Heterogeneous Attributed Network Embedding with Graph Convolutional Networks

Yueyang Wang, Ziheng Duan, Binbing Liao, Fei Wu, Yueting Zhuang

{yueyangw,duanziheng,bbliao,wufei,yzhuang}@zju.edu.cn
College of Computer Science and Technology,Zhejiang University
No.38 Zheda Road, Hangzhou, Zhejiang, China,310027

Abstract

Network embedding which assigns nodes in networks to low-dimensional representations has received increasing attention in recent years. However, most existing approaches, especially the spectral-based methods, only consider the attributes in homogeneous networks. They are weak for heterogeneous attributed networks that involve different node types as well as rich node attributes and are common in real-world scenarios. In this paper, we propose HANE, a novel network embedding method based on Graph Convolutional Networks, that leverages both the heterogeneity and the node attributes to generate high-quality embeddings. The experiments on the real-world dataset show the effectiveness of our method.

Introduction

Network embedding, which learns low-dimensional representations of nodes in networks, has been proved to be highly effective for a variety of network analysis tasks, such as recommendation, classification and clustering. Most network embedding approaches mainly focus on capturing structural correlations in networks. Recently, researchers proposed some methods based on Graph Convolutional Network (GCN) to encode both the network structure and the attributes of nodes.

In real-world scenarios, the majority of networks are heterogeneous attributed networks because there are different types of objects with various features and relationships. However, these networks cannot be handled efficiently by existing GCN-based models which only consider the homogeneous networks because of two reasons. On the one hand, different types of nodes have various attributes which cannot be directly applied due to the different dimensions of node attributes. For example, in academic networks, an "author" is associated with attributes like research interests and the number of published papers, while a "paper" has attributes like title and abstract. On the other hand, the influence of distinct relationships are different, which should not be treated equally. For instance, the correlation between "paper-writtenBy-author" is intuitively stronger than "paper-citedBy-paper".

In this paper, we study the problem of network embedding for heterogeneous attributed networks. Our proposed model,

HANE, leverages not only the heterogeneity but also the rich node attributes. Specifically, we design a GCN-based method and introduce the attention mechanism to assign different importance of different types of nodes in a neighborhood. We evaluate the effectiveness of our approach on the node classification task. The experimental results show that HANE outperforms state-of-the-art methods.

Methodology

Problem Definition. Given a heterogeneous attributed network $G = (V, E, \mathcal{A})$ where V is a set of nodes representing different types of objects, E is a set of edges representing relationships between two objects and $\mathcal{A}$ denotes the attributes of objects, heterogeneous attribute network embedding is to learn low-dimensional representations in heterogeneous attributed networks.

Learning Embeddings. We design a multi-layer GCN-based method for heterogeneous attributed network embedding, named HANE. The overall view of our model is shown in Figure 1.

We start by describing the first convolutional layer. To utilize the various types of nodes with different attributes uniformly, we first transform the input attributes to a common space. For each node $v_i \in V$ with its type $t_i \in \{1, 2, \cdots, n\}$ and its attributes $\vec{f_i} \in \mathbb{R}^{F_{t_i}}$ where n is the number of node types and F_{t_i} is the dimension of attributes for node type t_i, the transform function $g(\cdot)$ is defined as follow:

$$g(\vec{f_i}) = \vec{f_i} W_{t_i} \tag{1}$$

where $W_{t_i} \in \mathbb{R}^{F_{t_i} \times D}$ is the shared transform weight matrix for each node type and D is the dimension of the common space. Then, we introduce attention mechanism to distinguish the effects of different neighbors. For node v_i and its neighbor v_j, we define a score function as follow:

$$s_{i,j} = score(\vec{f_i}, \vec{f_j}) = g(\vec{f_i})g(\vec{f_j})^T \tag{2}$$

And we apply the softmax function to calculate the normalized impact coefficient:

$$c_{i,j} = \frac{\exp s_{i,j}}{\sum_{v_k \in \mathcal{N}_i} \exp s_{i,k}} \tag{3}$$

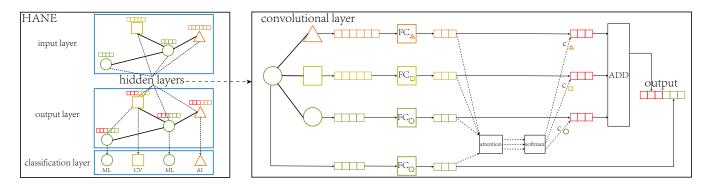where $\mathcal{N}_i$ is the neighbors of node v_i.

Figure 1: The overview of our HANE model.

After this, the output of this convolutional layer, which is the input of next hidden layer, can be written as follow:

$$\vec{f}_i^{(l+1)} = \sigma(g(\vec{f}_i^{(l)}) \oplus (\sum_{v_k \in \mathcal{N}_i} c_{i,k} g(\vec{f}_k^{(l)}))) \qquad (4)$$

where l is the number of layer, σ is the activation function (e.g. ReLU) and $\oplus$ is the concatenation operation. Finally, for the node classification task, we obtain predicted labels through a classification layer with cross-entropy loss function.

Experiments

Datasets. We derive a heterogeneous attributed network from the academic network of AMiner[1]. It consists of 28,431 nodes including 16,604 authors and 11,827 papers with 8 labels, and 122,219 edges representing 62,115 coauthor, 28,841 citation and 31,263 author-paper relationships. The text descriptions of authors and papers are treated as node attributes which are transformed into vectors by Doc2vec in the experiments.

Baseline Methods. Our method HANE, along with a degenerated variation, HANE-att (without attention mechanism), is compared with several state-of-the-art network embedding methods, including node2vec (Grover and Leskovec 2016), metapath2vec (Dong, Chawla, and Swami 2017) and GraphSAGE (Hamilton, Ying, and Leskovec 2017). In the experiments, the depth of model layers and the rule of neighbor sampling are the same as GraphSAGE.

Results and Discussion. We evaluate the performance of all the methods on the node classification task using Macro F1 (MaF1) and Micro F1 (MiF1) and set training ratio from 25% to 75%. The results are shown in Table 1. From this table, we have three observations as follows. First, HANE achieves significant improvement comparing to node2vec and metapath2vec, which demonstrates the effectiveness of incorporating node attributes in our model. Second, compared with GraphSAGE, HANE achieves $4 \sim 6$ percent point gain in terms of Macro-F1 and $3 \sim 5$ percent point gain in terms of Micro-F1. It proves that the performance can be improved by employing the heterogeneity of networks. Third, we notice that HANE shows better performance over HANE-att, which indicates that the attention

[1]https://aminer.org/aminernetwork

Table 1: Experimental Results of Node Classification.

Methods	25%		50%		75%	
	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1
node2vec	0.6146	0.8091	0.6165	0.8119	0.6188	0.8149
metapath2vec	0.6014	0.8024	0.6042	0.8013	0.6036	0.8015
GraphSAGE	0.6455	0.7918	0.6605	0.8036	0.7185	0.8502
HANE-att	0.6470	0.8129	0.7063	0.8299	0.7394	0.8594
HANE	**0.6821**	**0.8203**	**0.7204**	**0.8516**	**0.7651**	**0.8906**

mechanism can effectively distinguish the influence of different types of nodes on their neighbors. In summary, HANE achieves the best performance by effectively making full use of the heterogeneity and the attributes.

Conclusion

In this paper, we propose a novel network embedding method for heterogeneous attributed networks. Towards this end, we introduce an effective attention mechanism on graph convolutional networks to leverage not only the heterogeneity but also the node attributes. The experiments demonstrate the effectiveness and efficiency of our model.

Acknowledgments

References

Dong, Y.; Chawla, N. V.; and Swami, A. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 135–144. ACM.

Grover, A., and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864. ACM.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, 1024–1034.