Machine learning tasks and representations for heterogeneous information networks

Fang, Y.

Publication date
2023

6 Conclusions

In this chapter we summarize the thesis by formulating answers to the research questions formulated in Section 1.1. After that we identify some future research directions.

6.1 Results

RQ1 How to learn the representation of dynamic heterogeneous information networks (HINs)?

This question is answered in Chapter 2, where we introduce a dynamic representation model named M-DHIN. We regard a dynamic HIN as a sequence of snapshots with different time stamps. At the initial time stamp, M-DHIN adopts a metagraph-based complex embedding mechanism to learn the initial HIN representation. Then we build a change dataset that records the evolution process of the dynamic HIN. We only train the change set to avoid training the whole network, and in this way we make M-DHIN more scalable. Then we introduce an LSTM-based deep auto-encoder to predict future structure.

In experiments, M-DHIN significantly and consistently outperforms state-of-the-art models on real-life datasets on six downstream tasks. This confirms M-DHIN's ability to capture the features of dynamic HINs. We also conduct an ablation analysis to evaluate the effectiveness of different modules of M-DHIN. And lastly, we conduct a parameter analysis to evaluate the impact of parameters, i.e., the choice of dimension and the negative ratio when sampling.

RQ2 How to pre-train HINs?

We have answered this question in Chapter 3, where we introduce a self-supervision pre-training and fine-tuning framework named PF-HIN. During a pre-training stage, we first generate input sequences using rank-guided heterogeneous walks, and then group them into mini-sequences based on their types. We design two self-supervised pre-training tasks, i.e., masked node modeling (MNM) and adjacent node prediction (ANP). The MNM task captures the node-level structure while ANP captures edge-level structure. We adopt bi-directional transformer layers to realize the pre-train tasks. During the fine-tuning stage, four downstream tasks are chosen, i.e., link prediction, similarity search, node classification, and node clustering.

PF-HIN outperforms state-of-the-art models on the above tasks on four real-life datasets. The results confirm PF-HIN's ability to pre-train a large-scale unlabeled dataset and then quickly adapt to other datasets with different domains and different downstream tasks. We also conduct an ablation analysis to evaluate the effectiveness of different modules of PF-HIN. And, finally, we conduct a parameter analysis to evaluate the impact of parameters, i.e., the maximum length of the input sequence and the dimension of the node embedding.

RQ3 How to learn the representation of HINs in a few-shot setting?

Chapter 4 answered the above question. We propose a meta-learning framework based model named META-HIN. META-HIN is applicable to three tasks (i.e., node classification, link prediction, and anomaly detection) across different HINs. A structure module, a heterogeneous GNN module, and a GAN-based contrastive module are part of META-HIN. They allow us to make effective use of structural, heterogeneous, and unsupervised information in a network, respectively. For the link prediction and anomaly detection tasks, we choose a contrastive loss to train the meta-learner. Different tasks have different influences on the meta-learner, so we introduce a self-attention mechanism to assign different weights to different tasks.

In experiments, META-HIN consistently and significantly outperforms state-of-the-art methods on every task across datasets, which demonstrates META-HIN's ability to deal with a few-shot setting. We conduct an ablation analysis to evaluate the effect of different modules and strategies we adopt. The results confirm the effectiveness of META-HIN's design. META-HIN is able to handle both a single-graph scenario and multiple-graph scenarios. In a multi-graph scenario, aside from graphs from the same distribution, graphs from other domains can also be processed. Additionally, META-HIN is able to handle three tasks in a general framework, with each having a specific training loss. We also conduct parameter sensitivity analyses to demonstrate the stability of our model. The parameters chosen are the maximum number of nodes of a subgraph and the number of shots used for meta-learning.

RQ4 How to make use of the textual information when learning the representation of HINs?

We have answered the above question in Chapter 5, where we propose a novel prompt learning model, P-HIN, which allows one to employ textual information and handle the few-shot problems, simultaneously. It consists of a text encoder and a graph encoder. The generated textual embeddings and node embeddings are then aligned by a contrastive learning mechanism. We introduce masked node modeling and edge reconstruction tasks to make use of node-level and edge-level self-supervised information. Learnable continuous texts instead of manually-designed prompts are adopted to facilitate the transfer of the pre-trained model. We also introduce a residual connection is introduced to leverage the contextual information in graph to prompt the text model.

In experiments, P-HIN consistently and significantly outperforms state-of-the-art alternatives on every dataset and on all downstream tasks, which demonstrates the advantages of P-HIN. The results confirm that P-HIN is a powerful few-shot learner for textual HINs. We then conduct an ablation analysis to evaluate the effectiveness of different modules of P-HIN. We also conduct a parameter analysis to evaluate the impact of parameters, i.e., the number of the text tokens, and the number of shots used for training.

6.2 Future work

As with any research, this thesis has some limitations, and we believe that more questions have been generated than answered by our results. In addition to the suggestions for future work that we provided at the end of each of the research chapters, here we provide several general future research directions.

6.2.1 Model compression

For training of large-scale HINs, in addition to the cost of training time, the cost of space occupied by the model cannot be ignored. For example, just considering the parameters of the node embedding in the network representation learning model, the 200 dimensional embedding vector of 100 million nodes occupies nearly 200G of memory. If complex heterogeneous information is considered, or more complex neural network-based models are used, the parameter size of the model will be further increased. How can we effectively use the relationship between node representations and the long-tail distribution of nodes to compress the network? And how do such compressions affect the application of network representation learning in practical scenarios.

6.2.2 Multi-modal heterogeneous information network attributes

There are various attribute features contained in HINs. In this thesis, we only considered textual attributes as auxiliary information. However, a node may also be connected to other multi-modal attributes, such as images and videos. For example, in a social network, aside from textual introductions, a user node may have a photo as an attribute. Such unstructured information needs image or video encoders to be processed. How to leverage multi-modal attributes to further facilitate the representation learning of HINs remains to be addressed in future.

6.2.3 Application-oriented heterogeneous information network representation learning

The training objective of existing network representation learning models generally focuses on the ability to model and reconstruct the network structure, while ignoring the effect of the learned representation vectors in more practical application scenarios, e.g., recommendation, academic search, financial search, product search, and social media search. Therefore, how to combine the network representation learning model with specific application scenarios and improve the corresponding effect of node representations in specific applications remains to be an important challenge.

6.2.4 Interpretability of heterogeneous information network representation learning

As HINs incorporate a wealth of information, node representations based on semantic mining methods such as metapaths have strong explanatory properties. For example, a product recommendation engine based on HINs can give clear reasons for its results based on the attention weights of metapaths. This advantage has not been well analyzed or elaborated yet in many tasks. Similarly, we are optimistic that the possibility of misdiagnosis can be reduced if the underlying HIN-based diagnostic reasons are given for disease diagnosis. It is still a challenge how we can best enhance the interpretability of HIN representation learning.