中国科学: 信息科学 2023 年 第 53 卷 第 11 期: 2202–2213
SCIENTIA SINICA Informationis
论文

基于自监督的端到端图数据异常检测方法

张震, 刘美含, 李朝, 卜佳俊*

浙江大学浙江省服务机器人重点实验室, 浙江大学计算机学院, 杭州 310027
* 通信作者. E-mail: bjj@zju.edu.cn

收稿日期: 2022–05–07; 修回日期: 2022–11–07; 接受日期: 2023–01–06; 网络出版日期: 2023–11–08

国家自然科学基金 (批准号: 61972349) 资助项目

摘要 异常检测旨在发掘数据中异于寻常的模式, 它在金融欺诈以及网络入侵检测等领域有着广泛的应用前景. 本文主要研究了如何在结构复杂的图数据中进行异常检测, 这涉及到挖掘异常的图结构信息以及节点属性信息. 现有大部分工作通常采用一个两步的框架, 即先对结构复杂的图数据进行表征学习生成图表征向量, 然后再将该向量用于下游异常检测任务. 由于分开训练的图表征学习任务与下游异常检测任务存在一定的语义鸿沟, 这导致现有方法无法有效地挖掘出图中潜在的异常模式. 因此, 我们提出了一种基于自监督的端到端图数据异常检测框架 SGAD, 它可以有效地捕获图数据的语义信息并用于异常检测. 具体来说, SGAD 对无标签图数据进行了一系列变换用于构建自监督辅助任务, 然后该自监督任务的输出结果可以直接用于异常检测. 我们在多个公开数据集上进行了大量实验, 实验结果表明本文提出的 SGAD 与现有方法相比获得了显著的效果提升.

关键词 图结构数据, 异常检测, 自监督学习, 图神经网络, 图表征学习

1 引言

异常检测 (anomaly detection) 旨在挖掘数据中不同于寻常或者预期的模式 [1]. 近年来, 这类方法引起了学术界和工业界广泛的关注, 常见的应用场景包括但不限于金融欺诈检测 [2]、网络入侵检测 [3], 以及医学诊断 [4] 等. 此外, 异常检测的研究具有重大应用价值, 及时发觉异常数据并采取应对措施可以有效地降低经济损失. 大部分现有方法主要关注于图像、文本, 以及表格数据上的异常检测. 然而, 现实场景中的实体间往往存在着错综复杂的连接关系, 如社交网络、电子商务等. 这些复杂的连接关系构成了海量的、结构复杂的图数据. 因此, 基于图数据的异常检测任务也逐渐成为了关注的焦点 [5~7]. 但是, 在图数据上进行异常检测不是一个简单的任务, 一方面多样的图结构提供了额外的信息; 另一方面, 不规则的图结构数据使得提取有效信息的难度陡增. 图数据的复杂性为异常检测任务带来了新的机遇与挑战.

引用格式: 张震, 刘美含, 李朝, 等. 基于自监督的端到端图数据异常检测方法. 中国科学: 信息科学, 2023, 53: 2202–2213, doi: 10.1360/SSI-2022-0179
Zhang Z, Liu M H, Li Z, et al. Self-supervised end-to-end graph level anomaly detection (in Chinese). Sci Sin Inform, 2023, 53: 2202–2213, doi: 10.1360/SSI-2022-0179

一般而言, 为了挖掘异常的模式, 输入数据通常会被映射到低维向量空间中, 然后在向量空间中检测离群的数据点 [8~10]. 早期的方法严重依赖于特征工程来构建数据的特征向量, 其普适性不强且耗时耗力 [11,12]. 特别地, 图数据拥有复杂的结构信息和丰富的节点属性信息, 这加大了人工提取特征的难度, 从而进一步限制了传统异常检测方法在图数据上的应用. 随着深度学习的发展, 越来越多的方法被提出来进行自动的、端到端的数据特征提取. 例如, 通过卷积神经网络 (convolutional neural networks) 提取图片表征向量进行分类 [13,14] 以及通过循环神经网络 (recurrent neural networks) 提取文本表征向量进行翻译 [15,16] 等任务均取得了显著的进展. 紧接着, 各类图神经网络模型 (graph neural networks) [17~20] 相继被提了出来对结构复杂的图数据进行表征学习, 该类方法在各种图相关下游任务如分类、聚类, 以及链接预测中均表现出优异的性能.

基于此, 本文旨在研究基于图神经网络的复杂图数据异常检测. 具体来说, 本文聚焦于整图级异常检测 (graph-level anomaly detection), 即识别出图数据集合中模式显著不同的样本. 尽管诸多图神经网络模型被提出, 但是它们大都关注于如何设计有效的图卷积操作 [17~20] 以及图池化操作 [21,22], 忽略了在异常检测任务上的应用. 此外, 现有大部分图神经网络模型属于监督或者半监督方法, 无法直接用于无监督异常检测任务. 虽然有部分工作先采用无监督图神经网络模型进行图表征学习 [23,24], 然后再将其表征向量作为传统异常检测模型的输入来进行异常检测任务, 但是该类方法往往无法取得令人满意的效果. 其根本原因在于, 无监督图表征学习与下游异常检测任务存在一定语义鸿沟, 即无监督图表征学习提取的通用特征不一定适用于异常检测任务. 进一步来说, 现有图数据异常检测工作仅仅侧重于发掘单个结构复杂图数据中异常节点、异常连边或者异常子图 [25~27] 等局部异常模式. 而本文着重考虑整图级异常检测任务, 它需要从一系列结构多样、属性丰富的图集合中捕获到局部以及全局性的异常模式, 这对图神经网络在异常检测任务上模型的设计、特征的提取提出了更高的要求.

为了解决上述挑战及现有工作的不足, 本文提出了基于自监督的端到端图数据异常检测框架 (self-supervised graph anomaly detection, SGAD), 该框架引入了自监督学习来提升异常检测的能力. 自监督是一类特殊的无监督学习方法 [28~30], 它的核心思想是构建容易获取标签的辅助任务来帮助模型进行表征学习, 设计合理的辅助任务将有利于提升下游主任务的能力. 本文充分利用了自监督辅助任务的特点, 构造了适用于图数据异常检测的分类辅助任务, 实现了对结构复杂图数据端到端的异常检测. 需要注意的是, 这里端到端检测框架指的是它可以在一个优化框架中完成数据特征的提取、模型结果的预测. 而大部分现有工作都是两步骤的方法, 即先设计一个优化框架提取图数据的特征, 然后再设计另外一个异常检框架进行异常检测. 那么, 第 1 个步骤中特征提取的好坏将直接影响第 2 个步骤中异常检测的效果, 具有一定的局限性. 本文提出的 SGAD 框架由以下 3 个模块构成: (1) 伪标签构造模块, 该模块设计了合适的图数据增强方式, 并构造了相对应的伪标签. (2) 图神经网络编码模块, 该模块利用图神经网络模型将经过不同类型数据增强后的图数据映射到低维表征空间. (3) 异常检测模块, 基于自监督辅助任务的输出, 该模块为每个输入图数据计算出一个异常得分, 此分数越大则表示其为异常数据的可能性越大. 不同于现有两步骤的框架, SGAD 充分利用了自监督辅助任务的特点, 实现了对复杂图数据端到端的异常检测. 我们在 6 个常用公开数据集上对 SGAD 进行了充分验证, 实验结果表明该模型在不同数据集上均有不同程度性能的提升. 与最好的算法相比, SGAD 在 BZR 和 AIDS 数据集上分别取得了 7.2% 以及 11.3% AUC (area under curve) 效果的提升.

接下来, 第 2 节将介绍图异常检测以及自监督学习的相关工作; 第 3 节将给出整图级异常检测的问题定义; 第 4 节将详细介绍 SGAD 框架的 3 个主要模块; 第 5 节将给出实验设置以及实验结果的深入分析, 最后第 6 节总结了全文并展望了未来即将进行的工作.

2 相关工作

2.1 图异常检测

近年来, 图异常检测逐渐成为大家关注的焦点. 早期的方法通常将图数据异常检测转换为传统的异常检测问题[31, 32], 但是由于图数据拥有丰富多样的结构信息, 这些传统异常检测模型无法直接处理该类型数据. 因此, 一种简单的解决方案是针对图结构信息构造相关统计特征 (节点出入度以及最短路径等) 用于下游传统异常检测模型的输入. 例如, OddBall[33] 通过统计节点的一阶邻居个数以及相对应边权重信息来检测异常的图结构信息. 在很多现实场景中, 从大量图信息中抽取出最合适的特征往往比较困难, 这通常伴随着高昂人力成本且无法捕捉到有效的结构信息.

为了捕获图数据更深层次的语义信息, 越来越多的模型将图神经网络应用于异常检测任务[7, 25, 26]. 一般来说, 这些方法先将图数据映射到低维表征空间中, 然后在低维表征空间中进行数据分析以及异常检测任务. 例如, DOMINANT[25] 利用图卷积网络对结构复杂图数据进行表征学习, 并提出使用自编码器中的重构误差对模型进行优化, 该重构误差损失可用于判断样本是否异常. 类似的, GAAN[26] 在 DOMINANT 的基础之上做了进一步的扩展, 提出融入对抗生成网络[34] 的判别器来进行异常检测. 上述提到的各类模型主要还是聚焦于图数据中局部的异常模式, 比如异常节点、连边, 以及子图等[25~27], 无法直接应用于本文研究的整图级异常检测任务. 此外, 该类方法还严重依赖于模型对数据生成或者重构的程度, 这在一定程度上限制了其在整图级异常检测任务中的应用, 因为对结构复杂的图数据进行生成或者重构不是一件容易的事情.

2.2 自监督学习

自监督学习 (self-supervised learning) 是一类特殊的无监督方法, 它通过构建易于获得标签的辅助任务来进行表征学习, 常用的 BERT[35], word2vec[36] 以及 SimCLR[30] 均是自监督模型的典型代表. 目前, 基于自监督模型的效果已经可以超越相当一部分有监督模型[30, 37]. 近期, 越来越多的学者尝试将自监督方法应用于图神经网络. 比如, GCC[38] 通过区分不同子图结构来对图神经网络进行预训练; GraphCL[39] 提出了 4 种图数据增强的方式, 并通过优化不同图数据对之间的互信息进行图表征学习. 进一步的, JOAO[40] 以及 AutoGCL[41] 提出了自适应的图数据增强方式, 显著地提升了 GraphCL 模型的效果. 虽然这些方法在图分类、聚类等应用上有一定效果的提升, 但是它们都无法直接用于图异常检测任务. 基于此, 本文提出利用自监督学习的强大表征能力, 设计了端到端的图数据异常检测框架.

3 问题定义

本文将整图级异常检测问题定义为如下形式: 给定一系列图数据 $G = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_m\}$, 其中 $\mathcal{G} = (\boldsymbol{A}, \boldsymbol{X})$, $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ 表示图中节点和边构成的邻接矩阵, $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ 表示节点的属性信息. n 和 d 分别表示图中节点的个数以及节点属性信息的维度. 我们的目标是通过学习映射函数 $f : \mathcal{G}_i \mapsto s_i \in \mathbb{R}$ 鉴别图数据中异于寻常的样本, 它将输入图数据映射为一个异常得分, 该数值的大小可以作为判断样本是否异常的依据. 值得注意的是, 整个异常检测模型是一个端到端的过程.

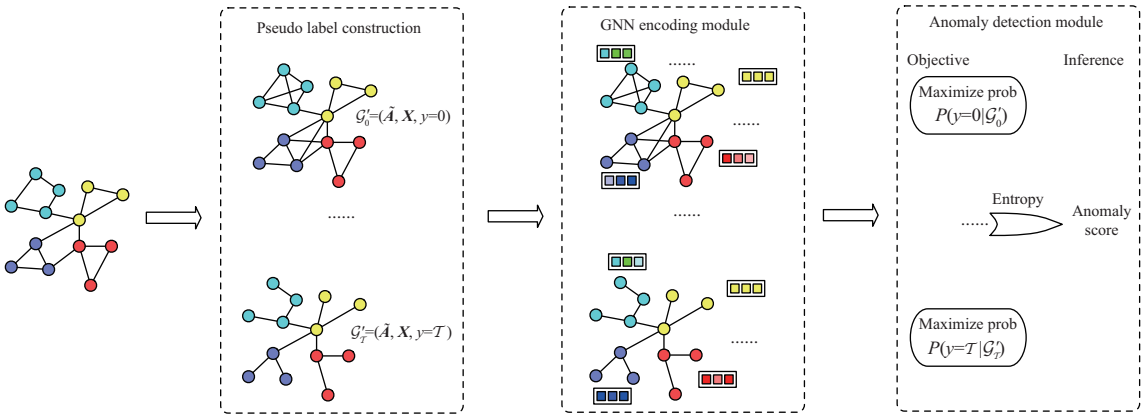

图 1 (网络版彩图) SGAD 的整体框架流程图

Figure 1 (Color online) Overall framework of SGAD

4 SGAD 框架

本节将对提出的 SGAD 框架进行详细的阐释. 我们首先对模型的整体框架做一个简单的概述, 然后对其 3 个核心组成部分一一展开介绍.

4.1 整体框架

图 1 展示了 SGAD 的整体框架, 主要由以下 3 个核心模块构成: 伪标签构造模块、图神经网络编码模块, 以及异常检测模块. 具体来说, 伪标签构造模块首先设计了一系列图数据增强操作, 不同的数据增强操作从不同的视角对图数据语义信息进行了编码. 在对图数据进行了一系列数据增强操作后, 我们构造了对图数据增强类型进行分类的自监督辅助任务, 从而帮助模型区分不同视角下的语义信息, 这将有助于对异常样本的判断. 然后, 使用图神经网络模块将图数据映射到低维表征空间中训练自监督辅助任务. 图神经网络模块经过多层图卷积以及图池化操作对结构复杂的图数据进行表征学习, 该模块可以使用多种常用图神经网络架构, 例如 GCN [17], GAT [18], GraphSAGE [19], 以及 GIN [20]. 在完成自监督辅助任务训练后, 该自监督任务的输出结果可以直接用于计算样本的异常得分进行异常检测. 通过这种自监督的方式, 该框架实现了一种端到端的异常检测方法. 接下来将对各个模块进行更加详细的介绍.

4.2 伪标签构造模块

由于本文重点研究的是无监督整图级异常检测, 因此现有大部分有监督以及半监督方法均无法适用于该场景. 与此同时, 图数据标签的获取尤其耗时耗力, 例如判断一个蛋白质图结构的类型需要大量的实验验证. 基于此, 我们通过自动构造伪标签来辅助模型的训练. 在该模块中, 构造了一种适用于异常检测的自监督辅助任务, 即对不同图数据增强的类型进行分类预测. 尽管数据增强已在计算机视觉、自然语言处理等领域得到了广泛的关注, 但常用的数据增强方式 (如裁剪、旋转等) 无法直接用于结构复杂的图数据. 图数据上数据增强方式的发展仍处于初级阶段, 且不合理的数据增强方式将会在不同程度上改变图数据的语义信息, 这会为模型带来负增益. 基于此, 我们利用图数据的结构以及节点属性信息, 从以下 3 个角度设计了如下几种图数据增强操作.

- 基于传播模型的图数据增强. 给定图数据的邻接矩阵 $\boldsymbol{A}$, 我们利用传播模型在邻接矩阵上进行

信息传播得到传播矩阵 $\tilde{\boldsymbol{A}}$. 当传播模型收敛后, 矩阵 $\tilde{\boldsymbol{A}}$ 表示的是全局视图下图结构信息的编码. 在这里, 我们采用了两种常用的传播模型, 即 Personalized PageRank [42] 以及 Heat Kernel [43], 具体公式如下所示:

$$\tilde{\boldsymbol{A}} = \alpha(\boldsymbol{I} - (1-\alpha)\boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2})^{-1}, \tag{1}$$

$$\tilde{\boldsymbol{A}} = \exp(t\boldsymbol{A}\boldsymbol{D}^{-1} - t), \tag{2}$$

其中 $\boldsymbol{A}$ 表示图数据的邻接矩阵, $\boldsymbol{D}$ 是由邻接矩阵构造的对角度矩阵, α 表示 Personalized PageRank 模型中随机游走在每一步终止的转移概率, t 描述的是 Heat Kernel 传播模型中的扩散时间.

• 基于结构扰动的图数据增强. 不同于考虑全局图结构信息的传播模型, 基于扰动的图数据增强主要关注于局部图结构信息的变换. 在这里, 使用随机删除以及添加图中节点或者连边的方式来数据增强 (p 为扰动数量),

$$\tilde{\boldsymbol{A}} = \mathrm{DELETE}(\boldsymbol{A}, p) \quad \text{or} \quad \tilde{\boldsymbol{A}} = \mathrm{ADD}(\boldsymbol{A}, p). \tag{3}$$

• 基于属性信息的图数据增强. 该方法利用节点的属性信息进行数据增强. 给定图数据的节点属性信息 $\boldsymbol{X}$, 我们通过节点属性信息计算两两之间余弦相似度, 然后利用 K 近邻 [44] 重构图结构 (K 为最近邻个数), 具体公式如下所示:

$$\tilde{\boldsymbol{A}} = \mathrm{KNN}(\boldsymbol{X}, K). \tag{4}$$

通过为每类数据增强操作定义不同的参数, 就可以得到多种数据增强操作. 分别为每种数据增强操作赋予一个自定义标签, 该模块便构造了基于数据增强类型分类的自监督辅助任务. 总的来说, 该模块构建了 $\mathcal{T}$ 种图数据增强的操作, 每种操作对应着一个伪标签. 对图数据 $\mathcal{G} = (\boldsymbol{A}, \boldsymbol{X})$ 进行第 $\mathcal{T}_i$ 种变换会生成新的数据 $\mathcal{G}' = (\tilde{\boldsymbol{A}}, \boldsymbol{X}, y_i)$, 且经过相同变换操作的图数据具有相同的伪标签. 因此, 该辅助任务的核心思想是通过区分不同图数据的变换类型来帮助模型进行表征学习以及用于下游异常检测任务.

4.3 图神经网络编码模块

图数据往往蕴含着复杂的图结构信息以及节点属性信息, 因此采用特征工程的方式提取特征将变得不可取. 在这里, 我们利用图神经网络的强大表达能力对整图数据进行编码. 在 SGAD 中, 图神经网络编码模块可以采用任意一种现有的图神经网络架构, 如 GCN [17], GAT [18], 以及 GIN [20] 等. 不同网络架构对异常检测性能会产生一定的影响, 我们将在实验部分具体分析网络架构对于异常检测性能的影响. 本小节将简单介绍一下图神经网络的运行机制, 现有的图神经网络模型基本遵循了表征的聚合与更新两个核心操作; 具体来说, 给定输入图数据, 图神经网络模型通过循环迭代的方式对邻居节点传播过来的表征进行融合并更新自己的表征. 例如在第 k 层图神经网络中, 节点 v 的表征学习过程如下所示:

$$\boldsymbol{h}_{\mathcal{N}(v)}^{k} = \mathrm{AGGREGATE}^{k}(\{\boldsymbol{h}_{u}^{k-1} | u \in \mathcal{N}(v)\}), \tag{5}$$

$$\boldsymbol{h}_{v}^{k} = \mathrm{UPDATE}^{k}(\boldsymbol{h}_{v}^{k-1}, \boldsymbol{h}_{\mathcal{N}(v)}^{k}), \tag{6}$$

其中 $\mathcal{N}(v)$ 表示的是节点 v 的邻居节点集合, AGGREGATE$(\cdot)$ 指的是聚合操作, 它将一系列邻居节点的表征聚合为一个表征, 且该聚合操作具有置换不变性的特点. 不同于聚合函数, UPDATE$(\cdot)$ 是另一个用来更新节点表征的函数, 它将聚合后的表征与当前节点的表征进行融合更新. 而不同的图神经网络架构设计了不同的更新方式, 例如拼接、求和等均是常见的操作.

在经过 K 层图神经网络得到节点的表征后, 整图级的表征可以通过图池化操作获得, 公式如下所示:

$$\boldsymbol{h}_{\mathcal{G}} = \mathrm{POOL}(\{\boldsymbol{h}_v^K | v \in \mathcal{V}\}), \tag{7}$$

其中 POOL$(\cdot)$ 表示图池化函数, 它既可以是简单的求和、求均值等操作, 也可以由神经网络模型构成的复杂的聚类或者抽样操作 [21, 22].

4.4 异常检测模块

在得到整图级表征后, 一个关键问题是如何计算其异常得分. 现有的两步骤方法往往依赖于上游表征的有效性, 通常难以取得令人满意的效果. 由于 SGAD 是端到端的框架, 因此在分类任务的学习过程中便可以进行异常检测, 并且该模型的参数可以通过自监督辅助任务所构造的伪标签进行优化. 具体来说, 在得到了整图级表征 $\boldsymbol{h}_{\mathcal{G}}$ 后, 我们将其输入到一个多层感知器 (multilayer perceptron, MLP) 中进行分类任务, 公式如下所示:

$$\hat{\boldsymbol{y}} = \mathrm{softmax}(\mathrm{MLP}(\boldsymbol{h}_{\mathcal{G}})), \tag{8}$$

$$\mathcal{L} = -\frac{1}{N \cdot |\mathcal{T}|} \sum_{n=1}^{N} \sum_{i=1}^{|\mathcal{T}|} y_{ni} \log(\hat{y}_{ni}), \tag{9}$$

其中 y_{ni} 和 $\hat{y}_{ni}$ 分别表示的是节点 n 属于第 i 类的伪标签和预测标签概率. N 指的是样本个数, $|\mathcal{T}|$ 表示的是伪标签类别个数.

当模型收敛后, 我们需要一个策略来计算每个样本的异常得分用于异常检测任务. 对于每个经过数据增强的样本, 根据式 (8) 可以得到一个预测其数据增强类型的概率分布, 该分布中属于某类的概率越大则其分布的信息熵越小, 这意味着模型预测的结果置信度越高. 由于正常样本的比例要远高于异常样本, 即自监督任务训练过程中正常样本模式出现的频率更高, 因此在不同视图下区分出正常样本模式的置信度会更高. 基于此, 我们提出利用其在不同数据增强方式下的信息熵作为样本的异常得分, 具体公式如下所示:

$$f(\mathcal{G}) = -\frac{1}{|\mathcal{T}|} \sum_{t=1}^{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \hat{y}_i^t \log(\hat{y}_i^t), \tag{10}$$

其中 $\hat{y}_i^t$ 表示的是该样本在第 t 类数据增强中预测为第 i 类的概率, 函数 $f(\cdot)$ 则计算的是 $|\mathcal{T}|$ 类信息熵的均值. 该数值越大, 则表明该样本为异常样本的可能性越大.

5 实验

5.1 数据集及实验设置

本文在 6 个公开数据集[1)]上对模型的效果进行了验证, 数据集的统计信息如表 1 所示. 类似于现有方法 [5, 6], 我们使用常见的图分类数据集进行整图级异常检测任务, 并将样本数量最少的那一类作为异常数据. 该实验设置也比较符合现实场景, 因为与正常数据相比较, 异常数据往往是占比较小的那一类. 从表 1 中可以看出, 这 6 个数据集涵盖了不同量级、不同大小, 以及不同异常占比的图数据, 有助于从多个角度测试本文所提出模型的有效性.

在整个实验中, 我们构建了 7 种不同类型的图数据增强方式用于自监督辅助任务: (1) 传播模型, 包括 Personalized PageRank 传播模型 ($\alpha = 0.2$) 和 Heat Kernel 传播模型 ($t = 5$); (2) 结构扰动, 包括随

1) https://ls11-www.cs.tu-dortmund.de/staff/morris/graphkerneldatasets.

表 1　数据集统计信息

Table 1　Datasets statistics

Dataset	#Graphs	Average #Nodes	Average #Edges	Anomaly ratio
BZR	405	35.75	38.36	0.21
COX2	467	41.22	43.45	0.22
PROTEINS	1113	39.06	72.82	0.40
AIDS	2000	15.69	16.20	0.20
NCI1	4110	29.87	32.30	0.49
NCI109	4127	29.68	32.13	0.49

机删除节点/连边以及增加节点/连边共 4 种变换 (扰动数量 $p = 5$); (3) 属性信息, 即基于节点属性相似度的 KNN 构图 ($K = 5$). 此外, 我们将图表征的维度设置为 128, 并使用 3 层 GIN 图神经网络架构对图数据进行编码, 不同超参数对模型性能的影响将在 5.6 小节进行详细讨论. ReLU [45] 和 Adam [46] 分别作为模型的激活函数以及优化器. 我们在 $[0.0001, 0.001, 0.01, 0.1]$ 的范围内搜索最优的学习率和衰减权重. 实验结果采用 5 折交叉验证, 使用 AUC [47] 作为评价指标并运行 5 次取均值和方差.

5.2　对比算法

我们将 SGAD 和如下 8 个最新的方法进行比较, 它们可以归纳为以下两大类.

• 两步骤方法 (two-step methods). 这类方法首先使用前沿的图表征学习模型得到图表征向量, 然后再利用现有的异常检测方法在图表征向量之上进行异常检测. 在这里, 我们采用了 3 种常用的模型用于获取图的表征向量, 包括基于核函数的方法 WL [48]、基于随机游走的方法 Sub2Vec [23], 以及基于无监督图神经网络的方法 InfoGraph [24], 随后 2 个经典的异常检测模型 OCSVM [49] 和 iForest [50] 被融入进来进行异常检测任务. 其中 InfoGraph 也使用了数据增强操作, 是一个比较有力的两步骤对比算法. 因此, 对 3 个图表征学习模型和 2 个异常检测模型进行组合可以得到 6 个对比方法.

• 端到端模型 (end-to-end models). 该类方法包含 2 个最新的端到端异常检测模型 OCGIN [6] 和 GLocalKD [5]. OCGIN 使用 GIN 作为图表征学习编码器, 并直接利用 SVDD 作为目标函数为异常检测任务进行优化. GLocalKD 提出使用随机蒸馏的方式进行异常检测, 即训练一个图神经网络预测另一个随机初始化的网络, 两个网络图表征的均方差大小可以作为样本的异常得分.

5.3　实验结果与分析

表 2 给出了 SGAD 框架与对比算法在不同数据集上性能的对比. 可以得到以下的观察结果.

• 首先, 本文提出的 SGAD 框架在各个数据集上均取得了最优的效果. 例如, 与效果最好的对比方法相比, SGAD 在 BZR 和 AIDS 数据集上分别取得了 7.2% 以及 11.3% AUC 性能的提升.

• 其次, 我们还观察到传统的基于核函数的方法无法取得令人满意的结果, 主要原因在于它的性能严重依赖于精心设计的核函数所提取的图特征. 而这通常需要涉及到大量专家领域知识, 并且难以推广到不同类型的图数据, 例如分子图数据和社交网络图数据往往具有不同的性质.

• 再者, 基于随机游走的图表征模型 Sub2Vec 在大部分数据集中没有基于图卷积操作的 InfoGraph 效果好. 这是由于 Sub2Vec 模型无法利用图中节点属性信息导致的, 且丢失了一定的图数据语义信息, 而 InfoGraph 则可以同时利用图数据的结构信息以及节点的属性信息进行表征学习.

• 对于两个常用的异常检测器 OCSVM 和 iForest, 它们在不同数据集上与不同表征模型的组合效

表 2 不同数据集上 AUC 的效果 [a)]

Table 2 AUC performance on different datasets[a)]

Dataset	WL		Sub2Vec		InfoGraph		OCGIN	GLocalKD	SGAD
	OCSVM	iForest	OCSVM	iForest	OCSVM	iForest			
BZR	56.44±1.71	59.12±1.15	57.04±1.61	56.70±1.29	59.45±0.56	56.95±1.17	66.54±1.41	67.89±1.39	**72.83±0.88**
COX2	57.16±0.59	56.54±0.96	56.89±1.07	56.38±1.52	63.28±1.07	59.29±0.60	65.66±0.39	60.70±1.32	**66.75±1.62**
PROTEINS	58.48±0.10	62.78±0.79	65.31±0.14	63.26±0.23	71.10±0.13	59.73±0.66	65.38±1.75	63.27±2.28	**76.30±0.18**
AIDS	82.81±0.04	84.51±0.66	83.89±0.04	84.25±0.21	66.93±0.81	88.09±0.20	89.56±0.36	75.86±2.17	**99.74±0.04**
NCI1	56.24±0.04	55.14±0.84	60.06±0.01	58.64±0.08	60.15±0.25	53.94±0.30	57.26±0.82	68.29±0.03	**68.43±0.17**
NCI109	56.62±0.02	54.26±0.45	60.28±0.03	58.52±0.13	62.63±0.25	53.08±0.43	58.08±0.34	67.96±0.02	**68.38±0.37**

a) Bold represents the best results and underline indicates the second best.

表 3 SGAD 在不同图神经网络架构下的效果

Table 3 SGAD performance with different GNN architectures

Dataset	SGAD_{GCN}	SGAD_{GAT}	$\text{SGAD}_{\text{SAGE}}$	SGAD_{GIN}
BZR	65.47±1.23	69.00±2.53	66.93±0.93	72.83±0.88
COX2	64.05±1.69	65.69±1.41	63.05±1.57	66.75±1.62
PROTEINS	75.68±0.29	66.48±0.44	73.27±0.36	76.30±0.18
AIDS	99.34±0.06	98.56±0.08	96.11±0.11	99.74±0.04
NCI1	65.16±0.17	64.22±0.27	66.31±0.15	68.43±0.17
NCI109	65.31±0.24	64.56±0.35	66.78±0.53	68.38±0.37

果都有一定的波动, 没有观测到这两个异常检测器存在绝对的优势. 这在很大程度上与上一步骤中表征学习结果的好坏息息相关.

• 值得注意的是, 端到端模型的效果总是能优于分两步骤的方法, 这进一步说明了端到端框架的优越性, 因为图表征向量直接为下游异常检测任务而优化. 而在基于两步骤的方法中, 图表征学习和异常检测是两个相互独立优化的任务, 具有一定的语义鸿沟.

• 最后, 在大部分情况下, GLocalKD 的效果优于 OCGIN. 一方面是因为 OCGIN 对超参数比较敏感, 另一方面是其目标函数 SVDD 存在着特征坍塌的风险. 而本文提出的 SGAD 在各个数据集上的性能均超越了 GLocalKD 模型, 这是由于 GLocalKD 高度依赖于其随机初始化的目标网络, 具有一定的随机性.

5.4 不同图神经网络架构的效果

正如前文所述, 我们提出的 SGAD 是一个通用的图神经网络异常检测框架, 它可以使用现有的任何一种图神经网络模型对图数据进行编码. 为了研究不同图神经网络架构对模型性能的影响, 我们测试了 4 种常用的图神经网络架构包括 GCN [17], GAT [18], GraphSAGE [19], 以及 GIN [20] 在不同数据集上的效果. 结果如表 3 所示, 从中可以观察到, 模型的性能随图神经网络架构的不同而不同, 这也与使用的数据集有关. 总的来说, GIN 模型的架构具有一定的优势, 因为它具有区分不同图结构的能力, 这在整图级表征学习中非常重要. 平均而言, SGAD_{GIN} 与 $\text{SGAD}_{\text{SAGE}}$ 相比在 BZR, COX2, 以及 AIDS 数据集中分别有 8.81%, 5.86%, 以及 3.77% 性能的提升. 尽管不同的图神经网络架构在不同数据集上的性能有所波动, 但它们都在合理的范围之内. 此外, 在这些数据集中, 不同的图神经网络架构在 NCI1

表 4　SGAD 在不同类型数据增强下的效果

Table 4　SGAD performance with different types of augmentations

Dataset	$\text{SGAD}_{\text{w/o Prop}}$	$\text{SGAD}_{\text{w/o Attr}}$	$\text{SGAD}_{\text{w/o Pert}}$	SGAD
BZR	71.91±1.26	72.11±1.96	72.09±1.76	72.83±0.88
COX2	65.64±1.38	66.38±1.51	66.03±1.54	66.75±1.62
PROTEINS	75.59±0.41	75.29±0.35	72.74±0.26	76.30±0.18
AIDS	99.11±0.06	99.61±0.07	99.51±0.05	99.74±0.04
NCI1	66.94±0.53	68.09±0.48	67.83±0.26	68.43±0.17
NCI109	66.09±0.72	68.11±0.25	67.10±0.58	68.38±0.37

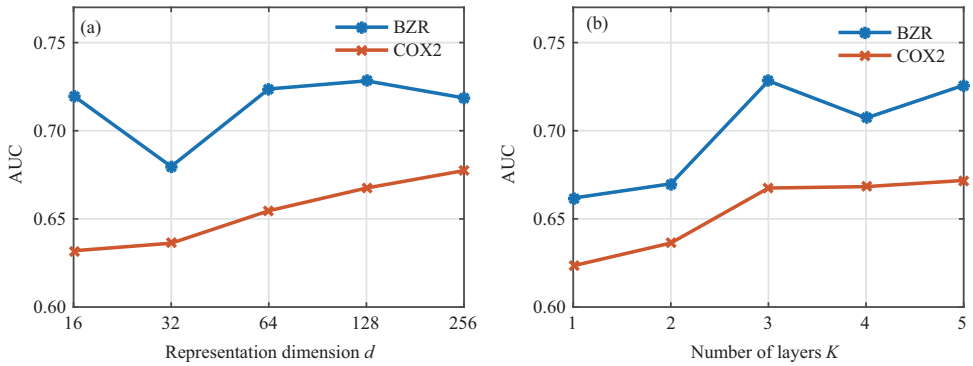

图 2　(网络版彩图) 不同超参数对模型性能的影响

Figure 2　(Color online) Impact of different hyper-parameters. (a) Representation dimension d; (b) number of layers K

和 NCI109 数据上效果的波动相对较小. 其中一个可能的原因是它们的数据样本较多, 提供了丰富的训练数据, 使得模型更加鲁棒; 另一个原因可能是由于正常样本和异常样本比例十分接近导致的.

5.5　不同数据增强方式的影响

我们还研究了不同数据增强方式对模型性能的影响, 结果如表 4 所示. $\text{SGAD}_{\text{w/o Prop}}$ 表示在辅助任务构造模块中不使用传播模型, $\text{SGAD}_{\text{w/o Attr}}$ 指不使用基于属性信息的数据增强, $\text{SGAD}_{\text{w/o Pert}}$ 表示不使用基于结构扰动的数据增强, 而 SGAD 描述的是使用所有的数据增强方式. 我们发现, 使用多种类型的数据增强方式有助于提升模型的性能, 去掉任何一类数据增强均会在一定程度上降低模型的效果. 而去掉不同类型的数据增强所带来的效果波动也不尽相同, 一个可能的原因在于不同类型的数据增强方式区分难度不同. 例如, 模型可以很容易地区分出增加节点与增加连边这两种不同的数据增强方式, 但是它却很难区分删除节点与删除连边这两种操作, 因为删除连边后也会同时删除度为零的节点.

5.6　超参敏感度分析

如图 2 所示, 我们研究了超参数表征学习维度 d 以及神经网络层数 K 对模型性能的影响. 我们将图表征学习的维度从 16 变化到 256, 然后记录模型在 BZR 和 COX2 数据集上 AUC 的变化. 从图 2(a) 中观察到, 随着维度 d 的增大, SGAD 框架在 BZR 数据集上先减小后增大, 然后再减小. 这可能是由于增加模型参数导致模型表达能力过强, 最终产生了过拟合现象. 同理, 图 2(b) 展示了随着神经网络层数的增加模型性能的变化趋势. 类似的, SGAD 在 BZR 数据集上的效果先增大后减小, 这也

是由于模型过拟合导致的. 而该过拟合现象在 COX2 数据集上有所缓解, 模型效果增大后逐渐趋于稳定, 可能的原因是 COX2 样本数多于 BZR 数据集中的样本. 同样的, 我们在剩余的数据集上也发现了类似的规律, 在这里不再进行赘述.

6 总结和未来工作

本文提出了一种端到端的整图级异常检测框架 SGAD, 它利用自监督学习构造的辅助任务对模型进行训练, 该任务的输出结果可以直接用于异常检测. 该框架是第 1 个基于自监督的整图异常检测工作. 同时, 在 SGAD 中我们设计了 3 大类共 7 种图数据增强的方式, 构建了基于数据增强类型判断的自监督分类任务, 其输出的概率分布信息熵可以作为样本的异常得分. 本文在 6 个公开数据集上验证了模型的有效性, 实验结果表明 SGAD 相比于最新的方法取得了显著性能的提升. 在未来的工作中, 我们打算将模型扩展到时序图数据上的异常检测. 在现实场景中, 随着时间的推移, 异常图数据的模式可能会发生变化, 比如从简单到复杂, 因此这也是一个亟需解决的问题.

参考文献

1 Chandola V, Banerjee A, Kumar V. Anomaly detection. ACM Comput Surv, 2009, 41: 1–58
2 Elliott A, Cucuringu M, Luaces M M, et al. Anomaly detection in networks with application to financial transaction networks. 2019. ArXiv:1901.00402
3 Krügel C, Toth T, Kirda E. Service specific anomaly detection for network intrusion detection. In: Proceedings of the ACM Symposium on Applied Computing, 2002. 201–208
4 Zhao H, Li Y X, He N J, et al. Anomaly detection for medical images using self-supervised and translation-consistent features. IEEE Trans Med Imag, 2021, 40: 3641–3651
5 Ma R R, Pang G S, Chen L, et al. Deep graph-level anomaly detection by glocal knowledge distillation. In: Proceedings of the 15th ACM International Conference on Web Search and Data Mining, 2022. 704–714
6 Zhao L X, Akoglu L. On using classification datasets to evaluate graph-level outlier detection: peculiar observations and new insights. 2020. ArXiv:2012.12931
7 Zheng L, Li Z P, Li J, et al. Addgraph: anomaly detection in dynamic graph using attention-based temporal GCN. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019. 4419–4425
8 Pang G S, Shen C H, Cao L B, et al. Deep learning for anomaly detection. ACM Comput Surv, 2021, 54: 1–38
9 Wang S Q, Zeng Y J, Liu X W, et al. Effective end-to-end unsupervised outlier detection via inlier priority of discriminative network. In: Proceedings of the Neural Information Processing Systems, 2019, 32
10 Qiu C, Pfrommer T, Kloft M, et al. Neural transformation learning for deep anomaly detection beyond images. In: Proceedings of International Conference on Machine Learning, 2021. 8703–8714
11 Eswaran D, Faloutsos C, Guha S, et al. SpotLight: detecting anomalies in streaming graphs. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, 2018. 1378–1386
12 Li N, Sun H, Chipman K, et al. A probabilistic approach to uncovering attributed graph anomalies. In: Proceedings of the SIAM International Conference on Data Mining, 2014. 82–90
13 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. Commun ACM, 2017, 60: 84–90
14 He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 770–778
15 Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput, 1997, 9: 1735–1780
16 Cho K, Van Merriënboer B, Bahdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches. 2014. ArXiv:1409.1259
17 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. 2016. ArXiv:1609.02907
18 Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. 2017. ArXiv:1710.10903

19 Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, 2017. 1025–1035
20 Xu K, Hu W H, Leskovec J, et al. How powerful are graph neural networks? 2018. ArXiv:1810.00826
21 Ying Z, You J X, Morris C, et al. Hierarchical graph representation learning with differentiable pooling. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Montréal, 2018
22 Zhang Z, Bu J J, Ester M, et al. Hierarchical graph pooling with structure learning. 2019. ArXiv:1911.05954
23 Adhikari B, Zhang Y, Ramakrishnan N, et al. Sub2Vec: feature learning for subgraphs. In: Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2018. 170–182
24 Sun F Y, Hoffmann J, Verma V, et al. InfoGraph: unsupervised and semi-supervised graph-level representation learning via mutual information maximization. 2019. ArXiv:1908.01000
25 Ding K Z, Li J D, Bhanushali R, et al. Deep anomaly detection on attributed networks. In: Proceedings of the SIAM International Conference on Data Mining, 2019. 594–602
26 Chen Z X, Liu B, Wang M Q, et al. Generative adversarial attributed network anomaly detection. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020. 1989–1992
27 Liu Z, Yu J X, Ke Y P, et al. Spotting significant changing subgraphs in evolving graphs. In: Proceedings of the 8th IEEE International Conference on Data Mining, Pisa, 2008. 917–922
28 Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations. 2018. ArXiv:1803.07728
29 Doersch C, Gupta A, Efros A A. Unsupervised visual representation learning by context prediction. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, 2015. 1422–1430
30 Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations. In: Proceedings of International Conference on Machine Learning, 2020. 1597–1607
31 Ding Q, Katenka N, Barford P, et al. Intrusion as (anti) social communication: characterization and detection. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, 2012. 886–894
32 Hooi B, Song H A, Beutel A, et al. Fraudar: bounding graph fraud in the face of camouflage. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, 2016. 895–904
33 Akoglu L, McGlohon M, Faloutsos C. OddBall: spotting anomalies in weighted graphs. In: Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, Hyderabad, 2010. 410–421
34 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, 2014
35 Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding. 2018. ArXiv:1810.04805
36 Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. 2013. ArXiv:1301.3781
37 Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning. In: Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, 2020. 21271–21284
38 Qiu J Z, Chen Q B, Dong Y X, et al. GCC: graph contrastive coding for graph neural network pre-training. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2020. 1150–1160
39 You Y N, Chen T L, Sui Y D, et al. Graph contrastive learning with augmentations. In: Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, 2020. 5812–5823
40 Suresh S, Li P, Hao C, et al. Adversarial graph augmentation to improve graph contrastive learning. In: Proceedings of the Advances in Neural Information Processing Systems, 2021
41 Yin Y H, Wang Q Z, Huang S Y, et al. AutoGCL: automated graph contrastive learning via learnable view generators. 2021. ArXiv:2109.10259
42 Bahmani B, Chowdhury A, Goel A. Fast incremental and personalized pagerank. 2010. ArXiv:1006.2880
43 Chung F. The heat kernel as the pagerank of a graph. Proc Natl Acad Sci USA, 2007, 104: 19735–19740

44 Bhatia N, Vandana. Survey of nearest neighbor techniques. 2010. ArXiv:1007.0085
45 Agarap A F. Deep learning using rectified linear units (ReLU). 2018. ArXiv:1803.08375
46 Kingma D P, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the International Conference on Learning Representations (ICLR), 2015
47 Bradley A P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn, 1997, 30: 1145–1159
48 Shervashidze N, Schweitzer P, Van Leeuwen E J, et al. Weisfeiler-lehman graph kernels. J Mach Learn Res, 2011, 12: 2539–2561
49 Schölkopf B, Williamson R C, Smola A, et al. Support vector method for novelty detection. In: Proceedings of the 12th International Conference on Neural Information Processing Systems, 1999
50 Liu F T, Ting K M, Zhou Z H. Isolation forest. In: Proceedings of the 8th IEEE International Conference on Data Mining, Pisa, 2008. 413–422

Self-supervised end-to-end graph level anomaly detection

Zhen ZHANG, Meihan LIU, Zhao LI & Jiajun BU*

Zhejiang Provincial Key Laboratory of Service Robot, College of Computer Science, Zhejiang University, Hangzhou 310027*, China*
* Corresponding author. E-mail: bjj@zju.edu.cn

Abstract Anomaly detection aims to identify unusual patterns deviating from the majorities, which is widely used in financial fraud detection, network intrusion detection, etc. This paper mainly focuses on anomaly detection of anomaly structural and node attribute information in graph-structured data. Most existing methods usually follow a two-step anomaly detection procedure. They first perform representation learning on the graph, and then the learned graph representations are fed into the downstream anomaly detection task. However, due to the separate training process of representation learning and anomaly detection, they cannot detect anomaly patterns effectively. Therefore, we propose a self-supervised graph-level anomaly detection (SGAD) framework, which can detect anomaly patterns in an end-to-end manner. Specifically, SGAD designs a self-supervised pretext task to perform representation learning, and then the output of this task can be used for explicit anomaly detection. We conduct extensive experiments on six public datasets, and the experimental results demonstrate that the proposed SGAD can achieve state-of-the-art performance on all the datasets.

Keywords graph-structured data, anomaly detection, self-supervised learning, graph neural network, graph representation learning