

自适应多头超图卷积网络的小样本回归模型

王梅^{1,2}, 闫祖嘉¹, 高雅田^{1,2}, 高俊涛^{1,2}

(1. 东北石油大学 计算机与信息技术学院, 黑龙江 大庆 163318; 2. 黑龙江省石油大数据与智能分析重点实验室 黑龙江, 大庆 163318)

摘要: 针对图神经网络基于二元边结构难以捕捉多节点间的高阶交互, 并且固定拓扑的图结构无法适应动态数据分布。为此, 提出一种基于自适应多头超图卷积网络的小样本回归模型(AM-HGCN)。首先, 通过动态超图构建方法融合特征相似性与拓扑结构, 结合 k 阶邻居(k -HopNeighbors)和 k 近邻(k -NN)策略生成多尺度超边, 自适应捕捉特征间的交互关系; 其次, 设计多头超图卷积网络, 利用并行注意力头提取异构特征, 并通过动态门控机制融合多粒度信息, 增强模型的表达能力; 最后, 引入模型无关元学习框架, 通过内外循环优化实现快速任务适应。在数据集 Boston Housing、Energy Efficiency、IMDB、MiniImageNet 上的实验表明: 对于结构化数据集, AM-HGCN 在评价指标上显著优于主流基线模型, 其中决定系数最高提升 1.1%, 验证了其对于复杂关系建模的有效性。显著性检验结果 p 值为 0.04, 从统计学角度有力证实了这一提升的可靠性。消融实验进一步证明, 动态超图与多头注意力机制的协同作用是小样本回归性能提升的关键, 实验结果验证了本方法的有效性。

关键词: 小样本学习; 超图; 超图卷积神经网络; 多头注意力机制; 元学习

中图分类号: TP301.6; O 1-0

文献标志码: A

doi: 10.13705/j.issn.1671-6833.2026.02.003

近年来, 小样本学习 (few-shot learning, FSL) 已经成为机器学习领域的一个研究热点^[1], 如何在小样本环境下有效建模特征关系, 并提升模型能力, 是当前小样本学习研究的核心挑战。

小样本学习在数据稀缺场景下的应用取得显著成功。其核心价值体现在跨领域的高效泛化能力与知识迁移。在医学诊断中^[2], 采用 smote-tomet links 方法对不平衡数据进行处理, 在脑小血管病小样本不平衡环境下, 能够显著提升准确率; 遥感领域^[3], 利用相关性原理设计基于对比语言-图像预训练 (CLIP) 模型的类别信息融合模块以增强查询图像与支持图像、文本间类别的适配;

基于图神经网络 (GNN) 的小样本学习方法融合了度量学习和元学习的优点^[3], 相较于基于元学习的方法, GNN 能够精准地捕捉样本的类间和类内关系。为了对传统的 GNN 进行改进, gidaris 等^[4]在 GNN 中加入了降噪自编码器, 以修正小样本类别的权重。velickovic 等^[5]进一步改进了这一方法, 他们

关注节点的邻域特征, 并通过自注意力机制为不同的邻居节点分配不同的权重, 从而增强信息聚合能力。赵晓等^[6]用注意力机制优化骨干网络提取到的基础特征并引入小波变换提取多频率域特征以捕捉细节信息。葛小三等^[7]决基于度量学习的遥感影像分类中小样本学习特征空间图像特征分布不明显问题, 提出 SER-GNN 进行遥感影像图像特征提取, 增强模型对关键区域的关注能力。沈瑜等^[8]提出一种基于图结构, 采用改进的 k -Means 聚类方法生成融合时序特征的图节点, 构建时空注意力图网络, 并创新性地引入方差差异损失函数, 在时间维度捕捉序列不同周期的演化规律。

然而, 图神经网络基于二元边结构, 难以捕捉多节点间的高阶交互^[9], 固定拓扑的图结构无法适应动态数据分布, 单一注意力机制难以捕捉异构特征的多粒度信息, 易产生冗余或信息丢失。超图作为一种泛化的图结构^[10], 通过超边连接任意数量节点的特性, 为高阶关系建模提供了更自然的数学框架。

收稿日期: 2025-12-21; 修订日期: 2026-02-18

基金项目: 国家自然科学基金项目 (51774090, 62076234); 黑龙江省科技创新基地项目 (JD24A009); 黑龙江省自然科学基金项目 (LH2024F005); 黑龙江省博士后科研启动金资助项目 (LBH-Q20080)

作者简介: 王梅 (1976—), 女, 河北安国人, 东北石油大学教授, 博士, 主要从事机器学习和智能油气田领域研究, E-mail:

wangmei@nepu.edu.cn。

超图卷积网络能有效提升特征交互建模能力。多头注意力机制增强异构特征表达能力。

为了解决上述问题,引入了超图卷积网络(HGCN)与多头注意力机制。传统图卷积网络受限于二元边关系建模,难以有效捕获群体交互特征。为此,Feng等^[11]率先提出超图卷积神经网络。Bai等^[12]在HGCN的卷积网络基础上加入了注意力机制,通过自注意力层进一步增强了表示学习能力。吴越等^[13]将HGCN的应用范围扩展至异质网络,使用网络模体构造超边,将异质网络转换为超图。该模型通过超图卷积网络和注意力机制聚合超图内部不同类型节点的信息,实现半监督超图节点分类任务。Jiang等^[14]认为构造的初始超图可能不适用于最终的数据表示,因此首次提出了一种动态超图神经网络DHGNN,通过特征重构动态调整超边关系,但未解决小样本场景下的任务适应问题。

基于上述思考,本文提出一种基于自适应多头超图卷积网络的小样本回归网络提出基于特征相似度的动态超图构建方法,利用余弦相似度构建初始特征,选择相似度高的节点集合作为超边,结合 k_hop 方法和 k_NN 方法从全局和局部的角度构建动态超图,使超图适用于更复杂的关系建模;提出多头超图卷积网络,通过多个卷积头并行地学习不同的图卷积特征,增强了模型的表达能力。在四个数据集上进行了实验,结果表明,相较于其他经典小样本学习模型,本文提出的模型在准确率上取得了不错的结果。

1 小样本学习回归模型

1.1 基于自适应多头注意力的超图卷积网络小样本回归模型框架

模型的主要思想:通过节点特征构建超图并生

成超图关联矩阵;定义多个卷积头,将关联矩阵和初始节点特征输入到多个超图卷积网络中,得到多个卷积后的节点特征;利用多头注意力机制对节点特征进行加权聚合,并基于聚合后的节点特征和超图动态构建与重建超图结构;这一过程重复进行,直至达到设定的模型层数;将最终得到的节点特征送入元学习训练循环进行训练,最后进行预测。多头注意力的超图卷积网络小样本回归模型主要包含动态超图构建模块,多头超图卷积网络模块,门控融合模块。模型整体框架结构如图1所示。

1.2 基于特征相似度的动态超图构建

本节介绍基于特征相似度的动态超图构建方法。该方法利用节点特征相似度自适应构建超边,增强超图对高阶关系的建模能力,实现超图结构自适应调整。具体而言,先计算节点特征的余弦相似度矩阵,再基于设定分位数阈值动态选取高相似度节点集作为超边。同时,结合 k_hop 与 k_NN 方法,分别从全局和局部角度构建超边以利用多尺度信息,最终通过等量融合策略合并两类超边,形成动态超图结构。

首先,通过计算节点特征之间的余弦相似度来度量节点之间的相似性,并基于设定的阈值来动态生成超边,通过余弦相似度计算,计算出特征之间的余弦相似度。

$$\text{similarity}(i,j) = \frac{\mathbf{X}_i \cdot \mathbf{X}_j}{\|\mathbf{X}_i\| \|\mathbf{X}_j\|} \quad (1)$$

式中: $\mathbf{X} \in \mathbf{R}^{N \times d}$ 为特征矩阵; $\|\mathbf{X}_i\|$ 为 \mathbf{X}_i 的 $L2$ 范数; N 为数据个数; d 为影响因素个数。

若节点 i 和节点 j 之间的相似度大于设定的阈值 τ ,就将节点放入一个超边。这样每个超边代表了一组相似的节点,构造出超图的初始超边集。

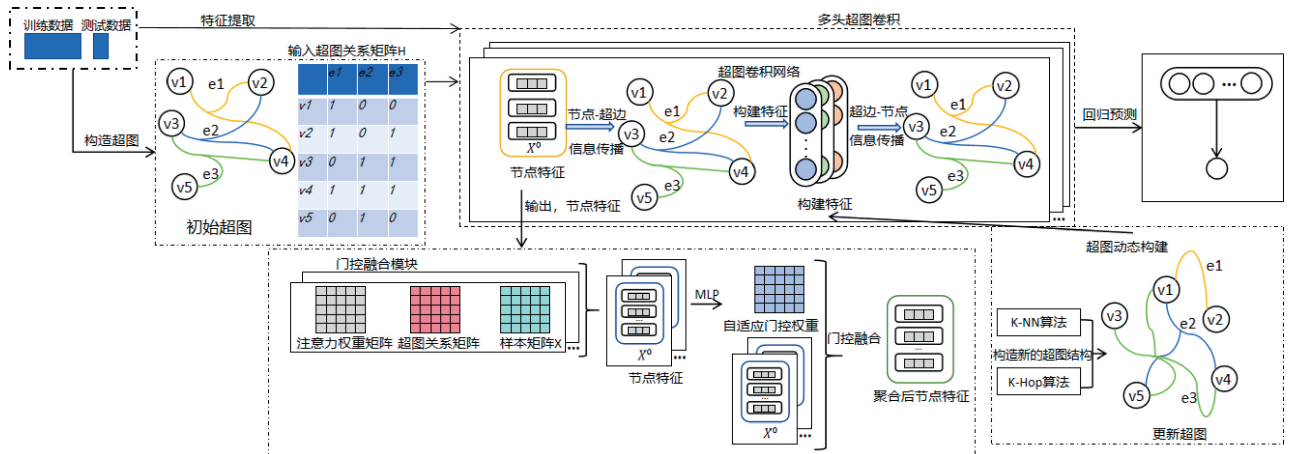


图1 基于自适应多头注意力的超图卷积网络小样本回归模型框架图

Figure 1 Framework diagram of a few-shot regression model based on adaptive multi-head attention hypergraph convolutional network

$$E_i = \{j \mid \text{similarity}(i, j) \geq \tau\}. \quad (2)$$

k_hop 方法基于图的拓扑结构构建超。选择中心点 v_i , 并确定 k_hop 邻居, 当 $k=1$ 时, 选择所有 1-hop 邻居形成超边 $E_{k-hop1}(v_i)$, 当 $k=2$ 时, 选择所有 2-hop 邻居形成超边 $E_{k-hop2}(v_i)$, 构建超边集: 对所有节点重复上述步骤, 形成完整的超边集; 构造超图关联矩阵 H_{k-hop} : 若节点 v_i 属于超边 E_k , 则 $H_{i,k}=1$, 否则 $H_{i,k}=0$ 。

k_NN 方法在局部范围内构建超边。计算特征相似度矩阵 S ; 选择每个节点的 k 个最近邻节点, 形成超边 $E_{k-NN}(v_i)$; 引入注意力权重: 利用注意力机制计算权重 α_{ij} , 定义为

$$\alpha_{ij} = \exp(f(S_{ij})) / \sum_{j \in N(i)} \exp(f(S_{ij})). \quad (3)$$

式中: $f(S_{ij})$ 为可学习的评分函数, $N(i)$ 为节点 v_i 的 k 个最近邻集合。

构建超边集: 将注意力加权后的邻居集合作为超边集, 构造超图关联矩阵 H_{k-NN} 。

1.3 多头超图卷积网络

本节介绍一种基于动态多头注意力机制的自适应超图卷积网络方法, 将多头注意力机制部分定义 k 个独立的注意力头, 每个头通过可学习参数生成节点特征与超边特征的交互权重, 具体为: 对输入节点特征 $X \in \mathbf{R}^{N \times d}$ (N 为节点数, d 为特征维度), 第 h 个头的注意力权重矩阵 $W^{(h)} \in \mathbf{R}^{N \times M}$ 计算为^[15]

$$W^{(h)} = \text{Softmax} \left(\frac{(XW_Q^{(h)})(HW_K^{(h)})^T}{\sqrt{d_k}} \right). \quad (4)$$

式中: $W_Q^{(h)} \in \mathbf{R}^{d \times d_k}$ 和 $W_K^{(h)} \in \mathbf{R}^{d \times d_k}$ 为第 h 个头的查询与键投影矩阵; $H \in \{0, 1\}^{N \times M}$ 为超图关联矩阵; d_k 为缩放因子。

注意力权重用于加权超边特征 $V^{(h)} = HW_V^{(h)} (W_V^{(h)} \in \mathbf{R}^{d \times d_v}$ 为值投影矩阵)。

通过动态构建的关联矩阵 H 和 X 作为输入实现特征传播, 得出多个节点特征矩阵, 其核心公式为

$$Z^{(h)} = D_v^{-1/2} H D_e^{-1/2} (D_v^{-1/2} H D_e^{-1/2})^T X W^{(h)}. \quad (5)$$

式中: $D_v = \text{diag}(H1_M)$ 和 $D_e = \text{diag}(1_N^T H)$ 分别为节点度矩阵和超边度矩阵; $W^{(h)} \in \mathbf{R}^{d \times d_k}$ 为第 h 个头的卷积参数。

自适应多头超图卷积将二者结合, 将式(6)得出的第 h 个节点特征矩阵作为注意力权重动态融合的输出, 使用式(7)得出融合后的节点特征矩阵:

$$Z_{\text{fused}} = \sum_{h=1}^k \alpha^{(h)} \odot Z^{(h)}; \quad (6)$$

$$\alpha^{(h)} = \text{Sigmoid}(\text{MLP}([\|Z^{(1)}\| \dots \|Z^{(h)}\|])). \quad (7)$$

式中: $\alpha^{(h)} \in \mathbf{R}^{N \times 1}$ 为自适应门控权重, MLP 为多层

感知机。门控融合就是对每个卷积头输出进行自适应加权融合。

1.4 MAML 模块

本节提出动态超图与元学习的小样本回归模型, 通过内外循环协同优化实现快速任务适应。模型输入包含多个任务集合, 每个任务 $T_i = (S_i, Q_i)$ 由支持集 S_i 和查询集 Q_i 构成, 其中, 支持集包含 K 个样本的特征矩阵 X_{supp} 与标签向量 $y_{\text{supp}} \in \mathbf{R}^K$, 查询集包含 M 个样本的 X_{query} 和 y_{query} 。

训练循环采用模型无关元学习(MAML)框架, 分为内循环任务适配与外循环元更新两阶段。内循环阶段, 对每个任务 T_i , 复制元参数 θ 为任务参数 θ'_i , 在支持集上通过梯度下降更新参数。

$$\theta'_i = \theta - \alpha \nabla_{\theta} L_{S_i}(f_{\theta}(X_{S_i})). \quad (8)$$

式中: α 为内循环学习率; L_{S_i} 为支持集上的均方误差损失; f_{θ} 为 AM-HGCN 模型。

$$L_{S_i} = \frac{1}{K} \sum_{k=1}^K (f_{\theta}(X_{S_i}^{(k)}) - y_{S_i}^{(k)})^2. \quad (9)$$

外循环阶段, 基于查询集计算元损失, 通过二阶梯度更新元参数 θ 。

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_i} L_{Q_i}(f_{\theta}(X_{Q_i})). \quad (10)$$

式中: β 为外循环学习率; L_{Q_i} 为查询集损失, 形式与 L_{S_i} 一致。

1.5 算法流程

AM-HGCN 模型分为动态超图构建与多头超图卷积训练。通过内外循环优化实现小样本回归任务。

算法 1 动态构建超图关联矩阵。

输入: 节点特征矩阵 $X \in \mathbf{R}^{N \times d}$, 阈值 τ , k_hop 参数 k , k_NN 参数 K ;

输出: 超图关联矩阵 H 。

- ① 初始化超边集 $E \leftarrow \emptyset$;
- ② 生成相似度矩阵 $S_{ij} = (X_i \cdot X_j) / (\|X_i\| \|X_j\|)$;
- ③ 生成初始超边 $E_{\text{sim}} = \{\{v_j \mid S_{ij} \geq \tau\} \mid \forall v_i \subset V\}$;
- ④ 对每个节点, 选择 k_hop 邻居节点集合, 构成超边 E_{global} ;
- ⑤ 对每个节点, 选择 K 个最近邻居点, 计算注意力权重 $\alpha_{ij} = \exp(f(S_{ij})) / \sum_{j \in N(i)} \exp(f(S_{ij}))$;
- ⑥ 定义局部超边 E_{local} ;
- ⑦ 合并超边 $E = E_{\text{global}} \cup E_{\text{local}}$;
- ⑧ 构造超图关联矩阵 H ;
- ⑨ 动态更新。根据当前隐藏层特征 $X^{(l)}$ 重新计算 S 并更新 E_{global} 和 E_{local} 。

算法 2 多头超图卷积与元学习训练

输入:支持集 S , 查询集 Q , 元学习初始参数 θ

输出:预测值 \hat{y}

- ① 通过算法一构造 H , 初始化 M 个独立注意力头, 第 h 个头的权重计算为;

$$W^{(h)} = \text{Softmax}\left(\frac{(XW_Q^{(h)})(HW_k^{(h)})^T}{\sqrt{d_k}}\right);$$

- ② 计算节点度矩阵 $D_v = \text{diag}(HI)$;
 ③ 计算超边度矩阵 $D_e = \text{diag}(I^T H)$;
 ④ 进行特征传播 $Z^{(h)} = D_v^{-1/2} H D_e^{-1} H^T D_v^{1/2} X W^{(h)}$;
 ⑤ 计算门控权重

$$\alpha^{(h)} = \text{Sigmoid}(\text{MLP}([Z^{(1)}, Z^{(2)}, \dots, Z^{(M)}]));$$

- ⑥ 多头融合 $Z_{\text{fused}} = \sum_{m=1}^M \alpha^{(m)} \odot Z^{(m)}$;
 ⑦ 内循环任务适配, 在支持集上更新任务参数 θ

$$\theta_i' = \theta - \alpha \nabla_{\theta} L_{S_i}(f_{\theta}(X_{S_i}));$$

- ⑧ 求出损失 $L_{S_i} = \frac{1}{K} \sum_{k=1}^K (f_{\theta}(X_{S_i}^{(k)}) - y_{S_i}^{(k)})^2$;

- ⑨ 外循环元更新: 在查询集上优化元参数 θ

$$\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{T_i} L_{Q_i}(f_{\theta_i'}(X_{Q_i}));$$

- ⑩ $\hat{y} = W_{\text{reg}} \cdot Z_{\text{fused}} + b_{\text{reg}}$

1.6 时间复杂度分析

本节将要对整个算法进行复杂度分析, 在余弦相似度计算中, 对 N 个节点的 d 维特征计算两两相似度, 时间复杂度为 $O(N^2 d)$, k_{hop} 超边生成时间复杂度为 $O(Nl)$, k_{NN} 超边生成时间复杂度为 $O(N^2)$, k 个注意力头的时间复杂度为 $O(kN^2 d_k)$, 经过超图卷积后, 总复杂度为 $O(k(N^2 d + Nd^2))$, 多头超图卷积的总复杂度为 $O(kN^2 d)$, 元学习内循环在支持集 K 个样本上迭代 T_{inner} 次时间复杂度为 $O(T_{\text{inner}} K d^2)$, 元学习外循环在查询集 M 个样本上迭代 T_{outer} 次时间复杂度为 $O(T_{\text{outer}} M d^2)$ 。综合所有模块:

$$O(N^2 d(1+k) + T_{\text{total}}(K+M)d^2)。 \quad (11)$$

2 实验

2.1 数据集介绍

小样本公共数据集: Boston Housing^[16]、Energy Efficiency^[17]、IMDB^[18]、MiniImageNet^[19]。

(1) Boston Housing: 数据集为房价, 任务类型为回归, 样本数量为 506, 特征为 13 个连续型数值特征。

(2) Energy Efficiency: 数据集为能源效率, 任务类型为回归, 样本数量为 768, 特征为 8 个连续型数

值特征。

(3) MiniImageNet: 粗粒度图像分类, 类别个数 100 个, 每类图像均含有 600 张图片, 尺寸为 84×84 。

(4) IMDB: IMDB 数据集是一个经典的文本分类数据集, 该数据集共有 25 000 条英文电影评论。

首先数据集数据量不多全部适用于小样本学习任务, 数据集采用两个回归任务, 一个图像任务, 一个文本任务, 来说明 AM-HGCN 的泛化性能。两个回归数据集具有明显差异, 房价数据集偏于线性低维, 存在一些噪声, 而能源效率数据集具有强非线性的特点, 特征交互显著。

表 1 公共数据集参数

Table 1 Public dataset parameters

数据集	个数	特征/类数	类型
Boston Housing	506	13	Regression
Energy Efficiency	768	8	Regression
IMDB	25 000	2	文本分类
MiniImageNet	60 000	100	粗粒度图像分类

2.2 实验设置

本文实验使用的操作系统为 Windows10, CPU 为 Intel Core i5-9300H, GPU 为 NVIDIA GeForce GTX 1660Ti, 在 PyCharm2022 平台上进行开发和调试。使用 PyTorch 框架实现 AM-HGCN 模型。

模型采用 14:3:3 的数据划分, 使用 4 头超图卷积, 内外循环学习率分别为 0.01 和 0.001。激活函数为 leaky_relu, Dropout 率为 0.6。所有实验均通过十折交叉验证取平均结果。

MiniImageNet 数据集进行了 N-way K-shot 分类任务的训练和测试, 采用 5-way 5-shot 的设置, 随机采样 5 个类, 每个类包含 5 个支持样本和 15 个查询样本。IMDB 在实验中小样本是从原始训练集中每个类别中抽取 k 个实例, $k \in (5, 10, 20)$, 形成小样本训练集, 再从每个类别中抽取 k 个实例形成验证集。

2.3 小样本学习基线模型和评价指标

本文采用基线模型包括: GCN^[20]、TabPFN^[21]、QPT^[18]、FHA-LP^[19]。

回归任务采用均方根误差 (RMSE)、决定系数 (R-squared)、平均绝对误差 (MAE), 其中 RMSE、MAE 越接近 0, R-squared 越接近 1, 模型越准确。文本分类采用准确率 (Accuracy) 和 Macro-F1 指标。

图卷积神经网络 (GCN) 与 AM-HGCN 同属图神经网络框架, 核心差异在于图表示与邻接关系建模: GCN 基于二元边图结构, AM-HGCN 则以超边连接多节点。TabPFN 是新颖模型, 可快速精准预测

小规模表格数据集。FHA-LP 与 AM-HGCN 均含注意力机制用于小样本学习,前者动态关注支持样本,后者动态构造特征关系。QPT 是小样本文本分类的突出模型,针对自然语言文本的稀疏性与语义离散性建模,领域适应度高且认可度广。

2.4 模型对比实验结果分析

本实验在 4 个公共数据集上对比了 AM-HGCN 与四种基线模型的性能,涵盖回归预测、文本分类、图像分类 3 部分。AM-HGCN 在回归数据集上表现较优,验证了其在回归任务中的有效性。

房价预测中,AM-HGCN 的 R-squared 达 0.775,显著优于其他模型,表明其在低维结构化数据中能更精准捕捉特征间非线性关系。能源效率数据集上,AM-HGCN (0.462) 较 TabPFN (0.451) 提升 2.4%,体现其在复杂数据中的鲁棒性,如图 2 所示。

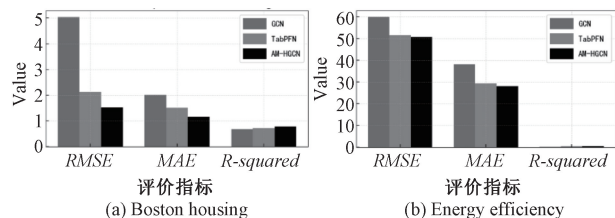


图 2 在回归数据集上模型对比结果

Figure 2 Model comparison results on a regression dataset

在 Boston Housing 和 Energy Efficiency 公开数据集上的实验表明,相较于基线模型 GCN 和 TabPFN,AM-HGCN 的预测准确性更优,如图 2 所示。

可以发现在 Energy Efficiency 数据集上 R-squared 值并不高,经过对数据集分布的查看如表 2 所示,可以发现样本 1 和 3 的预测误差较大,在样本 1 中,表面积为 700,外墙面积为 500,可能带来特征分布的异常,在样本 3 中外墙面积为 350,误差呈高估现象,这就可能由于外墙面积较小导致模型误判其能效较高。但样本 2 预测接近真实值,说明模型在常规结构的样本上具有良好的拟合能力。

表 2 Energy efficiency 样本表现

Table 2 Performance of energy efficiency samples

编号	表面积/m ²	真实值/kWh	预测值/kWh	绝对误差
1	0.667	15.6	12.10	3.5
2	0.710	23.8	23.50	0.3
3	0.775	18.2	21.50	3.3

为了证明本文章模型 AM-HGCN 在图像分类上是否具有泛化性,将 AM-HGCN 和前沿 FHA-LP 模型在 MiniImageNet 图像分类数据集上做对比,采用 5-way 1-shot 和 5-way 5-shot 设置。由于采样了 5 个类别,我们将 AM-HGCN 中元学习的任务数设置

为 5。

表 3 粗粒度实验结果

Table 3 Coarse-grained experimental results

模型	Accuracy/%	
	5-way 1-shot	5-way 5-shot
FHA-LP	55.02	74.04
AM-HGCN	54.13	75.11

结果显示,AM-HGCN 在 5-way1-shot 上优于 FHA-LP 模型,但在 5-way5-shot 上仅略有提升。二者虽均融合注意力机制,但 FHA-LP 专为小样本图像分类设计,含特征增强机制以强化支持集表示,通过局部原型建模更精细区分类间差异;而 AM-HGCN 核心是动态构建超图捕捉特征高阶关系,对图像数据缺乏天然空间感知能力,故在图像领域泛化能力有限,需结合视觉模型或设计适配性更强的跨模态结构以提升性能。为验证 AM-HGCN 在文本分类上的泛化性,将其与前沿 QPT 模型在 IMDB 情感分类数据集上对比。IMDB 采用手工模板引导预训练语言模型理解任务目标,模板需依据数据集标签构建提问句,以激发大语言模型结合上下文自主回答。5-shot、10-shot 和 20-shot 的实验结果如图 3 所示(与 QPT 模型对比)。

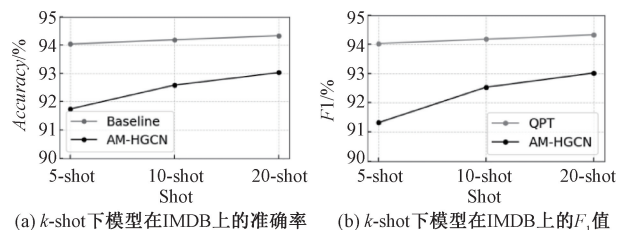


图 3 在 IMDB 上模型和 QPT 模型对比对比结果

Figure 3 Model comparison results on imdb and qpt

可以分析出 AM-HGCN 模型的上升趋势比 QPT 高,QPT 已接近平稳状态,预估样本数量再增加一些,AM-HGCN 模型的准确率会高于 QPT,但是这也违背了小样本学习的初衷,AM-HGCN 设计初衷是针对结构化特征进行建模,其超图卷积机制更适合处理特征之间的高阶关系,相比之下,QPT 是基于 Transformer 的轻量级参数高效模型,专为处理序列信息和上下文建模而设计,更适合自然语言文本分类。AM-HGCN 在文本分类中需要将文本先转换为结构化特征,而这一步可能损失了文本的上下文语义和结构信息。

2.5 参数对比实验

为了验证模型中关键参数对最终性能的影响,本节将每个参数的组合放入模型中进行实验分析,来证明模型参数对结果的影响,数据集采用 Boston

Housing,模型的各个参数如表 4 所示。

表 4 模型参数表

Table 4 Model parameters

参数	含义	取值范围	最优取值
k_NN	k_NN 超边的邻居数	1,3,5,7,9,11	9
k_hop	多跳邻居	1,2,3	1
num_heads	多头数量	1,2,4,8	8
hidden_dim	隐层维度	32,64,128	64

本实验使用控制变量法,先固定一组参数组合,然后通过改变其他参数的值,查看参数变化对结果的影响。

为分析各超参数对 AM-HGCN 模型性能的影响,我们在 Boston Housing 数据集上分别调整 k_NN , k_hop , num_heads, hidden_dim 四个关键参数,结果如图 4 所示。

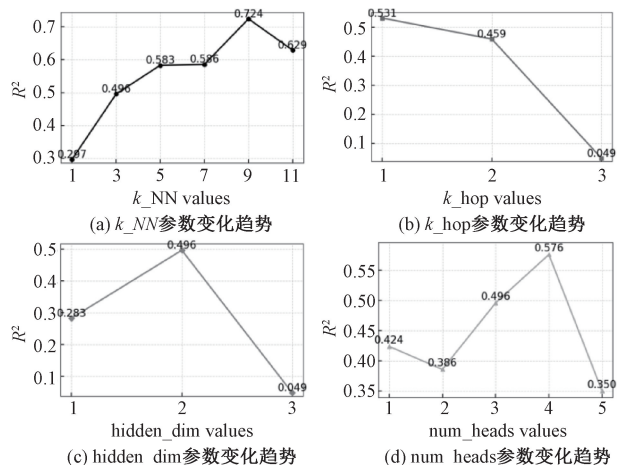


图 4 模型参数分析结果

Figure 4 Model parameter analysis results

当 k_NN 从 1 增加至 9 时, R^2 明显提升,说明合适的局部邻居数量有助于超边捕捉局部特征交互。但过大的邻居数反而导致轻微退化,可能是过多冗余连接引入噪声。类似地, num_heads 增加至 8 时效果最佳,过多头下降幅度过大,表明多头机制对特征增强有用,但过多也可能造成过拟合。综合各参数影响,最优组合如表 5 所示。

表 5 消融实验结果

Table 5 Results of ablation experiment

模型配置	R^2	RMSE	MAE
w/o DH	0.718	2.89	1.32
w/o MHA	0.681	2.81	1.21
w/o MAML	0.762	2.95	1.23
AM-HGCN	0.775	1.53	1.15

2.6 消融实验

为验证动态超图构建、多头注意力机制及元学习框架对模型性能的贡献,本节通过消融实验证明

模型有效性,结果如表 5 所示。w/o DH 指移除动态超图、使用固定超图;w/o MHA 指移除多头注意力、仅保留单一注意力头;w/o MAML 指移除元学习框架。实验在 Boston Housing 数据集上进行,评估指标为 R-squared、RMSE 和 MAE。

在 Boston Housing 数据集上,动态超图较固定超图提升 9.7% 性能。消融实验表明,动态超图构建、多头注意力机制与元学习框架是 AM-HGCN 性能优越的关键。三者协同,解决高阶关系建模、特征多样性表达及任务快速适应问题。

2.7 定性分析实验

为深入探究动态超边构建机制对模型性能的影响,我们在 Boston Housing 数据集上可视化了训练中动态生成的超图结构(图 7)。图中节点代表特征(如 nox, rm, tax 等),红色区域为基于互信息或相似性构成的超边,其透明度随训练进程变化,反映特征关联强度的演化。训练轮数增加后,可见部分特征的交互关系逐渐增强。



图 5 动态超图构建

Figure 5 Dynamic hypergraph construction

我们可以发现图中的 nox, tax, lstat, indus 连接了大量超边,这就说明这些特征具有高相关性或冗余性,zn, b, dis 节点连接较为稀疏,可能被模型边缘化,但在局部场景下依然关键。

表 6 Boston Housing 样本表现

Table 6 Performance of boston housing samples

编号	特征组合	真实值/ 预测值/	
		\$	\$
102	rm = 8.5, lstat = 4.0%, nox = 0.4 ppm	45	39.8
270	rm = 4.0, lstat = 25.0%, dis = 1.5	8	13.2
430	rm = 6.0, lstat = 15.0%, tax = 600 \$	20.5	18

通过对 Boston Housing 样本举例子,如表 6 所示,包含了 3 个样本和特征组合,查看样本 102 号可知,超图卷积没有是房价为 8,说明样本处于稀疏区域,可能没有形成足够的超边,查看样本 430 号可知,tax 异常高,代表交通不便,拉低预测值。

3 结论

(1) 本文设计了基于自适应多头超图卷积网络 (AM-HGCN) 的小样本回归模型,通过动态超图构建与多头注意力机制有效建特征间的高阶关系。

(2)动态超图构建方法突破了传统图结构对高阶交互捕捉能力不足的限制,使模型能够自适应地反映复杂特征关系。

(3)引入多头注意力机制与动态门控融合策略,在多粒度层面增强了异构特征的代表能力。

(4)在 Boston Housing、Energy Efficiency 等回归数据集上,AM-HGCN 在 $RMSE$ 、 R^2 等指标上显著优于主流基线模型;在 MiniImageNet 与 IMDB 数据集上的实验也体现出一定的泛化能力。

(5)通过消融实验,动态超图构建、多头注意力机制与元学习框架三者协同作用,是模型性能提升的关键因素。

(6)在图像分类与文本分类任务中,AM-HGCN 的表现尚不及专门设计的前沿模型,未来可探索与视觉模型、跨模态结构结合,以进一步提升跨领域泛化能力。

参考文献:

[1] Zhao Kailin, Jin Xiaolong, Wang Yuanzhuo. Survey on few-shot learning [J]. Journal of Software, 2021, 32(2): 349-369. [赵凯琳, 靳小龙, 王元卓. 小样本学习研究综述[J]. 软件学报, 2021, 32(2): 349-369.]

[2] Wang Benben, Yan Yuefei, Lu Huizhen, et al. A predictive approach for cerebral small vessel disease based on imbalanced limited-sample data[J]. Electro-Mechanical Engineering, 2025, 41(3): 87-92. [王奔奔, 严粤飞, 陆慧珍, 等. 基于小样本不平衡数据的脑小血管病预测方法[J]. 电子机械工程, 2025, 41(3): 87-92.]

[3] Zhou Wei, Wei Mangan, Xu Haixia, et al. A few-shot land cover classification model for remote sensing images based on multimodality[J]. Journal of Electronics & Information Technology, 2025, 47(6): 1747-1761. [周维, 魏名安, 许海霞, 等. 基于多模态的小样本遥感影像地物分类模型[J]. 电子与信息学报, 2025, 47(6): 1747-1761.]

[4] Gidaris S, Komodakis N. Generating classification weights with GNN denoising autoencoders for few-shot learning[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 21-30.

[5] Velickovic, P., Cucurull, G., Casanova, A., et al. Graph attention networks [PP/OL]. [PP/OL]. V1. ArXiv3 (2017-10-30) [2028-02-04]. <https://doi.org/10.48550/arXiv.1710.10903>.

[6] Zhao Xiao, (Yuan/Yun) Xiaosa, Liu Ruiling, et al. Small sample plant image classification based on wavelet multi-frequency feature fusion enhancement[J/OL]. Computer Engineering and Applications, 2025-05-21. <https://link.cnki.net/urlid/11.2127.tp.20250520.1517.003>.

[7] Ge Xiaosan, Zheng Mengmeng. Study on classification of few-shot remote sensing images based on SER-GNN[J]. Journal of Henan Polytechnic University (Natural Science), 2025, 44(5): 144-151. [葛小三, 郑猛猛. 基于 SER-GNN 的小样本遥感影像分类研究[J]. 河南理工大学学报(自然科学版), 2025, 44(5): 144-151.]

[8] Shen Yu, Wang Ruoxuan, Li Jiangcheng, et al. A Small-Sample Time Series Prediction Model Based on Graph Structure [J/OL]. Journal of Beihang University, 1-14 [2025-07-29]. <https://doi.org/10.13700/j.bh.1001-5965.2025.0198>. [沈瑜, 王若暄, 李江程, 等. 基于图结构的小样本时间序列预测模型[J/OL]. 北京航空航天大学学报, 1-14 [2025-07-29]. <https://doi.org/10.13700/j.bh.1001-5965.2025.0198>.]

[9] Liang Zhuomin, Bai Liang, Yang Xian, et al. Multi-channel disentangled graph neural networks with different types of self-constraints [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(9): 8001-8012.

[10] Berge, C. Hypergraphs: combinatorics of finite sets [M]. New York: Elsevier Science & Technology Books, 1984.

[11] Feng Yifan, You Haoxuan, Zhang Zizhao, et al. Hypergraph neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3558-3565.

[12] Bai Song, Zhang Feihu, Torr P H S. Hypergraph convolution and hypergraph attention[J]. Pattern Recognition, 2021, 110: 107637.

[13] Wu Yue, Wang Ying, Wang Xin, et al. Motif-based hypergraph convolution network for semi-supervised node classification on heterogeneous graph[J]. Chinese Journal of Computers, 2021, 44(11): 2248-2260. [吴越, 王英, 王鑫, 等. 基于超图卷积的异质网络半监督节点分类[J]. 计算机学报, 2021, 44(11): 2248-2260.]

[14] Jiang Jianwen, Wei Yuxuan, Feng Yifan, et al. Dynamic hypergraph neural networks [C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, 2019: 2635-2641.

[15] Feng Yifan, You Haoxuan, Zhang Zizhao, et al. Hypergraph neural networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3558-3565.

- [16] Harrison D, Rubinfeld D L. Hedonic housing prices and the demand for clean air[J]. *Journal of Environmental Economics and Management*, 1978, 5(1): 81-102.
- [17] Chen Jiandong, Liu Jialu, Qi Jie, et al. City- and county-level spatio-temporal energy consumption and efficiency datasets for China from 1997 to 2017[J]. *Scientific Data*, 2022, 9: 101.
- [18] Zhai Mengxin, Zhou Yanling, Yu Hang. Few-shot text classification based on question-oriented prompt-tuning [J]. *Application Research of Computers*, 2025, 42(3): 708-713. [翟梦鑫, 周艳玲, 余杭. 基于问题导向式提示调优小样本文本分类[J]. *计算机应用研究*, 2025, 42(3): 708-713.]
- [19] Zhou Bin, Xian Hao, Qin Yijia, et al. Local prototype small sample classification model integrating mixed attention mechanism[J/OL]. *Journal of Frontiers of Computer Science and Technology*, 2025-03-11. <https://kns.cnki.net/kcms/detail/11.5602.TP.20250311.1049.002.html>.
- [20] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks [PP/OL]. V4. arXiv (2017-02-22) [- -]. <https://doi.org/10.48550/arXiv.1609.02907>.
- [21] Li Peiwen, Li Feijiang, Wang Jieting, et al. Clustering method for tabular data based on pretrained foundation models with synthetic data[J/OL]. *Journal of Computer Research and Development*,. <https://link.cnki.net/urlid/11.1777.TP.20250723.1002.004>. [李培文, 李飞江, 王婕婷, 等. 基于合成数据预训练基础模型的表格数据聚类方法[J/OL]. *计算机研究与发展*,. <https://link.cnki.net/urlid/11.1777.TP.20250723.1002.004>.]

AM-HGCN: An Adaptive Multi-Head Hypergraph Convolutional Network for Few-Shot Regression

WangMei^{1,2}, YAN Zujia^{1,2}, GAO Yatian^{1,2}, GAO Juntao^{1,2}

(1. Heilongjiang Key Laboratory of Petroleum Big Data and Intelligent Analysis, Daqing 163318, China; 2. School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: High-order interactions among multiple nodes were difficult to be captured by graph neural networks based on binary-edge structures, and static graph topologies were unable to adapt to dynamic data distributions. To address these limitations, a few-shot regression model based on adaptive multi-head hypergraph convolutional networks (AM-HGCN) was proposed, in which feature similarity and topological structure were integrated through a dynamic hypergraph construction method, multi-scale hyperedges were generated using k -hop neighbors and k -nearest neighbors (k -NN) strategies to enable adaptive capture of feature interactions, a multi-head hypergraph convolutional network was designed to extract heterogeneous features via parallel attention heads and fuse multi-granularity information through a dynamic gating mechanism to enhance expressive capability, and a model-agnostic meta-learning framework was introduced to achieve rapid task adaptation through inner- and outer-loop optimization. Experiments were conducted on the Boston Housing, Energy Efficiency, IMDB, and MiniImageNet datasets, and for structured datasets, AM-HGCN was observed to outperform mainstream baseline models significantly in evaluation metrics, with the coefficient of determination (R^2) improved by up to 1.1%, validating the model's effectiveness in capturing complex relationships. Significance tests yielded a p-value of 0.04, statistically confirming the reliability of this improvement, and ablation studies further demonstrated that the collaborative effect of dynamic hypergraphs and multi-head attention mechanisms was crucial for the enhancement of few-shot regression performance, overall validating the effectiveness of the proposed method.

Keywords: Few-Shot Learning; Hypergraph; Hypergraph Convolutional Neural Network; Multi head attention mechanism; Meta le