Scientific Paper Classification Based on Graph Neural Network with Hypergraph Self-attention Mechanism

Jiashun Liu, Zhe Xue*, Ang Li

School of Computer Science, Beijing Key Laboratory of Intelligent Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: The number of scientific papers has increased rapidly in recent years. How to make good use of scientific papers for research is very important. Through the high-quality classification of scientific papers, researchers can quickly find the resource content they need from the massive scientific resources. The classification of scientific papers will effectively help researchers filter redundant information, obtain search results quickly and accurately, and improve the search quality, which is necessary for scientific resource management. This paper proposed a science-technique paper classification method based on hypergraph neural network(SPHNN). In the heterogeneous information network of scientific papers, the repeated high-order subgraphs are modeled as hyperedges composed of multiple related nodes. Then the whole heterogeneous information network is transformed into a hypergraph composed of different hyperedges. The graph convolution operation is carried out on the hypergraph structure, and the hyperedges self-attention mechanism is introduced to aggregate different types of nodes in the hypergraph, so that the final node representation can effectively maintain high-order nearest neighbor relationships and complex semantic information. Finally, by comparing with other methods, we proved that the model proposed in this paper has improved its performance.

Keywords: Hypergraph; Graph neural network; Graph attention;

1 Introduction

Graph neural network [1] has been well studied by researchers. The main methods only focus on homogeneous networks with only one kind of nodes and edges. However, many networks in the real world contain different types of nodes and edges, which constitutes heterogeneous information networks. Heterogeneous information network representation learning is usually based on meta paths [2]. We propose to build hyperedge relationships based on high-order subgraphs in the network. Compared with the original path, high-order subgraphs can better maintain the structural characteristics and semantic information of the network [3-6]. In addition, hyperedges can effectively explore the high-order adjacency relationship in the network, and distinguish different semantic roles in the subgraph, so as to obtain a more complete semantic representation of the nodes in the subgraph.

On the other hand, the relationship between nodes in heterogeneous information networks is complex, such as affiliation and high-order interaction, which are far more than the point-to-point relationship between nodes in homogeneous networks [7]. The advantage of the heterogeneous hypergraph network [8] is that the heterogeneous attributes in the network can be expressed by introducing high-order structure information and fusing hyperedge forms formed by different types of high-order subgraphs. After modeling heterogeneous information networks into hypergraph networks, graph neural networks are extended to heterogeneous hypergraph networks by using the basic properties of hypergraphs, so as to realize the node classification [9] of heterogeneous information networks. The main contributions of this paper are as follows:

With the help of high-order subgraphs, the heterogeneous information network is modeled as a heterogeneous hypergraph network, which retains the high-order nearest neighbor network information and complex semantic information in the heterogeneous information network.

A new attention mechanism for heterogeneous hypergraph networks of scientific papers is proposed, which can dynamically aggregate nodes with different semantic roles according to different importance, so that the model can obtain better node representation even in sparse large-scale heterogeneous information networks.

The experimental results show that this method has achieved good performance in science-technique paper classification task, and greatly improved the training speed.

2 Related Work

Scientific paper classification belongs to the field of text classification in natural language processing. The commonly used text feature selection methods include singular value decomposition, word frequency reverse file frequency, document topic model and so on. For example, Wei et al. [10] proposed a paper classification method based on SVD, which uses SVD to extract the features that can best represent the paper category; In the machine learning method based on Feature Engineering [11] the features obtained by SVD, TF-IDF and LDA are difficult to express the deep semantic features of this paper. With the development of deep learning technology, researchers gradually adopt the deep

*Corresponding author: Zhe Xue (xuezhe@bupt.edu.cn).

978-1-6654-7735-2/22/$31.00 ©2022 IEEE
learning model based on word vector to classify this paper. Compared with the traditional word2vec [16][17], these pre-training models consider the context and solve the problem of polysemy; On the other hand, different levels of semantic features are obtained through hierarchical learning, which provides rich feature selection for downstream tasks.

On homogeneous networks, Meng et al. [18] defined graph convolution based on spectral graph theory by introducing the Fourier transform of graphs. On this basis, Sui et al. [19] and Li et al. [20] used the Chebyshev polynomial of the eigenvalue diagonal matrix to approximate smooth the convolution kernel in the spectral domain. At the same time, in order to integrate high-order structural information when learning node representation, hypergraph representation learning has attracted attention in recent years. Kou et al. [21] proposed a hypergraph representation learning framework to capture the correlation between higher-order data by designing hyperedge convolution operations. Li et al. [22] proposed a method of training GCN on hypergraph based on the spectral theory of hypergraph, so that nodes on the same edge have similar representation. While defining hypergraph convolution, Liang et al. [23] used the attention mechanism to measure the interaction between nodes in the hyperedge to obtain a better node representation. The above work only focuses on homogeneous networks. The way to construct hypergraphs is to model the first-order nearest neighbors of nodes or K nearest neighbors based on KNN algorithm as hyperedges. For heterogeneous networks, how to effectively establish hyperedges is still a problem worth studying.

3 The Proposed Method

The proposed scientific paper classification method SPHNN includes three main modules. First, hypergraph construction is based on high-order subgraphs, and heterogeneous information networks are constructed as multiple hypergraphs. Second, the hyperedge attention mechanism is introduced into the hypergraph convolution process to aggregate nodes with different semantic roles. Finally, through the attention mechanism, the node representations with different semantics based on multiple hypergraphs are fused to obtain the final comprehensive semantic representation. The whole structure of the proposed method is shown in Figure 1.

![Figure 1 SPHNN structure](image)

3.1 Hypergraph construction based on higher-order subgraphs

Given heterogeneous information network G and subgraph structure set $M = \{M_1, M_2, ..., M_T\}$, we use the network subgraph detection algorithm to match all instances $S_M$ of M on G. Then, the network model instance $S_M$ should be determined. For a network subgraph structure $M_t \in M$, the construction is based on $M_t$ of hypergraph $H_{M_t} = (V_{M_t}, S_{M_t})$, where $S_{M_t}$ are all instances of $M_t$ on G (assuming that the matching instances are not randomly sampled), and $V_{M_t}$ are the set of $S_{M_t}$ covered nodes. Because heterogeneous networks contain rich semantic information, and a single subgraph structure contains only one kind of semantics, it is necessary to consider a variety of high-order subgraph structures to obtain a higher-quality network representation. Therefore, the heterogeneous information network G is transformed into multiple hypergraphs $\{H_{M_1}, H_{M_2}, ..., H_{M_T}\}$, each hypergraph $H_{M_t}$ can be expressed as a binary incidence matrix $H_{M_t} \in \mathbb{R}^{N \times |S_{M_t}|}$, $N$ is the number of all nodes in the heterogeneous network G, and $|S_{M_t}|$ is the number of hyperedges. Each element in $H_{M_t}$ is defined as:

$$h(v,e)=\begin{cases} 1, & \text{if } v \in e \\ 0, & \text{if } v \notin e \end{cases}$$ (1)

For a node V, when any super edge $e \in S_{M_t}$ is associated with node V, $h(v,e)=1$, otherwise it is 0. Each super edge $e \in S_{M_t}$ is assigned a weight $W_e$, and all the weights
are stored through the diagonal matrix \( W_m \in \mathbb{R}^{[SM_{i}] \times [SM_{i}]} \). The interaction between different super edges is not considered in this paper. Therefore, \( W_{m_i} \) is initialized as the identity matrix here, and all the super edges have the same weight. At the same time, the degree \( \delta(v) \) of node \( v \) and the degree \( \delta(e) \) of super edge \( e \) are stored by diagonal matrices \( D_v \in \mathbb{R}^{N_v \times N_v} \) and \( D_e \in \mathbb{R}^{[SM_{i}] \times [SM_{i}]} \) respectively. Then the degree of any node is defined as:

\[
\delta(v) = \sum_{e \in E} w_e \circ h(v,e)
\]

(2)

The number of nodes associated with any superedge \( e \), that is the degree of the superedge, is defined as:

\[
\delta(e) = \sum_{v \in V_o} h(v,e)
\]

(3)

For a given set of network subgraphs \( M \), construct different adjacency matrices \( A = \{A_{m_1}, A_{m_2}, \ldots, A_{m_T}\} \) based on network subgraphs corresponding to multiple hypergraphs. For any hypergraph \( H_m \), its corresponding adjacency matrix \( A_{m} \) is defined as:

\[
A_{m} = H_{m} W_{m} H_{m}^{T}
\]

(4)

### 3.2 Dynamic super edge attention mechanism

Graph convolution process is the process of information exchange between adjacent nodes, and the final information distribution reaches balance after several rounds of iterations. The transition probability is determined by the adjacency matrix \( A_{m} \) and degree matrix \( D_v \) and \( D_e \) of hypergraph. Here, four accumulation methods are considered: concat, mean, 1-norm, 2-norm, where concat is the feature \( \{x_{v_1}, x_{v_2}, \ldots, x_{v_m}\} \) of each node in \( e_i \), \( m \) is the number of nodes in the super edge, and the three methods can be calculated in the following ways:

\[
x_{e_i} = \begin{cases} 
\frac{1}{m} (x_{v_1} + x_{v_2} + \ldots + x_{v_m}) & \text{mean} \\
\left(\frac{1}{p} \left| x_{v_1} \right|^p + \frac{1}{p} \left| x_{v_2} \right|^p + \ldots + \frac{1}{p} \left| x_{v_m} \right|^p\right)^{\frac{1}{p}} & \text{p-norm}
\end{cases}
\]

(5)

where \( p \) is equal to 1 or 2, \( x_{v_i} \) and \( dx_{e_i} \) represent the nth column element of the node and super edge eigenvector respectively. Therefore, referring to the way of calculating attention in the homogeneous network, the super edge self-attention mechanism is introduced. For the ith node and its ith super edge, the calculation formula of \( a_{ij} \) is:

\[
a_{ij} = \frac{\exp(\text{LeakyReLU}(\alpha \left[ x_{ij} P_{m_i} \left| x_{ij} P_{m_i} \right| \right]))}{\sum_{k \neq i} \exp(\text{LeakyReLU}(\alpha \left[ x_{ij} P_{m_i} \left| x_{ij} P_{m_i} \right| \right]))}
\]

(6)

Among them, \( \text{LeakyReLU} \) is a nonlinear activation function. The weight matrix \( P_{m_i} \) is responsible for mapping the eigenvectors \( x_{ij} \) and \( x_{e_i} \) to the unified hidden space. \( \left| \right| \) represents the splicing operation, \( \alpha \) is a trainable attention vector, which is shared among different hypergraphs. \( N^{(i)} \) is the neighbor set (including \( i \)) of the node \( v_i \) in the super edge. Finally, by learning the self attention in the super edge, the interaction between different types of nodes in the super edge \( e_i \) can be described more accurately.

### 3.3 Hypergraph convolution network with integrated semantics

The representation of a single hypergraph \( H_m \) can be obtained based on the above hypergraph convolution network. For the different hypergraph \( \{H_{m_1}, H_{m_2}, \ldots, H_{m_T}\} \) constructed by the network subgraph set \( M \), the network representation \( \{Z_{M_1}, Z_{M_2}, \ldots, Z_{M_T}\} \) with specific semantics is obtained. Because different hypergraphs have different degrees of influence on the generation of the final heterogeneous network representation, the coefficients \( C_m \), calculated based on the attention mechanism are used to fuse the network representation with different semantics:

\[
C_m = \text{softmax}(\text{tanh}(W \cdot Z_m + b) \cdot a^T)
\]

(7)

Finally, the network representation of each specific semantic is fused to obtain the representation of heterogeneous network \( G \):

\[
Z = \sum_{i=1}^{q} C_m Z_{m_i}
\]

(8)

For multi-classification tasks, where \( Z \in \mathbb{R}^{N \times q} \), \( q \) is the category of nodes. Finally, the loss function of classification is defined as:

\[
L = \sum_{i \in V} \sum_{j \in V_i} Y_{ij} \ln Z_{ij}
\]

(9)

where \( V_i \) is the set of all labeled nodes, and \( V \) is the real label of the node. Under the guidance of label information, we can finally get the representation of all nodes and predict their labels based on the parameters in the back-propagation and gradient descent training model.

### 4 Experiments

#### 4.1 Datasets

In order to verify the proposed classification method of scientific papers based on Hypergraph neural network, the algorithm is compared with other related algorithms. In addition, we conducted the experiments on three datasets. The Summary is shown in the Table I

| Table I | Summary of the datasets used in experiments. |
|-------------------|-----------------------------------------|
| Number of nodes | \(18385\) \(31636\) \(13432\) |
| Number of edges  | \(85360\) \(162012\) \(90387\) |
| Category         | \(4\) \(3\) \(10\) |
### Table II Experimental results of scientific paper classification

|               | DBLP  |        |        |        |        |
|---------------|-------|--------|--------|--------|--------|
|               | Macro-F1 | Micro-F1 | Accuracy |
| DeepWalk      | 64.17  | 63.88  | 64.12  |
| ESim          | 63.01  | 62.89  | 65.11  |
| Metapath2Vec  | 63.81  | 63.49  | 64.34  |
| GCN           | 63.92  | 63.84  | 63.32  |
| GAT           | 76.29  | 76.35  | 75.33  |
| HAN           | 77.33  | 77.11  | 77.90  |
| HGNN          | 73.48  | 73.26  | 74.89  |
| DHGNN         | 78.17  | 78.95  | 78.32  |
| SPHNN         | 80.23  | 78.22  | **80.56** |

|               | DBLP  |        |        |        |        |
|---------------|-------|--------|--------|--------|--------|
|               | Macro-F1 | Micro-F1 | Accuracy |
| DeepWalk      | 64.81  | 66.26  | 67.12  |
| ESim          | 72.53  | 73.44  | 75.11  |
| Metapath2Vec  | 71.89  | 72.88  | 75.34  |
| GCN           | 72.34  | 73.72  | 76.39  |
| GAT           | 80.38  | 80.09  | 80.62  |
| HAN           | 79.73  | 79.55  | 79.90  |
| HGNN          | 80.81  | 80.69  | 80.59  |
| DHGNN         | 80.53  | 80.29  | 89.98  |
| SPHNN         | **80.91** | **80.77** | **81.61** |

According to our experimental results, our method is found to be the best, which proves that these components of the model are effective. By introducing super edge extraction, the correlation between vertex features is extracted to the greatest extent, and the node representation is related to adjacent nodes, that is, to nodes sharing edges. This can be directly applied to inductive learning without acquiring the entire graph. Then, through the attention mechanism of graph neural network, the characteristics of nodes are dynamically aggregated to make the final node representation more accurate.

### 5 Conclusions

This paper proposes a scientific paper classification model based on self-attention and hypergraph. The hypergraph structure is used to aggregate the non-labeled data in the citation relationship of the paper. The hypergraph neural network model based on self-attention realizes the adaptive weight assignment of different depths and different neighbors, which improves the learning ability of the graph neural network model. Experimental results on large-scale real data sets show that the performance of the proposed method is not only better than other traditional graph neural network models, but also better than the traditional supervised text classification model, which proves the effectiveness of the proposed method.

### Acknowledgements

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

### References

[1] Liang Xu, Junping Du, Qingping Li. Image fusion based on nonsampled contourlet transform and saliency-motivated pulse coupled neural networks. Mathematical Problems in Engineering, 2013.

[2] Shi, C., Li, Y., Yu, P.S. et al. Constrained-meta-path-based ranking in heterogeneous...
Proceedings of CCIS2022

information network. Knowl Inf Syst 49, 719–747 (2016).
[3] Jizhou Huang, Haifeng Wang, Yibo Sun, Miao Fan, Zhengjie Huang, Chunyuan Yuan, Yawen Li. HAGAMN: Heterogeneous Graph Attention Matching Network for Multilingual POI Retrieval at Baidu Maps. Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 3032-3040, 2021.
[4] Deyuan Meng, Yingmin Jia, and Junping Du. Consensus seeking via iterative learning for multi-agent systems with switching topologies and communication time-delays. International Journal of Robust and Nonlinear Control, 2016, 26(17): 3772-3790.
[5] Peng Lin, Yingmin Jia, Junping Du, Fashan Yu. Average consensus for networks of continuous-time agents with delayed information and jointly-connected topologies. 2009 American Control Conference, 2009: 3884-3889.
[6] Yawen Li, Ye Yuan, Yishu Wang, Xiang Lian, Yuilang Ma, Guoren Wang. Distributed Multimodal Path Queries. IEEE Transactions on Knowledge and Data Engineering, 34(7):3196-321, 2022.
[7] S. Bandypadhyay, K. Das and M. N. Murty, "Hypergraph Attention Isomorphism Network by Learning Line Graph Expansion," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 669-678, doi: 10.1109/BigData50022.2020.9378335.
[8] Jinxuan Li, Yawen Li, Zhe Xue. Keywords Extraction Algorithm of Financial Review Based on Dirichlet Multinomial Model. Chinese Intelligent Systems Conference, 107-116, 2020.
[9] Jianghi Lv, Yawen Li, Junping Du, Lei Shi. E-Product Recommendation Algorithm Based on Knowledge Graph and Collaborative Filtering. Chinese Intelligent Systems Conference, 38-47, 2020.
[10] Pengchao Cheng, Yawen Li, Zhe Xue. Mining Correlation Features of User Financial Behavior Based on Attention Mechanism and Dual Channel. Chinese Intelligent Systems Conference, 66-74, 2020.
[11] Ang Li, Junping Du, Feifei Kou, Zhe Xue, Xin Xu, Mingying Xu, Yang Jiang. Scientific and Technological Information Oriented Semantics-adversarial and Media-adversarial Cross-media Retrieval. arXiv preprint arXiv:2203.08615, 2022.
[12] Wenling Li, Yingmin Jia, and Junping Du. Tobit Kalman filter with time-correlated multiplicative measurement noise. IET Control Theory & Applications, 2016, 11(1): 122-128.
[13] 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.
[14] Xunpu Yuan, Yawen Li, Zhe Xue, Feifei Kou. Financial sentiment analysis based on pre-training and textcn. Chinese Intelligent Systems Conference, 48-56, 2020.
[15] Zeli Guan, Yawen Li, Zhe Xue, Yuxin Liu, Hongrui Gao, Yingxia Shao. Federated Graph Neural Network for Cross-graph Node Classification. In 2021 IEEE 7th International Conference on Cloud Computing and Intelligent Systems, 418-422, 2021.
[16] Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019.Entity-relation extraction as multi-turn question answering. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1340–1350.
[17] Wenling Li, Yingmin Jia, Junping Du. Distributed consensus extended Kalman filter: a variance-constrained approach. IET Control Theory & Applications, 11(3): 382-389, 2017.
[18] Deyuan Meng, Yingmin Jia, and Junping Du. Robust iterative learning protocols for finite-time consensus of multi-agent systems with interval uncertain topologies. International Journal of Systems Science, 2015, 46(5): 857-871.
[19] Dianbo Sui, Yubo Chen, Kang Liu, Jun Zhao, Xianyong Zeng, and Shengping Liu. 2020. Joint entity and relation extraction with set prediction networks. CoRR, abs/2011.01675.
[20] Wenling Li, Yingmin Jia, and Junping Du. Distributed extended Kalman filter with nonlinear consensus estimate. Journal of the Franklin Institute, 2017, 354(17): 7983-7995.
[21] Feifei Kou, Junping Du, Congxian Yang, Yansong Shi, Wanqiu Cui, Meiyu Liang, and Yue Geng. Hashtag recommendation based on multi-features of microblogs. Journal of Computer Science and Technology, 2018, 33(4): 711-726.
[22] Mingxing Li, Yinmin Jia, and Junping Du. LPV control with decoupling performance of 4WS vehicles under velocity-varying motion. IEEE Transactions on Control Systems Technology 2014, 22(5): 1708-1724.Bandypadhyay S, Das K, Murty M N. Line Hypergraph Convolution Network: Applying Graph Convolution for Hypergraphs[J]. 2020.
[23] Zeyu Liang, Junping Du, and Chaoyang Li. Abstructive social media text summarization using selective reinforced Seq2Seq attention model. Neurocomputing, 410 (2020): 432-440.