

基于对比预测的自监督动态图表示学习方法

蒋林浦, 陈可佳

引用本文

蒋林浦, 陈可佳. 基于对比预测的自监督动态图表示学习方法[J]. 计算机科学, 2023, 50(7): 207-212.

JIANG Linpu, CHEN Kejia. [Self-supervised Dynamic Graph Representation Learning Approach Based on Contrastive Prediction](#) [J]. Computer Science, 2023, 50(7): 207-212.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向纵向图联邦学习的数据重构攻击方法](#)

Data Reconstruction Attack for Vertical Graph Federated Learning

计算机科学, 2023, 50(7): 332-338. <https://doi.org/10.11896/jsjcx.220900038>

[基于对比学习的疾病诊断预测算法](#)

Disease Diagnosis Prediction Algorithm Based on Contrastive Learning

计算机科学, 2023, 50(7): 46-52. <https://doi.org/10.11896/jsjcx.230200216>

[基于对比学习的低光照图像增强](#)

Contrastive Learning for Low-light Image Enhancement

计算机科学, 2023, 50(6A): 220600171-6. <https://doi.org/10.11896/jsjcx.220600171>

[结合残差与自注意力机制的图卷积小样本图像分类网络](#)

Graph Neural Network Few Shot Image Classification Network Based on Residual and Self-attention Mechanism

计算机科学, 2023, 50(6A): 220500104-5. <https://doi.org/10.11896/jsjcx.220500104>

[基于增强AST的图神经网络函数级代码漏洞检测方法](#)

Function Level Code Vulnerability Detection Method of Graph Neural Network Based on Extended AST

计算机科学, 2023, 50(6): 283-290. <https://doi.org/10.11896/jsjcx.220600131>

基于对比预测的自监督动态图表示学习方法

蒋林浦¹ 陈可佳^{1,2}

1 南京邮电大学计算机学院 南京 210003

2 江苏省大数据安全与智能处理重点实验室(南京邮电大学) 南京 210023

(1020041111@njupt.edu.cn)

摘要 近年来,以图对比学习为代表的图自监督学习已成为图学习领域的热点研究问题,该类学习范式不依赖于节点的标签并具有良好的泛化能力。然而,大多数图自监督学习方法采用静态图结构设计学习任务,如对比图的结构学习节点级或者图级的表示等,而未考虑图随时间的动态变化信息。为此,文中提出了一种基于对比预测的自监督动态图表示学习方法(DGCP),利用对比损失引导嵌入空间捕获对预测未来图结构最有用的信息。首先,利用图神经网络对每个时间快照图编码,得到对应的节点表示矩阵;然后,使用自回归模型预测下一时间快照图中的节点表示;最后,利用对比损失和滑动窗口机制对模型进行端到端的训练。在真实图数据集上进行实验,结果表明,DGCP在链接预测任务上的表现优于基准方法。

关键词: 动态图表示学习; 对比学习; 图神经网络; 链接预测

中图法分类号 TP391

Self-supervised Dynamic Graph Representation Learning Approach Based on Contrastive Prediction

JIANG Linpu¹ and CHEN Kejia^{1,2}

1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2 Jiangsu Key Laboratory of Big Data Security & Intelligent Processing(Nanjing University of Posts and Telecommunications), Nanjing 210023, China

Abstract In recent years, graph self-supervised learning represented by graph contrastive learning has become a hot research topic in the field of graph learning. This learning paradigm does not depend on node labels and has good generalization ability. However, most of the existing graph self-supervised learning methods use static graph structures to design learning tasks, such as learning node-level or graph-level representations based on structural contrast, without considering the dynamic information of graph over time. To address this problem, the paper proposes a self-supervised dynamic graph representation learning method based on contrastive prediction(DGCP), which utilizes a contrastive loss inducing the embedding space to capture the most useful information for predicting future graph structures. Firstly, each temporal snapshot graph is encoded using a graph neural network to obtain its corresponding node representation matrix. Then, an autoregressive model is used to predict node representations in the next temporal snapshot graph. Finally, the model is trained end-to-end by using the contrastive loss and sliding window mechanism. Experimental results on real graph datasets show that DGCP outperforms baseline methods on the link prediction task.

Keywords Dynamic graph representation learning, Contrast learning, Graph neural network, Link prediction

1 引言

图结构能够自然地表示实体和实体间的联系,常用于形式化地表示现实中无处不在的网络系统,如社交网络^[1]、论文引用网络^[2]、蛋白质交互网络^[3]等。图表示学习是图学习的

核心内容,一般指将节点的属性和结构嵌入到低维稠密向量空间中,以更好地进行节点分类、链接预测等下游任务。基于图神经网络(Graph Neural Network, GNN)^[4-7]的图表示学习最近受到了大量的关注并表现出了优异的性能。然而,大多数 GNN 模型基于监督或半监督的学习范式,很大程度上

到稿日期:2022-05-11 返修日期:2022-10-09

基金项目:国家自然科学基金(61876091);南京大学计算机软件新技术国家重点实验室开放课题(KFKT2022B01);南京邮电大学校级科研基金(NY221071)

This work was supported by the National Natural Science Foundation of China(61876091), National Key Laboratory of New Technology of Computer Software of Nanjing University(KFKT2022B01) and Research Foundation of Nanjing University of Posts and Telecommunications(NY221071).

通信作者:陈可佳(chenkj@njupt.edu.cn)

依赖于节点标签的数量和质量,存在标签获取成本高以及在与标签相关的对抗性攻击下出现泛化能力差^[8]的问题。

自监督学习可以很好地解决上述问题,通过重构^[9-10]或对比的方法^[11-13]在不依赖数据标签的情况下学习数据的潜在分布。近年来,出现了一系列基于对比学习的图自监督表示学习方法,如 DGI^[11], GCA^[12], GMI^[13]等。这些方法的基本思想是在不使用标记信息的情况下利用静态图的结构信息设计前置(pretext)任务,训练 GNN 编码器。

然而,现实世界的图大多是动态变化的,即图结构随着时间的推移而不断演变。已有的图对比学习方法主要关注节点级或图级的结构对比,而未考虑节点和边生成的时间对比,后者有助于学习图的演化特征。因此,动态图对比学习是一个值得研究的课题。

动态图通常可建模为离散时间快照图序列,两个相邻的时间快照图具有较强的相关性。受数据压缩领域中预测编码思想^[14-15]的启发,本文提出了一种基于对比预测的自监督动态图表示学习方法(Self-supervised Dynamic Graph Representation Learning Approach via Contrastive Prediction, DGCP)。该方法将对比学习引入了动态图表示学习中,自监督地引导模型学习具有预测未来结构能力的节点表示。具体来说,首先对每个时间快照图进行结构编码,得到节点在各时间片上的结构嵌入;然后利用自回归模型聚合节点在各个时间片上的结构嵌入,得到节点的最终嵌入;最后利用 InfoNCE 损失和滑动窗口机制进行图对比预测,使节点在每个时间片上的最终嵌入能预测其在下一个时间片的结构嵌入。

综上所述,本文的主要贡献包括:

- (1)首次将对比学习引入离散时间图表示学习中,以更好地捕捉图的演化特征。
- (2)提出了图对比预测模块,通过训练引导嵌入空间捕获对预测未来图结构最有用的信息。
- (3)实验结果表明, DGCP 在链接预测的表现优于主流的静态图表示学习方法和离散时间图表示学习方法。消融实验进一步验证了图对比预测模块的有效性。

2 相关工作

本节简要回顾动态图表示学习和图对比学习的相关工作。

2.1 动态图表示学习

动态图表示学习旨在将图随时间的演化信息与结构信息一同嵌入到表示中。根据动态图的两种不同定义形式(即连续时间图和离散时间图),其表示学习方法可大致分为以下两类。

(1)连续时间图表示学习,将动态图视为一个边带有连续时间戳的图,在结构编码的同时考虑边的时间信息,代表性方法有 CTDNE^[16], HTNE^[17], DyRep^[18], TGN^[19]等。尽管该类方法能够捕捉图的局部连续变化特征,但较难捕捉图的全局演变规律。

(2)离散时间图表示学习,将动态图视为多个离散时间步长的快照图集合,然后通过时序建模学习节点的表示。与连续时间图表示学习相比,该类工作更关注图的

整体演化特征,代表性方法有 DynGEM^[20], DynamicTriad^[21], dyngraph2vec^[22], DySAT^[23]等。其中, DynGEM^[20]假设时间片之间具有平滑性,使用前一个时间片中学习的嵌入来初始化当前时间片的嵌入。DynamicTriad^[21]除了利用时间平滑假设之外,还使用了三元闭合的概念指导图来表示学习过程。dyngraph2vec^[22]使用循环神经网络(RNN)对图的动态演变过程进行时序建模。DySAT^[23]通过结构和时间两个维度的联合注意力机制来计算节点表示,取得了较优的结果。

本文方法沿用离散时间图表示学习范式,将当前时间片与后一时间片的节点嵌入进行对比预测,旨在学习图的全局演化特征。

2.2 图对比学习

对比学习(Contrastive Learning)是一种自监督学习范式,通过构造前置任务以及生成并对比正负样本来训练编码器^[24]。目前,对比学习被广泛应用于计算机视觉^[24]和自然语言处理^[25]领域,用于获得更好的表示。其中, Aaron 等^[26]首次提出了对比预测编码(Contrastive Predictive Coding, CPC)的思想,将未来的表示作为正样本、将过去的表示作为负样本,来学习动态全局特征。本文借鉴了这一思想,把对比学习引入离散时间图表示学习中。

目前,已有一些研究者尝试将对比学习引入图学习领域以解决标签依赖问题。DGI^[11]首次将 DIM(Deep InfoMax)^[27]思想引入图学习领域,通过最大化节点表示和相应图表示之间的互信息来学习节点的最终表示。GMI^[13]基于 MINE^[28]估计策略对 GNN 输入端和输出端之间节点表示的互信息进行优化。GRACE^[29]首先生成输入图的增强视图,然后优化视图间和视图内的节点表示的互信息。在此基础上, GCA^[12]利用度、特征向量、PageRank 中心性等改进了基于拓扑的图增强方法,其进一步提升了性能。

最近, Tian 等^[30]首次将对比学习引入动态图表示学习中,提出了 DDGCL 方法,通过采样并对比相同节点的两个附近时间视图进行学习。但是, DDGCL 是连续时间图表示学习方法,而本文方法是离散时间图表示学习方法,其根据相邻时间快照图之间的强相关性来设计前置任务,以捕捉图的全局演变特征。

3 DGCP 模型方法

3.1 问题定义

本文将动态图定义为一个时间快照图序列 $G = \{G^1, G^2, \dots, G^T\}$, 其中 $t \in \{1, 2, \dots, T\}$ 是时间步长。每个时间快照图是一个带权无向图 $G^t = (V, E^t, A^t)$, 其中 V 表示共享节点集合, E^t 表示 G^t 中边的集合, A^t 表示 G^t 的邻接矩阵。如果节点对 $(i, j) \in E^t$, 则 A_{ij}^t 表示该边的权重; 否则 $A_{ij}^t = 0$ 。本文定义的动态图不仅支持边的添加, 也支持边的删除。

给定动态图 G , 表示学习的目标是学习映射函数 $f_t: \{G^1, \dots, G^t\} \rightarrow \mathbb{R}^{|V| \times D}$, 从而得到每个节点 $v \in V$ 在时间步长 t 的潜在嵌入表示 $H_v^t \in \mathbb{R}^D$, 并且 H_v^t 应该包含图 G^t 的结构信息以及 $\{G^1, \dots, G^t\}$ 序列的时间演变信息。

3.2 DGCP 模型框架

DGCP 模型的总体框架如图 1 所示。它主要分为 3 个

模块:结构编码模块、自回归模块和图对比预测模块。首先,利用结构编码器 ϵ ,对一系列时间快照图 $\{G^1, G^2, \dots, G^T\}$ 分别编码,得到对应的节点表示矩阵 $\{S^1, S^2, \dots, S^T\}$;然后,将 $\{S^1, S^2, \dots, S^T\}$ 作为自回归器 \mathfrak{R} 的输入,经聚合捕捉动态图的演变特征,得到每个时间快照图最终的节点表示矩阵 $\{H^1, H^2, \dots, H^T\}$;然后,采用滑动窗口机制进行对比预测,即对于锚节点 v 的表示 H_v^{T-1} ,构造正样本 S_v^T 和负样本集 $\{S_v^{T-w}, S_v^{T-w+1}, \dots, S_v^{T-2}\}$ (w 为滑动窗口大小)并计算对比损失。

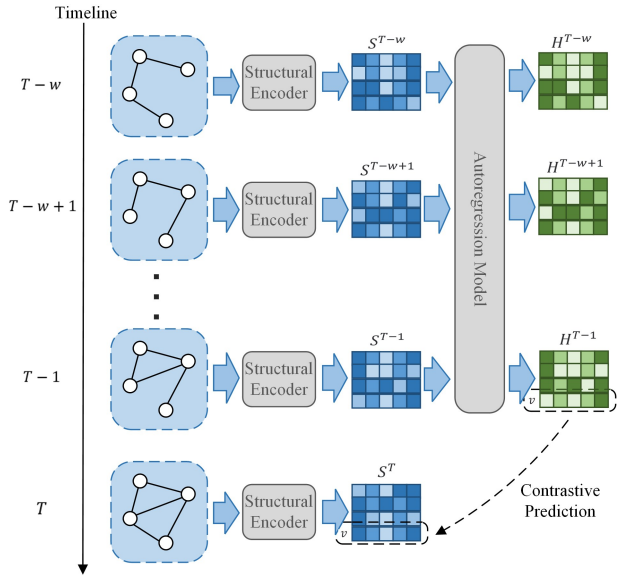


图1 DGCP 模型的框架图

Fig. 1 Framework of DGCP

3.3 结构编码模块

首先,采用结构编码器 ϵ 对 $\{G^1, G^2, \dots, G^T\}$ 中的每个时间快照图 G^t 分别编码,得到 $\{S^1, S^2, \dots, S^T\}$ 。其中, S^t ($S^t \in \mathbb{R}^{|V| \times F}$) 是 G^t 的点表示矩阵:

$$S^t = \epsilon(G^t) \quad (1)$$

为了更好地聚合邻居信息,本文采用图注意力网络(Graph Attention Network, GAT)^[4] 作为结构编码器。图注意力网络的输入是邻接矩阵 A 和节点表示矩阵 $X \in \mathbb{R}^{|V| \times d}$, 初始层的 X 由每个节点的 one-hot 编码构成,输出为更新的节点表示矩阵 $S \in \mathbb{R}^{|V| \times F}$ 。

图注意力网络通过计算中心节点与其一阶邻居节点的注意力权重来获得中心节点的表示,对于节点 v ,其传播规则为:

$$\begin{aligned} e_{uv} &= \sigma(\mathbf{A}_{uv} \cdot \mathbf{a}^T [\mathbf{W}^s \mathbf{x}_u \parallel \mathbf{W}^s \mathbf{x}_v]) \\ \alpha_{uv} &= \frac{\exp(e_{uv})}{\sum_{n \in N_v} \exp(e_{vn})} \\ s_v &= \sigma(\sum_{u \in N_v} \alpha_{uv} \mathbf{W}^s \mathbf{x}_u) \end{aligned} \quad (2)$$

其中, $s_v \in \mathbb{R}^{F'}$ 为网络的输出; N_v 是节点 v 的一阶邻居集合; $\mathbf{W}^s \in \mathbb{R}^{F' \times d}$ 是一个共享的参数矩阵; $\mathbf{a} \in \mathbb{R}^{2d}$ 是一个参数向量; \parallel 是拼接操作; $\sigma(\cdot)$ 是一个非线性激活函数。本文使用 LeakyReLU 函数来计算注意力权重,再用指数线性单元(ELU)激活输出表示。

本文采用多头注意力机制来计算节点的最终结构编码表示:

$$S_v = [s_v^1 \parallel s_v^2 \parallel \dots \parallel s_v^{h_n}] \quad (3)$$

其中, $S_v \in \mathbb{R}^F$ ($F = F' \times h_n$), h_n 是注意力头的个数。

3.4 自回归模块

经过结构编码,各个时间快照图的节点表示矩阵为 $\{S^1, S^2, \dots, S^T\}$, 每个 S^t 仅聚合了节点在 G^t 的邻居结构信息。然而,动态图表示学习还要捕捉节点表示随时间的演化信息。因此, DGCP 利用自回归器 \mathfrak{R} 聚合每个时间片之前的节点表示矩阵,得到各个时间片最终的节点表示矩阵:

$$\{H^1, H^2, \dots, H^T\} = \mathfrak{R}\{S^1, S^2, \dots, S^T\} \quad (4)$$

其中, $H^t \in \mathbb{R}^{|V| \times D}$ 。

具体来说,本文基于时间注意力网络^[23] 来设计自回归器 \mathfrak{R} 。首先,为了捕捉每个时间快照图的先后顺序,对每个时间步长 t 进行位置编码^[31], 得到位置嵌入矩阵序列 $\{P^1, P^2, \dots, P^T\}$ ($P^t \in \mathbb{R}^{|V| \times F}$), 然后,将位置嵌入矩阵序列和结构编码矩阵序列 $\{S^1, S^2, \dots, S^T\}$ 相加,得到时间注意力网络的输入 $\{P^1 + S^1, P^2 + S^2, \dots, P^T + S^T\}$; 最后,经过时间注意力网络得到各个时间快照图上的最终节点表示矩阵 $\{H^1, H^2, \dots, H^T\}$ 。

本文的时间注意力网络用于捕捉图结构的动态变化。对于节点 v , 将其表示序列 $\{P_v^1 + S_v^1, P_v^2 + S_v^2, \dots, P_v^T + S_v^T\}$ ($P_v^t + S_v^t \in \mathbb{R}^F$) 转为矩阵 $P_v \in \mathbb{R}^{T \times F}$, 作为网络的输入。利用放缩点积注意力^[32] 计算节点表示,其传播规则为:

$$\begin{aligned} e_v^{ij} &= \left(\frac{((X_v W_q)(X_v W_k)^T)_{ij}}{\sqrt{D^T}} + M_{ij} \right) \\ \beta_v^{ij} &= \frac{\exp(e_v^{ij})}{\sum_{k=1}^T e_v^{ik}}, \mathbf{Z}_v = \beta_v(X_v W_v) \end{aligned} \quad (5)$$

其中,矩阵 $Z_v \in \mathbb{R}^{T \times D'}$ 为网络的输出; $W_q \in \mathbb{R}^{F \times D'}$, $W_k \in \mathbb{R}^{F \times D'}$ 和 $W_v \in \mathbb{R}^{F \times D'}$ 分别为线性投影矩阵; $M \in \mathbb{R}^{T \times T}$ 是一个掩码矩阵,定义为:

$$M_{ij} = \begin{cases} 0, & i \leq j \\ -\infty, & \text{otherwise} \end{cases} \quad (6)$$

与 GAT 相同,时间注意力网络也采用多头注意力机制来计算节点的表示矩阵:

$$H_v = [z_v^1 \parallel z_v^2 \parallel \dots \parallel z_v^{h_m}] \quad (7)$$

其中, $H_v \in \mathbb{R}^{T \times D}$ ($D = D' \times h_m$), h_m 是注意力头的个数。

3.5 图对比预测模块

对比学习的核心思想是构造一个前置任务,并通过生成和对比正、负样本来训练编码器。一类常见的前置任务是预测未来、缺失或上下文信息。在离散时间图上,相邻的两个时间快照图具有很强的相关性,因此本文设计了一个图对比预测(Graph Contrastive Prediction)模块,把后一个时间快照图的节点表示作为正样本,将之前时间快照图的节点表示作为负样本,通过对比学习引导模型预测未来的图结构。

假设节点 v 在时间步长 t 上的最终表示 H_v^t 为锚点,将下一时间步长 $t+1$ 的结构嵌入表示 S_v^{t+1} 视为正样本,将 t 之前在滑动窗口 w 内的结构嵌入表示 $\{S_v^{t-w+1}, S_v^{t-w+2}, \dots, S_v^{t-1}\}$ 视为负样本。利用 InfoNCE 损失函数^[26] 进行对比学习,计算式如下:

$$\begin{aligned} l'_v &= -\log \frac{e^{\frac{\theta(H_v^t, S_v^{t+1})}{\tau}}}{\sum_{k=t-w+1}^{t+1} e^{\frac{\theta(H_v^t, S_v^k)}{\tau}}} \\ L_1 &= \frac{1}{|V|} \cdot \frac{1}{T-w} \sum_v \sum_{t=w-1}^T l'_v \end{aligned} \quad (8)$$

其中, $\theta(\cdot)$ 是余弦相似度函数, τ 是温度系数, 用于调节对困难样本的关注程度, τ 越小, 越能将该样本和最相似的其他样本分开。

除了学习图的动态性之外, 模型还要学习各个时间快照图的结构信息。DGCP 使用二元交叉熵损失函数, 使得在固定长度随机游走序列中共同出现的节点具有更相似表示, 定义如下:

$$l_v^t = \sum_{u \in N_{\text{walk}}^t(v)} -\log(\sigma(\langle \mathbf{H}_u^t, \mathbf{H}_v^t \rangle)) - \omega_n \cdot \sum_{u' \in P_n^t(v)} \log(1 - \sigma(\langle \mathbf{H}_u^t, \mathbf{H}_{v'}^t \rangle)) \quad (9)$$

$$L_2 = \sum_{t=1}^T \sum_{v \in V} l_v^t$$

其中, σ 是 sigmoid 函数; $\langle \cdot \rangle$ 表示内积操作; $N_{\text{walk}}^t(v)$ 表示在 G^t 中的固定长度随机游走序列上与 v 同时出现的节点集合; P_n^t 是时间快照图 G^t 的负采样分布; ω_n 是负采样率, 用于平衡正负样本的比例。

最后, 本文采用多任务学习策略^[33], 来获得最终的损失函数。

$$L = (1 - \lambda)L_1 + \lambda L_2 \quad (10)$$

其中, λ 是超参数, 用于调节两个损失的贡献度。

4 实验

本文实验从 3 个方面对 DGCP 模型进行评价。

(1) 采用链接预测任务, 比较 DGCP 和相关基线方法的表现学习能力;

(2) 对 DGCP 模型进行消融, 以评估每个模块的有效性及其对整体模型的影响;

(3) 分析模型对超参数的敏感度。

4.1 数据集

实验采用了 3 个广泛使用的真实动态网络数据集, 它们分别具有不同的类型和规模。数据集的详细统计信息如表 1 所列。

(1) Enron^[34] 数据集: 节点代表 Enron 公司的员工, 边代表员工之间的邮件互动。

(2) UCI^[35] 数据集: 节点代表加州大学欧文分校在线社交平台上的用户, 边代表用户之间发送过消息。

(3) ML-10M^[36] 数据集: 二分动态网络, 其中节点包括 MovieLens 上的用户和电影, 边代表用户对电影的打分行为。

表 1 数据集的统计信息

Table 1 Statistics of datasets

Dataset	Nodes	Edges	Timesteps
Enron	143	2347	16
UCI	1809	16822	13
ML-10M	20537	43760	13

4.2 实验设置

每个模型的输入均为时间快照图序列 $\{G^1, G^2, \dots, G^T\}$, 学得节点的表示矩阵 \mathbf{H}^t 后预测 $t+1$ 时刻的链接。链接预测采用逻辑回归分类器, 将每个节点对分为有链接或无链接。每个模型的节点表示的维数均设置为 128。与以往研究^[22-25]一致, 本文使用 AUC 指标评价链接预测的性能, 在每个数据

集上训练 10 次并报告各模型的平均 AUC 值。

DGCP 模型在 Tensorflow^[34] 框架下设计, 使用 Adam 优化器^[35] 训练, 初始学习率设置为 0.001。同时, 还加入正则化项来防止过拟合。温度系数 τ 设置为 0.4, 超参数 λ 和 ω 在每个数据集中均被设置为最优值。

4.3 比较方法

比较方法包括 5 种静态图表示学习方法和 4 种离散时间图表示学习方法。

静态图表示学习方法包括:

(1) Node2vec^[10], 先利用二阶随机游走采样得到随机游走序列, 再通过 skip-gram 模型学习节点的表示;

(2) GraphSAGE^[37], 先对节点的邻居进行采样, 再用多层聚合函数归纳式学习节点的表示;

(3) GraphSAGE + GAT^[4,37], GraphSAGE 的变体, 聚合函数使用图注意力网络^[4];

(4) GCN-AE^[38], 图自编码器模型, 使用 GCN 作为编码器;

(5) GAT-AE, 与 GCN-AE 类似, 但使用 GAT 作为编码器。

离散时间图表示学习方法包括:

(1) DynamicTriad^[21], 结合三元闭合和时间平滑性指导图表示学习过程;

(2) DynGEM^[20], 一种深度神经网络嵌入方法, 能自动选择图自编码器不同层的大小;

(3) DynAERNN^[22], dyngraph2vec 的变体, 使用 RNN 对图的动态演变过程进行时序建模;

(4) DySAT^[23], 使用结构编码和时间注意力机制学习节点的表示。

各比较方法均选用了其论文报告的最佳参数, 以确保比较的公平性。

4.4 链接预测比较

各方法的链接预测结果 (10 次实验的平均 AUC 和相对误差) 如表 2 所列。实验结果表明, DGCP 取得了最优的平均 AUC 值。

表 2 链接预测结果

Table 2 Link prediction results

Method	Enron	UCI	ML-10M
Node2vec	87.85 ± 0.8	82.95 ± 0.5	86.43 ± 0.2
GSAGE	85.93 ± 0.4	82.43 ± 0.3	86.73 ± 0.2
GSAGE+GAT	76.87 ± 0.6	78.34 ± 0.4	83.42 ± 0.4
GCN-AE	85.32 ± 1.3	81.38 ± 0.4	84.54 ± 0.2
GAT-AE	79.65 ± 1.4	80.76 ± 0.3	86.06 ± 0.3
DynamicTriad	84.93 ± 0.4	80.64 ± 0.5	89.12 ± 0.3
DynGEM	71.43 ± 0.6	79.32 ± 0.3	82.45 ± 0.4
DynAERNN	77.56 ± 0.3	80.94 ± 0.5	87.76 ± 0.2
DySAT	89.21 ± 0.3	83.13 ± 0.4	91.17 ± 0.2
DGCP	91.34 ± 0.4	84.82 ± 0.3	92.47 ± 0.1

通过进一步观察发现, Enron 和 UCI 数据集规模较小、动态变化的信息较少, 导致部分动态图表示学习方法的链接预测性能不如静态图表示学习方法。但是, DGCP 在不同规模数据集上的性能都显著优于静态图表示学习方法, 展现了较好的鲁棒性。相比其他数据集, DGCP 在 Enron 数据集上的

表现更好,这是因为 Enron 数据集相邻两个时间快照图的时间间隔更短、相关性更强,因此图对比预测方法起到了更好的作用。这表明,DGCP 更擅长处理细粒度划分的数据集。在所有对比的基线方法中,DySAT^[23]的性能最好,仅次于本文的 DGCP 模型,充分验证了自注意力机制在时序建模上的有效性,也说明了 DGCP 具有较优性能的部分原因来自于基于时间注意力网络的自回归模型。

4.5 消融实验

为了进一步研究 DGCP 的 3 个主要模块(结构编码模块、自回归模块和图对比预测模块)对整体模型性能的影响,本文对模型进行了一系列的消融。DGCP-S 代表用 GCN 取代 GAT 作为模型的结构编码器。DGCP-A 代表移除模型的自回归模块。值得注意的是,自回归模块的移除不会导致模型退化为静态图学习方法,因为模型的训练依然使用本文提出的损失函数。DGCP-C 代表移除模型的图对比预测模块。

各个消融模型的性能如表 3 所列,结果取 10 次实验的平均 AUC 和相对误差。结果表明,在所有数据集上移除或替换每个模块均在一定程度上影响了模型的性能,验证了各个模块的有效性。在 3 个模块中,自回归模块对 DGCP 性能的影响最大,表明对于动态图表示学习方法,时序建模方式是提高模型性能的关键性因素。此外,本文提出的图对比预测模块通过捕捉预测未来结构的全局特征,进一步提高了模型的性能。

表 3 消融实验结果

Table 3 Result of ablation study

Method	Enron	UCI	ML-10M
DGCP-S	90.87±0.3	83.32±0.2	90.84±0.3
DGCP-A	88.02±0.4	80.65±0.2	89.03±0.1
DGCP-C	89.21±0.3	83.13±0.4	91.17±0.2
DGCP	91.34±0.4	84.82±0.3	92.47±0.1

4.6 参数敏感度分析

首先,对超参数 λ 进行敏感度分析,观察对比预测损失函数对模型的影响。具体来说,将 λ 以 0.2 步长从 0 增加到 1 分别进行实验,此时超参数 ω 设置为相应数据集上的最佳值。实验结果如图 2 所示,随着 λ 的增加,AUC 值呈现先上升后缓慢下降的趋势。在不同的数据集上, λ 的最优值不同,这可能与数据集的划分粒度不同有关,即数据划分粒度越细,其相邻两个时间快照图的相关性更强,因此图对比预测损失函数对模型性能的影响更大。

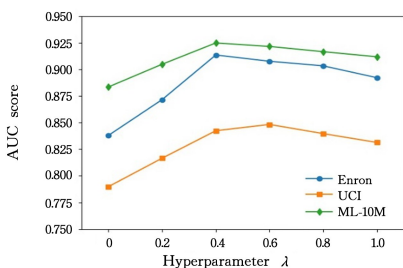


图 2 超参数 λ 对 DGCP 的影响

Fig. 2 Impact of hyperparameter λ on DGCP

滑动窗口大小 ω 决定了图对比预测模块中负样本的数量,因此本文对超参数 ω 也进行了敏感度分析。具体来说,

将 ω 以 1 步长从 3 增加到 13 分别进行实验。结果如图 3 所示,随着 ω 的增加,模型的 AUC 值先快速上升,但当 ω 达到一定阈值后,AUC 值趋向平缓并略微下降。这说明当滑动窗口 ω 设置过大时,离锚点太远的负样本对预测未来结构的作用较小,会降低模型的性能。因此,针对不同的数据集,其最优滑动窗口 ω 的选取需通过实验来获得。

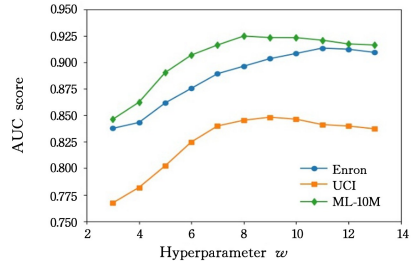


图 3 超参数 ω 对 DGCP 的影响

Fig. 3 Impact of hyperparameter ω on DGCP

结束语 本文提出了一种基于对比预测的自监督动态图表示学习方法(DGCP)。该方法假设相邻两个时间片具有强相关性,利用对比预测的思想学习具有预测未来结构能力的节点表示。在 3 个通用数据集上的实验结果表明,DGCP 比具有代表性的静态和动态图表示学习方法具有更优的链接预测性能,消融实验进一步验证了本文提出的图对比预测模块的有效性。未来工作将探索多种图对比预测策略,并进一步研究对比预测思想在连续图表示学习上的有效性。

参考文献

- [1] CHEN J, MA T, XIAO C. FastGCN: Fast Learning with Graph Convolutional Networks via Importance Sampling[C]// Proceedings of the 6th International Conference on Learning Representations, 2018.
- [2] HOU H, HILDRUN K, LIU Z. The Structure of Scientific Collaboration Networks in Scientometrics [J]. Scientometrics, 2008, 75(2): 189-202.
- [3] LIAO R, ZHAO Z, RAQUEL U, et al. LanczosNet: Multi-Scale Deep Graph Convolutional Networks[C]// Proceedings of the 7th International Conference on Learning Representations, 2019.
- [4] PETAR V, GUILLEM C, ARANTXA C, et al. Graph Attention Networks[C]// Proceedings of the 6th International Conference on Learning Representations, 2018.
- [5] WU F, AMAURI H S J, ZHANG T, et al. Simplifying Graph Convolutional Networks[C]// Proceedings of the 36th International Conference on Machine Learning, 2019: 6861-6871.
- [6] THOMAS N K, MAX W. Semi-Supervised Classification with Graph Convolutional Networks[C]// Proceedings of the 5th International Conference on Learning Representations, 2017.
- [7] QU M, YOSHUA B, TANG J. GMNN: Graph Markov Neural Networks[C]// Proceedings of the 36th International Conference on Machine Learning, 2019: 5241-5250.
- [8] LIU Y, PAN S, JIN M, et al. Graph Self-supervised Learning: a Survey [J]. arXiv:2103.00111, 2021.
- [9] BRYAN P, RAMI A, STEVEN S. DeepWalk: Online Learning of Social Representations[C]// Proceedings of the 20th ACM

- SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014:701-710.
- [10] ADITYA G, JURE L. Node2vec: Scalable Feature Learning for Networks[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016:855-864.
- [11] PETAR V, WILLIAM F, WILLIAM L H. Deep Graph Infomax [C]//Proceedings of the 7th International Conference on Learning Representations, 2019.
- [12] ZHU Y, XU Y, YU F, et al. Graph Contrastive Learning with Adaptive Augmentation[C]// Proceedings of the Web Conference, 2021:2069-2080.
- [13] PENG Z, HUANG W, LUO M, et al. Graph Representation Learning via Graphical Mutual Information Maximization[C]// Proceedings of the Web Conference, 2020:259-270.
- [14] PETER E. Predictive Coding-I [J]. IRE Transactions on Information Theory, 1955, 1(1):16-24.
- [15] BISHNU S A, MANFRED R S. Adaptive Predictive Coding of Speech Signals [J]. The Bell System Technical Journal, 1970, 49(8):1973-1986.
- [16] GIANG H N, JOHN B L, RYAN A R, et al. Continuous-Time Dynamic Network Embeddings[C]// Proceedings of the Web Conference, 2018:969-976.
- [17] ZUO Y, LIU G, LIN H, et al. Embedding Temporal Network via Neighborhood Formation[C]// Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018:2857-2866.
- [18] RAKSHIT T, MEGR DAD F, PRASENJEET B, et al. DyRep: Learning Representations over Dynamic Graphs[C]// Proceedings of the 7th International Conference on Learning Representations, 2019.
- [19] EMANUELE R, BEN C, FABRIZIO F, et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs [J]. arXiv: 2006.10637, 2020.
- [20] PALASH G, NITIN K, XINRAN H, et al. DynGEM: Deep Embedding Method for Dynamic Graphs [J]. arXiv:1805.11273, 2018.
- [21] ZHOU L K, YANG Y, REN X, et al. Dynamic Network Embedding by Modeling Triadic Closure Process[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018:571-578.
- [22] PALASH G, SUJIT R C, ARQUIMEDES C. Dyngraph2vec: Capturing Network Dynamics Using Dynamic Graph Representation Learning [J]. Knowledge Base System, 2020, 187:104816.1-104816.9.
- [23] ARAVIND S, WU Y, GOU L, et al. DySAT: Deep Neural Representation Learning on Dynamic Graphs via Self-Attention Networks[C]//Proceedings of the 13th ACM International Conference on Web Search and Data Mining, 2020:519-527.
- [24] JING L, TIAN Y. Self-supervised Visual Feature Learning with Deep Neural Networks: a Survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(11):4037-4058.
- [25] LIU X, ZHANG F, HOU Z, et al. Self-supervised Learning: Generative or Contrastive[J]. arXiv:2006.08218, 2020.
- [26] AARON V D O, LI Y, ORIOLO V. Representation Learning with Contrastive Predictive Coding [J]. arXiv:1807.03748, 2018.
- [27] HJELM R D, ALEX F, SAMUEL L, et al. Learning Deep Representations by Mutual Information Estimation and Maximization[C]// Proceedings of the 7th International Conference on Learning Representations, 2019.
- [28] ISHMAEL B, SAI R, ARISTIDE B, et al. MINE: Mutual Information Neural Estimation [J]. arXiv:1801.04062, 2018.
- [29] ZHU Y, XU Y, YU F, et al. Deep Graph Contrastive Representation Learning [J]. arXiv:2006.04131, 2020.
- [30] TIAN S, WU R, SHI L, et al. Self-supervised Representation Learning on Dynamic Graphs[C]// Proceedings of the 30th ACM International Conference on Information and Knowledge Management, 2021:1814-1823.
- [31] JONAS G, MICHAEL A, DAVID G, et al. Convolutional Sequence to Sequence Learning[C]// Proceedings of the 34th International Conference on Machine Learning, 2017:1243-1252.
- [32] ASHISH V, NOAM S, NIKI P, et al. Attention is All you Need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:6000-6010.
- [33] YOU Y, CHEN T, WANG Z, et al. When does Self-Supervision Help Graph Convolutional Networks? [C]// Proceedings of the 37th International Conference on Machine Learning, 2020:10871-10880.
- [34] MARTIN A, ASHISH A, PAUL B, et al. Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems [J]. arXiv:1603.04467, 2016.
- [35] DIEDERIK P K, JIMM Y B. Adam: A Method for Stochastic Optimization[C]// Proceedings of the 3rd International Conference on Learning Representations, 2015.
- [36] HARPER F M, JOSEPH A K. The MovieLens Datasets: History and Context [J]. ACM Transactions on Interactive Intelligent Systems, 2016, 5(4):1-19.
- [37] WILL H, YING Z, JURE L. Inductive Representation Learning on Large Graphs [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:1024-1034.
- [38] MARINKA Z, MONICA A, JURE L. Modeling Polypharmacy Side Effects with Graph Convolutional Networks [J]. Bioinformatics, 2018, 34(13):457-466.



JIANG Linpu, born in 1997, postgraduate. His main research interests include dynamic graph learning and contrast learning.



CHEN Kejia, born in 1980, Ph.D, associate professor. Her main research interests include complex network analysis and graph learning.