北京郵電大學

Beijing University of Posts and Telecommunications

异质图神经网络
模型、预训练与应用

石川

shichuan@bupt.edu.cn

北京邮电大学

Outline

- Basic concepts
- Models
- Pre-training
- Applications

√ Conclusion and future work

Technique summary

- Shallow model
  - Random walk-based
    - Semantic-aware random walk, e.g., metapath2vec, HHNE
    - Metagraph-guided random walks, e.g., metagraph2vec, spacey
  - Decomposition-based
    - Subgraph based, e.g., JERec, PME, PTE, HEBE
- Deep model
  - Message passing-based
    - Semantic-aware aggregation function, e.g., HAN, HetGNN, GTN
  - Encoder-decoder-based
    - Property-preserved autoencoder, e.g., HNE, Camel, DHNE
  - Adversarial-based
    - Relation-aware GAN, e.g., HeGAN
    - Adversarial completion, e.g., MV-ACM

Technique summary

TABLE 2: Typical heterogeneous graph embedding methods.

Method	Inductive	Label	Information	Task	Technique	Characteristic
mp2vec [8]			Strcuture	Embedding	Random walk (Shallow model)	• Easy to parallelize • Two-stage training • High memory cost Complexity: $\mathcal{O}(\tau \cdot l \cdot k \cdot n_s \cdot d \cdot \lvert\mathcal{V}\rvert)$
Spacey [59]						
JUST [60]						
BHIN2vec [61]						
HHNE [62]						
mg2vec [41]						
HeRec [2]		√	Strcuture+Task	Recommendation		
PME [17]			Strcuture	Embedding	Decomposition (Shallow model)	• Easy to parallelize • Two-stage training • High memory cost Complexity: $\mathcal{O}(\lvert\mathcal{E}\rvert \cdot d)$
EOE [50]						
HEER [53]						
MNE [57]						
PTE [17]						
RHINE [63]						
HAN [15]	√	√	Structure+Attribute		Message passing (Deep model)	• End-to-End training • Encoding structures and attributes • Semantic fusion • High training cost Complexity: $\mathcal{O}(\lvert\mathcal{V}\rvert \cdot d_1 + \lvert\mathcal{R}\rvert \cdot d_2)$
MAGNN [74]	√	√				
HetSANN [75]	√	√				
HGT [76]	√	√				
HetGNN [16]	√					
GATNE [72]	√					
GTN [79]		√				
RSHN [80]		√				
RGCN [81]	√	√				
IntentGC [20]	√	√	Strcuture +Attribute+Task	Recommendation		
MEIRec [19]	√	√				
GNUD [5]	√	√				
Player2vec [95]	√	√		Identification		
AHIN2vec [96]	√	√				
Vendor2vec [97]	√	√				
HIN2vec [9]			Strcuture	Embedding	Encoder-decoder (Deep model)	• End-to-End training • Flexible goal-orientation Complexity: $\mathcal{O}(\lvert\mathcal{V}\rvert \cdot d_1 + \lvert\mathcal{E}\rvert \cdot d_2)$
DHNE [65]						
HNE [69]	√	√	Structure+Attribute			
SHNE [70]		√				
NSHE [78]						
PAHNE [44]		√	Strcuture +Attribute+Task	Identification		
Camel [93]		√				
TaPEm [94]		√				
HeGAN [18]			Strcuture	Embedding	Adversarial (Deep model)	• Robustness • High complexity Complexity: $\mathcal{O}(\lvert\mathcal{V}\rvert \cdot \lvert\mathcal{R}\rvert \cdot n_s \cdot d)$
MV-ACM [120]						
Rad-HGC [24]		√	Strcuture+Task	Malware detection		

Future directions

- Preserving HG structures and properties
  - Motif, Network schema – beyond metapath and metagraph
  - Dynamic
  - Uncertainty e.g., Gaussian distribution
- Deep graph learning on HG data
  - Over-smoothing, e.g., deep HGNN
  - Self-supervised learning
  - Pre-training, e.g., transfer ability
- Making HG embedding reliable
  - Fairness or debias, e.g., age and gender
  - Robust
  - Explainable, e.g., disentangled learning

Future directions

- Real-world applications
  - Software engineering
  - Biological system
  - Large-scale industrial scenarios
- Others
  - Non-Euclidean space embedding
    - e.g., Hyperbolic embedding
  - Heterogeneous graph structure learning
  - Connections with knowledge graph

More materials

A Survey on Heterogeneous Graph Embedding: Methods, Techniques, Applications and Sources

Xiao Wang, Deyu Bo, Chuan Shi†, *Member, IEEE,* Shaohua Fan, Yanfang Ye, *Member, IEEE,* and Philip S. Yu, *Fellow, IEEE*

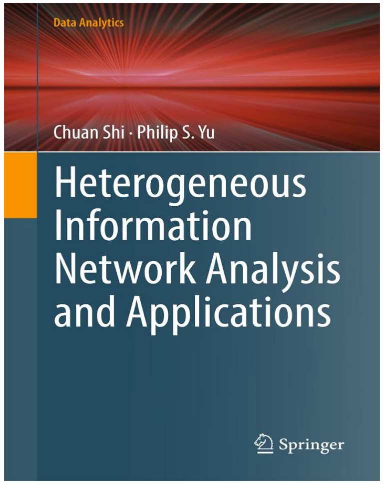
Abstract—Heterogeneous graphs (HGs) also known as heterogeneous information networks have become ubiquitous in real-world scenarios; therefore, HG embedding, which aims to learn representations in a lower-dimension space while preserving the heterogeneous structures and semantics for downstream tasks (e.g., node/graph classification, node clustering, link prediction), has drawn considerable attentions in recent years. In this survey, we perform a comprehensive review of the recent development on HG embedding methods and techniques. We first introduce the basic concepts of HG and discuss the unique challenges brought by the heterogeneity for HG embedding in comparison with homogeneous graph representation learning; and then we systemically survey and categorize the state-of-the-art HG embedding methods based on the information they used in the learning process to address the challenges posed by the HG heterogeneity. In particular, for each representative HG embedding method, we provide detailed introduction and further analyze its pros and cons; meanwhile, we also explore the transformativeness and applicability of different types of HG embedding methods in the real-world industrial environments for the first time. In addition, we further present several widely deployed systems that have demonstrated the success of HG embedding techniques in resolving real-world application problems with broader impacts. To facilitate future research and applications in this area, we also summarize the open-source code, existing graph learning platforms and benchmark datasets. Finally, we explore the additional issues and challenges of HG embedding and forecast the future research directions in this field.

A Survey of Heterogeneous Information Network Analysis

Chuan Shi, *Member, IEEE,* Yitong Li, Jiawei Zhang, Yizhou Sun, *Member, IEEE,* and Philip S. Yu, *Fellow, IEEE*

Abstract—Most real systems consist of a large number of interacting, multi-typed components, while most contemporary researches model them as homogeneous information networks, without distinguishing different types of objects and links in the networks. Recently, more and more researchers begin to consider these interconnected, multi-typed data as heterogeneous information networks, and develop structural analysis approaches by leveraging the rich semantic meaning of structural types of objects and links in the networks. Compared to widely studied homogeneous information network, the heterogeneous information network contains richer structure and semantic information, which provides plenty of opportunities as well as a lot of challenges for data mining. In this paper, we provide a survey of heterogeneous information network analysis. We will introduce basic concepts of heterogeneous information network analysis, examine its developments on different data mining tasks, discuss some advanced topics, and point out some future research directions.

Index Terms—Heterogeneous information network, data mining, semi-structural data, meta path



Heterogeneous Graph Representation Learning and Applications

Chuan Shi, Xiao Wang, Philip S. Yu

To be published by Springer(English) and HuaZhang(Chinese)

Graph Neural Network: Foundations and Frontiers

GAMMA Lab

Being processed

- More materials in my webpage: www.shichuan.org

Thanks！

GAMMA Lab:

Xiao Wang, Cheng Yang, Bai Wang, Binbin Hu, Houye Ji, Yuanfu Lu, Yugang Ji, Shaohua Fan, Jianan Zhao, Deyu Bo, Ruijia Wang, Meiqi Zhu, Tianchi Yang, Nian Liu, Xunqiang Jiang, Chen Li

Collaborators:

Philip Yu, Yuan Fang, Fanny Ye, Xin Zhao, Peng Cui, Guojie Song, Zhiyuan Liu, Zhiqiang Zhang, Ruobin Xie, Junxiong Zhu

图数据挖掘和机器学习

▶ 扫码关注我们 ◀

www.shichuan.org