



Research article

Patient multi-relational graph structure learning for diabetes clinical assistant diagnosis

Yong Li and Li Feng*

College of Computer Science and Engineering, Northwest Normal University, 967 Anning East Road, Lanzhou 730070, China

* **Correspondence:** Email: 2020211953@nwnu.edu.cn.

Abstract: The rapid accumulation of electronic health records (EHRs) and the advancements in data analysis technology have laid the foundation for research and clinical decision-making in the health-care community. Graph neural networks (GNNs), a deep learning model family for graph embedding representations, have been widely used in the field of smart healthcare. However, traditional GNNs rely on the basic assumption that the graph structure extracted from the complex interactions among the EHRs must be a real topology. Noisy connections or false topology in the graph structure leads to inefficient disease prediction. We devise a new model named PM-GSL to improve diabetes clinical assistant diagnosis based on patient multi-relational graph structure learning. Specifically, we first build a patient multi-relational graph based on patient demographics, diagnostic information, laboratory tests, and complex interactions between medicines in EHRs. Second, to fully consider the heterogeneity of the patient multi-relational graph, we consider the node characteristics and the higher-order semantics of nodes. Thus, three candidate graphs are generated in the PM-GSL model: original subgraph, overall feature graph, and higher-order semantic graph. Finally, we fuse the three candidate graphs into a new heterogeneous graph and jointly optimize the graph structure with GNNs in the disease prediction task. The experimental results indicate that PM-GSL outperforms other state-of-the-art models in diabetes clinical assistant diagnosis tasks.

Keywords: patient multi-relational graph; graph neural networks; graph structure learning; diabetes assistant diagnosis

1. Introduction

Electronic health records (EHRs) are generated by hospitals to maintain real-time digital records of patients, including clinical events, diagnosis, and treatment information of patients during the outpatient and inpatient periods. EHRs are composed of different structured [1] and unstructured data, in-

cluding demographic information, diagnosis information, laboratory results, medication prescriptions, medical imaging, and text descriptions, for example, as free clinical notes. Clinical assistant diagnosis is a fundamental task in smart healthcare, aiming to automatically extract valuable information from EHRs and make correct and intelligent diagnostic decisions, which improves the diagnostic efficiency of clinicians and reduces misdiagnosis. It is important in personalized medicine, disease prediction, and other tasks.

Although EHRs have been widely studied and have greatly improved clinical decision support [2], there are still some challenges for clinical assistant diagnosis based on EHRs. On the one hand, EHRs contain multiple types of data such as population information and laboratory tests, which make EHRs highly dimensional and heterogeneous, and it is challenging to create large-scale computational models from EHRs. On the other hand, an EHR is incomplete and noisy, because some hospitals are still further standardizing the unified writing standard of EHRs and patient follow-up records [3].

Recently, GNNs have been used extensively for clinical assistant diagnosis with significant research findings [1–4]. However, there are still three obstacles. First, most of the GNNs based on homogenous graph models focus on the co-occurrence of various medical concepts. Furthermore, although there are a few heterogeneous graph models based on EHRs, each relation in these models can represent only the local information of the heterogeneous graph. Combining different relationships forms a large amount of higher-order semantic information. Treating these heterogeneous relations uniformly is impossible, which limits the learning ability of GNNs. Finally, GNNs are highly sensitive to the original topology of information transmission in smart healthcare applications. An artificially designed heterogeneous graph model is usually extracted from structured and unstructured medical text data, which inevitably contain redundant, erroneous, and missing information. Consequently, it is difficult to deal with the challenges of structure learning and optimization in a heterogeneous graph model based on EHRs.

We propose a new model, PM-GSL, which introduces graph structure learning into diabetes clinical assistant diagnosis, and abstractly describes the semantic relations between entities in EHRs through patient multi-relational graph. PM-GSL can not only learn the structure of heterograph, but also learn the parameters of GNNs to overcome the weak structure problem of patient multi-relational graph. The contributions of this study are summarized as follows:

- 1) We apply graph structure learning to diabetes clinical assistant diagnosis and use EHRs to construct a patient multi-relational graph, which abstractly describes the relation between entities in EHRs from two perspectives: node attributes and meta-paths.

- 2) We propose a novel diabetes clinical assistant diagnosis model, PM-GSL. In this model, we design three candidate graphs, original subgraph, overall feature graph, and higher-order semantic graph, and fuse the three candidate graphs to generate a higher quality heterogeneous graph to obtain the complex interactive higher-order semantic information of the patient multi-relational graph.

- 3) We also conduct multiple experiments to compare the results with the existing state-of-the-art approach on two standard EHRs datasets: the large international public dataset MIMIC-IV and the Chinese hospital dataset P-EHRs. Our proposed PM-GSL model achieved a 9.62% improvement in Macro-F1, 9.17% improvement in Micro-F1 and 10.33% in AUC compared to the state-of-the-art models for diabetes clinical assistant diagnosis tasks.

The remainder of this paper is organized as follows. Section 2 describes related work, where we discuss deep learning models, such as GNNs, disease assistant diagnosis, and graph structure learning, which motivate this work. Section 3 describes the problem setting of diabetes clinical assistant diagno-

sis and presents the related definitions. Section 4 details the PM-GSL model. Section 5 demonstrates the results of our empirical evaluation of the PM-GSL on the MIMIC-IV and P-EHRs datasets. Finally, Section 6 provides concluding remarks and research directions.

2. Related work

2.1. Graph neural networks

GNNs were first proposed by Scarselli et al. [5] to further extend deep learning to graph-structured data, which mainly contains two types: spectral domain [6,7] and spatial domain [8,9]. The method based on the spectral domain adopts the spectral representation of the graph [10]. The spatial domain-based methods define convolution directly on the graph and aggregate feature information for each node from the spatial neighborhood, such as GraphSAGE [11] and GAT [12]. However, the above graph neural network can only be used to process a homogeneous graph. Some recent studies have attempted to extend GNNs to heterogeneous graphs, such as metapath2vec [13], which transforms heterogeneous graphs into homogeneous graph studies using random wandering based on meta-paths to learn graph representations.

The HAN [14] model utilizes the graph attention network method for the modeling and analysis of a heterogeneous graph, aggregating attribute information from meta-path-based neighbor nodes. Based on the HAN, the MAGNN [15] model combines the internal aggregation of meta-paths and aggregation between meta-paths. GNNs have powerful capabilities in learning node embedding representations and have achieved significant performance in numerous specific tasks [16]. However, almost all GNNs regard observed data containing noise or hypothetical data that are convenient for modeling as real information, which significantly depends on the quality of the original graph structure, greatly limiting their ability to handle the uncertainty in the graph structure.

2.2. Graph structure learning

Graph structure learning (GSL) can be traced back to several research achievements devoted to graph structure learning in network science [17,18]. The GSL handles the noise and incompleteness problems in the original graph data. Some existing methods learn graph structure from measurements of graph dynamical systems, such as coupled oscillators [19], and there are also studies [10] that set up attribute completion strategies for attribute missing problems. Nevertheless, the goal of these findings is different from graph representation learning. Recent studies have attempted to combine GSL with GNNs to improve the performance of downstream tasks. Franceschi et al. proposed a new method, LDS [20], to jointly learn and optimize the original graph and GNN parameters by approximating the solution.

The Pro-GNN [21] aims to extract the graph properties of sparsity, low rank and feature smoothness to design more robust GNNs that learn clean and efficient graph structures. The GTN [22] model can generate a new graph structure by identifying valuable edges that may exist as unconnected nodes on the original graph and can learn the effective embedded representation of nodes on the new graph from end to end. However, most GSL models are designed for homogeneous and smaller graphs. They modify the graph structure by removing noisy data from the graph topology and graph attribute similarity. Such models tend to optimize the entire graph [23], and it is difficult to analyze large-scale

graph data with complex structures and rich semantics in the medical field.

2.3. Disease assistant diagnosis

Graph-based methods have been widely used in auxiliary disease diagnosis tasks because of the powerful expression ability of graphs in relation modeling. In [3], a comprehensive disease self-diagnosis system based on the HealGCN model was proposed, which uses the predefined meta-paths inductive graph convolution operation to generate the embedded representation of new patients and solve the cold start problem by mining the complex interactions in EHRs data. The MVS-GCN [24] model, based on a graph neural network, learns the effective representation of the brain network in an end-to-end manner and combines graph structure learning and multitask graph embedding representation learning, thus improving the classification performance of brain disease diagnosis.

MM-STGNN [25] is a multimodal spatiotemporal graph neural network that uses graphs to represent the topological relationship between inpatients and predicts 30-day readmission according to patients' longitudinal imaging and EHRs data. In [26], the authors proposed an end-to-end multi-modal graph learning framework (MMGL) for disease prediction with multi-modality. And it aggregates the features of each modality by leveraging the correlation and complementarity between the modalities. Although GNN models have shown excellent predictive performance in numerous healthcare tasks, they are highly sensitive to the quality of graph structures. If a manually constructed graph is directly used in the GNNs and separated from the prediction module, it will lead to cumbersome tuning and poor generalization [27]. Simultaneously, these models failed to further explore the complementary information between multiple relations in the graph and the message propagation mechanism in the multi-relational subgraph.

3. Preliminaries

3.1. Clinical assistant diagnosis problem definition

Diabetes, glaucoma and central nervous system diseases (such as stroke) are common in middle-aged and elderly people. Their early symptoms are extremely similar and are difficult to distinguish clinically [28]. For example, in a diabetes patient with a wide range of blood glucose fluctuations, the osmotic pressure inside and outside the lens changes, causing a change in the curvature of the lens and affecting the patient's ability to focus the eye, thus producing blurred vision symptoms. Although blurred vision symptoms may be due to diabetes, glaucoma can also cause eye diopter changes in patients, resulting in lower intraocular pressure and causing blurred vision. In addition, stroke patients often show symptoms, such as blurred vision, sudden pins and needle sensations, and fatigue, which are similar to those of diabetic patients. This differs from many previous studies that modeled disease diagnosis as a link-prediction task. We use patient demographics, laboratory tests, physiological records, and medications in EHRs as nodes in the multi-relational graph to transform the disease prediction task into a classification task with fine-grained analysis.

3.2. Patient multi-relational graph

The patient multi-relational graph is defined as $G = (V, E, X, \{\xi_\gamma\}_{\gamma=1}^R)$, where V is the node set $\{v_1, v_2, \dots, v_N\}$. We let N denote the number of nodes, and the corresponding node type is M . E is the set

of edges, all edges form a primitive adjacency matrix $X \in \mathbb{R}^{N \times N}$, where X_{ij} indicates whether there is an edge between nodes v_i and v_j , edge relation type is ξ_r , $r \in \{1, 2, \dots, R\}$, and the multi-relational graph satisfies the inequality $|M| + |R| > 2$. The eigenvector of each node is $f_i \in \mathbb{R}^d$, and all the node feature matrices are represented as $X = \{f_1, f_2, \dots, f_N\} \in \mathbb{R}^{N \times d}$.

3.3. Meta-path

In the patient multi-relational graph, two nodes can be connected by different semantic patterns. The meta-path ϕ is formalized as $v_1 \xrightarrow{r_1} v_2 \xrightarrow{r_2} \dots \xrightarrow{r_{l-1}} v_l$, which is used to describe the composite relation $r_1 \circ r_2 \circ \dots \circ r_{l-1}$ between nodes v_1 and v_l , where \circ is a relational synthesis operation.

3.4. Patient multi-relational subgraph

The patient multi-relational subgraph G_r contains all node-pair triples with relation r , which is formalized as (v_i, r, v_j) . The adjacency matrix of a multi-relational subgraph is expressed as A_r , and if $(v_i, r, v_j) \in G_r$, then $A_{ij}^r = 1$; otherwise, $A_{ij}^r = 0$. $A = \{A_r, r \in R\}$ represents the set of all patient multi-relational subgraphs. For example, we define the type mapping functions $\varpi_h(r)$ and $\varpi_e(r)$ of the node to the triplet header and tail node in a patient medical record network. If $r = "pt"$ is interpreted as $\varpi_h(r) = "patient,"$ $\varpi_e(r) = "test."$ The adjacency matrix of the patient multi-relational subgraph is expressed as $A_r \in \mathbb{R}^{|\varpi_h(r)| \times |\varpi_e(r)|}$.

4. Patient multi-relational graph structure learning model

4.1. Constructing the patient multi-relational graph

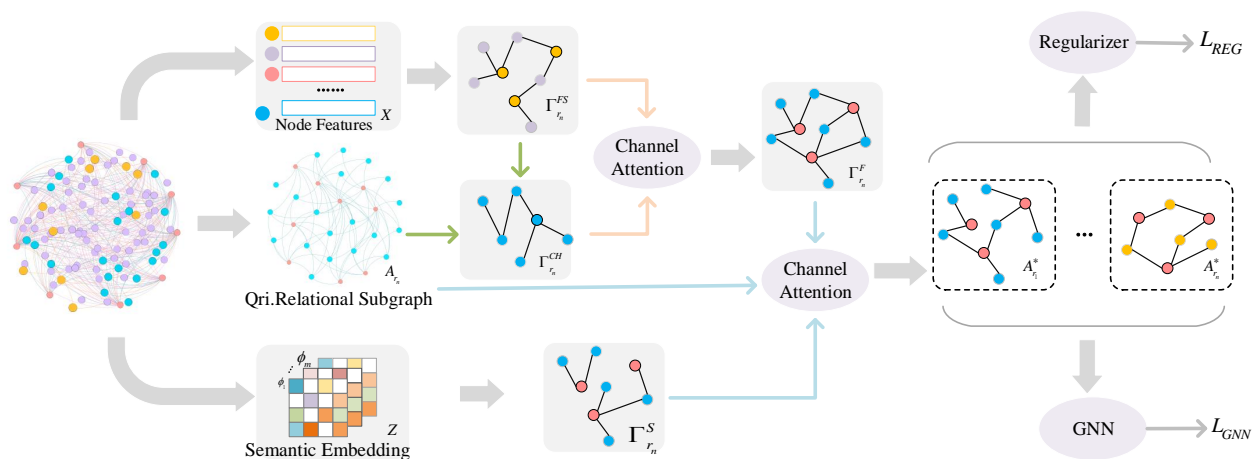
In this study, the EHRs are modeled as a patient multi-relational graph, where each patient is considered a node, and their EHRs data are associated with multiple medical concepts. For example, the following information is extracted from a record in the EHRs: if a patient's fasting blood glucose exceeded 7.0 mmol/L or postprandial blood glucose exceeded 11.1 mmol/L in the glucose tolerance test, the doctor identified abnormal blood glucose in the diagnosis conclusion, of which "fasting blood glucose" and "postprandial blood glucose" are medical concepts. We reduce the values corresponding to similar or identical medical concepts to smaller categories and set the nodes "normal" and "abnormal" for the patient's laboratory tests. Based on [28], patients' ages were grouped according to three thresholds of 15, 30 and 64, set as four nodes of representation. Over 100 medications are extracted from EHRs, which are divided into 20 types according to different abbreviations, formulations and their effects, with each type of medication considered as a node. The comorbidity diagnoses of the patients matched the appropriate ICD-10 codes.

In the diabetes diagnosis task, gender is not the most informative patient attribute, which is not conducive to the prediction task and may cause the problem of "over smoothing." According to the conclusion of [4], the PM-GSL model disregards gender attributes. We construct the multi-relational graph using different relations between five types of nodes, "Patient," "Age," "Medication," "Diagnoses" and "Laboratory Tests," and determine semantically meaningful meta-paths. For example, two patients with blurred vision, "patient1 \rightarrow metformin \leftarrow patient2," indicate that two patients had the same medication, and two patients were likely to have the same diagnosis. The meta-paths in Table 1 guide the random walk of the patient multi-relational graph.

Table 1. Meta-paths in the patient multi-relational graph.

Node types	Semantic	Meta-paths
Patient (P)		Patient \rightarrow Age \leftarrow Patient (PAP)
Age (A)	Two patients cite the	Patient \rightarrow Medicine \leftarrow Patient (PMP)
Medication (M)	same medical concept.	Patient \rightarrow Diagnosis \leftarrow Patient (PDP)
Diagnosis (D)		Patient \rightarrow Test \leftarrow Patient (PTP)
Laboratory Test (T)		

We propose the PM-GSL, a diabetes clinical assistant diagnosis model for patient multi-relational graph structure learning, as shown in Figure 1. First, we analyze the graph connectivity based on the PM-GSL to jointly train and optimize the patient multi-relational graph structure and GNNs parameters. Second, a semantic embedding matrix is constructed based on the node embedding method of meta-paths. Considering r_n as an example, we use the PM-GSL model to obtain a new overall feature graph $\Gamma_{r_n}^F$ and a higher-order semantic graph $\Gamma_{r_n}^S$ and fuse these two graphs with the original subgraph A_{r_n} to obtain a new subgraph $A_{r_n}^*$. Finally, the learned subgraphs are input into the GNN and regularizer to output the diagnosis and prediction of diabetes.

**Figure 1.** PM-GSL framework.

From the graph representation learning perspective, GNNs aim to learn node embedding by aggregating the information of neighboring nodes [29]. This iterative mechanism relies on the inherent information in the graph structure and attributes. However, artificially designed and constructed graphs have inevitable redundant noise and misinformation, which limits the predictive performance and interpretability of most GNNs, especially in smart healthcare, which requires higher graph quality. To optimize the patient multi-relational graph, the PM-GSL aims to fully use node features to enhance the original graph structure, which is mainly reflected in two aspects. First, we generate a feature similarity graph based on the similarity between node features. Then, node features are propagated to specific domains in the multi-relational graph to generate a message propagation graph, as shown in Figure 2.

4.2. Feature similarity graph

The pairwise connection of nodes is one of the most direct ways to represent information in the patient multi-relational graph. To preserve the pairwise proximity attribute between nodes in the patient multi-relational graph, the PM-GSL model determines the possibility of edges between two nodes with relation type $r_n \in R$ based on node features.

Specifically, for a node v_i with node type $\varphi(v_i)$ and node eigenvector $f_i \in \mathbb{R}^{1*d}$, the node features are projected into the d^* -dimensional common space using a type-specific mapping matrix $W_{\varphi(v_i)}$, which is denoted as $f_i^* \in \mathbb{R}^{1*d^*}$:

$$f_i^* = \sigma(W_{\varphi(v_i)} \cdot f_i + b_{\varphi(v_i)}) \quad (4.1)$$

$\sigma(\cdot)$ denotes the non-linear activation function, $W \in \mathbb{R}^{d \times d^*}$ denotes the mapping matrix, and $b \in \mathbb{R}^{1 \times d^*}$ denotes the bias vector. In the patient multi-relational graph, for relation r_n , we use weighted cosine similarity as the measurement function to measure node similarity and obtain a feature similarity graph $\Gamma_{r_n}^{FS} \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)|}$. The similarity between nodes v_i and v_j is expressed as follows:

$$\delta_{r_n}^{FS}(f_{v_i}^*, f_{v_j}^*) = \frac{1}{K} \sum_k^K \cos(Q_{k,r_n}^{FS} \odot f_{v_i}^*, Q_{k,r_n}^{FS} \odot f_{v_j}^*) \quad (4.2)$$

where $f_{v_i}^*$ and $f_{v_j}^*$ are the node features after projection, and Q_{k,r_n}^{FS} is the learnable parameter matrix. After the calculation, the node feature similarity matrix $\Gamma_{r_n}^{FS}$ is generated, and is symmetric, and the element range of the matrix is $[-1, 1]$. Therefore, we must extract a non-negative and sparse matrix from $\Gamma_{r_n}^{FS}$. Specifically, we defined a non-negative threshold λ^{FS} for automatic learning and set the elements that are smaller than λ^{FS} to 0. If λ^{FS} is larger, then $\Gamma_{r_n}^{FS}$ will be sparser, defined as follows:

$$\Gamma_{r_n}^{FS}(v_i, v_j) = \begin{cases} \delta_{r_n}^{FS}(f_{v_i}^*, f_{v_j}^*) & \delta_{r_n}^{FS}(f_{v_i}^*, f_{v_j}^*) \geq \lambda^{FS} \\ 0 & \text{otherwise} \end{cases} \quad (4.3)$$

4.3. Message propagation graph

The message propagation graph captures the interaction between the node features and topology, and it can find new effective edges more accurately. The key idea of the message propagation graph includes two aspects: First, many nodes of the same class have similar features and may be far apart in the feature space in the patient multi-relational graph structure data. Second, two nodes with similar node features are more likely to have similar neighbor nodes, and the feature similarity graph is propagated through the topology to generate new edges.

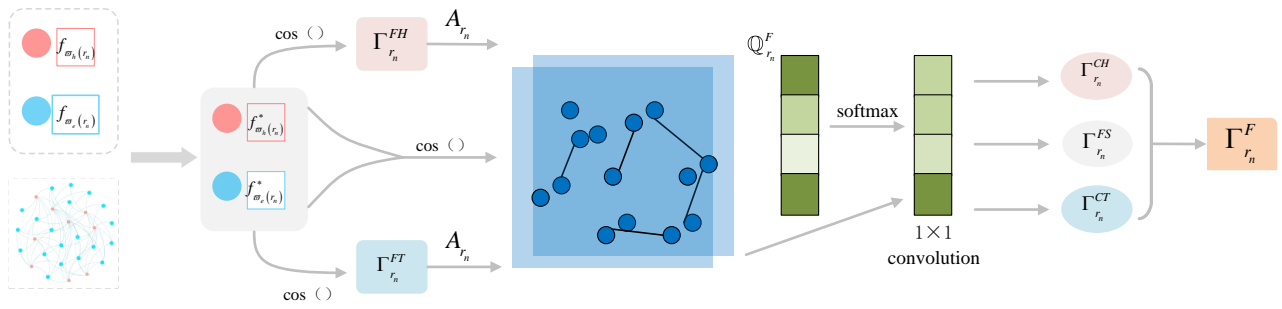


Figure 2. Message propagation graph.

The node types corresponding to the two nodes that satisfy the relation r_n in the patient multi-relational graph are $\varpi_h(r_n)$ and $\varpi_e(r_n)$, and the topology between them is $A_{r_n} \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)|}$. For two nodes $v_t, v_s \in \varpi_h(r_n)$, we calculate the feature similarity:

$$\delta_{r_n}^{FH}(f_{v_t}, f_{v_s}) = \cos(\mathbb{Q}_{r_n}^{FH} \odot f_{v_t}, \mathbb{Q}_{r_n}^{FH} \odot f_{v_s}) \quad (4.4)$$

Similar to Eq (4.3), a threshold λ^{FH} is set to control the sparsity of the feature similarity graph $\Gamma_{r_n}^{FH}$, as shown in Eq (4.5). After obtaining the similarity of $v_t, v_s \in \varpi_h(r_n)$, a potential message propagation graph $\Gamma_{r_n}^{CH} \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)|}$ is generated based on the feature similarity graph $\Gamma_{r_n}^{FH}$ and original subgraph A_{r_n} , as shown in Eq (4.6):

$$\Gamma_{r_n}^{FH}(v_t, v_s) = \begin{cases} \delta_{r_n}^{FH}(f_{v_t}, f_{v_s}) & \delta_{r_n}^{FH}(f_{v_t}, f_{v_s}) \geq \lambda^{FH} \\ 0 & \text{otherwise} \end{cases} \quad (4.5)$$

$$\Gamma_{r_n}^{CH} = \Gamma_{r_n}^{FH} \cdot A_{r_n} \quad (4.6)$$

Similarly, for the same type of $v_p, v_q \in \varpi_e(r_n)$, we calculate Eqs (4.4) and (4.5) to obtain another corresponding feature similarity graph $\Gamma_{r_n}^{FT}$ and the message propagation graph $\Gamma_{r_n}^{CT} \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)|}$:

$$\Gamma_{r_n}^{CT} = A_{r_n} \cdot \Gamma_{r_n}^{FT} \quad (4.7)$$

In summary, we fuse the generated feature similarity graph $\Gamma_{r_n}^{FS}$ and the message propagation graphs $\Gamma_{r_n}^{CH}$ and $\Gamma_{r_n}^{CT}$ through the channel attention layer [21] to obtain the overall feature graph $\Gamma_{r_n}^F \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)|}$ of relation r_n :

$$\Gamma_{r_n}^F = \Psi_{r_n}^F([\Gamma_{r_n}^{FS}, \Gamma_{r_n}^{CH}, \Gamma_{r_n}^{CT}]) \quad (4.8)$$

where $[\Gamma_{r_n}^{FS}, \Gamma_{r_n}^{CH}, \Gamma_{r_n}^{CT}] \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)| \times 3}$ denotes the stacking matrix of the candidate graphs that generate the overall feature graph. $\Psi_{r_n}^F$ represents the channel attention layer, and the corresponding parameter is $\mathbb{Q}_{r_n}^F \in \mathbb{R}^{1 \times 1 \times 3}$. The channel attention layer uses the function $\text{softmax}(\mathbb{Q}_{r_n}^F)$ to perform 1×1 convolution operations on the input. In this way, the PM-GSL model learns the different weights of the three candidate graphs of the overall feature graph respectively to measure the importance of each candidate graph to relation r_n .

4.4. Higher-order semantic graph

The semantic graph relies on the higher-order topology of the heterogeneous information network, which aims to describe the multi-hop structural interaction between two nodes determined by meta-paths. In the patient multi-relational graph, the overall feature graph describes the nodes from a microscopic perspective, whereas the higher-order semantic graph describes the nodes from a macroscopic perspective. The higher-order relations reflected by different meta-paths contain different semantic information. We obtain the higher-order semantic graph structure of the patient multi-relational graph from multiple meta-paths, as shown in Figure 3. Given a sequence r_1, r_2, \dots, r_n of the meta-path ϕ , a common method to generate a higher-order semantic graph is to multiply multiple adjacency matrices, such as the adjacency matrix $A_\phi = A_{r_n} \dots A_{r_2} A_{r_1}$. However, this method of obtaining high-order semantic graph by multiplying multiple adjacency matrices requires a considerable amount of time and space, and it discards the information of intermediate semantic nodes in the meta-paths.

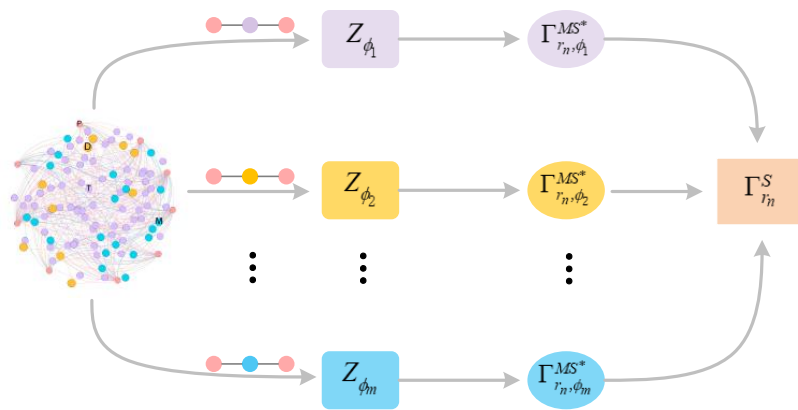


Figure 3. Higher-order semantic graph.

In the patient multi-relational graph, a set of meta-paths $\{\phi_1, \phi_2, \dots, \phi_m\}$ exists. Based on meta-path2vec [12], we obtained the meta-paths based node semantic embedding representation, $Z = \{Z_{\phi_1}, Z_{\phi_2}, \dots, Z_{\phi_m}\}$, to learn the potential higher-order semantic graph structure of the patient multi-relational graph. The entire training process adopts the meta-paths guidance random walk strategy and heterogeneous skip-gram model learning node embedding representation, which significantly retains the information of intermediate semantic nodes [30]. In addition, the training process of the meta-path2vec model does not involve adjacency matrix multiplication, which reduces the time and space complexity.

After obtaining the node semantic embedding representation Z , for each meta-path ϕ_m in the patient multi-relational graph, we generate the adjacency matrix $\Gamma_{r_n, \phi_m}^{MS*} \in \mathbb{R}^{|\varpi_h(r_n)| \times |\varpi_e(r_n)|}$, for the semantic subgraph, and the calculation method for each edge is

$$\Gamma_{r_n, \phi_m}^{MS*}(i, j) = \begin{cases} \delta_{r_n, \phi_m}^{MS*}(z_i^m, z_j^m) = \cos(\mathbb{Q}_{r_n}^{MS*} \odot z_i^m, \mathbb{Q}_{r_n}^{MS*} \odot z_j^m) & \delta_{r_n, \phi_m}^{MS*}(z_i^m, z_j^m) \geq \lambda^{MS*} \\ 0 & \text{otherwise} \end{cases} \quad (4.9)$$

where z_i^m denotes the i -th row of Z_{ϕ_m} . $\delta_{r_n, \phi_m}^{MS^*}$ is a metric function with parameter $\mathbb{Q}_{r_n}^{MS^*}$. We extend this calculation method to all meta-paths to generate m semantic subgraphs and aggregate them to obtain a higher-order semantic graph $\Gamma_{r_n}^S$:

$$\Gamma_{r_n}^S = \Psi_{r_n}^{MS^*} \left(\left[\Gamma_{r_n, \phi_1}^{MS^*}, \Gamma_{r_n, \phi_2}^{MS^*}, \dots, \Gamma_{r_n, \phi_m}^{MS^*} \right] \right) \quad (4.10)$$

where $\left[\Gamma_{r_n, \phi_1}^{MS^*}, \Gamma_{r_n, \phi_2}^{MS^*}, \dots, \Gamma_{r_n, \phi_m}^{MS^*} \right]$ is the stacking matrix of the corresponding m semantic subgraphs. $\Psi_{r_n}^{MS^*}$ represents the channel attention layer, and the corresponding weight matrix $\mathbb{Q}_{r_n}^{MS^*} \in \mathbb{R}^{1 \times 1 \times m}$ denotes the importance of semantic subgraphs based on different meta-paths of relation r_n . After obtaining $\Gamma_{r_n}^S$, we need to aggregate the learned overall feature graph $\Gamma_{r_n}^F$ and higher-order semantic graph $\Gamma_{r_n}^S$, with the original patient subgraph A_{r_n} to generate a new relational subgraph $A_{r_n}^*$ of relation r_n :

$$A_{r_n}^* = \Psi_{r_n} \left(\left[\Gamma_{r_n}^F, \Gamma_{r_n}^S, A_{r_n} \right] \right) \quad (4.11)$$

Similarly, $\left[\Gamma_{r_n}^F, \Gamma_{r_n}^S, A_{r_n} \right]$ is the stacking matrix of the generated new relational adjacency matrix. Weight $\mathbb{Q}_{r_n} \in \mathbb{R}^{1 \times 1 \times 3}$ of candidate graphs is learned from the channel attention layer Ψ_{r_n} , to evaluate the importance of the three candidate graphs in generating the relational subgraph $A_{r_n}^*$. Finally, we obtain a new patient multi-relational graph $A^* = \{A_{r_n}^*, r_n \in R\}$ in the PM-GSL model.

4.5. Optimization

In the PM-GSL model, we combine the GNN parameters to optimize the new graph structure A^* . Considering that the adjacency matrix of the original heterogeneous graph may provide misleading information for the aggregation process, we aggregate the feature representation of neighbor nodes through the obtained patient multi-relational graph structure and apply it to diabetes clinical assistant diagnosis. For the patient multi-relational graph structure A^* , the forward model of the two-layer GCN can be expressed as follows:

$$Y = f(X, F') = \text{softmax} \left(\hat{A} \text{ReLU} \left(\hat{A} X W^0 \right) W^1 \right) \quad (4.12)$$

where $X \in \mathbb{R}^{N \times d}$ is the original feature matrix of the node and the input to the first GCN layer. F' is an adjacency matrix constructed using A^* . $\hat{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$, $\tilde{D}_{ii} = 1 + \sum_j F'_{ij}$, $\tilde{A} = F' + I_N$, and I_N are unit matrices. The classification loss on the graph is expressed as follows:

$$L_{GNN} = \sum_{v_i \in V} l(Y_{v_i}, y_{v_i}) \quad (4.13)$$

where Y_{v_i} is the prediction label, and $l(\cdot, \cdot)$ indicates the cross-entropy loss function, which measures the error between the prediction label Y_{v_i} and the real label y_{v_i} .

The patient multi-relational graph constructed from multiple pieces of information in real-world EHRs is noisy and has task-independent edges. We propose the PM-GSL model, which makes the original patient multi-relational graph more adaptable to the prediction task but more prone to overfitting problems. Therefore, we impose sparsity constraints on the adjacency matrix F' of the graph A^* and apply the regularization term to the learned graph as follows:

$$L_{REG} = \beta \|F'\|_1 \quad (4.14)$$

The total loss of the PM-GSL model is described by Eq (4.15). Minimizing L can enable the joint optimization of the graph structure and GNN parameters to improve the prediction task performance.

$$L = L_{GNN} + L_{REG} \quad (4.15)$$

5. Experiment

We use two real EHRs datasets to evaluate the PM-GSL model: one from American hospitals and another from Chinese hospitals. Then, we compare the PM-GSL model with state-of-the-art methods. Specifically, we aim to answer the following four questions:

RQ1: Does the PM-GSL model improve the accuracy of a diabetes diagnosis?

RQ2: How important are the three types of candidate graphs in the PM-GSL model for generating new graph structures?

RQ3: Does PM-GSL adaptively assign greater channel attention values to important information?

RQ4: Is graph structure learning effective in improving the quality of the patient multi-relational graph?

5.1. Dataset

Our work is tested and analyzed by using two datasets. MIMIC-IV is a large international public EHRs dataset that collected clinical data of over 380,000 patients from 2008 to 2019. This dataset was collected at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, USA, and included patient centered clinical records, such as demographics, vital sign measurements, nursing notes, and laboratory tests.

Furthermore, we also use the EHRs of a tertiary care hospital located in a major metropolitan center in northwestern China called P-EHRs. Although the amount and completeness of the data are relatively small, the information included in this study is sufficient to support our work. More specifically, there was 1) demographic information, such as age and sex. 2) Diagnosis: The combination group is defined according to the coding information of ICD-9 and ICD-10, mainly including 14 diseases, such as myocardial infarction, congestive heart failure, hyperlipidemia, hypertension, diabetes, and chronic lung disease. 3) Laboratory tests: We extracted the laboratory test indicators obtained for the first time and the minimum, mean, and maximum values of laboratory tests during hospitalization. 4) Vital signs: We selected the minimum, mean, and maximum values of the patient's vital signs during the first day of admission, which included heart rate, systolic blood pressure, diastolic blood pressure, mean arterial pressure, respiratory rate, and percutaneous oxygen saturation. 5) Medications: We collected data on drugs that were used more than 300 times in total during the patients' hospitalization and classified them into 20 categories according to the functions of these drugs. Over all, we briefly summarize in Tables 2 and 3.

Table 2. Node statistics of the patient multi-relational graph.

Dataset	MIMIC-IV		P-EHRs	
	Quantity/Types	Nodes	Quantity/Types	Nodes
Patient	6000/3	6000	2000/3	2000
Age	4	4	4	4
Diagnosis	14	14	6	6
Medication	101	20	65	13
Laboratory test	38	76	26	52
Vital sign	6	12	6	12

Table 3. Statistics of the datasets.

Datasets	Nodes	Edges	Edge types	Features	Training	Validation	Test
MIMIC-IV	6122	118,717	4	107	2700	1200	2100
P-EHRs	2083	29,572	4	76	1700	400	900

5.2. Experimental settings and baselines

The experiment is based on Python 3.6.8 and TensorFlow 2.3.0. The calculation of a single model runs on NVIDIA RTX 6000 GPU. According to [30], in all the methods involved in this study, the node embedding dimension d and the common space dimension d^* were set to 64 and 16, respectively. Equation (4.2) in Section 4 represents a 2-head cosine similarity measure function, that is, $K = 2$. The learning rate and weight decay are set to 0.01 and 0.0005, respectively. The other hyperparameters are adjusted by using the grid search method, which is reflected in λ^{FS} , λ^{FH} , λ^{MS^*} and β for Eqs (4.3), (4.5), (4.9), and (4.14).

We compare the PM-GSL with six state-of-the-art baseline graph neural network models in recent years. They include two homogeneous graphs (GCN [7] and GAT [12]), two heterogeneous graphs (metapath2vec [13] and HAN [14]), and two graph structure learning methods (LDS [20] and Pro-GNN [21]). We conduct six ablation experiments to verify the effectiveness of the three candidate graphs in the PM-GSL for diabetes diagnosis.

5.3. Experimentation

5.3.1. RQ1: Overall comparison

In this section, we compare the overall performance of the PM-GSL and six baseline models, as shown in Figure 4. The classification process was repeated ten times to obtain the average value and ensure more stable and reliable prediction results. We use the evaluation metrics Macro-F1, Micro-F1, and AUC, which are commonly used in classification tasks.

The experimental findings are as follows:

(i) The PM-GSL model consistently outperforms the three types of baselines in predicting both EHRs datasets, which not only indicates that noisy data in the original EHRs prevent the GNN from aggregating effective feature information but also proves that PM-GSL can obtain a higher-quality

heterogeneous graph structure.

(ii) Compared with GCN and GAT, the performance of PM-GSL on the two EHRs datasets was significantly improved by 13.36–17.59%, which confirms that the fine-grained division of the patient multi-relational graph is extremely conducive to the diabetes diagnosis.

(iii) Although the metapath2vec and HAN models can be used to analyze heterogeneous relations, the prediction performance of these two models is 9.49–14.39% lower than that of the PM-GSL model proposed in this study. This is because the density of neighbor nodes in each relation in the patient multi relation graph is high. In particular, the HAN model directly uses the original graph structure as the input and cannot effectively filter out neighbor nodes with interference factors, which causes difficulties in diabetes diagnosis.

(iv) Compared with the LDS model and Pro-GNN model, the PM-GSL model proposed in this study can improve the prediction performance by 7.26–9.00%. In particular, the PM-GSL model can calculate node similarity and message propagation path and can make full use of multiple patient relations, which is helpful for learning better graph structure and more robust GNN parameters; thus, the new patient multi-relational graph has a stronger ability to adapt to prediction tasks.

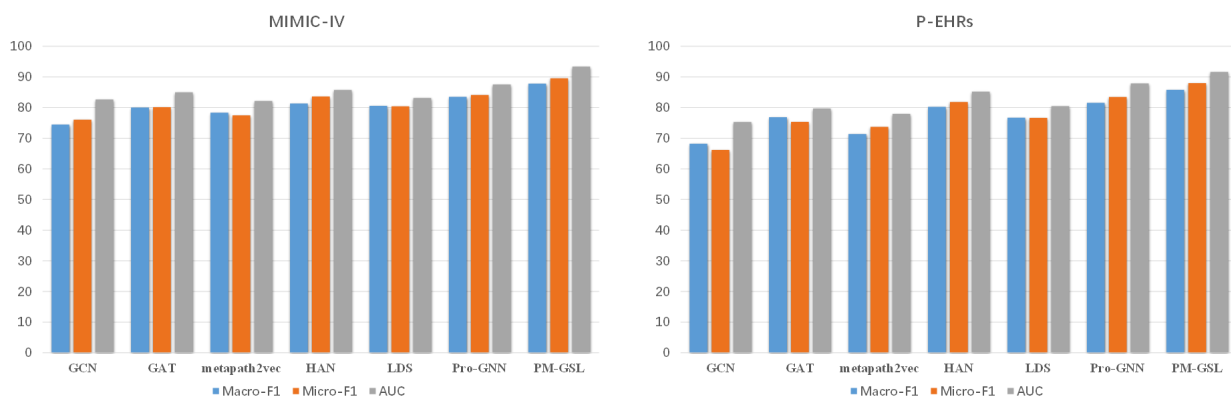


Figure 4. Performance evaluation of diabetes diagnosis.

5.3.2. RQ2: Ablation study

To further verify the impact of the three core candidate graphs in the PM-GSL model, we designed three variants of the PM-GSL model by deleting any type of candidate graphs, which are respectively represented as $PM-GSL^{FS}$, $PM-GSL^{HT}$ and $PM-GSL^{S^*}$. Table 4 shows the results of these variants in terms of the Macro-F1, Micro-F1 and AUC, and we can observe the following:

(i) The PM-GSL model outperforms the three ablation studies on diabetes diagnostic task, indicating the necessity of all candidate graphs in generating new graph structures.

(ii) Compared with the other two ablation studies, $PM-GSL^{S^*}$ significantly decreased in the three evaluation indicators, because the semantic information between different relations is more important in the patient multi-relational graph.

(iii) To verify the effectiveness of the candidate graph weights learned by the PM-GSL, we replace the channel attention layer in Eq (4.11) with an average aggregation layer that treats each type of candidate graph equally. The effect of the channel attention layer is better than that of the $PM-GSL^{avg}$.

If on average, three core candidate graphs are fused in the PM-GSL model, the influence of higher-order semantic information in the patient multi-relational graph will be weakened, and the prediction performance will be reduced.

Table 4. PM-GSL ablation experiment.

Dataset	MIMIC-IV			P-EHRs		
	Macro-F1	Micro-F1	AUC	Macro-F1	Micro-F1	AUC
PM-GSL ^{FS}	0.7714	0.7917	0.8538	0.7167	0.7045	0.7489
L-PM-GSL ^{HT}	0.8041	0.7933	0.8678	0.7383	0.7248	0.7744
R-PM-GSL ^{HT}	0.7920	0.7867	0.8346	0.7817	0.7823	0.8351
PM-GSL ^{HT}	0.7667	0.7333	0.7955	0.7367	0.7443	0.8195
PM-GSL ^{S*}	0.6963	0.7133	0.7679	0.6433	0.6729	0.7351
PM-GSL ^{avg}	0.8537	0.8469	0.8907	0.8117	0.8373	0.8796
PM-GSL	0.8783	0.8954	0.9337	0.8579	0.8800	0.9164

5.3.3. RQ3: Weight analysis of candidate graphs

We analyze the distribution of the channel attention weight, further studied the ability of the PM-GSL model to distinguish the importance of three core candidate graphs and fuse the three core candidate graphs to generate a new patient multi-relational subgraph. We trained the PM-GSL model 50 times and set all thresholds to 0.2. The distribution of channel attention values is shown in Figure 5.

Specifically, the original subgraph is an important structure for diabetes prediction. However, for P-T, the PM-GSL model assigns a large channel attention value to the learned higher-order semantic graph, indicating that the information in the higher-order semantic graph is more important. The above phenomenon is consistent with the conclusion (iii) from the ablation experiment, which further proves that the PM-GSL model can adaptively provide more channel attention to important information.

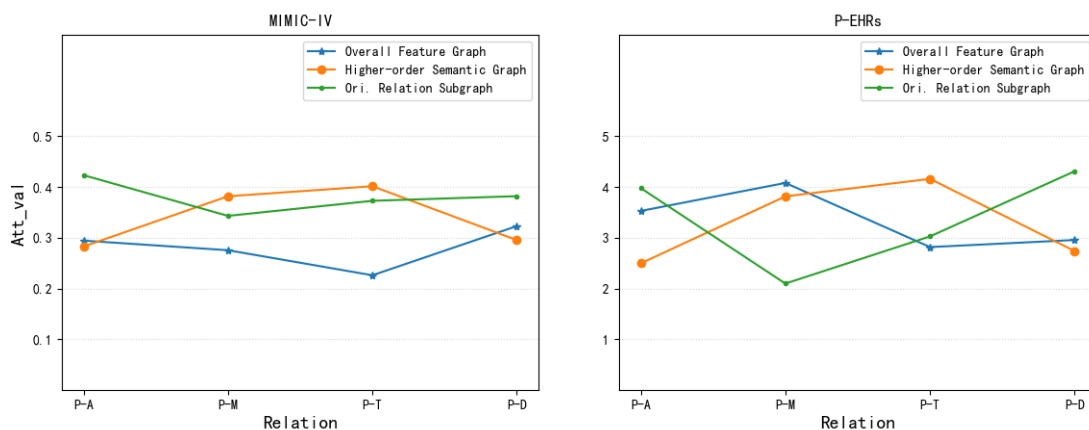


Figure 5. Performance evaluation of diabetes diagnosis.

5.3.4. RQ4: Parameter sensitivity analysis

The similarity thresholds defined in Eqs (4.3), (4.5) and (4.9) can be used to control the sparsity of generated graphs. To better compare the experimental results, we set up $\lambda^{FS} = \lambda^{FH}$ and conduct experiments on different λ^{FS} and λ^{MS^*} . Figure 6 shows the Macro-F1 and AUC values in the comparison experiment. On the two datasets, when both $\lambda^{FS} = \lambda^{FH}$ and λ^{MS^*} are set to 1, the prediction performance of the PM-GSL model is significantly reduced. This is because the PM-GSL model uses only the original graph structure of the patient multi-relational graph, which is similar to the general GNN model. This finding also proves the effectiveness of graph structure learning in improving the quality of the patient's multi-relational graph.

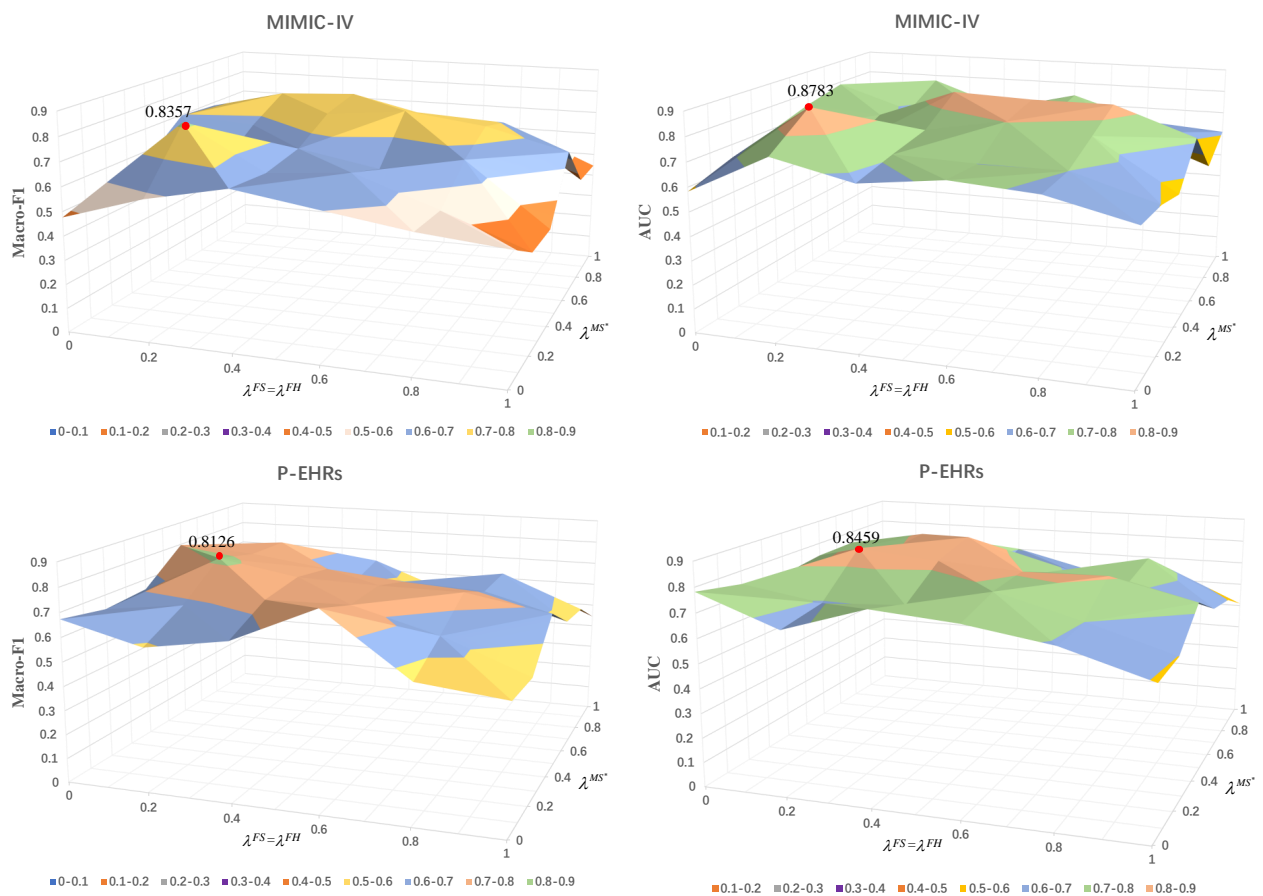


Figure 6. Parameter sensitivity.

6. Conclusions

In this study, we attempt to incorporate heterogeneous graph structure learning into the task of diabetes clinical assistant diagnosis and propose a model, PM-GSL, which jointly learns patient multi-relational graph structures and GNN parameters for diabetes prediction. In particular, we use the effective information in EHRs to build a patient multi-relational graph to simulate the complex inter-

actions between multiple medical entities in EHRs. The patient multi-relational graphs propagate the node characteristics through the topology to generate the underlying graph structure and describe the nodes from micro and macro perspectives, respectively. We designed three fine candidate graphs and fused them to generate clean and effective heterogeneous graphs to solve the weak structure problem of the EHRs.

Numerous experiments have proven the effectiveness of the PM-GSL model, which well proves that graph structure learning is helpful to improve the quality of patient multi-relational graph and the accuracy of disease prediction tasks. Our proposed PM-GSL model achieved a 9.62% improvement in Macro-F1, 9.17% improvement in Micro-F1 and 10.33% in AUC compared to the baselines. In the future, we aim to extend PM-GSL to the multi-view model and explicitly integrate the label information into graph structure learning to improve the performance of disease clinical assistant diagnosis and enhance the clinical interpretability of the model.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments. This research was funded by the National Natural Science Foundation of China (No. 61863032), Northwest Normal University Major Research Project Incubation Program, China (No. NWNLU-LKZD2021-06).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. Lin, S. Wang, Y. Ding, An empirical study of using radiology reports and images to improve ICU-mortality prediction, in *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, (2021), 497–498. <https://doi.org/10.1109/ICHI52183.2021.00088>
2. Z. Liang, Z. Zhang, H. Chen, Disease prediction based on multi-type data fusion from Chinese electronic health record, *Math. Biosci. Eng.*, **19** (2022), 13732–13746. <https://doi.org/10.3934/mbe.2022640>
3. Z. Wang, R. Wen, X. Chen, Online disease diagnosis with inductive heterogeneous graph convolutional networks, in *Proceedings of the Web Conference 2021*, (2021), 3349–3358. <https://doi.org/10.1145/3442381.3449795>
4. Z. Liu, X. Li, H. Peng, Heterogeneous similarity graph neural network on electronic health records, in *2020 IEEE International Conference on Big Data (Big Data)*, (2020), 1196–1205. <https://doi.org/10.1109/BigData50022.2020.9377795>
5. F. Scarselli, M. Gori, A. C. Tsoi, The graph neural network model, *IEEE Trans. Neural Networks*, **20** (2008), 61–80. <https://doi.org/10.1109/TNN.2008.2005605>
6. M. Defferrard, X. Bresson, P. Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in *Advances in Neural Information Processing Systems*, **29** (2016).

7. T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, preprint, arXiv:1609.02907.
8. F. Wu, A. Souza, T. Zhang, Simplifying graph convolutional networks, in *International Conference on Machine Learning*, (2019), 6861–6871.
9. J. Zhang, X. Shi, J. Xie, Gaan: Gated attention networks for learning on large and spatiotemporal graphs, preprint, arXiv:1803.07294.
10. K. Zhang, B. Hu, F. Zhou, Graph-based structural knowledge-aware network for diagnosis assistant, *Math. Biosci. Eng.*, **19** (2022), 10533–10549. <https://doi.org/10.3934/mbe.2022492>
11. W. L. Hamilton, R. Ying, J. Leskovec, Inductive representation learning on large graphs, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (2017), 1025–1035.
12. P. Veličković, G. Cucurull, A. Casanova, Graph attention networks, preprint, arXiv:1710.10903.
13. Y. Dong, N. V. Chawla, A. Swami, metapath2vec: Scalable representation learning for heterogeneous networks, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2017), 135–144. <https://doi.org/10.1145/3097983.3098036>
14. X. Wang, H. Ji, C. Shi, Heterogeneous graph attention network, in *The World Wide Web Conference*, (2019), 2022–2032. <https://doi.org/10.1145/3308558.3313562>
15. X. Fu, J. Zhang, Z. Meng, Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding, in *Proceedings of The Web Conference 2020*, (2020), 2331–2341. <https://doi.org/10.1145/3366423.3380297>
16. X. Wang, N. Liu, H. Han, Self-supervised heterogeneous graph neural network with co-contrastive learning, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, (2021), 1726–1736. <https://doi.org/10.1145/3447548.3467415>
17. M. E. J. Newman, Estimating network structure from unreliable measurements, *Phys. Rev. E*, **98** (2018), 62321. <https://doi.org/10.1103/PhysRevE.98.062321>
18. T. Martin, B. Ball, M. E. J. Newman, Structural inference for uncertain networks, *Phys. Rev. E*, **93** (2016), 12306. <https://doi.org/10.1103/PhysRevE.93.012306>
19. W. X. Wan, Y. C. Lai, C. Grebogi, Data based identification and prediction of nonlinear and complex dynamical systems, *Phys. Rep.*, **644** (2016), 1–76. <https://doi.org/10.1016/j.physrep.2016.06.004>
20. L. Franceschi, M. Niepert, M. Pontil, Learning discrete structures for graph neural networks, in *International Conference on Machine Learning*, (2019), 1972–1982.
21. W. Jin, Y. Ma, X. Liu, Graph structure learning for robust graph neural networks, in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, (2020), 66–74. <https://doi.org/10.1145/3394486.3403049>
22. S. Yun, M. Jeong, R. Kim, Graph transformer networks, in *Advances in Neural Information Processing Systems*, **32** (2019).
23. Y. Ji, G. Chu, X. Wang, Prohibited item detection via risk graph structure learning, in *Proceedings of the ACM Web Conference 2022*, (2022), 1434–1443. <https://doi.org/10.1145/3485447.3512190>

24. G. Wen, P. Cao, H. Bao, MVS-GCN: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis, *Comput. Biol. Med.*, **142** (2022), 105239. <https://doi.org/10.1016/j.compbiomed.2022.105239>
25. S. Tang, A. Tariq, J. Dunnmon, Multimodal spatiotemporal graph neural networks for improved prediction of 30-day all-cause hospital readmission, preprint, arXiv:2204.06766.
26. S. Zheng, Z. Zhu, Z. Liu, Multi-modal graph learning for disease prediction, *IEEE Trans. Med. Imaging*, **41** (2022), 2207–2216. <https://doi.org/10.1109/TMI.2022.3159264>
27. E. Choi, Z. Xu, Y. Li, Learning the graphical structure of electronic health records with graph convolutional transformer, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **34** (2020), 606–613. <https://doi.org/10.1609/aaai.v34i01.5400>
28. Y. Cao, H. Peng, P. S. Yu, Multi-information source hin for medical concept embedding, in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, Cham, (2020), 396–408. https://doi.org/10.1007/978-3-030-47436-2_30
29. J. Gilmer, S. S. Schoenholz, P. F. Riley, Neural message passing for quantum chemistry, in *International Conference on Machine Learning*, (2017), 1263–1272.
30. J. Zhao, X. Wang, C. Shi, Heterogeneous graph structure learning for graph neural networks, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 4697–4705. <https://doi.org/10.1609/aaai.v35i5.16600>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)