYSIS

To execute the relationship extraction experiment, use the real data set “Hospital Electronic Medical Record”. Patient rescue records, daily medical records, admission records, surgical records, preoperative discussions, preoperative summaries, and other few-shot data sets are all included in the data collection. Figure 5 illustrates the specific data format.

The patient’s rescue information, such as name, gender, age, rescue time, doctors and nurses who participated in the rescue, and the rescue process information, such as state of consciousness, respiratory status, blood pressure, and so on, are used in the experiment. Extract crucial information from this data and categorize it based on the contextual relationship and content. The findings of the text relation extraction are depicted in Figure 6.

V. CONCLUSION

In the extraction task of small-sample text relations, we propose a MetaGNN medical professional vocabulary feature representation learning method based on MetaGNN for the problem that it is difficult to obtain an accurate representation of medical professional vocabulary. Specifically, firstly, text information, sentence label information, and syntactic structure information of the sentences are extracted from the medical record text. Second, the contextual space information of the sentences is obtained. In addition, we propose a joint model based on LSTM and meta-learning for professional medical vocabulary and phrase extraction. The experimental results demonstrate the effectiveness of our method, which outperforms other text information extraction task methods, such as TreeLSTM, in terms of accuracy and precision.

ACKNOWLEDGMENT

The authors thank for the support and assistance from the Postdoctoral Station of Management Science and Engineering, Jiangxi University of Finance and Economics, China.

REFERENCES

[1] G. O. Young, “Synthetic structure of industrial plastics,” in *Plastics*, vol. 3, 2nd ed., J. Peters, Ed. New York, NY, USA: McGraw-Hill, 1964, pp. 15–64.

[2] D. Tang, B. Qin, X. Feng, and T. Liu, “Effective LSTMs for target-dependent sentiment classification,” in *Proc. COLING*, 2016, 3298–3307.

[3] X. Liang, X. Shen, J. Feng, L. Lin, and S. Yan, “Semantic object parsing with graph LSTM,” in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 125–143.

[4] K. Rusek and P. Cholda, “Message-passing neural networks learn little’s law,” *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 274–277, Feb. 2019.

[5] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proc. ICLR*, 2017, pp. 1–14.

[6] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, “Graph neural networks: A review of methods and applications,” *AI Open*, vol. 1, pp. 57–81, Jan. 2020.

[7] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, “Learning to compare: Relation network for few-shot learning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.

[8] B. Wang, Y. Sun, Y. Chu, D. Zhao, Z. Yang, and J. Wang, “Refining electronic medical records representation in manifold subspace,” *BMC Bioinf.*, vol. 23, no. 1, pp. 1–17, Dec. 2022.

[9] Q. Zhao, D. Xu, J. Li, L. Zhao, and F. A. Rajput, “Knowledge guided distance supervision for biomedical relation extraction in Chinese electronic medical records,” *Expert Syst. Appl.*, vol. 204, Oct. 2022, Art. no. 117606.

[10] J. Sanders, M. Gudala, K. Hamilton, N. Prasad, J. Stovall, E. Blanco, J. E. Hamilton, and K. Roberts, “Extracting adherence information from electronic health records,” in *Proc. 28th Int. Conf. Comput. Linguistics*, 2020, pp. 680–695.

[11] L. V. Lita, S. Yu, S. Niculescu, and J. Bi, “Large scale diagnostic code classification for medical patient records,” in *Proc. 3rd Int. Joint Conf. Natural Lang. Process.*, vol. 2, 2008, pp. 1–6.

[12] C. Wang and M. Niepert, “State-regularized recurrent neural networks,” in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6596–6606.

[13] Z. Huang, W. Xu, and K. Yu, “Bidirectional LSTM-CRF models for sequence tagging,” 2015, *arXiv:1508.01991*.

[14] E. Grefenstette and P. Blunsom, “A convolutional neural network for modelling sentences,” in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, Baltimore, MD, USA, 2014, pp. 1–16.

[15] B. Perozzi, R. Al-Rfou, and S. Skiena, “DeepWalk: Online learning of social representations,” in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 701–710.

[16] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 855–864.

[17] D. Wang, P. Cui, and W. Zhu, “Structural deep network embedding,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1225–1234.

[18] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, and S. Jaiswal, “Graph2vec: Learning distributed representations of graphs,” 2017, *arXiv:1707.05005*.

[19] B. Hu, Z. Hu, Z. Zhang, J. Zhou, and C. Shi, “KGNN: Distributed framework for graph neural knowledge representation,” 2022, *arXiv:2205.08285*.

[20] Y. Zhao, H. Zhou, R. Xie, F. Zhuang, Q. Li, and J. Liu, “Incorporating global information in local attention for knowledge representation learning,” in *Proc. Findings Assoc. Comput. Linguistics (ACL-IJCNLP)*, 2021, pp. 1341–1351.

[21] T. Eblisu and R. Ichise, “Generalized translation-based embedding of knowledge graph,” *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 5, pp. 941–951, May 2020.

[22] I. Balazevic, C. Allen, and T. Hospedales, “Multi-relational Poincaré graph embeddings,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[23] Y. Hou, J. Zhang, J. Cheng, K. Ma, R. T. B. Ma, H. Chen, and M.-C. Yang, “Measuring and improving the use of graph information in graph neural networks,” 2022, *arXiv:2206.13170*.

[24] H. Yao, Y. Zhou, M. Mahdavi, Z. J. Li, R. Socher, and C. Xiong, “Online structured meta-learning,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6779–6790.

[25] X. Yuan, C. Xu, P. Li, and Z. Chen, “Relational learning with hierarchical attention encoder and recoding validator for few-shot knowledge graph completion,” in *Proc. 37th ACM/SIGAPP Symp. Appl. Comput.*, Apr. 2022, pp. 786–794.

[26] Y. Ding, L.-P. Tian, X. Lei, B. Liao, and F.-X. Wu, “Variational graph auto-encoders for miRNA-disease association prediction,” *Methods*, vol. 192, pp. 25–34, Aug. 2021.

- [27] Y. Wada, N. Pecharanin, A. Taguchi, N. Iijima, Y. Akima, and M. Sone, "Application of recurrent neural network for active filter," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 1, 1995, pp. 488–491.
- [28] E. Triantafillou, R. Zemel, and R. Urtasun, "Few-shot learning through an information retrieval lens," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [29] L. Yan, Y. Zheng, and J. Cao, "Few-shot learning for short text classification," *Multimedia Tools Appl.*, vol. 77, no. 22, pp. 29799–29810, 2018.
- [30] A. Sankar, X. Zhang, and K. C.-C. Chang, "Meta-GNN: Metagraph neural network for semi-supervised learning in attributed heterogeneous information networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2019, pp. 137–144.
- [31] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Conf. Workshop Neural Inf. Process. Syst.*, 2016, pp. 1–9.
- [32] F. Zhou, C. Cao, K. Zhang, G. Trajcevski, T. Zhong, and J. Geng, "Meta-GNN: On few-shot node classification in graph meta-learning," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manag.*, Nov. 2019, pp. 2357–2360.
- [33] S. Flennerhag, A. A. Rusu, R. Pascanu, F. Visin, H. Yin, and R. Hadsell, "Meta-learning with warped gradient descent," in *Proc. 8th Int. Conf. Learn. Represent.*, Apr. 2020, pp. 1–28.
- [34] T. Lin, S. U. Stich, K. K. Patel, and M. Jaggi, "Don't use large mini-batches, use local SGD," in *Proc. 8th Int. Conf. Learn. Represent.*, 2018, pp. 1–40.
- [35] M. N. Wang, "Application of Banach's fixed point theorem," *J. Jilin Province Educ. College*, no. 1, pp. 153–154, 2010.
- [36] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 1025–1035.
- [37] X. Liang, C. Xu, X. Shen, J. Yang, S. Liu, J. Tang, L. Lin, and S. Yan, "Human parsing with contextualized convolutional neural network," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1386–1394.
- [38] X. Liang, X. Shen, D. Xiang, J. Feng, L. Lin, and S. Yan, "Semantic object parsing with local-global long short-term memory," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3185–3193.
- [39] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. NeurIPS*, 2017, pp. 1024–1034.
- [40] F. Wu, A. Souza, T. Zhang, C. Fifty, T. Yu, and K. Weinberger, "Simplifying graph convolutional networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6861–6871.
- [41] H. Cheng, J. T. Zhou, W. P. Tay, and B. Wen, "Attentive graph neural networks for few-shot learning," 2020, *arXiv:2007.06878*.
- [42] K. Ding, Q. Zhou, H. Tong, and H. Liu, "Few-shot network anomaly detection via cross-network meta-learning," in *Proc. Web Conf.*, 2021, pp. 2448–2456.
- [43] S. Fang, K. Zhao, G. Li, and J. Xu Yu, "Community search: Learn from small data," 2022, *arXiv:2201.00288*.
- [44] 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.



HONGXING LING received the Ph.D. degree in computer applications. His research interests include information system processing and machine learning.



GUANGSHENG LUO was born in Huangshi, Hubei, China, in 1982. He received the Ph.D. degree in computer application from Fudan University. He is currently a Postdoctoral Fellow at the Management Science and Engineering Research Station, Jiangxi University of Finance and Economics. His research interests include text processing and machine learning.



YU YANG received the bachelor's degree. He is currently the Vice President at AISHU Information Technology Company Ltd., Shanghai, China. His research interests include machine learning, deep learning, and smart finance.

• • •