Embedding Representation of Academic Heterogeneous Information Networks Based on Federated Learning

Junfu Wang\textsuperscript{1}, Yawen Li\textsuperscript{2*}, Meiyu Liang\textsuperscript{1}, Ang Li\textsuperscript{1}

\textsuperscript{1}Beijing Key Laboratory of Intelligent Communication Software and Multimedia, School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876
\textsuperscript{2}School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing

Abstract: Existing studies for representation learning of homogeneous information networks cannot be applicable to academic heterogeneous information networks (HINs) with multi-type nodes and multi-relations because of the lack of ability to issue heterogeneity. Meanwhile, due to the closeness and blocking of businesses among different enterprises, there is a serious phenomenon of data islands. To solve the above challenges, we proposed an academic heterogeneous information network embedding representation learning method based on federated learning (FedAHE), which utilizes node attention and meta path attention mechanism to learn low-dimensional, dense and real-valued vector representations while preserving the rich topological information and meta-path-based semantic information of nodes in network. Moreover, we combined federated learning with the representation learning of academic HINs and put forward a federal training mechanism based on dynamic weighted aggregation of parameters (FedDWA) to optimize the node embeddings of HINs. Through sufficient experiments, the efficiency, accuracy and feasibility of our proposed framework are demonstrated.

Keywords: Embedding representation, Heterogeneous information network, Federated learning

1 Introduction

Heterogeneous information network (HIN) \cite{1}\cite{2} is a directed graph with multi-type nodes and relationships. And network embedding\cite{3}, an effective method to represent large-scale networks\cite{4}, can well preserve the proximity of rich semantic information in original networks. The embedding results of HIN have also been proved to be useful as feature input and widely used in various graph analysis tasks, especially clustering\cite{5}, classification\cite{6}, link prediction\cite{7} and recommendation\cite{8}\cite{9}. Many researchers explore methods to combine the advantages of HIN with network embedding, but there still exists some unsolved challenges so far.

First, existing HIN embedding methods usually assume that the meta path weights of all nodes are the same, which cannot capture the individualized feature of nodes connected by path instances, leading to the insufficiency of degree of proximity preservation of network topology. In additional, while measuring node similarity, most of the previous methods only count the number of path instances, which lead to the ignorance of semantic differences between path instances.

At the same time, due to the competition and monopoly between industries, the business of different enterprises is limited by certain commercial factors\cite{10} and it is difficult to fully share data, which leads to a serious data island phenomenon\cite{11}. To solve the above problems, Google firstly proposed theory related to federated learning\cite{12}. As an emerging paradigm of machine learning, it provides a novel solution for user data sharing, enabling users to obtain a more optimized model without the need for local original data, so that "data does not move and model moves".

To tackle above challenges, we proposed an attention-based representation model for academic HIN embedding that can simultaneously preserve the rich node topology and meta-path-based semantic information in HINs. We combined federated learning with embedded representation learning of HINs composed of scientific research teams and focus on optimization method of representation learning model based on federated training mechanism. The main contributions of this paper are summarized as follows:

1. We constructed an academic heterogeneous information network to address the issue that large variety of scientific and technological data with complex structure and hidden semantics cannot be accurately represented by traditional homogeneous information network.

2. We proposed an academic heterogeneous information network embedding representation learning method based on federated learning (FedAHE), which utilizes node attention and meta path attention to preserve both the topological and meta-path-based semantic information of nodes in HINs.

3. We combined federated learning with the embedded representation learning of HINs to break the data islands, and designed a federal training mechanism based on dynamic weighted aggregation of parameters (FedDWA) to optimize the representation learning method of HINs.

2 Related Work

\textbf{Heterogeneous Information Network Embedding} As an effective paradigm for modeling complex relationships

\footnotesize{*Corresponding author: Yawen Li (warmly0716@126.com).}
and objects[13], heterogeneous information network has attracted extensive research in emerging research directions. Some scholars try to combine the representation method of homogeneous information network with the characteristics of heterogeneous information network. For example, GraphGan[14] et al. used the minimax game finally fits the true nodes distribution and obtains the vector representation of network nodes. This method can also be used for the representation of heterogeneous networks while ignoring the extra properties of nodes in heterogeneous networks, but the performance is mediocre. MeiLian[15] et al. proposed a vector element path model, which embeds different types of nodes into different spaces to deepen the differences between these multi-type nodes, and further proposed an improved Metapath2vec++ model. Liu[16] et al. combined various types of objects with their relationships under a unified framework to calculate the heterogeneous reachability probability using a meta-path-based outlier identification technique in the representation learning of HINs.

**Federated Learning** Traditional information network representation methods[17] usually construct high-dimensional sparse matrices, which require high computing and storage costs, and data requires centralized training and lacks flexibility[18]. Based on this, federated learning emerges as the times require. Most studies use secure multi-party computation[19], homomorphic encryption[20], and differential privacy[21] to improve security. The aggregation mechanism of parameters on federated learning server has also received extensive attention. The two primary categories of these works are exponential moving average based algorithm and federated average based algorithm. Taking an average aggregated weight of all trained weights in worker nodes participated in federated is what the original federated averaging algorithm[22] aims at. This means faster nodes must have to wait for the slower ones. In response to this problem, Li[23] dynamically adjusts local training rounds on various performance working nodes by a parameter server. However, this method of directly weighting the corresponding positions of the working node weights reduces the convergence[24]. Moreover, the exponential moving average aggregation update strategy[25] has high efficiency in the case of high communication delay and heterogeneity since it can be weighted and summed with the weight reserved by the parameter server with a fixed weight.

**3 The Proposed Method**

Our model is designed as a central parameter server and a set of local clients. Figure 1 shows the architecture of our framework.

![Figure 1]({filename})

**Figure 1** Federated Academic HIN Embedding Model Architecture (FedAHE)

**3.1 Client Update**

We designed the local HIN embedding model with two layers, it is shown in Figure 2.

![Figure 2]({filename})

**Figure 2** Local HIN Embedding Model

**3.1.1 Node Level Attention Layer**

We learn the embedding representation of node structural features based on adjacent vector $A^T_i$ and a multi-layer perceptron (MLP), parameterized by $W^T_i$, is applied to transfer the high-dimensional sparse adjacent vector into a d-dimensional space.

For the neighbors $V_j$ of node $V_i$, $V_j$ will be assigned a larger node-level attention weight if it has more similar structure to the node $V_i$, here cosine similarity is applied to measure the similarity of structural features between $V_i$ and $V_j$ and the node-level attention coefficient $c^\pi_{ij}$ can be computed as:

$$s^\pi_{ij} = \frac{(w^T_i A^T_i)^T w^T_j A^T_j}{\|w^T_i A^T_i\| \|w^T_j A^T_j\|}$$  \hspace{1cm} (1)

$$c^\pi_{ij} = \frac{\exp(s^\pi_{ij})}{\sum_{k\in N_i} \exp(s^\pi_{ik})}$$ \hspace{1cm} (2)

where $W^T_i$ is the parameter of MLP which denotes the transformation matrix for meta path $\pi$, $s^\pi_{ij}$ denotes the meta-path-based cosine similarity of structural feature between $V_i$ and $V_j$, $c^\pi_{ij}$ is the node-level attention
coefficients of node \( V_i \) and \( N_i \) denotes the collections of neighbors connecting to node \( V_i \).

We learn the aggregated structural feature embedding representation \( e^N_{N(i)} \) of nodes \( V_i \) by uniformly sampling nodes in the set of its neighbor nodes:

\[
e^N_{N(i)} = \sigma \left( \sum_{j \in N_i} c^N_{ij} W^N \cdot A^N_j \right)
\]  

(3)

where \( e^N_{N(i)} \) is the aggregated embedding of node \( V_i \), \( N_i \) is the neighborhood nodes connected to node \( V_i \) by meta path \( \pi \) and \( \sigma() \) is the activation function.

Finally, by combing the structural feature representation of node \( V_i \) itself and its neighbor nodes connected by path instances, the meta path based embedding \( e^\pi \) of \( V_i \) can be computed as follow:

\[
e^\pi = W^\pi_c [e^N_{N(i)} \cdot W^\pi] \]

(4)

where \( e^N_{N(i)} \) is the aggregated structural feature embedding of node \( V_i \), \( W^\pi \) is the structural feature representation of nodes \( V_i \), \([\cdot]\) indicates vector concatenation operation, and \( W^\pi_c \) denotes the dimensional transformation weight matrix.

### 3.1.2 Meta-path Level Attention Layer

We design the meta-path level attention mechanism based on a meta-path preference vector \( P_i \in \mathbb{R}^{1 \times k} \). Thereby, a comprehensive embedding representation of nodes in HINs can contain both structural and semantic features.

Similarly, a more similarity between the structural feature embedding \( e^\pi_i \) and the preference vector \( P_i \) will lead to larger attention coefficients. We first apply a line transformation parameterized by \( W_\rho \) to transform the structural feature embedding \( e^\pi_i \) into k-dimension space and denote the transformed embedding as \( e^\pi_i ' \), thus the meta path attention coefficients of node \( V_i \) can be denoted as:

\[
\delta^\pi_i = \frac{p_i \cdot e^\pi_i '}{||p_i|| \cdot ||e^\pi_i '||}
\]  

(5)

\[
\delta^\pi_i = \frac{\exp(\delta^\pi_i ')}{\sum_{m=1}^{M} \exp(\delta^\pi_m ')}
\]  

(6)

where \( p_i \) is the preference vector, \( \delta^\pi_i ' \) is the cosine similarity between the transformed structural feature embedding \( e^\pi_i ' \) and preference vector \( p_i \) based on meta path, \( || \cdot || \) denotes vector L2 normalization and \( \delta^\pi_i ' \) is the meta path attention coefficient of node \( V_i \).

Finally, given a HIN in which \( M \) meta paths exist, the ultimate comprehensive embedding representation \( e_i \) of node \( V_i \) which well preserves both rich semantic and structural feature in HIN can be denoted as:

\[
e_i = \sum_{\pi=1}^{M} \delta^\pi e^\pi_i
\]  

(7)

The objective of client updates is to minimize the cross-entropy loss between the predictions and ground-truth by back propagation and stochastic gradient descent.

### 3.2 Server Aggregation

We maintain two records on the federated parameter server. One is the latest weight parameters of all clients, which guarantees that while updating clients, all out-of-date weights of federated clients can be totally eliminated, and is denoted as \( S_w = (w^1, w^2, ..., w^n) \). Another is the record of the latest parameter version number of all worker nodes which is denoted as \( S_v = (v^1, v^2, ..., v^m) \).

Suppose there are \( n \) working nodes each identified by a unique \( id \) number in federation, the version number is initialized to 1 and increments by 1 after each update occurs. When each time uploading model parameters, each node sends its own node id, current weight \( w \) and version number \( v_{\text{latest}} \) to the parameter server and only the record \( S_w[id] \) of node with \( id \) is updated. The aggregated parameter weights \( w_{\text{latest}} \) of working node with \( id \) are updated as follows:

\[
w_{\text{latest}} = \sum^n_{i=1} (v_{\text{latest}} - S_v[id] + 1)^{-a} \cdot S_w[id]
\]  

(8)

where \( v_{\text{latest}} \) is the latest version provided by working node with \( id \), and \( S_v[id] \) is the corresponding version record on the parameter server.

Moreover, the normalization aggregated parameter weights \( w_{\text{latest normal}} \) is denoted as:

\[
w_{\text{latest normal}} = \frac{w_{\text{latest}}}{\sum^n_{i=1} (v_{\text{latest}} - S_v[id] + 1)^{-a}}
\]  

(9)

Each time the aggregation is completed, it is checked whether the version of all nodes is behind the latest version, and whether it exceeds the threshold limit of the version gap. When the version gap limit is exceeded, the parameter server will send the latest weight to all nodes, otherwise the updated aggregation parameters will be only sent to this client with \( id = id \).

### 4 Experiments

We conduct experiments on three datasets(DBLP[26], Aminer[27] and ACM[28]) oriented scientific research, which are typical representatives of academic HINs. For experimental implementation, we apply Adam optimizer with initial learning rate as 0.001 for local HIN Embedding model optimization, and respectively set the client local batch size and epoch to 256 and 1. Moreover, the embedding vector dimension is 128 and the number of clients chosen is 3.

We compare to Node2Vec[29] and LINE[30] which loss network heterogeneity since regarding all nodes and relationships as the same type. Also, we compare to Metapath2vec[31], HIN2vec[32] and GAT[33] for providing evidence about the superiority in heterogeneity exploring of our proposed attention mechanism. In this paper we only compare FedAHE with previously mentioned non-federated-based baseline, leaving the comparison with federated methods for future work.
4.1 Node Classification

We choose Macro-F1 and Micro-F1 as the evaluation metric for node classification, and the results of node classification in datasets mentioned above are shown in Table 1. According to the first two rows of the table, the metrics of FedAHE is higher than that of all these homogeneous network embedding methods (Node2Vec and LINE), which indicates that FedAHE has a more efficient ability and better performance to correctly capture rich heterogeneous information in HINs. For the next two rows representing methods (MetaPath2Vec and HIN2Vec) that only consider meta-paths, but ignore the preference weights of these different meta-paths, FedAHE almost performs best. Furthermore, FedAHE also performs better than GAT which takes the weights of different meta path in account, demonstrating the efficiency, correctness and feasibility of proposed node level attention and meta-path attention mechanism.

| Dataset | DBLP Aminer ACM |
|---------|----------------|
| Node2Vec | Micro-F1 Micro-F1 Micro-F1 Micro-F1 |
| LINE | 0.836 0.837 0.841 0.826 0.874 0.865 |
| MetaPath2Vec | 0.841 0.843 0.899 0.881 0.913 0.896 |
| HIN2Vec | 0.783 0.776 0.923 0.878 0.921 0.887 |
| GAT | 0.850 0.852 0.927 0.902 0.927 0.904 |
| FedAHE (ours) | 0.867 0.859 0.932 0.901 0.931 0.913 |

4.2 Parameter Aggregation

We also compare our proposed FedDWA with FedAvg and AFed, two typical federated parameter aggregation algorithms. Figure 3 shows the rate of gradient descent with different parameter aggregation methods on DBLP. We can find that the gradient decreases faster on the same communication round, and the model also converges faster than the traditional methods, demonstrating that FedDWA works more effectively by dynamically aggregating and updating parameter weights uploaded by federated participants.

Figure 3 The Rate of Gradient Descent with Different Parameter Aggregation Methods on DBLP.

5 Conclusions

We proposed a federated embedding representation learning method of academic heterogeneous information network (FedAHE) based meta path attention and node attention mechanism, which can not only learn low-dimensional, dense vector representations of nodes in HINs, but also preserve the rich network topological information and meta-path-based semantic information of nodes. In additional, we combined federated learning with embedding learning of HINs, proposed a federal training mechanism based on dynamical weighted aggregation of parameters (FedDWA) to optimize the local HINs embedding learning. Through experiments, the proposed method has been proved to have much better correctness, feasibility and efficiency than the traditional methods.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No.62192784, No.62172056, No.61877006).

References

[1] Li A, Du. J, Kou F, et al. Scientific and Technological Information Oriented Semantic-adversarial and Media-adversarial Cross-media Retrieval[J]. arXiv preprint arXiv:2203.08615, 2022.
[2] Zhou H, Zhao Z Y, Li C. Survey on Representation Learning Methods Oriented to Heterogeneous Information Network [J]. Journal of Frontiers of Computer Science and Technology, 2019, 13(7): 1081-1093.
[3] H. Peng et al., ”Lime: Low-Cost and Incremental Learning for Dynamic Heterogeneous Information Networks,” IEEE Transactions on Computers[J], vol. 71, no. 3, pp. 628-642, 1 March 2022, doi: 10.1109/TC.2021.3057082.
[4] He Y, Song Y, Li J, et al. Hetspace walk: A heterogeneous space random walk for heterogeneous information network embedding[C]/Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019: 639-648.
[5] Xue Z Z, Meng X D, Li X Y, et al. Luminescent thermochromism and white-light emission of a 3d [Ag4Br6] cluster-based coordination framework with both adamantane-like node and linker[J]. Inorganic Chemistry, 2021, 60(7): 4375-4379.
[6] Guan Z, Li Y, Xue Z, et al. Federated Graph Neural Network for Cross-graph Node Classification[C]/2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems (CCIS). IEEE, 2021: 418-422.
[7] Wang Z, Chen C, Li W. Predictive network representation learning for link prediction[C]/Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval. 2017: 969-972.
[8] Xiao S, Shao Y, Li Y, et al. LECLF: recommendation via learnable edge collaborative filtering[J]. Science China Information Sciences, 2022, 65(1): 1-15.
[9] Kou F, Du J, Yang C, et al. Hashtag recommendation based on multi-features of microblogs[J]. Journal of Computer Science and Technology, 2018, 33(4): 711-726.
[10] Hunt T, Song C, Shokri R, et al. Privacy-preserving Machine Learning as a Service[J]. Proceedings on Privacy Enhancing Technologies, 2018, 2018(3): 123-142.
Proceedings of CCIS2022

[11] Isabella Yunfei Zeng, Shiqi Tan, Jianliang Xiong, Xuesong Ding, Yawen Li, Tian Wu. Estimation of real-world fuel consumption rate of light-duty vehicles based on the records reported by vehicle owners. Energies, 14(23): 7915, 2021.

[12] Yang Q, Liu Y, Chen T, et al. Federated Machine Learning: Concept and Applications[J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19.

[13] Li W, Sun J, Jia Y, et al. Variance-constrained state estimation for nonlinear complex networks with uncertain coupling strength[J]. Digital Signal Processing, 2017, 67: 107-115.

[14] Wang H, Wang J, et al. Graphgan: Graph Representation Learning with Generative Adversarial Nets[C]//Thirty-second AAAI Conference on Artificial Intelligence. 2018.

[15] Meilian L U , Danna Y E . HIN_DRL: A Random Walk Based Dynamic Network Representation Learning Method For Heterogeneous Information Networks[J]. Expert Systems with Applications, 2020, 158(2): 113427.

[16] Liu L, Wang S. Meta-path-based Outlier Detection in Heterogeneous Information Network[J]. Frontiers of Computer Science, 2020, 14(2): 388-403.

[17] Yawen Li, Isabella Yunfei Zeng, Ziheng Niu, Jiahao Shi, Ziyang Wang and Zeli Guan, Predicting vehicle fuel consumption based on multi-view deep neural network, Neurocomputing, 502:140-147, 2022.

[18] Yawen Li, Di Jiang, Rongzhong Lian, Xueyang Wu, Conghui Tan, Yi Xu, Zhiyang Su. Heterogeneous Latent Topic Discovery for Semantic Text Mining. IEEE Transactions on Knowledge and Data Engineering, 2021.

[19] Zhu H, Wang L, Wang C. Privacy-Enhanced Multi-User Quantum Private Data Query Using Partial Quantum Homomorphic Encryption[J].International Journal of Theoretical Physics, 2021, 60(6): 2090-2101.

[20] Mohassel P, Zhang Y. SecureML: A System for Scalable Privacy-Preserving Machine Learning[C]//Security & Privacy. IEEE, 2017: 19-38.

[21] Mcglinchey A, Mason O. Some Novel Aspects of The Positive Linear Observer Problem: Differential Privacy and Optimal L1 Sensitivity[J]. Journal of the Franklin Institute, 2020, 357(18): 13923-13940.

[22] Bonawitz Keith, Eichner Hubert, Grieskamp Wolfgang, et al. Towards Federated Learning at Scale: System Design[J]. arXiv Preprint arXiv:1902.01046,2019.

[23] Li Tian, SahuAnit Kumar, Zaheer Manzil, et al. Federated Optimization in Heterogeneous Networks[C]//Proceedings of Machine Learning and Systems,2020.

[24] Wang H, Yurochkin M, Sun Y, et al. Federated Learning with Matched Averaging [C]//International Conference on Learning Representations,2020.

[25] Lian Xiang-ru, Zhang Wei, Zhang Ce, et al. Asynchronous Decentralized Parallel Stochastic Gradient Descent[C]// International Conference on Machine Learning,2018: 3043-3052.

[26] Chikazawa Y, Katsurai M, Ohmukai I. Multilingual author matching across different academic databases: a case study on KAKEN, DBLP, and PubMed[J]. Scientometrics, 2021, 126(3): 2131-2127.

[27] Stergiopoulos V, Tsianaka T, Tousidou E. AMiner Citation-Data Preprocessing for Recommender Systems on Scientific Publications[C]//25th Pan-Hellenic Conference on Informatics. 2021: 23-27.

[28] Li J, Chen C, Tong H, et al. Multi-layered network embedding[C]//Proceedings of the 2018 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2018: 684-692.

[29] Shi C, Kong X, Huang Y, et al. Hetesim: A general framework for relevance measure in heterogeneous networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(10): 2479-2492.

[30] Grover A, Leskovec J. node2vec: Scalable feature learning for networks[C]//Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. 2016: 855-864.

[31] Dong Y, Chawla N V, Swami A. Metapath2vec: Scalable Representation Learning for Heterogeneous Networks[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 135-144.

[32] Fu T, Lee W C, Lei Z. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning[C]//Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 1797-1806.

[33] Brody S, Alon U, Yahav E. How attentive are graph attention networks?[J]. arXiv preprint arXiv:2105.14491, 2021.