

Deep Graph Contrastive Representation Learning

[ICML 2020]

[论文地址](#)

[代码地址](#)

abstract

图表示学习现在成为分析图结构数据的基础。受最近对比方法成功的启发，在本文中，我们提出了一种新的无监督图表示学习框架，通过在节点级别上使用对比目标。具体来说，我们通过腐败生成两个图视图，并通过最大化这两个视图中节点表示的一致性来学习节点表示。为了给对比目标提供多样化的节点上下文，我们提出了一种在结构和属性层面生成图视图的混合方案。此外，我们从互信息和经典三重态损失两个角度提供了我们动机背后的理论解释。我们使用多种真实数据集在直推式和归纳式学习任务上进行了实证实验。实验结果表明，尽管该方法简单，但我们提出的方法始终优于现有的最先进的方法。此外，我们的无监督方法在直推式任务上甚至超过了有监督的方法，展示了其在现实世界应用中的巨大潜力。

1. introduction

在过去的几年中，图表示学习作为一种分析图结构数据的有力策略应运而生。图表示学习旨在学习一种编码函数，将节点转换为保留图属性和结构特征的低维稠密嵌入。

近年来，基于图神经网络(Graph Neural Networks, GNN)的图表示学习受到了广泛的关注。然而，现有的 GNN 模型大多是以监督的方式建立的，需要大量的标记节点进行训练。尽管有一些尝试将之前的无监督目标(即矩阵重构)连接到 GNN 模型，但这些方法仍然严重依赖于预先设定的图邻近矩阵。

到目前为止，已经提出了一系列的对比学习方法，通过对比正样本对和负样本对来最

大化输入(即图像)与其表示(即图像嵌入)之间的互信息。

在本文中，我们介绍了一个简单而强大的无监督图表示学习对比框架 deep **GRA**ph **C**ontrastive **r**epresentation learning (GRACE)。在 GRACE 中，我们首先通过随机执行腐败操作生成两个相关的图视图。然后，我们使用对比损失来训练模型，以最大化这两个视图中节点嵌入之间的一致性。

本文贡献：

- 首先，本文提出了一个通用的通用的对比学习框架用于无监督图表征学习框架。提出的 GRACE 框架简化了以前的工作，并通过最大化两个图视图之间节点嵌入的一致性来工作
- 其次，我们提出了两种具体的方案，去除边和掩盖特征来生成图的视图。
- 实验结果表现突出，使用 6 个流行的公共基准数据集对直推式和归纳式节点分类进行了全面的实证研究

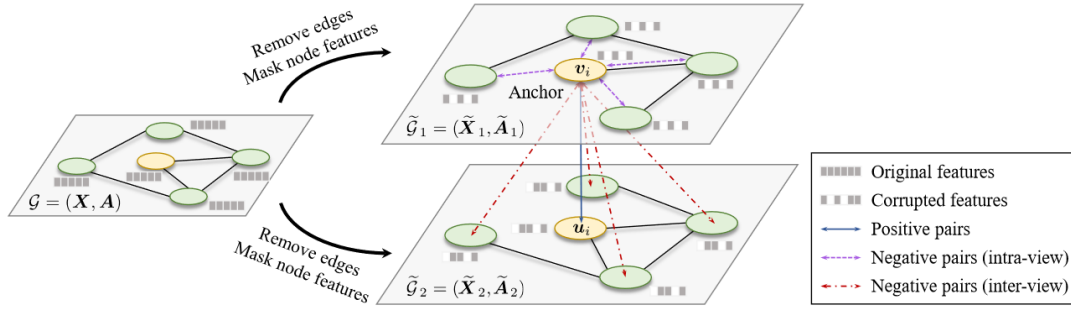


Figure 1: Our proposed deep GRAph Contrastive rEpresentation learning (GRACE) model.

2. Preliminaries

In unsupervised graph representation learning, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a graph, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ represent the node set and the edge set respectively. We denote the feature matrix and the adjacency matrix as $\mathbf{X} \in \mathbb{R}^{N \times F}$ and $\mathbf{A} \in \{0, 1\}^{N \times N}$, where $x_i \in \mathbb{R}^F$ is the feature of v_i , and $A_{ij} = 1$ iff $(v_i, v_j) \in \mathcal{E}$. There is no given class information of nodes in \mathcal{G} during training. Our objective is to learn a GNN encoder $f(\mathbf{X}, \mathbf{A}) \in \mathbb{R}^{N \times F'}$ receiving the graph features and structure as input, that produces node embeddings in low dimensionality, i.e., $F' \ll F$.

We denote $\mathbf{H} = f(\mathbf{X}, \mathbf{A})$ as the learned representations of nodes, where \mathbf{h}_i is the embedding of node v_i . These representations can be used in downstream tasks, such as node classification.

3. Deep Graph Contrastive Representation Learning

在这一部分中，我们详细介绍了我们提出的 GRACE 框架，从对比目标的整体框架开始，然后是具体的图视图生成方法。在本节的最后，我们从两个角度提供了理论证明，即与 infomax 算法原则和经典三元组损失的联系。

我们生成了两个图视图，分别记为'G1 和'G2'，生成的两个视图中的节点嵌入分别为 $U = f(X_1, A_1)$ 和 $V = f(X_2, A_2)$ 。

$$\ell(\mathbf{u}_i, \mathbf{v}_i) = \log \frac{e^{\theta(\mathbf{u}_i, \mathbf{v}_i)/\tau}}{\underbrace{e^{\theta(\mathbf{u}_i, \mathbf{v}_i)/\tau}}_{\text{the positive pair}} + \underbrace{\sum_{k=1}^N \mathbb{1}_{[k \neq i]} e^{\theta(\mathbf{u}_i, \mathbf{v}_k)/\tau}}_{\text{inter-view negative pairs}} + \underbrace{\sum_{k=1}^N \mathbb{1}_{[k \neq i]} e^{\theta(\mathbf{u}_i, \mathbf{u}_k)/\tau}}_{\text{intra-view negative pairs}}},$$

然后将要最大化的总体目标定义为所有正对上的平均值，形式上为

$$\mathcal{J} = \frac{1}{2N} \sum_{i=1}^N [\ell(\mathbf{u}_i, \mathbf{v}_i) + \ell(\mathbf{v}_i, \mathbf{u}_i)].$$

综上所述，在每个训练历元，GRACE 首先生成图 G 的两个图视图" G 1 "和" G 2 "，然后使用 GNN 编码器 f 得到" G 1 "和" G 2 "的节点表示 U 和 V。最后，通过最大化式中的目标来更新 f 和 g 的参数。Algorithm1 进行了总结：

Algorithm 1: GRACE training algorithm

- 1 **for** $epoch \leftarrow 1, 2, \dots$ **do**
 - 2 Generate two graph views $\tilde{\mathcal{G}}_1$ and $\tilde{\mathcal{G}}_2$ by performing corruption on \mathcal{G}
 - 3 Obtain node embeddings U of $\tilde{\mathcal{G}}_1$ using the encoder f
 - 4 Obtain node embeddings V of $\tilde{\mathcal{G}}_2$ using the encoder f
 - 5 Compute the contrastive objective \mathcal{J} with Eq. (2)
 - 6 Update parameters by applying stochastic gradient ascent to maximize \mathcal{J}
-

Graph View Generation

Removing edges (RE).

在原图上随机移除部分的边。处理邻接矩阵 A：在有边的地方放置伯努利分布的数据，

在没有边的地方置零。

Masking node features (MF).

在节点特征中随机地用零掩盖一部分维数。

4. Experiment

transductive and inductive node classification

Dataset: Cora、Citeseer、Pubmed、DBLP、PPI 等

(a) <i>Transductive</i>					
Method	Training Data	Cora	Citeseer	Pubmed	DBLP
Raw features	X	64.8	64.6	84.8	71.6
node2vec	A	74.8	52.3	80.3	78.8
DeepWalk	A	75.7	50.5	80.5	75.9
DeepWalk + features	X, A	73.1	47.6	83.7	78.1
GAE	X, A	76.9	60.6	82.9	81.2
VGAE	X, A	78.9	61.2	83.0	81.7
DGI	X, A	82.6±0.4	68.8±0.7	86.0±0.1	83.2±0.1
GRACE	X, A	83.3±0.4	72.1±0.5	86.7±0.1	84.2±0.1
SGC	X, A, Y	80.6	69.1	84.8	81.7
GCN	X, A, Y	82.8	72.0	84.9	82.7

(b) <i>Inductive</i>			
Method	Training Data	Reddit	PPI
Raw features	X	58.5	42.2
DeepWalk	A	32.4	—
DeepWalk + features	X, A	69.1	—
GraphSAGE-GCN	X, A	90.8	46.5
GraphSAGE-mean	X, A	89.7	48.6
GraphSAGE-LSTM	X, A	90.7	48.2
GraphSAGE-pool	X, A	89.2	50.2
DGI	X, A	94.0±0.1	63.8±0.2
GRACE	X, A	94.2±0.0	66.2±0.1
FastGCN	X, A, Y	93.7	—
GaAN-mean	X, A, Y	95.8±0.1	96.9±0.2

5. Conclusion

在本文中，本文提出了一种新的基于节点级别一致性最大化的图对比表示学习框架。我们

的模型通过使用两种方案首先生成图视图来学习表示，去除边和掩盖节点特征，然后应用

对比损失来最大化这两个视图中节点嵌入的一致性。理论分析揭示了从我们的对比目标到互信息最大化和经典三元组损失之间的联系，从而证明了我们的动机。我们在 transductive 和 inductive 设置下使用各种真实数据集进行了全面的实验。实验结果表明，我们提出的方法能够以较大的优势持续地超过现有的先进方法，甚至在直推式任务上超过有监督的方法。