

摘要

在大数据时代,各种实体之间的相互联系形成了复杂的属性图,也被称为属性网络。节点分类任务作为分析属性网络的核心任务已经引起了研究人员的广泛关注。由于数据收集和标注成本的限制,很大一部分节点类别只包含有限的标记数据。小样本节点分类旨在让机器能够快速从有限的标记数据中学习识别和分类的能力。并且在实际场景中,随着时间的推移可能会出现新的类别,图是随着新节点和边的出现而动态演化的。这些新的节点类别通常只有少量的标签节点。为了处理连续的小样本节点分类任务,研究人员研究了图小样本类增量节点分类问题,其中每个阶段出现的新类别都只有有限的标记数据。然而,在类增量学习中常常会出现灾难性遗忘现象,即模型在学习了新类别数据后,在旧类别数据的分类性能会大幅下降。

近年来,基于深度度量学习的图小样本节点分类方法逐渐成为研究热点。这些方法通常包括三个步骤:提取属性图中节点的特征、根据每个类别仅有的有限标记节点聚合该类别的原型表示、通过计算查询节点和所有类原型的距离度量来完成分类。查询节点的预测类别是距离它最近的类原型的类别。本文基于深度度量学习,对属性图的小样本节点分类方法以及小样本类增量节点分类方法展开了研究,主要贡献如下:

(1) 针对图小样本节点分类方法中存在的类原型不准确以及不同度量函数造成的分类性能差距等问题,本文提出了一种基于深度度量学习的图加权原型缩放网络(Graph Weighted Prototype Scaling Network, GWPSN)。首先, GWPSN 利用简化的图卷积网络快速地学习到属性图中节点的代表。其次,基于节点和其邻居的度信息以及通过神经网络学习到的每个节点的分数,可以得到属于同一类中的节点的权重,再对节点表示进行加权求和得到此类的类原型表示。然后,通过对距离度量进行缩放可以减少因不同度量函数带来的性能差距并提高小样本学习的性能。最后,通过计算查询节点和所有类原型之间的缩放距离度量完成分类。查询节点的预测类别就是距离它最近的类原型的类别。在三个真实世界的数据集上的大量实验表明, GWPSN 在图小样本节点分类任务的分类准确率比最佳基线提高了 1.2%到 10.6%,实现了卓越的性能。

(2) 针对图小样本类增量节点分类方法中存在的特征提取不充分、类原型不准确以及灾难性遗忘等问题,本文提出了一种基于深度度量学习的缓解遗忘的图融合网络

(Graph Fusion Network with Easing Forgetting, GFNEF)。首先, GFNEF 利用融合网络为属性图中的所有节点学习到具有强表征能力的节点表示。该融合网络利用已有的图卷积网络和图注意力网络来学习属性图中节点的表示, 然后学习它们对应的注意力系数, 并通过加权求和来生成最终的节点表示。其次, GFNEF 基于节点和其邻居的度信息以及多头注意力机制来聚合类原型表示。最后, 通过知识蒸馏以及原型相似约束, 可以有效缓解增量学习中的灾难性遗忘现象。在两个公共数据集上的实验结果表明, GFNEF 在图小样本类增量节点分类任务的分类准确率比最佳基线提高了 0.31%到 4.97%。

关键词: 小样本学习, 类增量学习, 深度度量学习, 属性网络, 节点分类

目 录

摘 要.....	I
ABSTRACT.....	III
第 1 章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	3
1.2.1 图小样本学习.....	3
1.2.2 图小样本类增量学习.....	6
1.3 主要研究内容.....	8
1.4 论文组织结构.....	9
第 2 章 相关技术介绍.....	11
2.1 属性图.....	11
2.2 图神经网络.....	11
2.2.1 卷积机制.....	12
2.2.2 注意力机制.....	13
2.3 度量学习相关技术.....	14
2.3.1 距离度量.....	14
2.3.2 度量缩放.....	15
2.4 本章小结.....	16
第 3 章 基于深度度量学习的属性图小样本节点分类方法.....	17
3.1 属性图小样本节点分类问题.....	17
3.2 图小样本节点分类方法.....	19
3.2.1 基于 SGC 的节点特征提取.....	19
3.2.2 基于度信息的类原型学习.....	20
3.2.3 基于随机变分方法的度量缩放.....	21
3.3 算法流程.....	22
3.4 实验.....	23
3.4.1 数据集.....	23
3.4.2 评估指标.....	23
3.4.3 对比方法.....	24
3.4.4 实验设置.....	25
3.4.5 实验结果与分析.....	26
3.4.6 消融实验.....	27
3.5 本章小结.....	30
第 4 章 基于深度度量学习的图小样本类增量节点分类方法.....	31
4.1 属性图小样本类增量节点分类问题.....	31
4.2 图小样本类增量节点分类方法.....	32
4.2.1 基于融合网络的节点特征提取.....	32
4.2.2 基于度信息和多头注意力机制的类原型学习.....	34
4.2.3 基于知识蒸馏和原型相似约束的类增量学习.....	35

4.3 算法流程.....	36
4.4 实验.....	37
4.4.1 数据集.....	37
4.4.2 评估指标.....	38
4.4.3 对比方法.....	38
4.4.4 实验设置.....	38
4.4.5 实验结果与分析.....	39
4.4.6 消融实验.....	40
4.5 本章小结.....	42
第5章 总结与展望.....	45
5.1 总结.....	45
5.2 展望.....	45
参考文献.....	47
致 谢.....	51
攻读学位期间发表的学术论文目录.....	53

第1章 绪论

1.1 研究背景及意义

随着大数据时代的来临，各种实体之间的交互越来越密切。图，也被称为网络，被广泛用于表示不同实体之间的联系^[1]。在图数据中，每个节点表示一个实体，边代表不同实体之间的关系。由含有额外的描述性信息的节点或边组成的网络，被称为属性网络（Attributed Network）。属性网络在现实世界中非常常见，比如交通网络^[2]、电子商务图^[3]、引文网络^[4]以及社交网络^[5]等等。

图学习（在图上的机器学习）试图挖掘和分析图数据，以发现有价值的信息^[6]。图学习在许多任务上是有效的，如节点分类^[7]、链路预测^[8]、聚类^[9]等等。节点分类是图分析中最重要的任务之一。它的目的是根据图中节点的属性以及节点之间的关系，为图中未标记的节点预测一个特定的类。例如，如图 1-1 所示，黄色和蓝色节点表示已标记节点，灰色节点为未标记节点。节点分类任务的目标就是根据已标记节点的信息，为每个灰色节点预测一个标签。节点分类任务具有广泛的应用场景，例如，在引文网络中，每篇论文都可以表示一个节点，两篇论文之间的引用关系构成了两个节点的边。每篇论文的摘要可以构成节点的属性特征。通过对节点进行分类，就可以识别出具有相似主题或研究方向的论文^[10]。在社交网络中，节点可以表示用户，边可以表示它们之间的关系。通过对节点进行分类，可以识别具有相似兴趣和行为的用户，从而进行社交推荐^[11]。因此，图上的节点分类问题引起了广泛的兴趣和关注^[12]。



图 1-1 节点分类示意图

随着图神经网络（Graph Neural Network, GNN）的出现，图上的节点分类任务取得了显著的进展^[13]。GNN 可以将属性网络中的节点映射成低维密集的向量，称为节点表示。然后通过计算向量之间的相似性，来判断节点之间的相似性，进而完成节点分类。尽管基于图神经网络的模型在处理图上的节点分类任务时很有效，但它们通常依赖于使用大量标记数据的训练才能获得令人满意的分类性能。图神经网络主要针对在具有丰富

标记的固定类别中对未标记节点进行分类。而在现实的属性网络中,呈现长尾(long-tail)节点类分布^[14],即小部分节点类包含大量标记样本,而大部分节点类只包含有限的标记样本。当用于训练的标记节点数量有限时,基于 GNN 的模型会面临过拟合的情况,其节点分类性能将严重下降。

小样本学习可以通过学习先验知识,用有限的标记样本来解决新任务,从而解决了数据不足的问题^[15]。例如,只需要给孩子几张不同角度的长颈鹿的图片,他就很容易地学会如何辨别长颈鹿这个类别。因此,结合了图神经网络和小样本学习的图小样本学习已经成为了一个研究的热点^[16]。

目前,基于度量学习的小样本学习方法由于其运行效率最快且分类效果较好,被广泛应用。度量学习主要是学习样本间的相似性度量,相同类别的样本之间的相似性评分应该相对较高,不同类别的样本之间的相似性评分应该较低。传统度量学习处理原始数据的能力有限,一般都是使用特征工程的知识对数据进行预处理,然后再用度量学习的算法进行学习。随着深度学习的出现,它能够从原始数据中学出高质量的特征。因此,产生了深度度量学习。基于深度度量学习的小样本学习方法主要包括特征提取和距离度量两个模块。特征提取模块将支持集和查询集的样本映射到同一特征空间中,然后通过距离度量模块得到查询样本和标记样本之间的距离或者查询样本和由有限的标记样本聚合得到的类原型之间的距离,在嵌入空间中距离相近的样本被归为同一类。在以往的研究工作中,基于深度度量学习的小样本分类方法主要集中在图像分类上,在图上进行小样本节点分类的研究还相对较少。图小样本节点分类任务高度依赖于节点的编码质量、类原型的表示和度量的选择。但现存的工作只考虑了前两个方面,并且在节点分类准确率上还有很大的提升空间。

此外,在现实场景中,随着时间的推移,真实世界的图可能会在不同的时间段内逐渐出现新的节点类,同时会产生新的节点和新的边。在理想情况下,期望的小样本节点分类模型应该能够准确识别新会话中引入的那些新类,同时保留先前会话中所有“看到”的节点类的分类记忆。随着新出现的类别不断加入,模型需要处理一系列用于节点分类的增量学习会话,其中每个会话都会引入一组新类别,并且那些新出现的类只有少量的标记样本可用。而现存的图小样本节点分类模型,都存在灾难性遗忘现象,也就是在学习完某一阶段的新类别之后,模型对之前旧类别的分类性能将大大下降。图小样本类增量节点分类模型考虑了在已有的分类器中增加新的节点类别的情况,通过少量的标记数

据来学习一个新的分类器。它可以在已有分类器不被破坏的前提下，快速适应新的类别，并在保持原有分类性能的基础上，准确地分类新的节点。

因此，本文基于深度度量学习，针对属性图中只有少量标记节点的新类的节点分类以及类增量节点分类问题展开相关研究。

1.2 国内外研究现状

1.2.1 图小样本学习

小样本学习旨在从含有大量样本的基类中学习泛化经验，以形成可转移的先验知识，然后利用新类中有限的标记数据识别新类中未标记的节点。在图上的小样本学习方法通常分为两类^[16]：（1）基于度量的方法，学习查询集中的数据和支持集中的数据之间的相似性度量，按照最近邻的思想完成分类；（2）基于优化的方法，为模型学习到一组良好的初始化参数，这些参数可以使得模型在一次或几次的梯度下降后，就可以对具有少量标记样本的新类完成分类。

（1）基于度量的小样本学习

基于度量的小样本学习方法^[17]通过计算查询集中的样本和支持集中的样本之间的距离度量，然后借助最近邻的思想完成分类。具体来说，基于度量的小样本学习方法通常包含了两个模块：特征提取模块和距离度量模块。特征提取模块将任务中支持集和查询集的样本映射到特定空间中。距离度量模块将学习查询集和支持集样本之间的通用匹配度量。其中度量可以采用固定度量（如欧氏距离或余弦距离）或通过神经网络学习到样本之间的相似性度量。

Vinyals 等人在 2016 年提出了匹配网络^[18]。匹配网络利用编码器对支持集和查询集样本进行特征提取，然后通过注意力机制计算查询集样本和支持集样本之间的相似度权重，最后通过对支持集中样本标签的加权求和得到查询集样本的预测标签。并且 Vinyals 等人提出了情节式训练（Episodic Training）策略。该策略遵循传统机器学习的一个重要原则：训练和测试应当在相同的条件下进行。因此，在训练时仅使用每一类别的少量样本，这与最终测试的过程保持一致。

Snell 等人于 2017 年提出了原型网络^[19]。他们首次提出了“原型”的概念。他们认为在数据集里，每个类别都存在一个原型点，样本距离某个原型点越近，其类别与该原型点对应的类别相同的概率就越大。原型网络通过计算支持集中每个类别所有样本的特征向量的平均值，作为类原型的向量表示。通过度量模块可以计算出查询样本与所有类原

型的距离，并根据距离预测查询样本的类别。例如，如图 1-2 所示，存在三个类，并且每类存在 5 个标记样本。通过每个类的标记样本可以聚合类原型（黑点 p_1 、 p_2 、 p_3 ），然后得到查询样本（灰点）到 p_1 、 p_2 、 p_3 的距离为 d_1 、 d_2 、 d_3 。因为 d_1 最小，所以认为查询样本和 p_1 的类别一样。

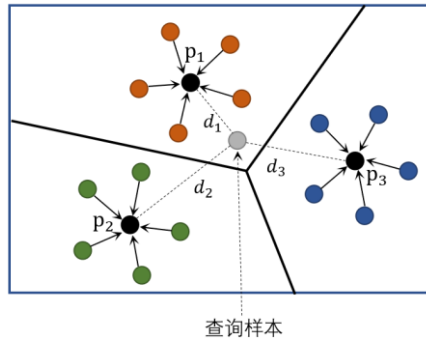


图 1-2 原型网络计算查询样本与类原型距离

Sung 等人于 2018 年提出了关系网络^[20]。相比于匹配网络中采用的余弦距离以及原型网络中采用的欧氏距离，关系网络是通过构建神经网络来计算查询样本和支持样本之间的距离度量。通过将查询样本的嵌入向量和支撑样本的嵌入向量进行串联，然后送到关系网络中学习相似性度量，来完成小样本分类。

GPN^[14]在原型网络的基础上，使用了图卷积网络（Graph Convolutional Network, GCN）^[21]聚合邻居节点的信息来更新当前节点的表示，然后通过自注意力机制来学习类原型表示，最后通过查询节点和支持节点之间的欧氏距离完成分类。Yao 等人^[22]使用图自动编码器来学习每个节点的表示，并利用辅助图的先验知识，进一步将结构知识转移到目标图中。MetaTNE^[23]在不考虑节点内容的情况下，根据图的结构信息和节点标签之间的关系，获得特定于任务的节点的潜在表示。图自适应原型网络^[24]计算支持集中节点的全局重要度和局部重要度，并将它们进行融合得到用于计算类原型的节点权重。

基于度量的小样本学习方法由于其简单性和有效性，被广泛应用。但这种方法高度依赖于样本的编码质量、类原型聚合方式以及度量的选择。

（2）基于优化的小样本学习

基于优化的小样本学习方法将小样本学习问题转化为一个优化问题。它主要通过微调实现小样本学习，避免了对新类的过度拟合。它的核心思想是为模型学习到一组良好的初始化参数，使得模型在一次或几次的梯度下降后，就可以对具有少量标记样本的新类完成分类。

Finn 等人在 2017 年提出了模型无关的元学习算法（Model-Agnostic Meta-Learning,

MAML) [25]。之所以称它为模型无关,是因为它可以应用于任何神经网络模型。MAML 算法示意图如图 1-3 可示。

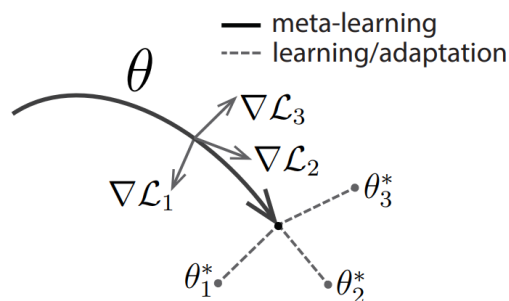


图 1-3 MAML 算法示意图[25]

假设模型为 $f(\cdot; \theta)$, 它的参数为 θ , 新任务从分布 $P(\mathcal{T})$ 进行采样。在三个新任务 \mathcal{T}_1 、 \mathcal{T}_2 、 \mathcal{T}_3 上的梯度分别为 $\nabla \mathcal{L}_1$ 、 $\nabla \mathcal{L}_2$ 、 $\nabla \mathcal{L}_3$ 。图中的 θ_1^* 、 θ_2^* 和 θ_3^* 是 \mathcal{T}_1 、 \mathcal{T}_2 、 \mathcal{T}_3 上的最优模型参数。MAML 模型参数 θ 的选取并不是某个任务的最优解,而是希望考虑所有的任务,从而使得模型在只进行少量的梯度更新后,就能指向各自任务的最优解。而这个模型参数位置的选择,就是利用不同任务的梯度下降方向,模型参数向着任务空间的平均位置进行更新。例如, θ 的更新并不是直接更新为 θ_1^* 、 θ_2^* 或 θ_3^* ,因为这样会很容易对某个任务过拟合。 θ 的更新会考虑所有任务,变为实线箭头所指向的位置。MAML 包含内循环与外循环两个部分。内循环主要是针对每个新任务,得到其对应的模型参数。外循环主要是利用内循环中得到的模型参数在对应新任务的损失之和来更新模型的初始化参数。

MAML 是一个典型的框架,很多基于优化的小样本学习方法都是 MAML 的变体。Nichol 等人提出 Reptile 模型[26],进一步简化了 MAML。Reptile 对 MAML 中初始化参数的梯度更新方向进行了改进,改善了计算复杂度高的问题。Finn 等人提出了 Meta-SGD 算法[27],可以同时学习初始化参数、梯度更新方向以及学习率。MetaOptNet[28]通过用 SVM 替换线性分类器来增强 MAML。

Meta-GNN[29]是首个将 MAML 和图神经网络结合的模型,它提出了用于处理小样本节点分类的通用框架,但它没有考虑到不同的采样任务的特征分布会存在差异。AMM-GNN[30]使用了一个属性级注意力机制,来捕获不同任务之间的特征分布差异。G-META[31]是第一个使用局部子图来进行元学习的模型,它利用子图来学习可转移的知识。Meta-GHN[32]可以从少量弱标记的节点中学习性能良好的节点表示。RALE[33]通过捕捉任务级和图级的依赖关系,从基础类中学习可转移的先验知识。HG-Meta[34]提出通过对图结构异质性和任务多样性进行建模来解决异构图上的小样本节点分类问题。

但基于优化的小样本学习方法的计算复杂度高，并且对超参数比较敏感，因此需要仔细调参才能获得较好的分类性能。

以上两种主流方法中，由于基于度量的小样本学习方法简单、有效且复杂度低，因此本研究也遵循这种范式。然而，当前大多数基于度量学习的图小样本节点分类方法的研究还相对较少，并且普遍还存在分类准确率低的问题，因此对于图小样本节点分类方法还有待进一步的研究。

1.2.2 图小样本类增量学习

增量学习^[35]，也称为持续学习或终身学习。增量学习是指在已有知识基础上不断地学习新知识的过程。本文研究是类增量学习^[36]，它是增量学习的一种形式，它的目标是在学习新类别的同时保持对以前学习的类别的知识。在类增量学习中缓解灾难性遗忘的方法主要可以分为三类：基于样本重放的方法^[37,38]、基于正则化的方法^[39-41]以及基于知识蒸馏的方法^[42-45]。

(1) 基于样本重放的方法

基于样本重放的方法是指在训练新类别时，保留一部分旧类别的数据并与新类别数据一起训练模型。这样就可以让模型在增量学习的过程中回顾旧类别知识，防止已经学习到的知识丢失，以缓解灾难性遗忘。iCaRL^[37]使用 herding 样本挑选算法从旧类别中选择一部分样本进行保留，以保证模型在学习新类别时能够保留对旧类别的知识。它采用交叉熵损失函数和知识蒸馏损失函数来训练模型，使其在新类别和旧类别上都具有较好的分类能力。Aljundi 等人^[38]通过对样本的梯度进行分析，选择对当前任务具有最大信息量的样本进行训练。

因为保存了一部分旧类别的数据，这类方法在所有方法中效果最好，但它需要额外的计算资源和存储空间并存在数据隐私泄露的隐患。

(2) 基于正则化的方法

基于正则化的方法通过对神经网络学习的参数施加限制来缓解灾难性遗忘。EWC^[39]使用 Fisher 信息矩阵来评估先前任务对新任务的影响，并加入对先前任务的正则化项来限制参数的变化。EWC 方法是一种离线学习方法，需要知道先前任务的数据集来计算 Fisher 信息矩阵。而 SI^[40]方法是一种在线学习方法，能够处理不断更新的数据流。SI 中每个参数的权重是通过反向传播算法中的梯度信息计算得到的，以此来衡量其对先前任务的影响。EWC 会为每个历史任务都维护一个惩罚项，所以惩罚项数量会随任务数量

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/217032106123010005>