

论文分享

Node Feature Extraction by Self-Supervised Multi-scale Neighborhood Prediction

王云帆

2022 年 4 月 22 日

① Introduction

② Method

③ 理论分析

④ 实验

① Introduction

② Method

③ 理论分析

④ 实验

Introduction

- 自监督学习 SSL 被应该在了计算机视觉, NLP 等诸多方面
- 图自监督学习的 common idea 是 explore the correlated information provided by the numerical node features and graph topology
- 问题: 如何从 raw data (文本, 图片, 音频) 中获得向量化的数字特征
 - 目前可以通过 bag of the words, word2vec, BERT。但这些方法都是 graph-agnostic (图不可知的), 没有在最开始就使用 graph 信息。“none of these works leverage graph information for actual self-supervision”
 - 只能达到次优解

Introduction

- 作者提出 GIANT，一个自监督学习框架，将 raw node attribute 作为输入，生成 numerical node feature
- 为了将图的拓扑信息整合进入语言模型，如 BERT 等，提出了一个全新的自监督学习任务——邻居预测。对同构图和异构图均适用
- 将邻居预测任务建模成为一个多标签学习任务 (XMC)，每个节点的邻居相当于一个二进制的多标签。
- 通过这种多标签预测，对 BERT 模型进行微调
- 这种模型，不仅借用了 XMC 问题的先进方法，解决了 graph-agnostic 的特征提取；还对 XMC 问题进行了理论分析，研究其对于图自监督学习的作用

Contributions

- 明确了 graph-agnostic 特征提取问题，提出了 GIANT 自监督框架里解决这个问题
- 提出邻居预测任务，可以用它来 fine-tune BERT model。不同于链接预测，邻居预测可以解决异构图问题。
- 将邻居预测建模为 XMC 问题，使用 XR-transformer 解决。
- 实验表明 GIANT 提升了基于 GNN 下游任务。(相当于自监督的 GIANT 为下游 GNN 初始化了特征)
- 理论分析了 XR-transformer 的对邻居预测问题的优势

① Introduction

② Method

③ 理论分析

④ 实验

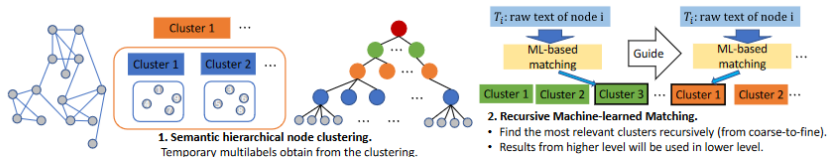
problem formulation

- Graph $G = (V, E)$, V for nodes, E for edges. Adjacency matrix $A \in \{0, 1\}^{n \times n}$
- For each node i , T_i is the raw text
- 语言模型训练一个 encoder Φ , map T_i to numerical node feature $X_i \in R^d$
- 邻接矩阵的每一行 A_i 就是节点 i 的邻居们，可以被视为目标多标签。那么我们就有 $L=n$ 个标签

XR-transformer

- XR-transformer is the advanced approach for XMC problem.
- For basic one-versus-all (OVA) model
$$f(T, l) = w_l^T \Phi(T_l), \text{ where } l \in [L]$$
- 对于这种 OVA 模型，其中 Φ 的选择可以是词袋模型，TFIDF，BERT 模型。对图形数据，可以是 CNN。格局不同数据进行调整
- 对于邻居预测问题，一个图可能有百万个节点。XR-transformer 的层次化标签聚类思路很合适。
 - 标签特征学习：Positive Instance Feature Aggregation(PIFA):
$$Z_l = \frac{v_l}{\|v_l\|}, v_l \text{ 代表属于 } l \text{ 标签的节点特征}$$
这也恰恰是一次图卷积操作。
 - 标签聚类

示意图



XR-Transformer can train the model on multi-resolution objectives.

① Introduction

② Method

③ 理论分析

④ 实验

链接预测 vs 邻居预测

- 图自监督学习的一个标准任务是链接预测，其基于图同构假设。可以认为是相似的节点，应该有边相连。

$$P(A_{ij} = 1) = \text{Similarity}(\Phi(T_i), \Phi(T_j))$$

- 但是现实的很多情况不符合图同构假设。



Figure 3: A counter-example for standard link prediction methodology.

- 作者提出假设：有相似特征的节点扮演相似的“structural roles”，在本研究中，将这种结构建构为其 1 阶邻居。
- 这也和 XMC 问题中的假设相符，有轻微差别的节点应该有相似的多标签。

① Introduction

② Method

③ 理论分析

④ 实验

dataset

Table 1: Basic statistics of the OGB benchmark datasets (Hu et al., 2020a).

	#Nodes	#Edges	Avg. Node Degree	Split ratio (%)	Metric
ogbn-arxiv	169,343	1,166,243	13.7	54/18/28	Accuracy
ogbn-products	2,449,029	61,859,140	50.5	8/2/90	Accuracy
ogbn-papers100M	111,059,956	1,615,685,872	29.1	78/8/14	Accuracy

Result

Dataset		ogbn-arxiv				ogbn-products		
		MLP	GraphSAGE	RevGAT	RevGAT+SelfKD	MLP	GraphSAINT	SAGN+SLE
X_{plain}	OGB-feat [†]	55.50 ± 0.23	71.49 ± 0.27	74.02 ± 0.18	74.26 ± 0.17	61.06 ± 0.08	79.08 ± 0.24	84.28 ± 0.14
	BERT*	62.91 ± 0.60	70.97 ± 0.33	73.59 ± 0.10	73.55 ± 0.41	60.90 ± 1.09	79.55 ± 0.85	83.11 ± 0.18
X_{SSLGNN}	OGB-feat+GZ	70.95 ± 0.38	71.41 ± 0.09	72.42 ± 0.16	72.50 ± 0.08	74.19 ± 0.55	78.38 ± 0.21	79.78 ± 0.11
	BERT*+GZ	70.46 ± 0.21	71.24 ± 0.19	72.33 ± 0.06	72.30 ± 0.20	OOM	OOM	OOM
	OGB-feat+DGI	56.02 ± 0.16	71.72 ± 0.26	73.48 ± 0.14	73.90 ± 0.26	70.54 ± 0.13	79.26 ± 0.16	81.59 ± 0.14
	BERT*+DGI	59.42 ± 0.38	72.15 ± 0.06	73.24 ± 0.25	73.60 ± 0.21	73.62 ± 0.23	81.29 ± 0.41	82.90 ± 0.21
	OGB-feat+GAE	56.47 ± 0.08	72.00 ± 0.27	73.70 ± 0.28	74.06 ± 0.10	74.81 ± 0.22	78.23 ± 0.10	82.85 ± 0.11
	BERT*+GAE	62.11 ± 0.32	72.72 ± 0.17	74.26 ± 0.20	74.48 ± 0.15	78.42 ± 0.14	82.74 ± 0.16	84.42 ± 0.04
	OGB-feat+VGAE	56.70 ± 0.20	72.04 ± 0.29	73.59 ± 0.17	73.95 ± 0.09	74.66 ± 0.10	78.65 ± 0.20	83.06 ± 0.06
	BERT*+VGAE	62.48 ± 0.14	72.92 ± 0.02	74.21 ± 0.01	74.44 ± 0.09	78.81 ± 0.25	82.80 ± 0.11	84.40 ± 0.09
X_{GIANT}	BERT+LP	67.33 ± 0.54	66.61 ± 2.86	75.50 ± 0.11	75.75 ± 0.04	73.83 ± 0.06	81.66 ± 0.08	82.33 ± 0.16
	NO TFIDF+ NO PIFA	69.33 ± 0.19	73.41 ± 0.34	74.95 ± 0.07	75.16 ± 0.06	74.16 ± 0.22	80.70 ± 0.51	81.63 ± 0.28
	NO TFIDF+PIFA	72.74 ± 0.17	74.43 ± 0.20	75.88 ± 0.05	76.06 ± 0.02	78.91 ± 0.28	81.54 ± 0.14	82.22 ± 0.15
	TFIDF+NO PIFA	71.74 ± 0.15	74.09 ± 0.33	75.56 ± 0.09	75.85 ± 0.05	79.37 ± 0.15	83.83 ± 0.14	85.01 ± 0.10
	GIANT-XRT	73.08 ± 0.06	74.59 ± 0.28	75.96 ± 0.09	76.12 ± 0.16	79.82 ± 0.07	84.40 ± 0.17	85.47 ± 0.29

Thanks!