持续学习中图神经网络的局限性与经验回放

Junwei Su
香港大学计算机科学系 香港
香港 jwsu@cs.hku.hk

Difan Zou
香港大学计算机科学系 香港
dzou@cs.hku.hk

Chuan Wu
计算机科学系 香港大学 香港
cwu@cs.hku.hk

arXiv:2302.03534v2 [cs.LG] 9 Jul 2024

摘要

持续学习旨在使模型能够从一系列任务中逐步获取信息。这种方法对于许多现实世界中的系统至关重要，因为这些系统会随时间动态变化并不断演进。近期研究见证了图神经网络（GNN）在节点级图持续学习（NGCL）领域的探索热潮。节点级图持续学习是一种实用但颇具挑战的范式，涉及在与节点相关的任务上持续训练图神经网络。尽管近年来针对 NGCL 中图神经网络的持续学习策略取得了进展，但人们仍缺乏全面的理论认识，尤其是对其可学习性的理解。可学习性关注的是，是否存在一种学习算法，能够从假设/权重空间中找到一个良好的候选模型，这对 NGCL 开发中的模型选择至关重要。本文首次从理论角度探究了 NGCL 中图神经网络的可学习性，并揭示出，由于图数据具有相互连接的特性，可学习性受到结构偏移的显著影响。具体而言，当结构变化显著时，图神经网络可能不适用于 NGCL，这凸显了管理结构偏移的必要性。为减轻结构偏移的影响，我们提出了一种新颖的经验回放方法，称为结构演化感知经验回放（SEA-ER）。SEA-ER 采用创新的经验选择策略，利用图神经网络的拓扑感知能力；同时，它还采用独特的回放策略，通过结构对齐有效应对灾难性遗忘，并降低结构偏移对 NGCL 中图神经网络的影响。大量实验验证了我们的理论见解以及 SEA-ER 的有效性。

1 引言

持续学习，也称为增量学习或终身学习，研究能够使模型在持续获取新知识的同时保留先前所获知识的机器学习方法。这种能力至关重要，因为它使模型能够适应新信息，而不会遗忘过去的信息。在持续学习的一般设定中，任务流按顺序到来，模型通过多轮训练阶段积累特定目标（如分类）所需的知识。其主要目标是找到一种学习算法，使模型能够根据新任务持续更新参数，同时避免灾难性遗忘 (Delange 等人,2021)，即在学习新任务时无法保留先前学到的信息。持续学习对于机器学习系统的实用性至关重要，因为它使模型能够适应新信息，无需频繁重新训练；而在当今的深度学习领域，频繁重新训练可能代价高昂。

另一方面，图神经网络（GNNs）作为建模图结构和关系数据结构的有效工具，已广受欢迎（Wu 等人，2020；Su等人，2024a）。然而，大多数关于图神经网络的研究主要聚焦于静态设置，并假设底层图和学习任务保持不变。这种方法虽然对某些应用很有价值，但无法体现现实网络的动态演化特性。例如，引文网络（Zliobaite，2010）和金融网络（Gama 等人，2014）

会随时间自然演化，并伴随着新任务的出现。因此，为进一步提升图神经网络的实用性和有效性，研究其在图持续学习（GCL）中的效果与适用性至关重要。

鉴于其重要的实际意义，近期研究对图神经网络在图持续学习中的应用展开了越来越多的探讨（Febrinanto等人，2023年；Yuan等人，2023年）。然而，理论基础方面仍存在明显不足，尤其是在节点级预测任务方面，这类任务通常称为节点级图持续学习（NGCL）（Su等人，2023年；Zhang 等，2022）。在NGCL中，图神经网络模型会经历多个训练轮次（任务），以积累特定目标所需的知识，例如节点分类。在这种情形下，不同任务中的顶点可能相互连接，而新任务的引入可能会改变现有图的结构，如图1所示。由于图神经网络以图结构作为输入，这些结构变化可能导致先前任务的数据分布发生偏移，这种现象称为结构偏移。Su等人（2023年）指出，图神经网络在NGCL中的性能风险（上界）与结构偏移密切相关。尽管如此，结构变化是否以及如何影响图神经网络在NGCL中的可学习性，目前仍不明确。可学习性关乎是否存在一种学习算法，能够从假设/权重空间中生成优质的候选模型；图神经网络在NGCL中的可学习性对NGCL的发展至关重要。这促使我们提出本研究的核心问题：

在结构偏移下，*GNN* 对 *NGCL* 而言是否总是可学习的（即模型选择是否合适）？

在这项工作中，我们通过对 NGCL 中图神经网络的可学习性进行正式建模与分析，来回答上述问题。据我们所知，本工作首次从理论上研究了 NGCL 中图神经网络的可学习性，我们的首要贡献概述如下：

• 我们提出了一种数学表述，用于研究 NGCL 中图神经网络的可学习性。这一表述对于严谨地探索可学习性问题至关重要。具体而言，我们证明，大幅度的结构偏移会使图神经网络在 NGCL 中变得不可学习，正如定理 3.1 所述。这意味着，在易变网络系统（图结构快速变化的系统）中，图神经网络可能并不适合用于 NGCL，从而回答了本研究的核心问题。此外，这一发现对 NGCL 的实践开发尤为重要，因为它能为模型选择提供指导，帮助判断何时应选择或排除用于 NGCL 的类图神经网络模型架构。

鉴于结构变化在可学习性中发挥着关键作用，我们认为，即使在 NGCL 中图神经网络具有可学习性的情形下，控制结构变化也至关重要。为应对这一挑战，我们聚焦于持续学习的一种主要方法——经验回放。已有实证证明，该方法能够有效缓解 NGCL 中的灾难性遗忘（Ahrabian 等，2021；Zhou & Cao，2021a）。我们提出一种新颖的经验回放方法，旨在减轻 NGCL 中结构变化对图神经网络的影响。我们的相关贡献如下：

• 我们提出结构演化感知经验回放（SEA-ER），这是一种新颖的经验回放方法，旨在同时应对灾难性遗忘以及 NGCL 中结构变化这一独特挑战。SEA-ER 利用图神经网络的两个关键原则——结构对齐和拓扑感知——来选择经验样本，并根据这些样本与其余图的结构相似性为其分配不同的回放权重。这种方法能够有效缓解 NGCL 中图神经网络的结构变化和灾难性遗忘，并为长期性能提供理论保证（命题4.1）。

为进一步验证上述理论结果以及我们提出的经验回放方法的有效性，我们在真实世界数据集和合成数据集上进行了实证评估。结果与我们的理论发现一致，并证明了所提方法的有效性。我们提出的新型经验回放方法能够有效应对结构偏移和灾难性遗忘，在 NGCL 中优于当前的经验回放方法（Zhou & Cao，2021a；Ahrabian 等，2021；Kim 等，2022）。

2 相关工作

2.1 持续学习

持续学习，也称为增量学习或终身学习，近年来受到越来越多的关注，并已在欧几里得数据上得到广泛研究。有关这些工作的更全面综述，请参阅相关综述（Delange 等人，2021；Parisi 等人，2019年；Biesialska 等人，2020）。持续学习面临的主要挑战是解决灾难性遗忘问题，即模型在新任务上训练后，其在先前任务上的性能会下降。

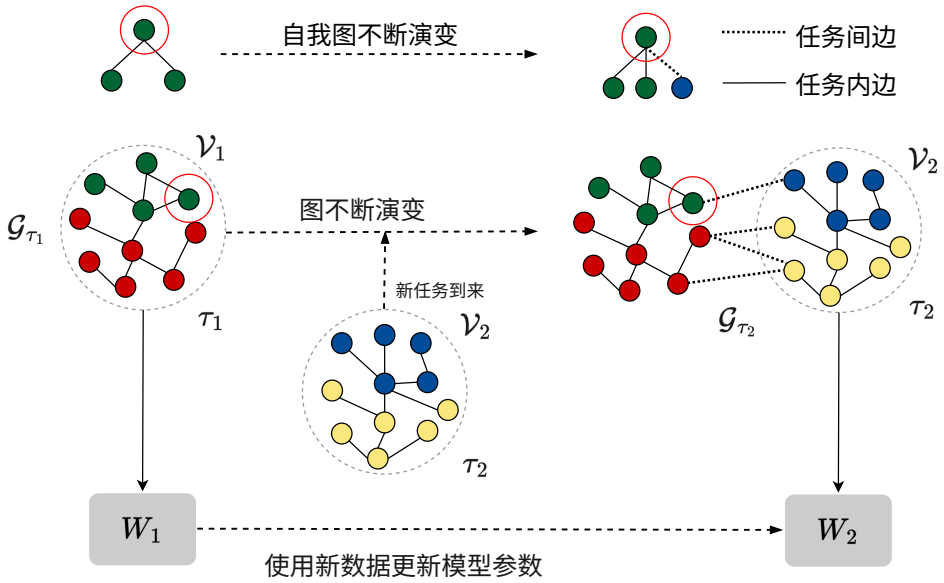


图 1：NGCL 的发展过程示意图。任务 2 引入一批新的顶点，并利用新任务的数据，将模型参数从 W_2 更新为 W_1。随着与新任务 τ_2 相关的新顶点批次不断引入，图结构会发生变化，这可能改变现有顶点的图结构（图神经网络的输入），其变化可通过这些顶点自我图的改变体现出来。这称为结构偏移。

现有解决这一问题的方法大致可分为三类：基于正则化的方法、基于经验回放的方法和基于参数隔离的方法。基于正则化的方法通过惩罚模型参数的大幅变动，力求维持模型在先前任务上的性能（Jung 等人，2016；Li & Hoiem，2017；Kirkpatrick 等人，2017；Farajtabar 等人，2020；Saha 等人，2021）。基于参数隔离的方法则通过不断为新任务引入新参数，避免对先前任务重要的参数发生剧烈变化（Rusu 等人，2016；Yoon 等人，2017；2019；Wortsman 等人，2020；Wu 等人，2019）。受人类学习理论中的间隔重复启发，基于经验回放的方法可视为神经网络的一种记忆巩固方式（Zhou & Cao，2021年）。经验回放方法包含两个关键组成部分：1）经验选择；2）回放策略。经验选择旨在从先前数据中挑选少量具有代表性或较强的样本，并将其存储在经验缓冲区中。随后，回放策略会有针对性地利用经验缓冲区中的数据重新训练模型，以巩固已有知识并遏制灾难性遗忘。

本研究深入探讨了 NGCL 中图神经网络的理论基础与实际应用。在理论方面，我们借鉴了此前广泛应用于领域自适应（David 等人，2010）、多任务学习（Ben-David & Borbely，2008）和持续学习（Benavides- Prado & Riddle，2022）等领域的成熟统计学习框架。具体而言，我们运用这些框架探究结构偏移的本质及其对 NGCL 场景中图神经网络可学习性的影响。在方法方面，我们着重研究经验回放方法，因为该方法简单有效（Lopez-Paz & Ranzato，2017；Shin 等人，2017；Aljundi 等人，2019；Caccia 等人，2020；Chrysakis & Moens，2020；Knoblauch 等人，2020；Su等人，2024年b）。由于图结构数据具有非独立同分布特性，NGCL 中的经验回放需要采用不同的处理方式（Zhou & Cao，2021年）。本文提出了一种新颖的经验回放方法，专门用于应对 NGCL 中图神经网络的灾难性遗忘和结构偏移。

2.2 图持续学习（GCL）

近年来，由于图持续学习在各种应用中具有重要的实际意义，人们对其兴趣日益高涨（Wang 等，2022；Xu 等，2020；Daruna 等，2021；Kou 等，2020；Ahrabian 等，2021；Cai 等，2022；Wang 等，2020年；Liu 等，2021；Zhang 等，2021；Zhou & Cao，2021年；Carta 等，2021；Zhang 等，2022；Kim 等，2022；Tan 等，2022）。关于图持续学习方法及其效果的详细综述，请参阅近期的综述和基准（Zhang 等，2022；Yuan 等，2023；Febrinanto 等，2023）。然而，现有的大多数节点级图持续学习研究主要关注整张图结构预先定义好的情形，或忽略任务间互连的情形。随着新任务的引入、图结构不断演变的节点级图持续学习设定，仍然

相对缺乏探索，亟待更深入的理论研究。据我们所知，Su等人，2023年）是目前唯一一项针对图结构不断演变的节点级图持续学习的理论研究。Su等人（2023年）证明，图神经网络的灾难性遗忘风险（上界）取决于底层图结构的结构偏移。相比之下，我们研究结构偏移下节点级图持续学习中图神经网络的可学习性，这可以解释为一种“下界”。我们首次提出了关于节点级图持续学习中图神经网络可学习性的不可能性结果，凸显了结构变化的关键作用。此外，我们提出了一种新颖的经验回放方法，能够有效应对节点级图持续学习中的灾难性遗忘和结构变化。

2.3 其他相关工作

尽管我们的主要关注点是图持续学习（GCL），但值得注意的是，动态图学习和图遗忘领域也存在与我们的工作在背景和术语方面相交的研究方向。动态图学习旨在使图神经网络能够捕捉不断变化的图结构（Galke等人，2021；Wang 等，2020年；Han等人，2020；Yu等人，2018；Nguyen等人，2018；Ma 等人，2020；Feng等人，2020；Bielak等人，2022；Su等人，2024年b）。动态图学习的目标是在能够访问所有先前信息的同时，将图的时间动态融入表示向量。相比之下，图持续学习关注的是灾难性遗忘问题，即模型在学习新任务后，在先前任务上的表现会下降。在评估方面，动态图学习算法仅在最新数据上进行测试，而图持续学习模型也会在过去的数据上接受评估。另一方面，图遗忘旨在“遗忘”图神经网络在给定数据集或一组任务上的训练过程中习得的特定模式、关联或信息。因此，动态图学习、图遗忘和图持续学习是彼此独立且不同的研究方向，关注重点各异，应当分别考量。

3 NGCL 中图神经网络的可学习性

3.1 NGCL 预备知识

我们使用小写字母表示标量和与图相关的目标。此外，我们使用粗体小写字母和粗体大写字母分别表示向量和矩阵。NGCL 假设存在一个训练任务流 τ_1，τ_2，……，τ_m，其特征是观测到的顶点批次 $\mathcal{V}_1$，$\mathcal{V}_2$，……，$\mathcal{V}_m$。每个顶点 v 都与一个节点特征向量 $\mathbf{x}_v \in \mathcal{X}$ 和一个目标标签 $y_v \in \mathcal{Y}$相关联。训练任务 τ_i 中的观测图结构由累积的顶点诱导得出，表示为 $\mathcal{G}_{\tau_i} = \mathcal{G}[\bigcup_{j=1}^{i} \mathcal{V}_j]$。在这种设定下，随着学习在不同训练任务间推进，图结构也会不断演变，如图 1所示。

基于图神经网络的节点级学习任务依赖于在节点的 k 跳邻域内进行信息聚合，并将其作为输入。为适应这一特点，我们在学习问题表述中采用局部视角。我们将 $\mathcal{N}_k(v)$ 记作顶点 v 的 k 跳邻域，其中的节点 $\mathcal{N}_k(v)$ 构成自我图 $\mathbf{G}_v$，该图由一个（局部）节点特征矩阵 $\mathbf{X}_v = \{x_u | u \in \mathcal{N}_k(v)\}$ 和一个（局部）邻接矩阵 $\mathbf{A}_v = \{a_{uw} | u, w \in \mathcal{N}_k(v)\}$组成。我们将 $\mathbb{G}$ 记作可能的自我图构成的空间，并将 $\mathbf{G}_v = (\mathbf{A}_v, \mathbf{X}_v)$ 记作目标顶点 v 的自我图。令 $\mathbf{G}_\mathcal{V} = \{\mathbf{G}_v | v \in \mathcal{V}\}$表示与顶点集 $\mathcal{V}$ 相关联的自我图集合。从图神经网络的角度来看，自我图 $\mathbf{G}_v$ 是马尔可夫毯，其中包含根顶点 v 的预测问题所需的全部信息。因此，我们可以将来自训练会话 τ_i 的数据 $\{(\mathbf{G}_v, y_v)\}_{v \in \mathcal{V}_i}$ 相关的预测问题视为从联合分布 $\mathbb{P}(\mathbf{y_v}, \mathbf{G}_v | \mathcal{V}_i)$ 中抽取。上述表述通过底层图结构捕捉并体现了顶点之间的依赖关系。

令 $\mathcal{F}$表示假设空间，$F \in \mathcal{F}$表示一个分类器，其 $\widehat{y}_v = F(\mathbf{G}_v)$；给定损失函数 $L(.,.)$，我们用 $R^L_{\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | \mathcal{V})}(F)$ 表示分类器 f 关于 L 和 $\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | \mathcal{V})$ 的泛化性能，其定义如下：

$$R^L_{\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | \mathcal{V})}(F) = \mathbb{E}_{\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | \mathcal{V})}[L(F(\mathbf{G}_v), y_v)]. \tag{3.1}$$

灾难性遗忘。根据上述表述，分类器 F 在 τ_m 上训练后的灾难性遗忘（**灾难性遗忘**），可通过其在先前顶点/任务上的性能保持来刻画，定义如下：

$$\mathbf{CF}(F) := R^L_{\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | \mathcal{V}_1, \ldots, \mathcal{V}_{m-1})}(F), \tag{3.2}$$

这意味着分类器 F在先前任务 τ_1、……、τ_{m-1}上从 τ_m 获得的性能保持。

3.2 结构偏移与可学习性

本文深入探讨了 NGCL 中结构变化与图神经网络可学习性之间的关系。我们首先定义假设空间的表达能力，并刻画结构偏移。

定义 1（假设空间的表达能力）。对于一组分布 $\{\mathbb{P}_1, \mathbb{P}_2, ..., \mathbb{P}_k\}$ 和一个假设空间 $\mathcal{F}$，我们将该假设空间在给定分布上的联合预测误差定义为：

$$\lambda_{\mathcal{F}}(\{\mathbb{P}_1, \mathbb{P}_2, ..., \mathbb{P}_k\}) = \inf_{F \in \mathcal{F}} \left[\sum_{\mathbb{P} \in \{\mathbb{P}_1, \mathbb{P}_2, ..., \mathbb{P}_k\}} R^{\mathcal{L}}_{\mathbb{P}}(F) \right].$$

联合预测误差刻画了假设空间同时拟合给定一组分布的能力（表达能力）。显然，假设空间的表达能力是特定设定具备可学习性的一个平凡必要条件。如果假设空间的表达能力不足，即使是该假设空间中最好的分类器也会产生较差的表现。本文始终聚焦于结构变化，并假设每个任务的训练样本以及假设空间的表达能力都足够。这意味着，当所有任务同时呈现时，存在一种学习算法，能够在假设空间内生成一个误差最小的分类器。关于表达能力和样本复杂度的进一步讨论见附录A。

接下来，我们将形式化定义不同顶点批次（任务）之间结构偏移的概念。

定义 2。设 $\mathbb{P}_1$ 和 $\mathbb{P}_2$ 是定义在某个域 $\mathcal{X}$ 上的两个分布，则 $\mathcal{F}\Delta\mathcal{F}$ 距离定义为 $\mathbb{P}_1$ 与 $\mathbb{P}_2$ 之间的距离，

$$d_{\mathcal{F}\Delta\mathcal{F}}(\mathbb{P}_1, \mathbb{P}_2) = 2 \sup_{A \in \mathcal{F}\Delta\mathcal{F}} \|\mathbb{P}_1(A) - \mathbb{P}_2(A)\|,$$

其中 $\mathcal{F}\Delta\mathcal{F} = \{\{\mathbf{x} \in \mathcal{X} \colon F(\mathbf{x}) \neq F'(\mathbf{x})\} \colon F, F' \in \mathcal{F}\}$。

$\mathcal{F}\Delta\mathcal{F}$ 距离通常用于衡量分布相对于假设空间的相关性（距离）$\mathcal{F}$（Redko 等人，2020）。在此，我们用它来刻画新任务的出现对先前顶点任务分布的影响。如果不同任务之间的结构偏移较大，那么新顶点批次的出现就会在更大程度上改变先前顶点批次的分布。因此，更新后的分布与先前分布之间的 $\mathcal{F}\Delta\mathcal{F}$ 距离会较大

不失一般性，我们可以将可学习性分析聚焦于两个训练阶段的情形，称为 NGCL-2 问题。如果 NGCL-2 问题不可学习，那么显然，一般的 NGCL-k 也不可学习。此外，通过递归应用该分析并将连续任务视为一个联合任务，相应的见解也可以轻松推广到多个训练任务。NGCL-2 问题由三个分布刻画：$\mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_1})$、$\mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_2})$和 $\mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_2, \mathcal{G}_{\tau_2})$，它们分别是图 $\mathcal{G}_{\tau_1}$ 和 $\mathcal{G}_{\tau_2}$中的 $\mathcal{V}_1$ 分布，以及图 $\mathcal{G}_{\tau_2}$中的 $\mathcal{V}_2$ 分布。接下来，以下定理刻画了图神经网络在 NGCL-2 中的可学习性。

定理 3.1（可控结构偏移的必要性，非正式表述）。设 $\mathcal{F}$ 是由图神经网络所刻画的假设空间。如果无法控制

$$d_{\mathcal{F}\Delta\mathcal{F}}\big(\mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_1}), \mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_2, \mathcal{G}_{\tau_2})\big) \quad \text{and} \quad d_{\mathcal{F}\Delta\mathcal{F}}\big(\mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_1}), \mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_2})\big)$$

那么，对于每个 $c > 0$，都存在一个满足以下条件的 *NGCL-2* 问题实例：

1. $\lambda_{\mathcal{F}}(\{\mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_1}), \mathbb{P}(y, \mathbf{G}_v|\mathcal{V}_1, \mathcal{G}_{\tau_2})\}) \leq c$，
2. 对于任意给定的持续学习算法，它至少有 $1/2$ 概率会产生一个分类器 F，该分类器来自 $\mathcal{F}$，且其灾难性遗忘（F）至少为 $1/2$。

定理 3.1 的正式表述和证明见附录A。定理 3.1指出，即使假设空间具有足够的表达能力（也就是说，每个单独的任务及其组合都可以由图神经网络学习），如果不控制不同顶点批次之间的结构偏移，NGCL 问题仍可能无法学习。我们提出以下几点说明。

说明 3.2。定理*3.1* 表明，在考虑易变网络（结构偏移较大）的持续学习时，类图神经网络模型并非合适的选择，因为结构偏移会严重损害其可学习性。这需要采用不同的处理方式。这是一个有趣的未来研究方向，但超出了本文的讨论范围。

备注 3.3。结构偏移会损害 *NGCL* 中图神经网络的可学习性。认识到结构偏移所发挥的关键作用后，我们得出结论：即使在图神经网络于 *NGCL* 中具有可学习性的场景下，控制结构偏移也十分重要。这促使我们设计了下一节所述的经验回放方法，并且还可以对其他用于 *NGCL* 中图神经网络的持续学习方法进行改进。

4 结构演化感知经验回放（SEA-ER）

在前几节中，我们探讨了 NGCL 背景下图神经网络的可学习性，并强调了结构偏移对这种可学习性的重要影响。在本节中，我们将重点从理论层面转向实际应用，考察经验回放方法在 NGCL 中的应用。我们提出了 SEA-ER，将图神经网络的拓扑感知能力与结构对齐相结合，以减轻灾难性遗忘并解决 NGCL 问题中的结构偏移。经验回放包含两个关键组成部分：1）经验缓冲区选择，即从过去的数据中选取具有代表性的数据样本；2）回放策略，即确定如何在使用新任务进行训练时利用经验缓冲区。我们方法的伪代码见附录 E。

4.1 拓扑感知经验缓冲区选择

我们将任务 τ_i 的经验缓冲区记为 $\mathcal{P}_i = \{P_1, P_2, ..., P_{i-1}\}$，其中 P_j 表示从任务 τ_j 中选出的样本。图神经网络在结构特征密切相关的顶点上表现出了优异的泛化性能（Ma 等人，2021）。因此，我们提出一种经验缓冲区选择策略，选取与其余顶点结构相似性最高的样本。该策略可表述为以下优化问题：

$$\begin{aligned} &\underset{P_j \subset \mathcal{V}_j}{\operatorname{argmin}} \max_{u \in \mathcal{V}_j \setminus P_j} \min_{v \in P_j} d_{\text{spd}}(u, v), \\ &s.t. \quad |P_j| = b. \end{aligned} \tag{4.1}$$

这里，$d_{\text{spd}}(。)$表示图中的最短路径距离，b 是每个任务中要选取的样本数量（预算）。上述优化问题会返回集合 P_j，其中的样本与同一任务内其余顶点具有较高的结构相似性。在我们的实验中，我们采用经典的最远优先遍历算法来求解。

4.2 结构对齐回放策略

标准回放策略将经验缓冲区纳入最新任务的训练集，并将缓冲区中的样本视为普通训练数据。然而，从 $\mathcal{G}_{\tau_{i-1}}$ 到 $\mathcal{G}_{\tau_i}$ 的演化图结构所引发的结构变化可能改变前一任务中顶点的分布，从而导致学习效果不佳。为减轻结构偏移的影响，我们提出了一种结构对齐回放策略。该策略为与结构偏移前的分布更相似的回放样本赋予更高权重，确保模型优先利用相关信息。采用结构对齐的学习目标如下：

$$L = \frac{1}{|\mathcal{V}_i^{\text{train}}|} \sum_{v \in \mathcal{V}_i^{\text{train}}} L(v) + \sum_{P_j \in \mathcal{P}_i} \frac{1}{|P_j|} \sum_{v \in P_j} \beta_v L(v), \tag{4.2}$$

其中 β_v（蓝色突出显示）是与标准回放策略的主要区别，表示经验缓冲区中顶点 v 的权重。为确定 β（由 β_v构成），我们针对经验缓冲区中顶点的嵌入使用核均值匹配（KMM）Gretton 等人（2009）进行计算。这需要求解以下凸二次规划问题：

$$\begin{aligned} &\min_{\beta} \left\| \sum_{P_j \in \mathcal{P}_i} \sum_{v \in P_j} \beta_v \phi(h_v) - \phi(h'_v) \right\|^2, \\ &s.t. \quad B_l \leq \beta_v < B_u, \end{aligned} \tag{4.3}$$

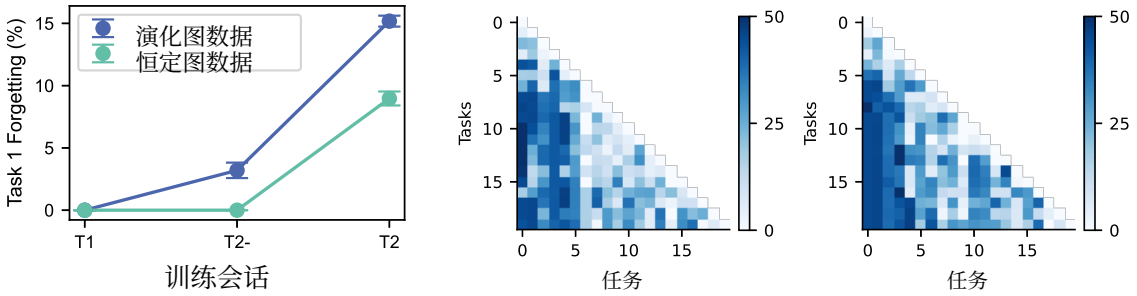
其中，h_v 是 v（GNN 的输出）相对于 $\mathcal{G}_{\tau_i}$ 的节点嵌入，h'_v 是 v 相对于 $\mathcal{G}_{\tau_{i-1}}$的节点嵌入。式 4.3 等价于匹配核空间 $\mathcal{K}$中的均值元素，而 $\phi(.)$ 是由核 $\mathcal{K}(.,.)$诱导的再生核希尔伯特空间的特征映射。在我们的实验中，我们使用高斯核的混合 $\mathcal{K}(x, y) = \sum_{\alpha_i} e^{-\alpha_i ||x-y||_2}$, $\alpha_i = 1, 0.1, 0.01$。下界 B_l 和上界 B_u 约束旨在确保大多数实例具有合理的权重。

4.3 SEA-ER 的理论分析

本节通过引入失真率的概念，为我们提出的方法提供理论基础，该概念有助于刻画图神经网络的拓扑感知能力。

定义 3（失真率）。给定两个度量空间 $(\mathcal{Q}, d)$ 和 $(\mathcal{Q}', d')$，若存在常数 $r > 0$，使得 $\forall u, v \in \mathcal{E}$，则称映射 $\gamma: \mathcal{Q} \mapsto \mathcal{Q}'$, γ 具有畸变 $\alpha \geq 1$,

$$rd(u, v) \leq d'(\gamma(u), \gamma(v)) \leq \alpha r d(u, v).$$



(a) 任务 1 性能的演化　(b) Arxiv，恒定图　(c) Arxiv，演化图

图 2：恒定图与演化图设置下 Arxiv 上裸模型的遗忘动态。(a) 展示了在恒定图与演化图设置下，从任务 1 转换到任务 2 时，任务 1 灾难性遗忘的变化。(b) 和 (c) 分别为裸模型在归纳式和传导式设置下的完整灾难性遗忘矩阵（x，y 轴分别表示 i，j 在 $r_{i,\,j} - r_{i,\,i}$）。颜色密度表示遗忘程度（颜色越深，遗忘越严重）。

失真率量化了嵌入空间中的距离与原始图空间中距离的比值。接近 1 的低失真率表明，嵌入有效保留了原始图的结构信息。这一度量使我们能够评估图神经网络的拓扑感知能力，并为我们分析经验选择策略与底层图拓扑之间的关系提供理论基础。

命题 4.1。给定集合任务 $\tau_1, ..., \tau_m$ 及其对应的顶点集 $\mathcal{V}_1, ..., \mathcal{V}_m$。设 $\mathcal{P}_m = \{P_1, ..., P_{m-1}\}$ 为经验缓冲区，其中每个 $P_i \in \mathcal{P}_m$ 的基数相同，且覆盖τ_i中的所有类别。设 F 为失真率为 α的已学习图神经网络模型。假设损失函数 L 有界、连续且光滑，则有，

$$R^L_{\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | \mathcal{V}_i \setminus P_i)}(F) \leq R^L_{\mathbb{P}(\mathbf{y}_v, \mathbf{G}_v | P_i)}(F) + \mathcal{O}(\alpha D_s(\mathcal{V}_i \setminus P_i, P_i)),$$

其中

$$D_s(\mathcal{V}_i \setminus P_i, P_i) = \max_{u \in \mathcal{V}_i \setminus P_i} \min_{v \in P_i} d_{\text{spd}}(v, u).$$

命题 4.1 的证明见附录 C。该命题表明，模型的灾难性遗忘程度（以其在过去数据上保持性能的能力衡量）受到经验缓冲区上的训练损失的约束，并进一步受到多种因素的共同影响，包括模型的失真率以及经验缓冲区内部的结构距离。这凸显了选择合适经验缓冲区时图结构的关键作用，因为结构距离会显著影响性能界。因此，命题 4.1验证了用于选择经验缓冲区的优化目标，即式 4.1，并为我们提出的经验选择策略提供了性能保证。

5 实验

本节将对理论结果进行实证验证，并评估我们为 NGCL 提出的经验回放方法。在这项实证研究中，我们旨在回答以下问题：(1) 结构偏移是否确实会影响 NGCL 中图神经网络的可学习性和灾难性遗忘？(2) 命题 4.1中讨论的图神经网络失真率在实践中是否成立？(3) 我们提出的经验回放方法与现有的图神经网络经验回放方法（Zhou & Cao，2021a；Ahrabian 等，2021；Kim 等，2022）相比，在 NGCL 中表现如何？受篇幅限制，正文仅总结实验的关键细节；关于数据集、实验设置及更多结果的详细说明则见附录 D。

数据集与一般设置。总体而言，我们严格遵循（Su等人，2023年；Zhang 等，2022）中的实验设置。我们在真实世界数据集和合成数据集上进行实验。真实世界数据集包括 OGB-Arxiv（Hu 等人，2020）、Reddit（Hamilton 等人，2017）和 CoraFull（Bojchevski & G¨unnemann，2017）。为生成合成数据集，我们采用广为使用的上下文随机块模型（cSBM）（Deshpande 等人，2018）（cSBM 的更多详情见附录 D.4）。实验设置采用广泛使用的任务持续学习（task-CL）设置（Su等人，2023年；Zhang 等，2022），从数据集中提取一个 k 类分类任务

表1：SEA-ER与现有经验回放NGCL框架的性能比较（↑ 越高越好）。结果取五次试验的平均值。**粗体字母**表示该条目在相应数据集上取得了最佳性能。所有经验回放方法均使用训练集大小的5%作为经验回放缓冲区的大小。

数据集	CoraFull		OGB-Arxiv		Reddit	
性能指标	FAP (%) ↑	FAF (%) ↑	FAP (%) ↑	FAF (%) ↑	FAP(%) ↑	FAF (%) ↑
裸模型	61.41 ± 1.2	-30.36 ± 1.3	59.24 ± 3.6	-33.72 ± 3.3	68.60 ± 1.8	-23.87 ± 1.7
联合训练	92.34 ± 0.6	N.A.	94.36 ± 0.4	N.A.	95.27 ± 1.2	N.A.
ER:IID-Random	77.24 ± 5.8	-16.73 ± 5.6	79.72 ± 4.6	-14.89 ± 4.5	80.08 ± 3.3	-10.89 ± 2.5
ER-deg	86.24 ± 0.7	-8.73 ± 0.7	85.28 ± 0.8	-10.89 ± 1.9	85.08 ± 0.9	-8.89 ± 0.9
ER-infl	87.99 ± 0.6	-8.31 ± 0.5	86.42 ± 1.2	-7.65 ± 1.3	86.98 ± 1.4	-6.82 ± 1.6
ER-rep	88.95 ± 0.9	-7.55 ± 0.8	88.12 ± 0.9	-8.23 ± 0.9	86.02 ± 1.2	-7.21 ± 1.3
SEA-ER（不含 sa）（我们的方法）	89.13 ± 0.8	-7.21 ±0.9	89.19 ± 0.7	-7.13 ± 0.5	88.48 ± 1.7	-6.10 ± 1.6
ER:IID-Random，使用 sa	80.24 ± 6.9	-14.73 ± 4.8	81.72 ± 5.2	-13.23 ± 5.2	83.43 ± 2.8	-10.32 ± 3.8
ER-deg w. sa	87.32 ± 0.5	-7.89 ± 0.5	86.88 ± 0.6	-9.32 ± 2.0	87.11 ± 1.0	-6.89 ± 1.0
ER-infl w. sa	88.88 ± 0.5	-8.01 ± 0.5	89.42 ± 1.3	-5.65 ± 1.7	90.64 ± 1.6	-3.63 ± 1.3
ER-rep w. sa	89.06 ± 0.7	-7.00 ± 0.9	88.92 ± 0.8	-8.00 ± 1.1	88.99 ± 1.4	-4.21 ± 1.1
SEA-ER（我们的方法）	**91.67** ± 1.3	**-4.01** ± 1.4	**92.88** ± 1.1	**-3.08** ± 0.7	**92.89** ± 1.5	**-2.72** ± 1.4

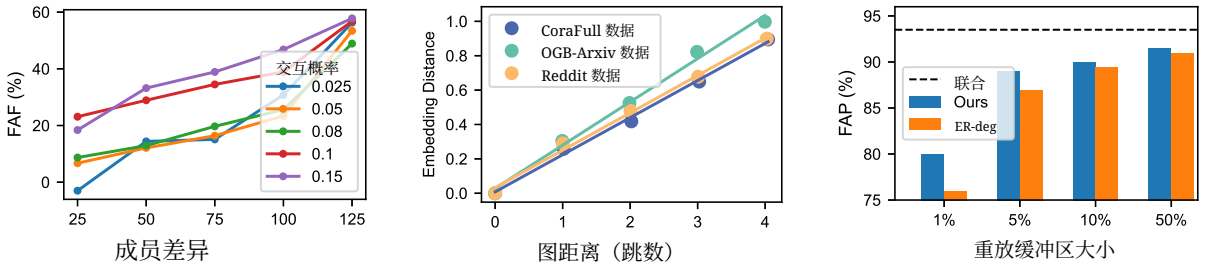
数据集	OGB-Arxiv	Reddit	CoraFull
# 顶点	169,343	227,853	19,793
# 边	1,166,243	114,615,892	130,622
# 类别	40	40	70
# 任务	20	20	35
# 顶点 / # 任务	8,467	11,393	660
# 边 / # 任务	58,312	5,730,794	4,354

表2：各数据集的持续学习设置。

供每个训练会话使用。例如，在包含 40 个类别的 OGB-Arxiv 数据集中，我们将其划分为 20 个任务：任务 1 是类别 0 和 1 之间的二分类任务，任务 2 是类别 2 和 3 之间的二分类任务，依此类推。在每个任务中，系统只能访问由当前及更早学习阶段的顶点诱导出的图，这与第 3.1 节中的表述一致。表 2 简要介绍了这些数据集，以及它们如何划分为不同的节点分类任务。我们采用近期 NGCL基准（Zhang 等，2022）中的实现来构建任务持续学习设置，并严格遵循其设置（例如训练集/验证集/测试集划分）。

评估指标。设 $r_{i,j}$ 表示模型依次经过从 1 到 i 的一系列任务训练后，在任务 j 上的性能（准确率）。任务 j 在模型依次经过从 1 到 i 的一系列任务训练后的遗忘程度由 $r_{i,j} - r_{j,j}$衡量。我们使用最终平均性能（FAP）：$= \sum_j^m r_{m,j}/m$ 以及最终平均遗忘（FAF）：$= (\sum_j^m r_{m,j} - r_{j,j})/m$ 来衡量 NGCL 框架的整体有效性。这些指标用于衡量模型在接触所有任务后的性能。对于所选的这两项指标，任务排列的顺序不应影响经验回放的性能（关于这一点的更多结果和讨论见附录 D

基线方法。我们将 SEA-ER 的性能与以下经验回放基线方法进行比较。所有基线方法默认采用标准回放策略，即不含 β的式 4.2。ER-IID-Random 使用从每个任务中随机选取顶点的经验缓冲区来更新各任务。
ER-IID-Random 用作基线，以展示不同经验缓冲区选择策略的有效性。我们采用当前最先进的面向图神经网络的经验回放框架进行比较。ER-rep（Zhou & Cao，2021a）根据顶点的表示选择经验缓冲区。ER-deg（Ahrabian 等，2021）根据顶点的度选择经验缓冲区。ER-infl（Kim 等，2022）根据顶点表示的影响力分数选择经验缓冲区。我们采用基准论文（Zhang 等，2022）中两个基线方法（Ahrabian 等，2021；Zhou & Cao，2021a）的实现，并根据我们的理解实现了（Kazemi 等，2020）的方法，因为其代码未公开。SEA-ER（不含 sa）是我们提出的、采用标准回放策略的经验回放选择策略；SEA-ER 则是我们提出的、采用学习目标式 4.2 的经验回放选择策略。后缀 w. sa 表示使用式 4.2。除上述框架外，我们还纳入



(a) cSBM 的 FAF　(b) 畸变　(c) 性能与缓冲区大小的关系

图 3：图 3(a) 展示了对 cSBM 模型进行不同配置的实验。图3(b)展示了图距离与嵌入距离之间的畸变。图 3(c) 展示了关于回放缓冲区大小影响的实验（不进行结构对齐）。

NGCL 的两个自然基线：**裸模型** 和 **联合训练**。**裸模型** 指未采用任何持续学习技术的骨干图神经网络，作为持续学习性能的下界。另一方面，**联合训练** 会在所有任务上同时训练骨干图神经网络，因此不会出现遗忘问题，并作为持续学习性能的上界。

5.1 实验结果总结

结构偏移的影响。 我们首先展示演化图结构（结构偏移）对持续学习系统的影响。为此，我们比较裸模型的遗忘动态（$r_{i,\,j} - r_{i,\,i}$ 在不同 i、j下的变化）。OGB-Arxiv数据集的结果如图2(a)所示。在图2(a)中，x 轴上的 T_1 和 T_2 分别表示第一个和第二个任务（训练阶段），而 T_2- 表示任务 2 已经到来、但模型尚未使用任务 2 数据进行训练的状态。如图所示，在演化图的情况下，进入阶段 2 时，任务 1 会发生两次灾难性遗忘：一次是在任务 2 的顶点加入图中时（结构偏移），另一次是在模型使用任务 2 数据进行训练时。图2(b) 和图2(c)分别展示了两种设置下完整的灾难性遗忘矩阵（以 $r_{i,\,j} - r_{i,\,i}$ 为矩阵元素）。两个性能矩阵的颜色密度差异表明，在演化图设置下，现有任务会遭受更严重的灾难性遗忘。该实验结果符合我们的预期，说明结构偏移的存在会导致更严重的灾难性遗忘。

接下来，我们进一步量化结构偏移与持续学习系统性能之间的关系。我们使用 cSBM 创建一个包含两个训练阶段的 NGCL，每个训练阶段（任务）均包含一个规模为 300 的双社区图，学习目标是区分并分类两个社区中的顶点。社区成员配置由一个四元组描述，$(c_0^{(1)},\ c_1^{(1)},\ c_0^{(2)},\ c_1^{(2)}) = (150+\delta,\ 150-\delta,\ 150-\delta,\ 150+\delta)$，其中 $c_i^{(j)}$ 表示会话 j中来自社区 i 的顶点数量 j。然后，我们通过改变（1）任务/会话之间的连通性 p_{inter} （任务/会话之间存在边的概率）以及（2）不同阶段中每个社区成员的差异（δ），来改变结构偏移。直观而言，随着我们增大 δ 和任务间连通性 p_{inter}，两个训练阶段之间的结构偏移会增大，因此，正如理论结果所预测的，我们应当观察到更大的 FAF 收敛至随机猜测。如图3(a)所示，实际情况确实如此，从而验证了定理 3.1。

失真率。接下来，我们验证图神经网络的失真率确实较小。我们使用每个数据集的默认设置训练一个原始 GraphSAGE 模型。我们提取训练集 5跳邻域内的顶点，并根据这些顶点到训练集的距离对其分组。随后，我们计算每组顶点在嵌入空间中到训练集的平均嵌入距离。在图 3(b) 中，图神经网络的失真率和缩放因子体现在嵌入距离与图距离之间近似线性关系的斜率上。我们可以看到，命题 4.1 中的性能界的失真因子确实较小（接近 1），这验证了我们提出的经验选择策略。

SEA-ER 的有效性。 在表 1中，我们展示了各方法在学习完整个任务序列后的 FAP 和 FAF。平均而言，未采用任何持续学习技术的裸模型表现最差，而联合训练表现最佳。由于联合训练模型不遵循持续学习设定，而是同时对所有任务进行训练，因此 FAF 不适用于这类模型。我们的经验缓冲区选择策略 SEA-ERw.o. sa 在所提出的设定下优于 ER-

rep、ER-infl 和 ER-deg。此外，经过重要性重加权的修正学习目标公式 4.2 可以通过调整演化图结构引起的结构变化，提升所有经验回放框架的性能。

经验回放缓冲区大小的消融研究。如图3(c)所示，随着经验回放缓冲区大小的增加，SEA-ER-re 的性能可预见地逐渐收敛于联合训练设置。然而，经验回放缓冲区的大小通常受系统约束的影响，例如可用存储容量。实验表明，即使经验回放缓冲区的大小受到限制，SEA-ER 仍能非常有效地保持系统性能（由最终平均任务性能 FAP 表示）。

6 结论

本文在演化图背景下，对图持续学习中图神经网络的可学习性进行了全面的理论探索。我们的形式化分析揭示了一个关键见解：当结构变化不受控制时，图神经网络在图持续学习中可能并不总是具有可学习性。此外，我们提出了一种新颖的经验回放框架 SEA-ER，能够有效应对该问题中固有的灾难性遗忘和结构变化挑战。实证结果验证了我们的理论发现，并凸显了所提出方法的有效性。

6.1 局限性与未来工作

本文提出的 NGCL 中图神经网络灾难性遗忘下界，是以存在某种学习算法为前提的。尽管这一结果揭示了可能导致图神经网络发生灾难性遗忘的结构变化程度，但进一步定量刻画 NGCL 中结构变化与图神经网络灾难性遗忘之间的关系，将是值得深入研究的方向。这有助于加深我们对 NGCL 中图神经网络的理解，并推动改进方法的开发。
本研究主要考察了广泛使用的图神经网络架构，包括 GCN、GAT 和 SAGE。然而，在异配性显著的 NGCL 场景中，亟须探索针对异配性设计的图神经网络模型；这类模型可能采用不同的聚合与更新机制，并从根本上改变问题的解决思路。这为图持续学习领域的未来研究提供了一个重要方向，但超出了本研究的范围。

致谢

我们感谢匿名审稿人和领域主席提出的宝贵意见。JS 和 CW 得到了香港研究资助局资助的项目 17207621、17203522 和 C7004-22 G（协作研究基金）的部分支持。DZ 得到了国家自然科学基金委员会项目 62306252 和香港大学创新与发展研究院中央基金的支持。

参考文献

Kian Ahrabian、Yishi Xu、Yingxue Zhang、Jiapeng Wu、Yuening Wang 和 Mark Coates。《面向基于图的推荐系统增量学习的结构感知经验回放》。发表于*第30届 ACM 信息与知识管理国际会议论文集*，第 2832–2836 页，2021 年。

Rahaf Aljundi、Min Lin、Baptiste Goujaud 和 Yoshua Bengio。面向在线持续学习的基于梯度的样本选择学习。进展神经信息处理系统，32，2019年。

Shai Ben-David 和 Reba Schuller Borbely。一个任务相关性的概念，可为多任务学习提供可证明的保证。机器学习，73:273–287，2008年。

Shai Ben-David 和 Ruth Urner。论领域自适应的困难性以及无标签目标样本的效用。载于国际算法学习理论会议，第139–153页。Springer，2012年。

Diana Benavides-Prado 和 Patricia Riddle。持续学习中的知识迁移理论。载于终身学习智能体会议，第647–660页。PMLR，2022。

Piotr Bielak、Kamil Tagowski、Maciej Falkiewicz、Tomasz Kajdanowicz 和 Nitesh V Chawla。Fildne：一种用于动态网络嵌入增量学习的框架。基于知识的系统，236:107453，2022。

Magdalena Biesialska、Katarzyna Biesialska 和 Marta R Costa-juss`a。自然语言处理中的持续终身学习：一项综述。*arXiv* 预印本 *arXiv:2012.09823*，2020。Aleksandar Bojchevski 和 Stephan G¨unnemann。图的深度高斯嵌入：通过排序进行无监督归纳式学习。*arXiv* 预印本 *arXiv:1707.03815*，2017。Lucas Caccia、Eugene Belilovsky、Massimo Caccia 和 Joelle Pineau。采用自适应量化模块的在线学习持续压缩。载于国际机器学习会议，第1240–1250页。PMLR，2020。Jie Cai、Xin Wang、Chaoyu Guan、Yateng Tang、Jin Xu、Bin Zhong 和 Wenwu Zhu。采用神经架构搜索的多模态持续图学习。载于*2022*年*ACM Web* 会议论文集，第1292–1300页，2022。Antonio Carta、Andrea Cossu、Federico Errica 和 Davide Bacciu。深度图网络中的灾难性遗忘：图分类的入门基准。*arXiv* 预印本 *arXiv:2103.11750*，2021。Aristotelis Chrysakis 和 Marie-Francine Moens。来自不平衡数据的在线持续学习。载于国际机器学习会议，第1952–1961页。PMLR，2020。Angel Daruna、Mehul Gupta、Mohan Sridharan 和 Sonia Chernova。知识图谱嵌入的持续学习。*IEEE*机器人与自动化快报，6(2):1128–1135，2021。Shai Ben David、Tyler Lu、Teresa Luu 和 D´avid P´al。领域自适应的不可能性定理。载于第十三届国际人工智能与统计会议论文集，第129–136页。JMLR研讨会与会议论文集，2010。Matthias Delange、Rahaf Aljundi、Marc Masana、Sarah Parisot、Xu Jia、Ales Leonardis、Greg Slabaugh 和 Tinne Tuytelaars。持续学习综述：在分类任务中克服遗忘。*IEEE*模式分析与机器智能汇刊，2021。Yash Deshpande、Subhabrata Sen、Andrea Montanari 和 Elchanan Mossel。上下文随机块模型。神经信息处理系统进展，31，2018。Mehrdad Farajtabar、Navid Azizan、Alex Mott 和 Ang Li。用于持续学习的正交梯度下降。载于国际人工智能与统计会议，第3762–3773页。PMLR，2020。Falih Gozi Febrinanto、Feng Xia、Kristen Moore、Chandra Thapa 和 Charu Aggarwal。图终身学习：一项综述。*IEEE*计算智能杂志，18(1):32–51，2023年。Yutong Feng、Jianwen Jiang 和 Yue Gao。增长图上的增量学习。*openReview* 预印本 *https://openreview.net/forum?id=nySHNUlKTVw*，2020。Lukas Galke、Benedikt Franke、Tobias Zielke 和 Ansgar Scherp。用于开放世界节点分类的图神经网络终身学习。载于*2021*年国际神经网络联合会议（*IJCNN*），第1–8页。IEEE，2021。Jo˜ao Gama、Indr˙e ˇZliobait˙e、Albert Bifet、Mykola Pechenizkiy 和 Abdelhamid Bouchachia。概念漂移适应综述。*ACM*计算调查（*CSUR*），46(4):1–37，2014。Arthur Gretton、Alex Smola、Jiayuan Huang、Marcel Schmittfull、Karsten Borgwardt 和 Bernhard Sch¨olkopf。通过核均值匹配处理协变量偏移。机器学习中的数据集偏移，3(4):5，2009。William L Hamilton、Rex Ying 和 Jure Leskovec。大型图上的归纳式表示学习。载于第*31*届神经信息处理系统国际会议论文集，第1025–1035页，2017。William L. Hamilton、Rex Ying 和 Jure Leskovec。大型图上的归纳式表示学习，2018。Yi Han、Shanika Karunasekera 和 Christopher Leckie。用于检测社交媒体假新闻的持续学习图神经网络。*arXiv* 预印本 *arXiv:2007.03316*，2020。Dorit S Hochbaum。覆盖与装箱问题的近似：集合覆盖、顶点覆盖、独立集及相关问题。载于*NP*难问题的近似算法，第94–143页。1996。Weihua Hu、Matthias Fey、Marinka Zitnik、Yuxiao Dong、Hongyu Ren、Bowen Liu、Michele Catasta 和 Jure Leskovec。开放图基准：图上机器学习的数据集。*arXiv* 预印本 *arXiv:2005.00687*，2020。

Heechul Jung、Jeongwoo Ju、Minju Jung 和 Junmo Kim。深度神经网络中的少遗忘学习。*arXiv* 预印本 *arXiv:1607.00122*，2016。Seyed Mehran Kazemi、Rishab Goel、Kshitij Jain、Ivan Kobyzev、Akshay Sethi、Peter Forsyth 和 Pascal Poupart。动态图的表示学习：一项综述。机器学习研究杂志，21(70):1–73，2020。Seoyoon Kim、Seongjun Yun 和 Jaewoo Kang。Dygrain：一种动态图增量学习框架。载于第*31*届国际人工智能联合会议，*IJCAI 2022*，第3157–3163页。国际人工智能联合会议，2022。Thomas N. Kipf 和 Max Welling。基于图卷积网络的半监督分类，2017。James Kirkpatrick、Razvan Pascanu、Neil Rabinowitz、Joel Veness、Guillaume Desjardins、Andrei A Rusu、Kieran Milan、John Quan、Tiago Ramalho、Agnieszka Grabska-Barwinska 等。克服神经网络中的灾难性遗忘。美国国家科学院院刊，114(13):3521–3526，2017。Jeremias Knoblauch、Hisham Husain 和 Tom Diethe。最优持续学习拥有完美记忆且是 NP难问题。载于国际机器学习会议，第5327–5337页。PMLR，2020。Xiaoyu Kou、Yankai Lin、Shaobo Liu、Peng Li、Jie Zhou 和 Yan Zhang。基于解耦的持续图表示学习。*arXiv* 预印本 *arXiv:2010.02565*，2020。Zhizhong Li 和 Derek Hoiem。无遗忘学习。*IEEE*模式分析与机器智能汇刊，40(12):2935–2947，2017。Huihui Liu、Yiding Yang 和 Xinchao Wang。克服图神经网络中的灾难性遗忘。载于*AAAI* 人工智能会议论文集，第35卷，第8653–8661页，2021。David Lopez-Paz 和 Marc'Aurelio Ranzato。用于持续学习的梯度情景记忆。神经信息处理系统进展，第30卷，2017。Jiaqi Ma、Junwei Deng 和 Qiaozhu Mei。图神经网络的子群泛化与公平性，2021。Yao Ma、Ziyi Guo、Zhaocun Ren、Jiliang Tang 和 Dawei Yin。流式图神经网络。载于第*43*届*ACM SIGIR*信息检索研究与发展国际会议，第719–728页，2020。Giang Hoang Nguyen、John Boaz Lee、Ryan A Rossi、Nesreen K Ahmed、Eunyee Koh 和 Sungchul Kim。连续时间动态图嵌入。载于*2018*年万维网大会伴随论文集，第969–976页，2018。German I Parisi、Ronald Kemker、Jose L Part、Christopher Kanan 和 Stefan Wermter。基于神经网络的持续终身学习：综述。神经网络，113:54–71，2019年。Ievgen Redko、Emilie Morvant、Amaury Habrard、Marc Sebban 和 Youn`es Bennani。领域自适应理论综述：学习界与理论保证。*arXiv* 预印本 *arXiv:2004.11829*，2020。Andrei A Rusu、Neil C Rabinowitz、Guillaume Desjardins、Hubert Soyer、James Kirkpatrick、Koray Kavukcuoglu、Razvan Pascanu 和 Raia Hadsell。渐进式神经网络。*arXiv* 预印本 *arXiv:1606.04671*，2016。Gobinda Saha、Isha Garg 和 Kaushik Roy。用于持续学习的梯度投影记忆。*arXiv* 预印本 *arXiv:2103.09762*，2021。Hanul Shin、Jung Kwon Lee、Jaehong Kim 和 Jiwon Kim。基于深度生成式回放的持续学习。神经信息处理系统进展，第30卷，2017。Junwei Su、Difan Zou、Zijun Zhang 和 Chuan Wu。迈向演化图上的鲁棒图增量学习。2023年。Junwei Su、Lingjun Mao 和 Chuan Wu。Bg-hgnn：迈向可扩展且高效的异构图神经网络。*arXiv* 预印本 *arXiv:2403.08207*，2024a。Junwei Su、Difan Zou 和 Chuan Wu。Pres：迈向可扩展的基于记忆的动态图神经网络。*arXiv* 预印本 *arXiv:2402.04284*，2024年b。

Zhen Tan、Kaize Ding、Ruocheng Guo 和 Huan Liu。图少样本类增量学习。载于第十五届 *ACM* 国际网络搜索与数据挖掘会议，第987–996页，2022。

Chen Wang、Yuheng Qiu 和 Sebastian Scherer。通过特征交互连接图网络与终身学习。2020年。Chen Wang、Yuheng Qiu、Dasong Gao 和 Sebastian Scherer。终身图学习。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第 13719–13728 页，2022 年。Junshan Wang、Guojie Song、Yi Wu 和 Liang Wang。通过持续学习实现流式图神经网络。载于第*29*届 *ACM* 国际信息与知识管理会议论文集，第 1515–1524 页，2020年。Mitchell Wortsman、Vivek Ramanujan、Rosanne Liu、Aniruddha Kembhavi、Mohammad Rastegari、Jason Yosinski 和 Ali Farhadi。叠加中的超级掩码。神经信息处理系统进展，33:15173– 15184，2020 年。Yue Wu、Yinpeng Chen、Lijuan Wang、Yuancheng Ye、Zicheng Liu、Yandong Guo 和 Yun Fu。大规模增-量学习。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第 374–382 页，2019年。Zonghan Wu、Shirui Pan、Fengwen Chen、Guodong Long、Chengqi Zhang 和 S Yu Philip。图神经网络综述。*IEEE* 神经网络与学习系统汇刊，32(1):4–24，2020 年。Yishi Xu、Yingxue Zhang、Wei Guo、Huifeng Guo、Ruiming Tang 和 Mark Coates。Graphsail：面向推荐系统的图结构感知增量学习。载于第*29*届 *ACM* 国际信息与知识管理会议论文集，第 2861–2868 页，2020 年。Jaehong Yoon、Eunho Yang、Jeongtae Lee 和 Sung Ju Hwang。基于动态可扩展网络的终身学习。*arXiv* 预印本 *arXiv:1708.01547*，2017 年。Jaehong Yoon、Saehoon Kim、Eunho Yang 和 Sung Ju Hwang。基于加性参数分解的可扩展且顺序鲁棒的持续学习。*arXiv* 预印本 *arXiv:1902.09432*，2019年。Wenchao Yu、Wei Cheng、Charu C Aggarwal、Kai Zhang、Haifeng Chen 和 Wei Wang。Netwalk：一种用于动态网络异常检测的灵活深度嵌入方法。载于第*24*届 *ACM SIGKDD* 国际知识发现与数据挖掘会议论文集，第 2672–2681 页，2018 年。Qiao Yuan、Sheng-Uei Guan、Pin Ni、Tianlun Luo、Ka Lok Man、Prudence Wong 和 Victor Chang。持续图学习：综述。*arXiv* 预印本 *arXiv:2301.12230*，2023年。Xikun Zhang、Dongjin Song 和 Dacheng Tao。用于持续图表示学习的分层原型网络。*arXiv* 预印本 *arXiv:2111.15422*，2021 年。Xikun Zhang、Dongjin Song 和 Dacheng Tao。CGLB：持续图学习基准任务。载于第三十六届神经信息处理系统会议数据集与基准赛道，2022 年。Fan Zhou 和 Chengtai Cao。通过经验回放克服图神经网络中的灾难性遗忘。载于*AAAI* 人工智能会议论文集，第 35 卷，第 4714–4722 页，2021a。Fan Zhou 和 Chengtai Cao。通过经验回放克服图神经网络中的灾难性遗忘。载于*AAAI* 人工智能会议论文集，第 35 卷，第 4714–4722 页，2021年。Fan Zhou 和 Chengtai Cao。通过经验回放克服图神经网络中的灾难性遗忘。载于*AAAI* 人工智能会议论文集，第 35 卷，第 4714–4722 页，2021年。Indre Zliobaite。概念漂移下的学习：概述。*arXiv* 预印本 *arXiv:1010.4784*，41，2010。

关于可学习性与定理的进一步讨论3.1

在本附录中，我们将证明定理3.1。证明的核心思路是建立 NGCL-2 问题与领域自适应问题之间的形式化联系（下文将其称为 DA 问题）。我们首先简要介绍领域自适应（DA）问题。不失一般性，我们将分析重点放在二分类任务上。显然，如果二分类不可能，那么任意 k 路分类也不可能。

A.1 领域自适应

设 $\mathcal{X}$和 $\mathcal{Y}=\{0,1\}$ 分别为输入空间和输出空间，l：$\mathcal{X}\mapsto\mathcal{Y}$ 为标记函数。在领域自适应问题中，$\mathcal{X}$上存在两个不同的分布：源分布 $\mathcal{S}$和目标分布 $\mathcal{T}$。领域自适应的目标是利用源领域$\mathcal{S}$中的知识，学习一个分类器 h：$\mathcal{X}\mapsto\mathcal{Y}$，使其在目标分布上表现良好，即 $R^l_{\mathcal{T}}(h)$ 较小。与 NGCL-2 问题类似，领域自适应问题的可学习性可定义如下。

定义 4 (DA 可学习性)。 设 $\mathcal{S}$和 $\mathcal{T}$为 $\mathcal{X}$上的分布，$\mathcal{H}$为假设类，$\epsilon>0,\delta>0$。如果存在一个学习算法，能够以至少 $1-\delta$的概率，根据 $\mathcal{H}$ 生成一个分类器 h，且 h 的误差小于 ϵ，即 $Pr[R^l_{\mathcal{T}}(h)\leq\epsilon]\geq1-\delta$，那么我们称领域自适应问题是可学习的。该算法可访问来自 $\mathcal{S}$ 的大小为 m 的有标签样本L，以及来自 $\mathcal{T}$的大小为 n 的无标签样本 U。

A.2 领域自适应与 NGC 之间的联系 L-2

回顾一下，NGCL-2 问题由三个分布表征： $\mathbb{P}(\mathrm{y},\mathrm{g}_v|\mathcal{V}_1,\mathcal{G}_{\tau_1})$、 $\mathbb{P}(\mathrm{y},\mathrm{g}_v|\mathcal{V}_1,\mathcal{G}_{\tau_2})$ 和$\mathbb{P}(\mathrm{y},\mathrm{g}_v|\mathcal{V}_2,\mathcal{G}_{\tau_2})$。为简化记号，我们将 $\mathbb{P}(\mathrm{y},\mathrm{g}_v|\mathcal{V}_1,\mathcal{G}_{\tau_1})$、 $\mathbb{P}(\mathrm{y},\mathrm{g}_v|\mathcal{V}_1,\mathcal{G}_{\tau_2})$、 $\mathbb{P}(\mathrm{y},\mathrm{g}_v|\mathcal{V}_2,\mathcal{G}_{\tau_2})$ 重写为$\mathbb{P}_1^{(1)}$、$\mathbb{P}_1^{(2)}$、$\mathbb{P}_2^{(2)}$。然后，NGCL-2 问题的可学习性可正式定义如下：

定义 5 (NGCL-2 可学习性)。 给定一个具有分布 $\mathbb{P}_1^{(1)}$、$\mathbb{P}_1^{(2)}$、$\mathbb{P}_2^{(2)}$ 和标记函数 l 的 *NGCL-2*，设$\mathcal{H}$ 为一个假设类，$\epsilon>0$、$\delta>0$。如果存在一个学习算法，能够输出一个分类器 h，其来自 $\mathcal{H}$，且 h 的误差小于 ϵ的概率至少为 $1-\delta$，即$Pr[R^l_{\mathbb{P}}(h)\leq\epsilon]\geq1-\delta$，那么我们称 *NGCL-2* 问题相对于 $\mathcal{H}$ 是可学习的。该算法可访问来自 $\mathbb{P}$ 的大小为 $m+n$ 的带标签样本 L （$\mathbb{P}$ 是 $\mathbb{P}_1^{(1)}$ 与 $\mathbb{P}_2^{(2)}$ 的混合分布），以及大小为 k、来自 $\mathbb{P}_1^{(2)}$ 的无标签样本 U

为了在 NGCL-2 与领域自适应之间建立形式化联系，我们考虑 NGCL-2 问题的一个宽松版本：在该版本中，学习算法可以按照任意顺序访问来自分布 $\mathbb{P}_1^{(1)}$ and$\mathbb{P}_2^{(2)}$ 的等量数据，且只需在更新后的分布 $\mathbb{P}_1^{(2)}$ 上表现良好。我们将这个宽松版本的 NGCL-2 问题称为 NGCL-2-weak，其可学习性定义如下。

定义 6 (NGCL-2-弱可学习性)。 给定一个具有分布 $\mathbb{P}_1^{(1)},\mathbb{P}_1^{(2)},\mathbb{P}_2^{(2)}$ 和标记函数 l 的 *NGCL-2*，设 $\mathcal{H}$ 为一个假设类，$\epsilon>0,\delta>0$。我们称 *NGCL-2* 问题相对于 $\mathcal{H}$ 可学习，如果存在一种学习算法，以至少 $1-\delta$ 的概率从 $\mathcal{H}$ 输出一个分类器 h，且 h 的误差小于 ϵ，即 $Pr[R^l_{\mathbb{P}_1^{(2)}}(h)\leq\epsilon]\geq1-\delta$，其中算法可访问来自 P 的、大小为 $m+n$ 的带标签样本 L （P 是 $\mathbb{P}_1^{(1)}$ 和 $\mathbb{P}_2^{(2)}$ 的混合分布），以及来自 $\mathbb{P}_1^{(2)}$ 的、大小为 k 的无标签样本 U

显然，NGCL-2-弱可学习性的定义涵盖了 NGCL-2 可学习性。换言之，我们可以直接从定义得出以下推论。

推论 A.1. 如果 *NGCL-2*-弱问题不可学习，那么 *NGCL-2* 问题也不可学习 .

接下来，我们将建立领域自适应问题与 NGCL-2-弱问题之间的形式化联系。

定理 *A.2*（领域自适应与 *NGCL-2*-弱的联系）。令 $\mathcal{X}$ 为一个域，$\mathcal{H}$ 为 $\mathcal{X}\times\{0,1\}$ 上的假设类。假设存在一个领域自适应问题实例，其在 $\mathcal{X}$ 上的分布为 $\mathcal{S}$ 和 $\mathcal{T}$，并且可以访问来自 $\mathcal{S}$ 的 m' 个有标签样本，以及来自 $\mathcal{T}$ 的 k' 个无标签样本。类似地，假设存在一个 *NGCL-2*- 弱问题实例，其在 $\mathcal{X}$ 上的分布为 $\mathbb{P}_1^{(1)}$、$\mathbb{P}_1^{(2)}$、$\mathbb{P}_2^{(2)}$，并且可以访问来自 $\mathbb{P}_1^{(1)}$、$\mathbb{P}_1^{(2)}$ 的 $m+n$ 个有标签样本，以及来自 $\mathbb{P}_1^{(2)}$ 的 k 个无标签样本。对于相同的假设类 $\mathcal{H}$，假设 $m'=m+n$、$k'=k$、$\lambda_{\mathcal{H}}(\{\mathcal{S},\mathcal{T}\})=\lambda_{\mathcal{H}}(\{P,\mathbb{P}_1^{(2)}\})$，且 $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}、\mathcal{T})=d_{\mathcal{H}\Delta\mathcal{H}}(P、\mathbb{P}_1^{(2)})$，则领域自适应问题可学习当且仅当 *NGCL-2*-弱问题可学习。

证明。我们首先证明，在定理的前提下，如果 NGCL-2-弱问题可学习，那么领域自适应问题也可学习。

假设给定一个由分布 $\mathcal{S}$、$\mathcal{T}$ 和 $\mathcal{H}$ 构成的领域自适应问题实例。令 $\mathcal{S}$ 为 $\mathcal{S}_1$ 和 $\mathcal{S}_2$ 的混合分布。接下来，我们可以将领域自适应映射为一个 NGCL-2-弱问题，具体如下。

$$\mathcal{S}_1 = \mathbb{P}_1^{(1)}, \mathcal{S}_2 = \mathbb{P}_1^{(2)}, \mathcal{T} = \mathbb{P}_1^{(2)}$$

根据上述映射，我们有 $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}, \mathcal{T}) = d_{\mathcal{H}\Delta\mathcal{H}}(P, \mathbb{P}_1^{(2)})$，且 $\lambda_{\mathcal{H}}(\{\mathcal{S}, \mathcal{T}\}) = \lambda_{\mathcal{H}}(\{P, \mathbb{P}_1^{(2)}\})$，其中 P 是分布 $\mathbb{P}_1^{(1)}$ 和 $\mathbb{P}_1^{(2)}$ 的混合分布。

根据假设，NGCL-2-weak 问题是可学习的。根据定义 6，这意味着存在一个学习算法，该算法接收来自 $m+n$ 的带标签数据 P 以及 k 来自 $\mathbb{P}_1^{(2)}$ 的无标签数据，并以至少 $1-\delta$ 的概率输出一个分类器 h，其误差至多为 ϵ，相对于 $\mathbb{P}_1^{(2)}$ 而言。根据上述映射以及定理的前提 $m+n=m'$，$k'=k$，同一个学习算法将满足定义 4。这表明，NGCL-2-weak 的可学习性蕴含领域自适应问题的可学习性。

现在我们证明相反的结论：在定理的前提下，如果领域自适应问题是可学习的，那么 NGCL-2-弱问题也是可学习的。

我们反向构造。假设给定一个由分布$\mathbb{P}_1^{(1)}$、$\mathbb{P}_1^{(2)}$、$\mathbb{P}_1^{(2)}$ 构成的 NGCL-2-弱问题实例，以及一个 $\mathcal{H}$。令 $\mathcal{P}$ 为 $\mathbb{P}_1^{(1)}$ 和 $\mathbb{P}_1^{(2)}$ 的混合分布。然后，我们可以将 NGCL-2- 弱问题映射到领域自适应问题，如下所示。

$$P = \mathcal{S}, \mathbb{P}_1^{(2)} = \mathcal{T}$$

类似地，在上述映射下，我们有 $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}, \mathcal{T}) = d_{\mathcal{H}\Delta\mathcal{H}}(P, \mathbb{P}_1^{(2)})$，且 $\lambda_{\mathcal{H}}(\{\mathcal{S}, \mathcal{T}\}) = \lambda_{\mathcal{H}}(\{P, \mathbb{P}_1^{(2)}\})$。

根据假设，领域自适应问题是可学习的。根据定义4，这意味着存在一种学习算法，它接收来自 m'的带标签数据，以及来自 $\mathcal{S}$的 k'无标签数据（来自 $\mathcal{T}$），并以至少 $1-\delta$的概率输出一个分类器 h，其相对于 ϵ 的误差至多为 $\mathcal{T}$。根据上述映射以及定理前提中的 $m+n=m'$，$k'=k$，同一个学习算法将满足定义 6。这表明，领域自适应问题的可学习性蕴含着 NGCL-2-弱问题的可学习性。

这就完成了定理A.2的证明。 □

既然我们已经证明了领域自适应问题的可学习性等价于 NGCL-2-弱问题的可学习性。结合定理 A.2 和推论 A.1，我们立即得到以下关于 NGCL-2 问题与定义5 以及领域自适应问题之间联系的推论。

推论 *A.3*（领域自适应与 *NGCL-2* 的联系）。设 $\mathcal{X}$ 是一个域，$\mathcal{H}$ 是 $\mathcal{X} \times \{0,1\}$ 上的假设类。假设存在一个领域自适应问题实例，其在 $\mathcal{X}$ 上的分布为 $\mathcal{S}$ 和 $\mathcal{T}$，并且可以访问来自 $\mathcal{S}$ 的 m' 个有标签样本以及来自 $\mathcal{T}$ 的 k' 个无标签样本。类似地，假设存在一个 *NGCL-2* 问题实例，其在 $\mathcal{X}$ 上的分布为 $\mathbb{P}_1^{(1)}$、$\mathbb{P}_1^{(2)}$、$\mathbb{P}_1^{(2)}$，并且可以顺序访问来自 $\mathbb{P}_1^{(1)}$ 的 m 个有标签样本、来自 $\mathbb{P}_1^{(2)}$ 的 n 个有标签样本，以及来自 $\mathbb{P}_1^{(2)}$ 的 k 个无标签样本。对于相同的假设空间 $\mathcal{H}$，假设 $m'=m+n$、$k'=k$、$\lambda_{\mathcal{H}}(\{\mathcal{S}, \mathcal{T}\}) = \lambda_{\mathcal{H}}(\{P, \mathbb{P}_1^{(2)}\})$ 以及 $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}, \mathcal{T}) = d_{\mathcal{H}\Delta\mathcal{H}}(P, \mathbb{P}_1^{(2)})$，则有：如果领域自适应不可学习，那么 *NGCL-2* 问题也不可学习。

A.3 定理 3.1

推论 A.3上文确立了领域自适应问题与 NGCL-2 问题之间的形式联系。在所提出的同一映射下，领域自适应问题不可行意味着 NGCL-2 问题也不可行。领域自适应是一个经过深入研究的问题，其可学习性已有成熟的认识。

David 等人 (2010) 已证明，源域与目标领域之间较小的散度距离是领域自适应成功的“必要”条件。我们在下文重述该定理。

定理 A.4 ((David 等人，2010))。设 $\mathcal{X}$ 为某个域集合，且 $\mathcal{H}$ 为定义在 $\mathcal{X} \times \{0,1\}$上的函数类。对于每个 $c>0$，都存在概率分布 $\mathcal{S}$、$\mathcal{T}$ 定义在 $\mathcal{X}$ 上，使得对于任意领域自适应学习器以及任意整数 m、$n>0$，都存在一个标记函数 l：$\mathcal{X} \mapsto 0, 1$，使得 $\lambda_{\mathcal{H}}(P, Q) \leq c$，且该领域自适应问题不可学习。

定理 A.4 的证明见（David 等人，2010）。接着，定理 3.1的证明可由定理 A.4 和推论A.3立即得出。

B 表达能力与样本复杂度

在本附录中，基于前一附录中建立的联系，我们进一步讨论样本复杂度和表达能力。

定理 B.1（假设空间表达能力的必要性）。设 $\mathcal{X}$ 是某个域集合，$\mathcal{H}$ 是定义在 $\mathcal{X}\times\{0,1\}$ 上的函数类，其 *VC* 维数远小于 $|\mathcal{X}|$。那么，对任意 $c>0$，都存在一个 *NGCL-2* 问题，使得 $d_{\mathcal{H}\Delta\mathcal{H}}(\mathbb{P},\ \mathbb{P}_1^{(2)}) \leq c$，其中 $\mathbb{P}$ 是 $\mathbb{P}_1^{(1)}$ 和 $\mathbb{P}_2^{(2)}$ 的混合分布，且无论从 $\mathbb{P}_1^{(1)}$、$\mathbb{P}_2^{(2)}$ 获取多少带标签数据，以及从 $\mathbb{P}_1^{(2)}$ 获取多少无标签数据，该 *NGCL-2* 问题都不可学习。

定理 B.1 形式化了这一直觉：假设类（图神经网络模型）需要具备足够的表达能力，以捕捉 NGCL 中的不同分布，才能实现可学习。

定理 B.2 (样本复杂度)。设 $\mathcal{X}$ 是 $\mathbf{R}^d$中的某个紧域集合，$\mathcal{H}$ 是定义在$\mathcal{X}\times\{0,1\}$上的函数类。假设标记函数 l：$\mathcal{X}\mapsto\{0,1\}$ 是 α-利普希茨连续的。那么，若 $m+n+k<\sqrt{(1-2(\epsilon+\delta))(\alpha+1)^d}$，对于每个 $c>0$，都存在一个 *NGCL-2* 问题，使得 $\lambda_{\mathcal{H}}(\{\mathbb{P}_1^{(1)},\mathbb{P}_2^{(2)},\mathbb{P}_1^{(2)}\}) \leq c, d_{\mathcal{H}\Delta\mathcal{H}}(P,\mathbb{P}_1^{(2)}) \leq c$，其中 $\mathbb{P}$ 是 $\mathbb{P}_1^{(1)}$ 和 $\mathbb{P}_2^{(2)}$ 的混合分布，并且该 *NGCL-2* 问题不可学习。

定理 B.2 指出，即使定理 3.1 和定理 B.1 中的两个“必要”条件都得到满足，仍然存在一项采样要求，该要求取决于输入域的维数和标记函数的平滑性。这表明，对于 NGCL 而言，能够访问先前数据至关重要。

B.1 定理的证明 （定理 B.1 和定理 B.2

证明思路与前文所述类似。（David 等人，2010）证明了，假设空间的表达能力是领域自适应成功的一个“必要”条件。下面重述该定理。

定理 B.3. 设 $\mathcal{X}$ 为某个域集合，$\mathcal{H}$ 为定义在 $\mathcal{X}\times\{0,1\}$ 上的函数类，其 *VC* 维数远小于 $|\mathcal{X}|$。那么，对于每个 $c>0$，都存在定义在 $\mathcal{X}$上的概率分布 Q、P，使得对于任意领域自适应学习器以及任意整数 m、$n>0$，都存在一个标记函数 l： $\mathcal{X}\mapsto 0,1$ 使得 $d_{\mathcal{H}\Delta\mathcal{H}}(P、Q)\leq c$，且该领域自适应问题不可学习。

定理 B.3 的证明可见（David 等人，2010）。然后，定理 B.1可由定理 B.3 和推论 A.3立即得出。

（Ben-David 和 Urner，2012）证明了，即使领域自适应问题的散度距离很小，且假设空间具有足够强的表达能力，领域自适应问题仍需满足一定的采样要求才能实现可学习。我们在下面重述该定理。

定理 *B.4.* 设 $\mathcal{X}\subset R^d$ 和 $\mathcal{H}$ 是 $\mathcal{X}\times\{0,1\}$ 上的一个函数类。假设标记函数 l 是 α-利普希茨连续的。对于每个 $c>0$，都存在 $\mathcal{S}$、$\mathcal{T}$，它们属于 $\mathcal{X}$，使得 $\lambda_{\mathcal{H}}\leq c$ 且 $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S},\mathcal{T})\leq c$；如果 $m+n\geq\sqrt{(\alpha+1)^d(1-2(\epsilon+\delta))}$，则该领域自适应问题不可学习，其中 m、n、ϵ、δ 由定义 *4* 给出。

定理 B.4 的证明见（Ben-David &Urner，2012）。因此，定理 B.2 的证明可直接由定理 B.4和推论 A.3。

注 B.5. 除了对可学习性的研究外，还有大量研究探讨了领域自适应的不同技术。形式化的联系*A.3* 使我们能够严谨地关联 *NGCL-2* 问题与领域自适应问题，并谨慎地将领域自适应中的认识和技术迁移或拓展到 *NGCL-2*。

C 证明命题 4.1

在本附录 中，我们给出命题 4.1的证明。首先重述畸变的定义 .

定义 7 (失真率)。给定两个度量空间 $(\mathcal{Q},d)$ 和$(\mathcal{Q}',d')$，以及一个映射 $\gamma:\ \mathcal{Q}\mapsto\mathcal{Q}',\gamma$，若存在常数 $r>0$，使得 $\forall u, v\in\mathcal{E}, rd(u,v)\leq d'(\gamma(u)\ \gamma(v))\leq\alpha rd(u,v)$，则称其具有畸变 $\alpha\geq 1$。

证明。设 $P_i\subset\mathcal{V}_i$ 为通过求解方程4.1所选出的经验回放集。设 s 为具有距离度量 d_s的目标结构。设 g 为一个畸变为 α、缩放因子为 r 且预测函数为 g的给定图神经网络模型。

令 $\sigma: \mathcal{V}_i \mapsto P_i$ 表示一个映射，它将 v 从 $\mathcal{V}_i$ 映射到 P_i中距离最近的顶点，即，

$$\sigma(v) = \underset{u \in P_i}{\operatorname{argmin}}\, d_s(u, v). \tag{C.1}$$

考虑顶点 v, $L(g \circ g(v)\ y_v)$ 的损失。根据假设， L 是平滑的，并记 B_L^{up1} 为关于 $g \circ g(v)$ 的一阶导数上界，记 B_L^{up2} 为关于 y_v 的一阶导数上界。接下来，考虑 $L(g \circ g(v)\ y_v)$ 关于 $\sigma(v)$ 的泰勒展开，如下所示

$$\begin{aligned} L(g \circ g(v), y_v) \leq\ & L(g \circ g(\sigma(v)), y_{\sigma(v)}) + \\ & B_L^{up1} ||g \circ g(v) - g \circ g(\sigma(v))|| \\ & + B_L^{up2} ||y_v - y_{\sigma(v)}|| \end{aligned} \tag{C.2}$$

接下来，我们逐项考察上面的不等式，先看 $||g \circ g(v) - g \circ g(\sigma(v))||$。记 h_v 和 $h_{\sigma(v)}$ 分别为顶点 v 和 $\sigma(v)$ 的嵌入。那么，我们有：

$$||g \circ g(v) - g \circ g(\sigma(v))|| = ||g(h_v) - g(h_{\sigma(v)})|| \tag{C.3}$$

根据定义 7 中给出的畸变定义，我们有：

$$||h_v - h_{\sigma(v)}|| \leq r\alpha d_s(v, \sigma(v)) \tag{C.4}$$

根据假设，预测函数 g 是平滑的。令 B_g 表示预测函数 g 的一阶导数上界。那么，我们有

$$||g(h_v) - g(h_{\sigma(v)})|| \leq B_g^{up} ||h_v - h_{\sigma(v)}|| \tag{C.5}$$

将所有这些代回，可得

$$||g \circ g(v) - g \circ g(\sigma(v))|| \leq B_g^{up} r\alpha d_s(v, \sigma(v)) \tag{C.6}$$

接下来，让我们考虑 $||y_v - y_{\sigma(v)}||$。类似地，根据数据平滑性假设和畸变，我们有

$$||y_v - y_{\sigma(v)}|| \leq B_l ||h_v - h_{\sigma(v)}|| \leq B_l^{up} r\alpha d_s(v, \sigma(v)) \tag{C.7}$$

将这些代回最初的不等式，可得

$$\begin{aligned} & L(g \circ g(v), y_v) \leq L(g \circ g(\sigma(v)), y_{\sigma(v)}) + \\ & \quad B_L^{up1} B_g^{up} r\alpha d_s(v, \sigma(v)) + B_L^{up2} B_l^{up} r\alpha d_s(v, \sigma(v)) \\ & \leq L(g \circ g(\sigma(v)), y_{\sigma(v)}) + \\ & (B_l^{up} B_L^{up1} + B_g^{up} B_L^{up2}) r\alpha \max_{v' \in \mathcal{V}_i} d_s(v', \sigma(v')) \\ & = L(g \circ g(\sigma(v)), y_{\sigma(v)}) + \\ & (B_l^{up} B_L^{up1} + B_g^{l} B_L^{up2}) r\alpha D_s(\mathcal{V}_i, P_i). \end{aligned} \tag{C.8}$$

由于上述不等式对每个顶点 v都成立，因此我们有

$$\begin{aligned} R^L(g \circ g, \mathcal{V}_i) \leq\ & R^L(g \circ g, P_i) + \\ & (B_l^{up} B_L^{up1} + B_g^{up} B_L^{up2}) r\alpha D_s(\mathcal{V}_i, P_i) \\ & = R^L(g \circ g, P_i) + \mathcal{O}(\alpha D_s(\mathcal{V}_i, P_i)) \end{aligned} \tag{C.9}$$

□

D 附加实验细节

在本附录中，我们提供了更多实验结果，并详细介绍实验设置，以确保可复现性。

D.1 硬件与系统

本文所有实验均在以下机器上进行：CPU：两颗 Intel Xeon Gold 6230 2.1G，20C/40T，10.4GT/s，27.5M Cache，Turbo，HT（125W）；DDR4-2933。GPU：四块用于 NVLink 的 NVIDIA Tesla V100 SXM2 32G GPU 加速卡。内存：256GB（8 x 32GB）RDIMM，3200 MT/s，双列。

操作系统：Ubuntu 18.04LTS

D.2 数据集与处理

D.2.1 数据集描述

OGB-Arxiv. OGB-Arxiv数据集 Hu 等人 (2020）是一个用于节点分类的基准数据集。该数据集基于 arXiv 电子预印本库构建，这是一个广受欢迎的平台，供研究人员分享预印本。图结构通过连接相互引用的论文构建。节点特征包括论文摘要、标题和作者姓名的文本。每个节点被分配到 40 个类别之一，这些类别对应论文的主要学科领域。

Cora-Full。 Cora-FullBojchevski & G¨unnemann （2017）是用于节点分类的基准数据集。与 OGB-Arxiv 类似，它是一个包含 70 个类别的引文网络。

Reddit。 Reddit 数据集Hamilton 等人 （2017）是用于节点分类的基准数据集，包含来自 Reddit.com 网站的帖子和评论。每个节点代表一篇帖子或一条评论，每条边代表帖子或评论之间的回复关系。

表 3：各数据集的持续学习设置。

数据集	OGB-Arxiv	Reddit	CoraFull
# 顶点	169,343	227,853	19,793
# 边	1,166,243	114,615,892	130,622
# 类别	40	40	70
# 任务	20	20	35
# 顶点 / # 任务	8,467	11,393	660
# 边 / # 任务	58,312	5,730,794	4,354

D.2.2 许可证

本文使用的数据集整理自现有的公共数据源，并遵循其许可协议。OGB-Arxiv 采用开放数据共用署名许可协议（ODC-BY）。Cora-Full 数据集和 Reddit 数据集是根据公开来源（公开论文和 Reddit 帖子）构建的两个数据集，作者未为其附加许可协议。

D.2.3 数据处理

对于这些数据集，我们按照Zhang 等（2022）的方法，移除了 Reddit-CL 的第 41 类。此举旨在确保每个数据集的类别数量为偶数，以便将其划分为一系列二分类任务。对于所有数据集，训练-验证-测试集划分比例分别为 60%、20% 和 20%。训练集、验证集和测试集通过随机抽样划分，因此，不同轮次随机抽样所得的划分可能会导致性能略有差异。

D.3 持续训练过程

假设 τ_1， τ_2， ...τ_m 是根据数据集创建的分类任务集合。那么，标准的持续训练流程如算法所示。1 ，图示说明见图7。

D.4 上下文随机块模型

我们使用上下文随机块模型（Deshpande 等人，2018）为每个学习阶段创建一个图。对于每个图，我们创建 300 个顶点，并为它们分配社区标签 $\{0,1\}$。我们将不同社区中顶点比例的特定取值称为一种配置，用于改变社区比例的公式如下：配置 $i = (150 + (i-1) * 25, 150 - (i-1) * 25, 150 - (i-1) * 25, 150 + (i-1) * 25)$。随着 i 增大，两个学习阶段的比例差异也随之增大。直观而言，这应当使我们能够控制不同学习阶段之间的分布差异。

接下来，我们采用标准的随机块模型。我们使用 $p_{intra} = 0.15$ 在同一社区的顶点之间随机创建边，并使用 $p_{inter} = 0.1$ 在不同社区的顶点之间随机创建边。然后，我们使用 $p_{stage} = [0.025, 0.05, 0.08, 0.1, 0.15]$ 在不同学习阶段的顶点之间创建边。

接下来，每个顶点的节点特征 x_v 根据以下公式生成，

$$x_v = \sqrt{\frac{\mu}{n}} y_v u + \frac{Z_v}{p},$$

其中，$y_v \in \{0,1\}$ 是顶点 v的社区标签，u，$Z_v \sim \mathcal{N}(0,\ I)$ 是维度为 p的随机高斯分布，n 是图中的顶点数，μ 是模型的一个超参数，用于控制不同社区顶点的特征之间的可分性。在我们的实验中，每个阶段的图均使用 $\mu = 5$，$n = 300$，$p = 500$ 。

D.5 图神经网络详情

在我们的实验中，我们使用两种流行的图神经网络模型：GCN（Kipf 和 Welling，2017）和 GraphSAGE（Hamilton 等人，2018）。在实现方面，我们使用 DGL 提供的默认实现来构建这些模型。在所有实验中，我们使用两个图神经网络层，并在其后添加一个多层感知机层作为读出函数。最后一个隐藏层使用 Sigmoid 作为激活函数，其余层使用 ReLU。每项实验中，我们仅调整图神经网络层的隐藏维度。

D.6 归纳式与直推式的区别

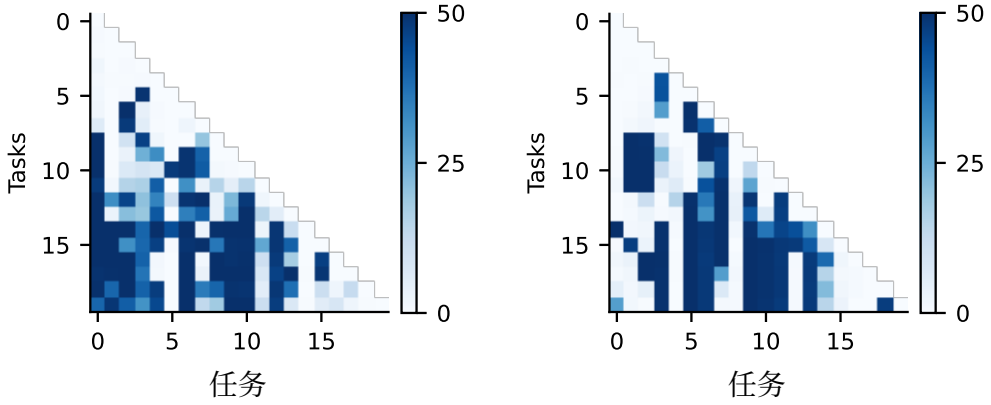
在本小节中，我们补充报告了演化图和恒定图设置下其余数据集的结果，并进一步展示了直推式与归纳式 NGCL（存在结构偏移）之间区别的示意图。补充的实验结果见图4 和5。直推式与归纳式 NGCL 之间区别的图示见图8

D.7 任务顺序排列

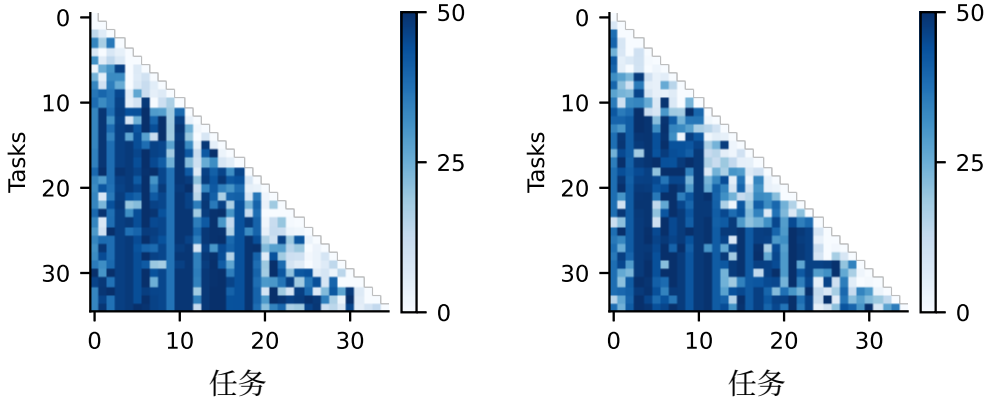
在实证研究中，每个数据集的任务顺序只构建一次。原因如下：本文重点研究的持续学习策略是经验回放，即无论先前任务以何种顺序出现，模型都会接触到（先前任务的一个小子集）。此外，本文使用的性能指标是 FAP（最终平均性能）和 FAF（最终平均遗忘）。这些指标衡量模型接触所有任务后的性能。因此，任务排列的顺序可能会影响中间的学习动态，但不应影响采用这两个指标评估的经验回放性能。图6 进一步验证了任务顺序排列对经验回放在所选指标上的性能影响极小。

算法1 持续训练

输入： τ_1，τ_2，$...\tau_m$ //从数据集创建的分类任务。对于$_{\tau_i}\tau$ 在 τ_1，τ_2，$...\tau_m$：1. 使用来自 τ_i的数据更新图结构；2. 获取3 的 V_i^{train}。更新现有的图神经网络模型，并在 V 上训练新的任务预测头$_{\text{训练}i}$4. 如果需要评估，则在当前任务和先前任务上评估模型



(a) Reddit，遗忘，演化图　(b) Reddit，遗忘，恒定图



(c) CoraFull，遗忘，演化图　(d) CoraFull，遗忘，恒定图

图 4：裸模型在 CoraFull 和 Reddit 数据集上的直推式与归纳式设置下的性能矩阵。

E 完整框架与伪代码

E.1 总体框架

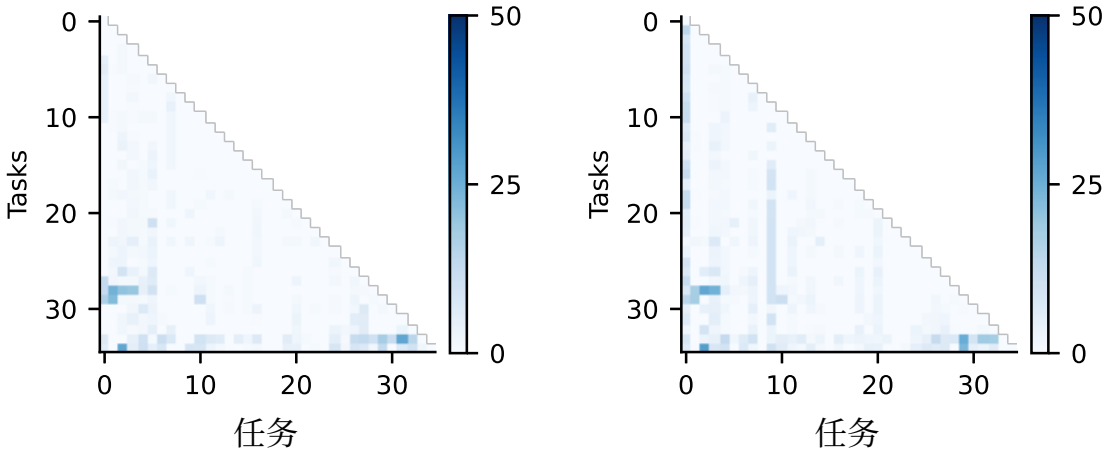
我们框架的持续学习流程总结于算法2和图9 展示了该框架在两阶段持续学习中的训练过程。

E.2 选择回放集

回放集选择策略如公式所示 4.1。显然，当 $D_i = V_i$ 时，该问题可简化为 k-中心问题（Hochbaum，1996），即选择一组由 k 个顶点构成的集合，使其余顶点到这 k 个选定顶点的距离最小化。尽管 k-中心问题是 NP 难问题，但已有高效的贪心近似算法，例如，选择与已确定集合距离最远的顶点（Hochbaum，1996）。然而，这种启发式方法可能会选中位于长路径末端、与其他顶点几乎没有连接的顶点。为缓解这一问题，我们提出根据顶点的度对距离加权。此外，与其他顶点采样算法类似，我们改进了上述选择过程：根据贪心准则和顶点的度为每个顶点分配一个概率，为算法留出一定的探索空间。

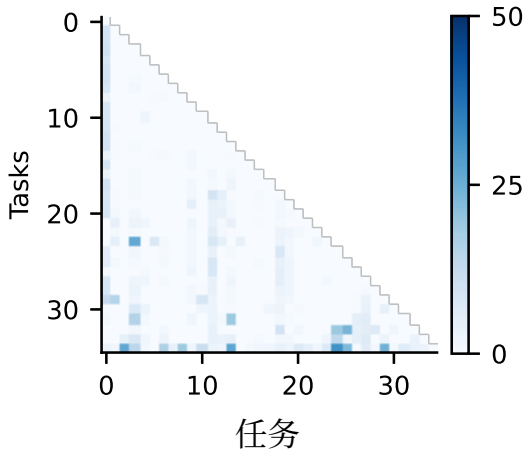
E.3 复杂度分析

从算法的步骤可以很容易看出其时间复杂度；其计算复杂度为 O(N)，其中 N 表示图中的顶点数。值得注意的是，我们的方法与基线方法具有一个共同的



(a) SEA-ER，Coral-Full

(b) SEA-ER，OGB-Arixv



(c) SEA-ER，Reddit

图 5：SEA-ER 在 CoraFull 和 Reddit数据集上的遗忘性能矩阵。

特征，即问题的形式化可能固有地属于 NP难问题。然而，我们的方法和基线方法都利用高效启发式方法来获得实际可行的解。因此，我们所提出方法与基线方法的时间复杂度相近，均为 O(N) 阶。这凸显了我们的方法在应对该问题所带来挑战时的计算效率和可行性。

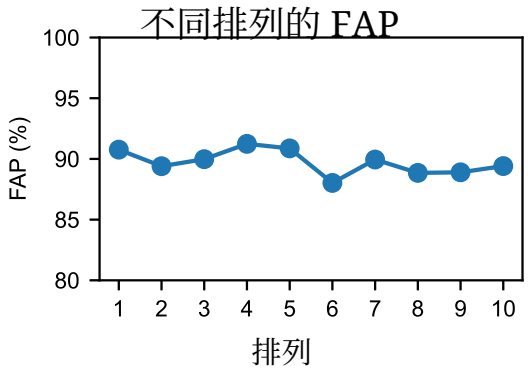


图 6：不同排列下我们方法的 FAP 性能。横轴表示排列 1–10。图中显示，不同排列对性能的影响较小（1%）。

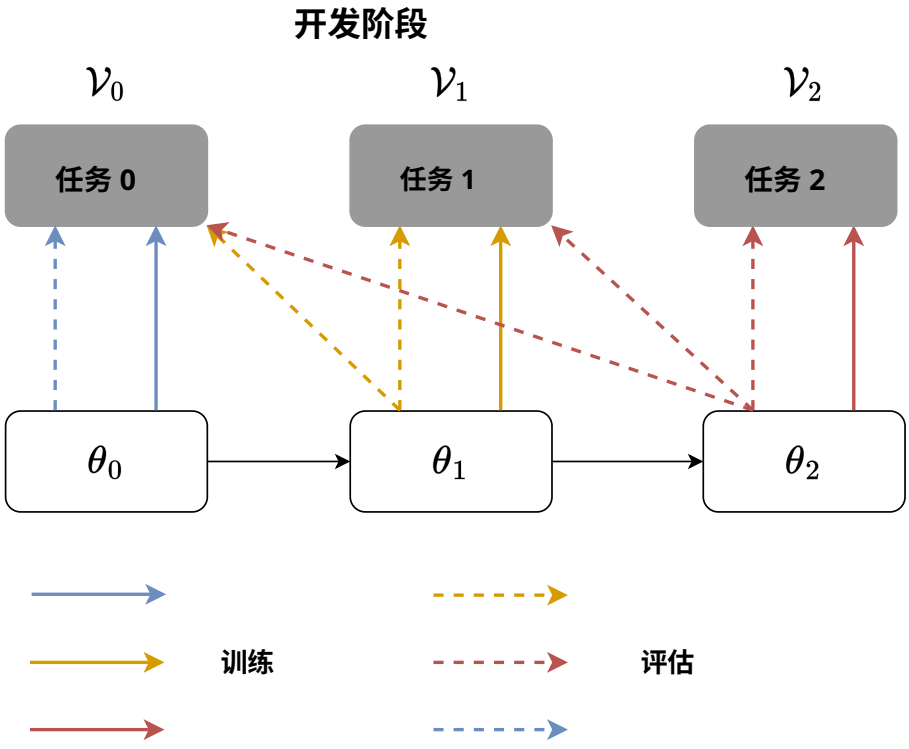


图 7：开发阶段的持续训练。上图展示了模型使用哪些任务数据进行训练和评估。

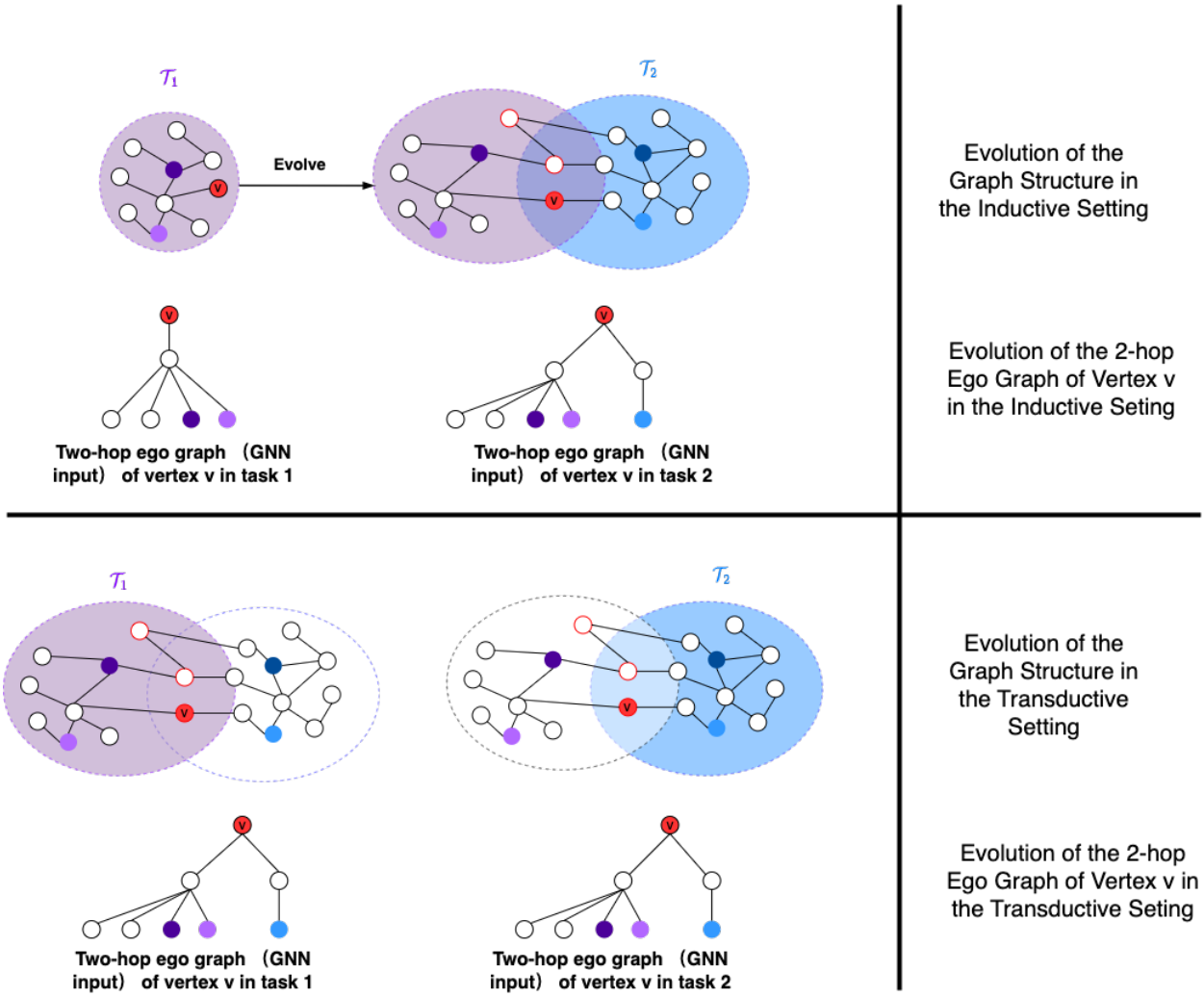


图 8：归纳式（演化图）图持续学习与直推式（恒定图）图持续学习的区别。如图的上半部分所示，在演化图设置中，新出现的任务会扩展或改变现有图结构，导致部分先前顶点的邻域（图神经网络的输入）发生变化。正如论文正文所述，这种结构变化／依赖会导致模型先前学到的信息发生分布偏移。另一方面，在恒定图设置中，从一开始就可以获得完整的图结构。所有顶点的邻域在不同学习阶段始终保持静态／不变。因此，结构依赖是演化图持续学习中特有的挑战。

算法 2 SEA-ER-GNN

输入： τ_i //新的节点分类任务
要求： $\mathcal{P}_{i-1} = \{P_1, P_2, ..., P_{i-1}\}$ //回放缓冲区
要求： g //当前图神经网络模型
要求： $\mathcal{G}_{\tau_{i-}\ _1}$ //由 τ_{i-1}的顶点诱导的图
要求： β_l, β_u
为此学习阶段创建训练集

$$\widehat{\mathcal{V}}_i = \mathcal{V}^{\text{train}} \cup \mathcal{P}_{i-1};$$

Comp用 E 执行 βq° [4.3]
Ru使用损失函数进行训练过程（前向计算和反向传播）

$$L = \frac{1}{|\mathcal{V}_i^{\text{train}}|} \sum_{v \in \mathcal{V}_i^{\text{train}}} L(v) + \sum_{P_j \in \mathcal{P}_i} \frac{1}{|P_j|} \sum_{v \in P_j} \beta_v L(v),$$

使用来自 τ_i 的回放集 P_i 更新回放缓冲区

$$\mathcal{P}_i = \mathcal{P}_{i-1} \cup P_i;$$

算法3 回放集

输入： $D_i \subset V_i$ //任务 i 的标注集（τ_i）**要求：** $\mathcal{G}_{i-1} = \mathcal{G}[\mathcal{V}_{i-1}]$//现有图 按照以下步骤采样/选择回放集 P_i： 1. 使用 p_v 及其度初始化每个顶点 $v \in D_i$，并根据与度成正比的概率采样一个初始顶点 u（例如，将每个顶点的度除以总度进行归一化）2. 将 u 分配到回放集 $P_i = \{u\}$3. 使用公式 $d(v, P_i)$ =min$\{d(v, u)$ $|u \in P_i\}$计算每个顶点 $v \in D_i \setminus P_i$ 到 P_i 的最短路径距离 4. 根据距离 P_i 与其度的乘积计算每个顶点 $v \in D_i \setminus P_i$ 的权重，即$p_v =$ 度$(v) * d(v, P_i)$ 5. 根据与 p_v成正比的概率，从 D_i 中采样 v，例如，将其除以 D_i中剩余顶点对应数值之和进行归一化 6. 重复步骤 2-5，直到 $|P_i| = k$

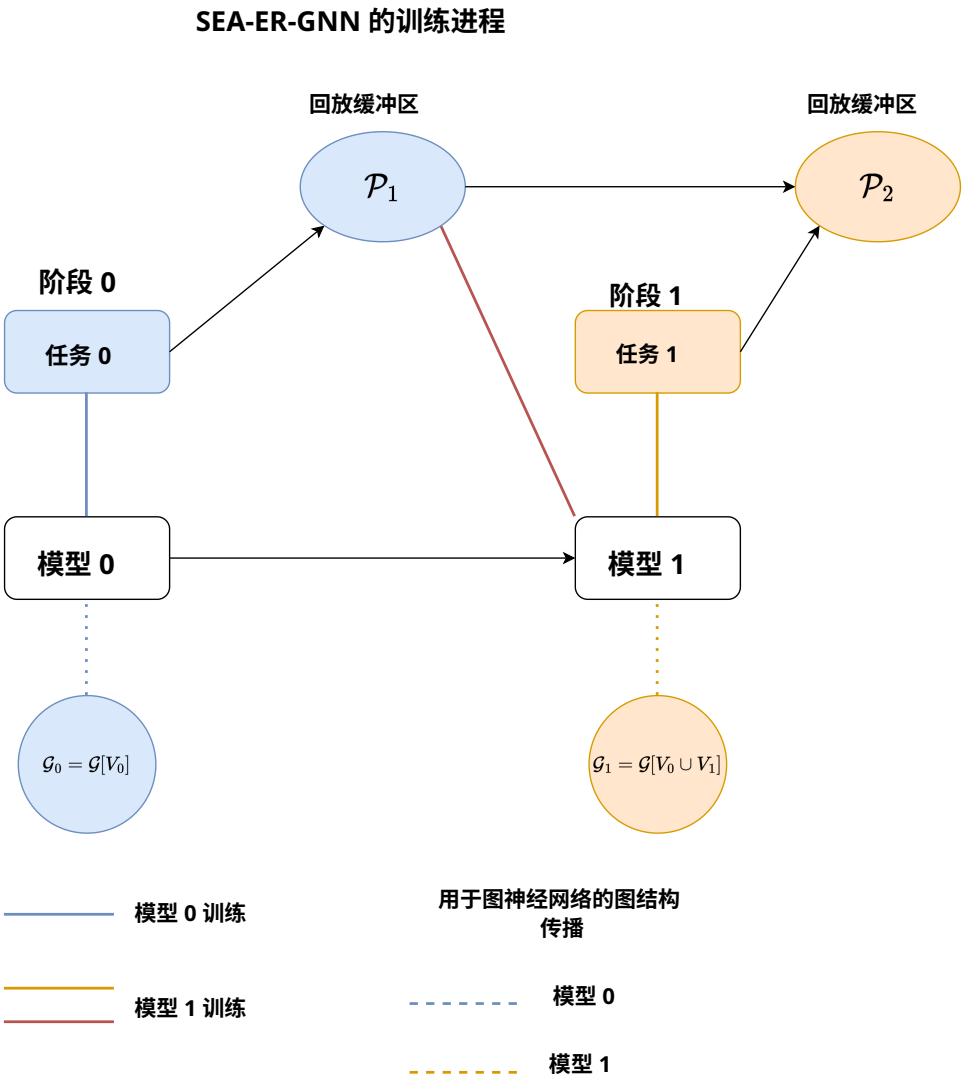


图 9：两阶段 SEA-ER-GNN 训练