속성 네트워크에서 엣지없는 네트워크 임베딩

Edgeless Network Embedding in Attributed Networks

Yong-Min Shin[1], Cong Tran[1,2], Won-Yong Shin[1]
[1]Yonsei University, [2]Dankook University

jordan3414@yonsei.ac.kr, trancong208@gmail.com, wy.shin@yonsei.ac.kr

Abstract

In this study, we present the embedding of nodes that newly enter an underlying graph without any edges by using a graph neural network (GNN), a widely used network architecture for graph data. Motivated by the fact that GNN only operates on nodes with connections to other nodes, we construct a k-nearest-neighbor graph (kNNG) used for the computation graph. Experimental results using the Cora dataset demonstrate that the embedding generated by our proposed method leads to higher performance for various downstream graph mining tasks.

Ⅰ. Introduction

Network embedding has gained its popularity as a tool to effectively solve various downstream graph mining tasks. Graph neural networks (GNNs) have been widely used for network embedding as it can naturally integrate node attributes with graph topology [1]. In real-world graphs, new nodes can enter without any connections. An example includes co-authorship networks in which a paper was written by a single author. However, GNN cannot calculate embeddings for such edgeless nodes as it inherently requires edges for the computation. In this study, we propose a new method that enables us to perform GNN for such new edgeless nodes.

Ⅱ. Methodology

We denote the set of nodes as V_o, set of edges as E_o, and the attribute matrix as X_o, for an unweighted and undirected graph $G_o = (V_o, E_o, X_o)$. We assume that a set of new nodes V_n and their attributes X_n enter G_o, without any edges to V_o. The embedding space calculated by the GNN model is expressed as $Z = GNN_\theta(G_c, X_o)$, where θ is the parameter and G_c is computation the graph used in the GNN model. In our method, we construct a k-nearest-neighbor graph (kNNG) G_{kNNG} from X_o and set $G_c = G_{kNNG}$. To train the model, we design a loss function $\mathcal{L}(Z, G_o) = \omega_{ij} cos(Z_i, Z_j)$, where the parameter ω_{ij} is 1 when there is an edge (i,j) in G_o; it is positive and proportional to the portion of common neighbors when the edge is two-hop apart; and it is negative and inversely proportional to the shortest path when not connected.

Ⅲ. Experimental results

We adopt the GraphSAGE model [1], a well-known GNN architecture, to generate embedding. We evaluate our method using the Cora dataset, which is a citation network where the nodes are papers, node attributes contain words in the abstract, and edges are formed when a citation occurs. First, we mask a portion of nodes in the underlying graph to represent V_n. Then, we construct kNNG G_{kNNG} and train the model with $\mathcal{L}$. After training, we generate links for V_n by selecting k-nearest-neighbors in V_o and run the model. We perform link prediction, node classification, and community detection to see the superiority of our method. For comparison, we consider two benchmark methods: 1) DEAL [2] and 2) inference with node attributes only (Att-Only).

Table 1: Experimental results

Method	AUC	Macro-F1	NMI
Proposed	0.8930	0.6783	0.5109
DEAL	0.8550	0.6410	0.4321
Att-Only	0.7546	0.4923	0.2213

Table 1 shows the experimental results of the various downstream tasks on V_n. We observe that the proposed method achieves higher performance compared to the two benchmark methods for all three tasks, measured in AUC for link prediction, Macro-F1 for node classification, and NMI for community detection. This is because our method exploits the high representation capability of the designed model.

ACKNOWLEDGMENT

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C3004345), by the Republic of Korea's MSIT (Ministry of Science and ICT), under the High-Potential Individuals Global Training Program (No. 2020-0-01463) supervised by the IITP (Institute of Information and Communications Technology Planning Evaluation), and by the Yonsei University, Republic of Korea Research Fund of 2021 (2021-22-0083).

REFERENCES

[1] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," In Proc. 31st Conference on Neural Information Processing Systems (NeurIPS), Long Beach, CA, December 2017, pp. 1024-1034.

[2] Hao, Y., Cao, X., Fang, Y., Xie, X. and Wang, S., "Inductive Link Prediction for Nodes Having Only Attribute Information." In Proc. 29th International Joint Conference on Artificial Intelligence (IJCAI), Virtual Event, January 2021, pp. 1209-1215.