HGCL: Heterogeneous Graph Contrastive Learning for Traditional Chinese Medicine Prescription Generation

Chapter · October 2022

DOI: 10.1007/978-3-031-20627-6_9

CITATIONS
0

READS
15

3 authors, including:

Zecheng Yin
The Chinese University of Hong Kong (Shenzhen)
2 PUBLICATIONS **1** CITATION
SEE PROFILE

Yanchun Zhang
Victoria University Melbourne
316 PUBLICATIONS **6,672** CITATIONS
SEE PROFILE

HGCL: Heterogeneous Graph Contrastive Learning for Traditional Chinese Medicine Prescription Generation

Zecheng Yin[1(✉)], Yingpei Wu[2], and Yanchun Zhang[1,3]

[1] Guangzhou University, Guangzhou, China
yinzecheng@e.gzhu.edu.cn
[2] Fudan University, Shanghai, China
ypwu16@fdu.edu.cn
[3] Pengcheng Lab, Shenzhen, China

Abstract. Traditional Chinese Medicine (TCM) is a highly empirical, subjective and practical discipline. Generating an appropriate prescription has been one of the most crucial components in building intelligent diagnosis systems that provide clinical decision support to physicians. While various machine learning models for prescription generation have been created, they suffer from specific limitations (e.g., data complexity and semantic ambiguity, lack of syndrome differentiation thinking, etc.). For handling these limitations, we propose a novel Heterogeneous Graph Contrastive Learning (HGCL) based model to conduct prescription generation with the idea of syndrome differentiation and treatment. Specifically, we first model the TCM clinical prescriptions as a Heterogeneous Information Network (THIN), and then explore node- and semantic-level contrastive learning on THIN, so as to enhance the quality of node representations for several downstream tasks such as node classification, prescription generation, etc. We conduct extensive experiments on three real TCM clinical datasets, demonstrating significant improvement over state-of-the-art methods, even though some of which are fully unsupervised.

Keywords: Heterogeneous graph · Data mining · Contrasitve learning · Representation learning

1 Introduction

Syndrome differentiation and treatment (Bian Zheng Lun Zhi) is one of the most important principles in the clinical practice of TCM. In TCM treatment, physicians first obtain the symptoms through the four diagnoses (i.e., observation, listening, questioning, and pulse analyses). By carefully analyzing these symptoms, physicians deduce the syndromes of the patient and then choose a set of suitable herbs as a formula to cure the syndromes [7].

The syndrome differentiation process, however, is particularly sophisticated. A symptom can show in a variety of syndromes, while a syndrome contains

A. Traina et al. (Eds.): HIS 2022, LNCS 13705, pp. 88–99, 2022.
https://doi.org/10.1007/978-3-031-20627-6_9

multiple symptoms [3,24]. Even for the same specific set of symptoms, different physicians may reach different syndromes, rendering it difficult to find criteria for syndrome differentiation [8]. These make the clinical diagnosis and treatment of TCM very difficult for physicians. Consequently, a study of the prescription generation with syndrome reasoning is needed to help TCM physicians prepare prescriptions and provide clinical decision support.

Several recent studies on TCM prescription generation have achieved encouraging results [10]. Topic-model based methods consider each prescription as a document with several latent topics where symptoms and herbs are viewed as a group of words respectively [26,27]. Unfortunately, the performance of these methods is limited by the sparsity of prescriptions [14]. Graph representation learning-based methods focus on learning low-dimensional embeddings of TCM entities, and then generating prescription based on these embeddings [8,14,15]. Because of being based on random walk-based skip-gram model or graph convolution network (GCN) [9], most of these methods fail to capture the global information of a graph and thus overlook much of high-order properties [13]. In summary, current TCM prescription generation models suffer from following limitations:

(1) Data complexity and semantic ambiguity. A TCM clinical prescription consists of three parts (i.e., a set of symptoms, herbs, and syndromes), implying the syndrome differentiation and treatment procedure. There are rich semantic relations among TCM entities, such as the herb compatibility, the generalization relation between symptoms and syndromes, and the evolutionary relation between symptoms. Considering entities of prescriptions in weak order, modeling the prescription text and extracting the semantic relation information between entities is a critical issue. **(2) Lack of syndrome differentiation thinking.** Most of the above-mentioned methods simply deliver a combination of herbs for the symptoms. The generation process is a black box lacking an explicit syndrome induction and not following the philosophy of syndrome differentiation and treatment.

To address these limitations, we propose a novel Heterogeneous Graph Contrastive Learning (HGCL) based model to conduct prescription generation with the philosophy of syndrome differentiation and treatment. Based on the constructed THIN, our HGCL framework is designed to optimize two objectives, namely node-level and semantic-level contrastiveness. Firstly, considering the skewed data distribution in THIN, we perform topology and feature augmentations on THIN and generate two correlated graph views. Then, a novel node-level contrastive loss tailored for THIN is proposed to maximize the agreement of node embeddings of these two views. The intuition behind is that modifying the graph structure and reducing the influence of high-degree nodes can help alleviate the degree biases, which guide the model to capture the structure information of THIN without the influence of skewed data distribution. Secondly, to capture high-order semantic information in heterogeneous graph, the semantic-level contrastive learning is conducted on THIN. We first generate meta-path instances including the positive and negative samples which contain rich and subtle

semantic information in THIN, and then distinguish these samples with a innovative semantic-level contrastive learning. Our main contributions are summarized as follows:

- We model the TCM prescription generation problem as a graph representation problem in THIN, which is constituted by herbs, symptoms, and syndromes and their multiple relations extracted from clinical prescriptions.
- We propose a novel heterogeneous graph contrastive representation learning model, named HGCL, which combines the node- and semantic-level contrastive learning. To the best of our knowledge, this is the first attempt to take advantage of the heterogeneous graph contrastive learning for TCM prescription generation.
- Extensive experimental results have demonstrated that our proposed HGCL framework significantly outperforms seven state-of-the-art baselines, even though it is fully unsupervised.

2 Related Work

2.1 TCM Prescription Generation

TCM prescription generation has been a popular and challenging research issue in recent years. Wang et al. [23] identified the issue of detecting the relationship between symptoms and herbs as a machine translation problem and developed a transformer-based model to convert symptoms to herbs. Based on the idea of Transformer [18], Li et al. [11] presented a TCM knowledge enhanced Seq2Seq model to generate prescription. Although effective in prescription generation, Seq2Seq-based models have a weakness in dealing with short-text-like prescriptions, making it hard to capture the semantic information included in TCM prescriptions. Furthermore, due to lacking an explicit syndrome induction, most of the above methods do not follow TCM's philosophy of syndrome differentiation and treatment, which further limits their performance and applicability.

2.2 Graph Contrastive Learning

Lately, contrastive methods adopted for computer vision have also been adapted to graph domain and achieved great success. Velickovic et al. proposed DGI [20] which extended deep InfoMax [5] to graphs and contrasted node-local patches against global graph representations. Following this line of study, Sun et al. developed InfoGraph [17] which learned graph representations by maximizing the MI between the graph representations and sub-structural representations. For multiplex graph, Park et al. presented DMGI [13] which divided the original graph into several homogeneous ones and followed infomax objective of DGI for each relation graph. Inspired by SimCLR [1], MoCo [4] Zhu et al. presented GRACE [29] which generated two graph views by corruption and learned node representations by maximizing the agreement of node representations in these two views. Motivated by the above models, our proposed HGCL is tailored for THIN representation learning.

3 Preliminary

Definition 1. TCM Heterogeneous Information Network (THIN). As a HIN, THIN is defined as a heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ and $\mathcal{E}$ represent the sets of nodes and edges respectively. A THIN is also associated with a node type mapping function $\phi(v) : \mathcal{V} \rightarrow \mathcal{A}$, $\forall v \in \mathcal{V}$ and a relation type mapping function $\varphi(e) : \mathcal{E} \rightarrow \mathcal{R}$, $\forall e \in \mathcal{E}$. $\mathcal{A}$ and $\mathcal{R}$ denote the sets of predefined node types and relation types. Specifically, Fig. 1(a) shows an example of THIN. In the THIN, we define three types of nodes corresponding to symptom, syndrome, herb, and two types of links representing various relationships between them. Symptom-Syndrome edges represent the generalization of syndrome to symptom in a prescription. Syndrome-Herb edges indicate the therapeutic or palliative effect of prescribed herbs on specific symptoms.

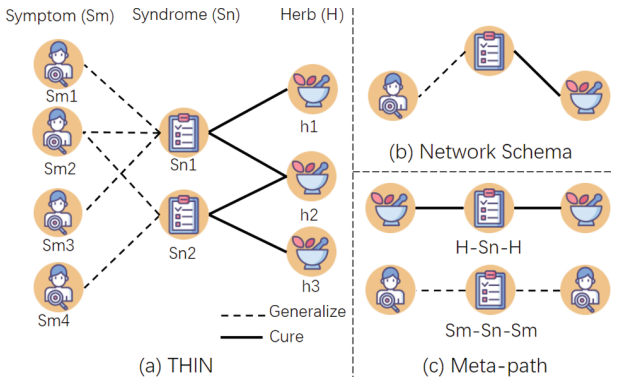

Fig. 1. A example of THIN and associated demonstration of network schema and meta-path.

Definition 2. TCM Meta-path. Given a THIN $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, meta-path M is a path which is defined as $A_1 \xrightarrow{r_1} A_2 \xrightarrow{r_2} \cdots \xrightarrow{r_l} A_{l+1}$, where $r_i \in \mathcal{R}$, $A_i \in \mathcal{A}$. For the THIN $\mathcal{G}$, let $\mathcal{M} = \{M_1, M_2, \cdots, M_{|\mathcal{M}|}\}$. It is worth noting that a sequential nodes $m = \{v_1, v_2, \cdots, v_n\}$ conforming with the rule of meta-path M is a path instance of meta-path M. For example, Fig. 1(c) illustrates two meta-paths extracted from THIN in Fig. 1(a). Symptom-Syndrome-Symptom describes that a syndrome summarizes two interrelated symptoms, and Herb-Syndrome-Herb denotes that both herbs are effective for a specific syndrome.

4 Our Proposed Model: HGCL

Figure 2 illustrates the overall framework. To tackle the skewed data distribution problem, in Fig. 2(b), we design the node-level contrastive learning strategy, in which we first generate two augmented graph views by randomly corrupting

the topology and feature of the THIN. Then, a novel node-level contrastive loss tailored for THIN is proposed to maximize the agreement of node embeddings in these two views. To capture high-order semantic information in THIN, in Fig. 2(c), we propose the semantic-level contrastive learning strategy based on meta-path instances. Finally, we jointly optimize the node- and semantic-level contrastive tasks to train the parameters of GNN.

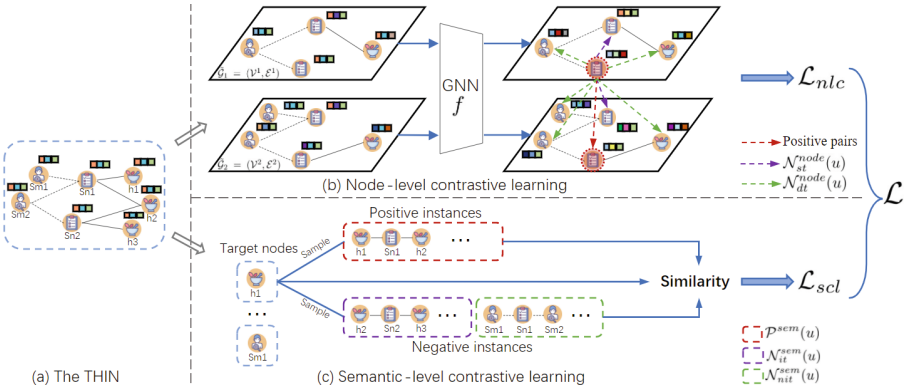

Fig. 2. An overview of our proposed HGCL framework.

4.1 Data Augmentation on THIN

Considering node-level contrastive learning relies on contrasting between node embeddings in different views, we propose to corrupt the THIN at both topology and attribute levels, thus constructing different node contexts for the model to be compared.

Node Dropping (ND) and Edge Remving (ER) randomly drop and remove a portion of nodes and edges in THIN. Formally, we sample two random vectors $\boldsymbol{N} \in \{0,1\}^{|\mathcal{V}|}$ and $\boldsymbol{E} \in \{0,1\}^{|\mathcal{E}|}$, where each element is sampled from a Bernoulli distribution $\boldsymbol{N}_i \sim Bernoulli(p_{nd})$, $\boldsymbol{E}_j \sim Bernoulli(p_{er})$ to determine whether to drop the i-th node and remove the j-th edge. Here p_{nd} and p_{er} are the probabilities for ND and ER. These operators are formulated as:

$$\boldsymbol{\mathcal{V}}' = \boldsymbol{\mathcal{V}} \odot \boldsymbol{N}, \tag{1}$$

$$\boldsymbol{\mathcal{E}}' = \boldsymbol{\mathcal{E}} \odot \boldsymbol{E}, \tag{2}$$

where $\odot$ denotes the Hadamard product (the same as below).

Feature Masking (FM) randomly masks a fraction of dimensions with zeros in node features. Formally, we first sample a random matrix $\boldsymbol{F} \in \{0,1\}^{|\mathcal{V}| \times d}$ where each element is sampled from a Bernoulli distribution $\widetilde{\boldsymbol{F}}_{ij} \sim Bernoulli(p_{fm})$ to determine whether to mask the feature matrix at position (i,j). Here p_{fm} is the

masking rate for THIN features. The generated node features matrix $\widetilde{\boldsymbol{X}}$ can be computed as:

$$\widetilde{\boldsymbol{X}} = \boldsymbol{X} \odot \boldsymbol{F}. \tag{3}$$

By applying these augmentations on topology and feature in THIN $\mathcal{G}$, we generate two correlated graph views, denoted as $\tilde{\mathcal{G}}_1 = (\mathcal{V}^1, \mathcal{E}^1)$ and $\tilde{\mathcal{G}}_2 = (\mathcal{V}^2, \mathcal{E}^2)$ then the node-level contrastive learning will be conducted on these two views.

4.2 Node-level Contrastive Learning (NCL)

Having generated two augmented graphs for the THIN, as shown in Fig. 2(b), we consider the representations of the same node in two views as the positive pairs, and any different nodes in these two views as the negative pairs. In previous graph contrastive learning, however, the negative sampling is based only on homogeneous graph structures [28,29], which ignores the heterogeneity in the THIN. It is undesirable to graft it onto our scenario directly.

According to the alignment principle [22], similar samples have similar representations, and negative samples require more differences than positive ones. To make NCL exploit the semantics of node types in THIN, we need to characterize the differences between different types of negative samples. In specific, for a given anchor node u, we define two types of negative samples as:

$$\mathcal{N}_{st}^{node}(u) = \{v \mid \phi(u) = \phi(v), u \neq v, v \in \mathcal{V}^1 \cup \mathcal{V}^2\}, \tag{4}$$

$$\mathcal{N}_{dt}^{node}(u) = \{v \mid \phi(u) \neq \phi(v), v \in \mathcal{V}^1 \cup \mathcal{V}^2\}, \tag{5}$$

where $\mathcal{N}_{st}^{node}(u)$ and $\mathcal{N}_{dt}^{node}(u)$ represent the set of negative samples of the same and different type of the anchor node u respectively, as depicted by the purple and green dash lines in Fig. 2(b). Formally, we follow SimCLR [1]/MoCo [4] and utilize the normalized temperature-scaled cross entropy loss (NT-Xent) [12,25] to maximize the agreement of positive pairs and minimize it of negative pairs:

$$\begin{aligned} \mathcal{L}_{nlc}^1(u) &= -\log \frac{\exp(s(h_u^1, h_u^2)/\tau)}{\Phi_1 + \Phi_2} \\ \Phi_1 &= \sum_{v \in \mathcal{N}_{dt}^{node}(u)} \exp(s(h_u^1, h_v)/\tau) \\ \Phi_2 &= \sum_{v \in \mathcal{N}_{st}^{node}(u)} \exp(s(h_u^1, h_v)/\tau) \end{aligned} \tag{6}$$

where $s(\cdot,\cdot)$ denotes dot similarity and τ is a temperature parameter. Here, h_u^* indicates the representation of node u in the *-th view. Similarly, we can calculate the symmetrical loss $\mathcal{L}_{nlc}^2(u)$ for view 2.

Finally, by integrating above two losses, we obtain the final NCL as follows:

$$\mathcal{L}_{nlc} = \frac{1}{2|\mathcal{V}|} \sum_{u \in \mathcal{V}} (\mathcal{L}_{nlc}^1(u) + \mathcal{L}_{nlc}^2(u)). \tag{7}$$

4.3 Semantic-Level Contrastive Learning (SCL)

Meta-path is well used to explore the semantics and extract relations among nodes [16]. Consequently, to exploit the high-order semantic information on the THIN, we generate meta-path instances to generate the positive and negative samples, which capture the TCM semantic information on the THIN and are used to conduct SCL.

Sampling Positive Meta-path Instances. Given a meta-path $M \in \mathcal{M}$ and a target node u, we define a positive meta-path instance of u as a sequential nodes $m = \{v_1, v_2, \cdots, v_n\}$ conforming with the rule of meta-path M and including the target node u. For example, in Fig. 2(c), we have a meta-path instance as $m = \{h1, sn1, h2\}$. Formally, we generate the positive meta-path samples $w.r.t.$ the target node u as

$$\mathcal{P}^{sem}(u) = \{m \mid m \in \Psi(M), u \in m, M \in \mathcal{M}\}, \tag{8}$$

where $\mathcal{M}$ is all of the pre-defined meta-paths, Ψ(M) denotes the set of all instances of M. Intuitively, sampling positive meta-path instances can capture not only local structural information, but also TCM semantic context.

Sampling Negative Meta-path Instances. To achieve fine-grained SCL and improve its semantic capturing capability, like NCL, we also define two types of negative meta-path instances of a target node u as:

$$\mathcal{N}_{it}^{sem}(u) = \{m \mid m \in \Psi(M), u \notin m, \phi(u) \in T(m), M \in \mathcal{M}\} \tag{9}$$

$$\mathcal{N}_{nit}^{sem}(u) = \{m \mid m \in \Psi(M), \phi(u) \notin T(m), M \in \mathcal{M}\} \tag{10}$$

where $T(m)$ denotes the set of all node types in the meta-path m, $\mathcal{N}_{it}^{sem}(u)$ represents the set of negative meta-path instances of the target node u, which contains the type of u, but not the u, $\mathcal{N}_{nit}^{sem}(u)$ denotes the set of meta-path instances that do not contain the u node type. These two types of negative instances are depicted by the purple and green dashed boxes in Fig. 2(c).

Given $\mathcal{N}_{it}^{sem}(u)$ and $\mathcal{N}_{nit}^{sem}(u)$, we then model the likelihood that the target node u is related to the positive samples while not being relevant to negative samples as follows:

$$\begin{aligned}
\mathcal{L}_{scl} &= -\sum_{u \in \mathcal{V}} \sum_{I^+ \in \mathcal{P}^{sem}(u)} \log \frac{\exp(\mathbf{h}_u^\top f(I^+))}{\Phi_1 + \Phi_2} \\
\Phi_1 &= \sum_{I^- \in \mathcal{N}_{nit}^{sem}(u)} \exp(\mathbf{h}_u^\top f(I^-)) \\
\Phi_2 &= \sum_{I^- \in \mathcal{N}_{it}^{sem}(u)} \exp(\mathbf{h}_u^\top f(I^-))
\end{aligned} \tag{11}$$

where I^* represents a positive(+) or negative(-) meta-path instance to the target node u, $f(\cdot)$ is a mean pooling function which generates the representation of nodes in the meta-path instance I^*.

4.4 Model Training

We jointly optimize the both of the NCL and SCL contrastive losses (cf. Eq. (7) and Eq. (11)). Specifically, the final objective function is defined as:

$$\mathcal{L} = \mathcal{L}_{nlc} + \mathcal{L}_{scl}, \tag{12}$$

Our graph encoders f are 2-layer GCN and augmentation probabilities p_{nd}, p_{er}, p_{fm} are all set to be 0.5.

5 Experiments

In this section, we first describe our experimental settings, including datasets, baseline methods and detailed configuration, then extensive experiments are conducted to validate the effectiveness of the proposed framework.

5.1 Datasets

- **TCMRel** [21] is constructed based on the corpus of TCM literature, with herb, symptom, disease included.
- **ChP** [2] is built from the Pharmacopoeia of the People's Republic of China 2015 Edition, with formula, herb, symptom, and function included.
- **LuCa** is a real-world lung cancer clinical prescription dataset, collected from our cooperative hospital, which includes symptoms, syndromes, and herbs. We conduct the prescription generation task **on this dataset**.

5.2 Baseline Methods

We compare HGCL with two categories of baselines. For supervised methods, we select GCN [9], GAT [19], and HGT [6]. DGI [20], GRACE [29], and DMGI [13] are chosen as the self-supervised baselines.

5.3 Herb Classification and Similarity Search

For herb classification task, we implement a logistic regression model on the learned embeddings in the training set, and then evaluate the test set, 10 times repeat, 8:2 split rate, to get Micro-F1 and Macro-F1. For the similarity search task, we first calculate the cosine similarity scores of the learned embedding between all pairs of herbal nodes. For each node, we then rank the nodes based on the similarity score. Finally, we compute the percentage of nodes belonging to the same class among the top 5 nodes (Sim@5).

The results reported in Table 1 show that HGCL performs the best on LuCa and ChP, even compared with several supervised models. Furthermore, HGCL performs the second best on TCMRel in Macro F1, its performance is still very competitive with that of the best baseline HGT. For supervised methods, the method designed for heterogeneous graphs, HGT, obtains better results than

Table 1. Experiment results for the herbal classification and similarity search task.

Method	TCMRel			ChP			LuCa		
	MaF1	MiF1	Sim@5	MaF1	MiF1	Sim@5	MaF1	MiF1	Sim@5
GCN	0.1128	0.1304	0.0965	0.0488	0.1351	0.0885	0.0316	0.1321	0.0438
GAT	0.1128	0.1739	0.1246	0.02	0.1622	0.1027	**0.1238**	**0.2407**	0.2189
HGT	0.1818	0.2353	0.1069	0.0394	0.1538	0.0964	0.0314	0.1286	0.0957
DGI	0.0896	0.2877	0.1204	0.2	0.3415	0.1148	0.0194	0.1124	0.0713
GRACE	0.0514	0.2055	0.2459	0.0663	0.2033	0.1121	0.0069	0.1	0.22
DMGI	0.169	0.274	0.2633	0.07	0.187	0.1944	0.0258	0.1573	0.0582
HGCL	**0.1871**	**0.3132**	**0.2694**	**0.2169**	**0.4878**	**0.3124**	0.0408	0.1667	**0.339**

ones for homogeneous graphs, e.g., GCN and GAT. Through the above analysis, we can observe that our proposed node- and semantic-level contrastive learning on THIN can indeed explore the semantic information about the efficacy of herbs and be used in the downstream tasks of herb classification and similarity search.

5.4 Prescription Generation

For generating effective prescriptions, we start from the symptoms, first find the related syndromes by similarity search, then combine these syndromes with the corresponding syndromes, and finally get the target herbal collection.

This procedure highly conforms to the principle of syndrome differentiation and treatment. The precision@k and recall@k are employed as the evaluation metrics which denote the hit ratio of top-k herbs to true herbs and the coverage of true herbs in top-k herbs, respectively. The results are reported in Table 2.

Table 2. Performance of the prescription generation task on LuCa.

Method	p@10	p@20	r@10	r@20
GCN	0.2404	0.2969	0.1222	0.3017
GAT	0.0869	0.0434	0.0438	0.0438
HGT	0.0526	0.0611	0.0299	0.0322
DGI	0.0028	0.0202	0.0013	0.0197
GRACE	0.0485	0.0453	0.0250	0.0453
DMGI	0.0088	0.0368	0.0043	0.0359
HGCL	**0.3227**	**0.3485**	**0.1703**	**0.3372**

As shown in Table 2, our HGCL performs consistently much better than all baselines. For the dmgi, it splits the heterogeneous graph into muti-view network with the help of meta-path, where it loses the intermediate nodes of meta-path and a lot of semantics in it, so it does not perform well.

5.5 Ablation Study

To better understand our proposed HGCL model, we conduct experiments on TCMRel to answer the following two questions. ($Q1$) Is two-level contrastive learning superior to single-level comparative learning? ($Q2$) Can GCN, as an encoder of THIN, capture heterogeneous information? Here, we use HGCL_{ncl} and HGCL_{scl} to denote the ablated model with NCL loss $\mathcal{L}_{nlc}$ or SCL loss $\mathcal{L}_{scl}$ being masked. HGCL_{gat} and HGCL_{hgt} switch to adopting GAT and HGT as the encoder f to learn embeddings respectively.

Table 3. Ablation study on TCMRel dataset.

Method	Macro-F1	Micro-F1
HGCL_{ncl}	0.1111	0.3151
HGCL_{scl}	0.076	0.2877
HGCL_{gat}	0.1189	0.3699
HGCL_{hgt}	0.0222	0.2192
HGCL	**0.1871**	**0.4932**

For $Q1$, comparing HGCL with HGCL_{ncl} and HGCL_{scl}, we can see two-level contrastive learning, that benefits from capturing both local node and global semantic information, clearly contribute the superior performance over two single-level comparative learning. For $Q2$, the results of HGCL_{gat} and HGCL_{hgt} suggest that although GCN is used as encoder, the proposed model well captures the structural and semantic information of the heterogeneity in THIN via NCL and SCL (Table 3).

5.6 Parameters Study

Here we analyze the dimension of the final embedding H. We report herb classification (Macro F1 and Micro F1) on the TCMRel and ChP datasets in Fig. 3. We can find that herb classification performance increases with the growth of d', which shows that the high-dimensional representations successfully learn much semantic information of TCM formula. However, when d' reaches 32, the performance decreases. This may be caused by overfitting or additional redundancies. As $d' = 32$ shows relatively good performance, we set d' as 32 in the experiments.

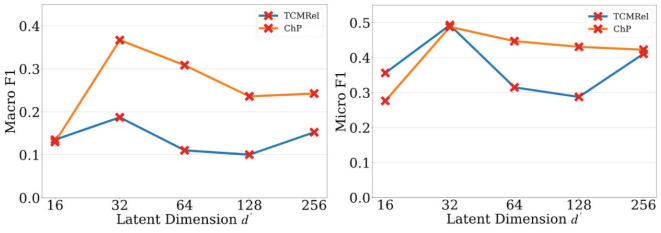

Fig. 3. Performance on various dimension of HGCL.

6 Conclusion

In this paper, we proposed a novel contrastive learning model, HGCL, to generate prescriptions following the philosophy of syndrome differentiation and treatment. For modeling rich semantics and complex relation in TCM literature and

clinical prescription, we first model the TCM clinical prescriptions as a THIN, and then corrupt it at both topology and attribute aspect to support node-level contrastive learning. By sampling meta-path instances, we also implement semantic-level contrastive learning on THIN, so as to enhance the quality of node representations for several downstream tasks such as node classification, prescription generation, etc. In experiments, HGCL achieves state-of-the-art results on three real-world datasets in the node classification, similarity search, and prescription generation tasks. The ablation study also demonstrates the rationality of the design of the HGCL's components.

References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607. PMLR (2020)
2. Chen, X., Ruan, C., Zhang, Y., Chen, H.: Heterogeneous information network based clustering for categorizations of traditional Chinese medicine formula. In: BIBM, pp. 839–846. IEEE (2018)
3. Guo, L., Wang, Y.Y.: Study thoughts on complex phenomena in syndrome of Chinese medicine. Chin. J. Basic Med. Tradit. Chin. Med. **10**(2), 3–12 (2004)
4. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: CVPR, pp. 9729–9738 (2020)
5. Hjelm, R.D., et al.: Learning deep representations by mutual information estimation and maximization. arXiv preprint arXiv:1808.06670 (2018)
6. Hu, Z., Dong, Y., Wang, K., Sun, Y.: Heterogeneous graph transformer. In: 2020 Proceedings of The Web Conference, pp. 2704–2710 (2020)
7. Jiang, M., et al.: Syndrome differentiation in modern research of traditional Chinese medicine. J. Ethnopharmacol. **140**(3), 634–642 (2012)
8. Jin, Y., Zhang, W., He, X., Wang, X., Wang, X.: Syndrome-aware herb recommendation with multi-graph convolution network. In: ICDE, pp. 145–156. IEEE (2020)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: 5th ICLR (2017). arXiv preprint arXiv:1609.02907
10. Lee, D., Xu, H., Liu, H., Miao, Y.: Cognitive modelling of Chinese herbal medicine's effect on breast cancer. Health Inf. Sci. Syst. **7**(1) (2019). Article number: 20. https://doi.org/10.1007/s13755-019-0083-3
11. Li, C., et al.: Herb-know: knowledge enhanced prescription generation for traditional Chinese medicine. In: BIBM, pp. 1560–1567. IEEE (2020)
12. van den Oord, A., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
13. Park, C., Kim, D., Han, J., Yu, H.: Unsupervised attributed multiplex network embedding. In: AAAI, vol. 34, pp. 5371–5378 (2020)
14. Ruan, C., Ma, J., Wang, Y., Zhang, Y., Yang, Y., Kraus, S.: Discovering regularities from traditional Chinese medicine prescriptions via bipartite embedding model. In: IJCAI, pp. 3346–3352 (2019)
15. Ruan, C., Wang, Y., Zhang, Y., Yang, Y.: Exploring regularity in traditional Chinese medicine clinical data using heterogeneous weighted networks embedding. In: Li, G., Yang, J., Gama, J., Natwichai, J., Tong, Y. (eds.) DASFAA 2019. LNCS, vol. 11448, pp. 310–313. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-18590-9_35

16. Shi, C., Li, Y., Zhang, J., Sun, Y., Philip, S.Y.: A survey of heterogeneous information network analysis. IEEE Trans. Knowl. Data Eng. **29**(1), 17–37 (2016)
17. Sun, F.Y., Hoffmann, J., Verma, V., Tang, J.: InfoGraph: unsupervised and semi-supervised graph-level representation learning via mutual information maximization. arXiv preprint arXiv:1908.01000 (2019)
18. Vaswani, A., et al.: Attention is all you need. In: NeurIPS, pp. 5998–6008 (2017)
19. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
20. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. In: ICLR (Poster) (2019)
21. Wan, H., et al.: Extracting relations from traditional Chinese medicine literature via heterogeneous entity networks. JAMIA **23**(2), 356–365 (2016)
22. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: ICML, pp. 9929–9939. PMLR (2020)
23. Wang, Z., Poon, J., Poon, S.: TCM translator: a sequence generation approach for prescribing herbal medicines. In: BIBM, pp. 2474–2480. IEEE (2019)
24. Wu, Y., et al.: A hybrid-scales graph contrastive learning framework for discovering regularities in traditional Chinese medicine formula. In: BIBM, pp. 1104–1111 (2021)
25. Wu, Z., et al.: Unsupervised feature learning via non-parametric instance discrimination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3733–3742 (2018)
26. Yao, L., et al.: Discovering treatment pattern in traditional Chinese medicine clinical cases by exploiting supervised topic model and domain knowledge. J. Biomed. Inform. **58**, 260–267 (2015)
27. Yao, L., Zhang, Y., Wei, B., Zhang, W., Jin, Z.: A topic modeling approach for traditional Chinese medicine prescriptions. IEEE Trans. Knowl. Data Eng. **30**(6), 1007–1021 (2018)
28. Zhu, Y., Xu, Y., Liu, Q., Wu, S.: An empirical study of graph contrastive learning (2021). https://doi.org/10.48550/ARXIV.2109.01116. arXiv:2109.01116
29. Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., Wang, L.: Deep graph contrastive representation learning. arXiv preprint arXiv:2006.04131 (2020)