

【动手做】 利用HGT异构图转换器 预测节点属性

AchieveFun

讨论内容

0 工具准备

- Python和Anaconda
- DGL-Deep Graph Library

1 知识准备: 异构图转换器的整体架构

2 案例演示

- 代码运行流程
- 代码结构
- 代码演示

3 参考资料

4 回顾【动手做】图神经网络模型

AchieveFun

HGT: 异构图转化器

- Recent years have witnessed the emerging success of graph neural networks (GNNs) for modeling structured data. However, most GNNs are designed for homogeneous graphs, in which all nodes and edges belong to the same types, making them infeasible to represent heterogeneous structures. In this paper, we present the Heterogeneous Graph Transformer (HGT) architecture for modeling Web-scale heterogeneous graphs. To model heterogeneity, we design node- and edge-type dependent parameters to characterize the heterogeneous attention over each edge, empowering HGT to maintain dedicated representations for different types of nodes and edges. To handle dynamic heterogeneous graphs, we introduce the relative temporal encoding technique into HGT, which is able to capture the dynamic structural dependency with arbitrary durations. To handle Web-scale graph data, we design the heterogeneous mini-batch graph sampling algorithm---HGSampling---for efficient and scalable training. Extensive experiments on the Open Academic Graph of 179 million nodes and 2 billion edges show that the proposed HGT model consistently outperforms all the state-of-the-art GNN baselines by 9%--21% on various downstream tasks.

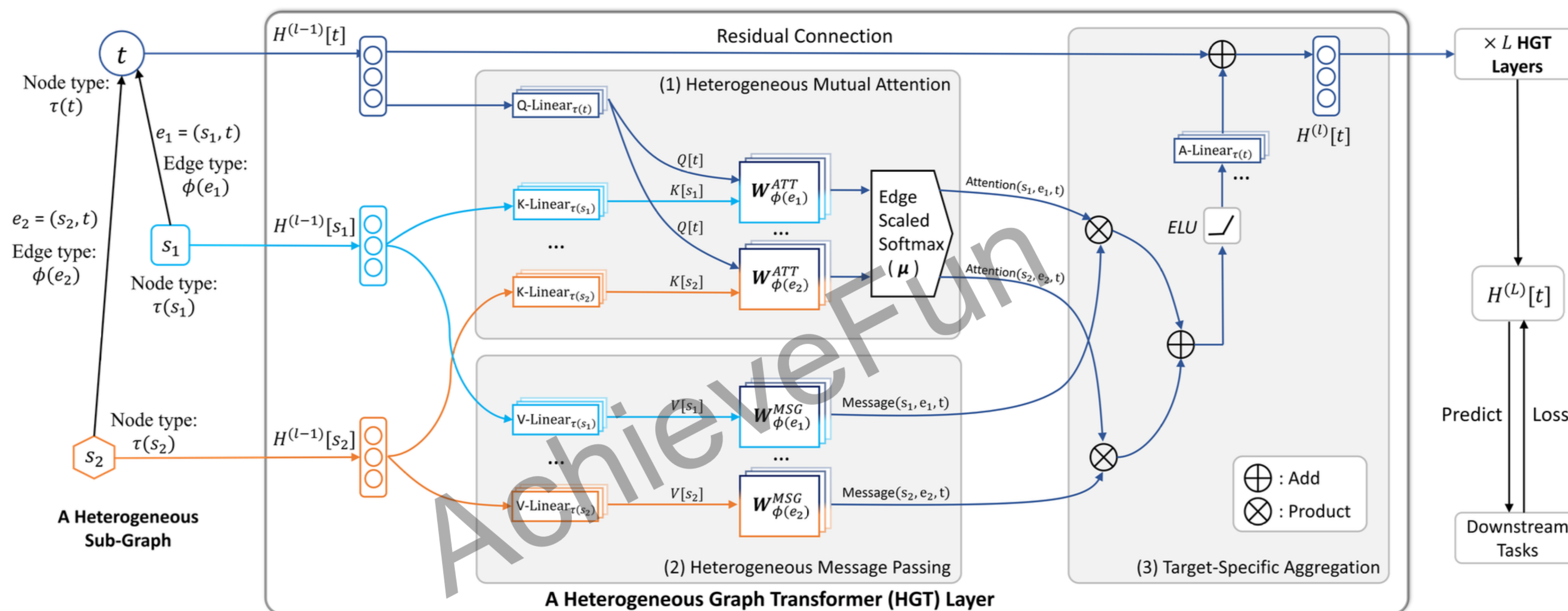
- 下游任务包括：论文-领域预测、论文-地点预测、作者消歧任务

- 来源：<https://arxiv.org/pdf/2003.01332v1.pdf>

- 中文参考 <https://blog.csdn.net/TgqDT3gGaMdkHasLZv/article/details/120755172>

机器翻译：近年来，图神经网络（GNN）在结构化数据建模方面取得了巨大成功。然而，大多数图神经网络都是针对同构图设计的，在同构图中，所有节点和边都属于相同的类型，因此无法表示异构结构。在本文中，我们提出了异构图转换器（HGT）架构，用于对网络规模的异构图建模。为了对异构性进行建模，我们设计了与节点和边缘类型相关的参数来描述每条边缘上的**异构注意力**，从而使 HGT 能够为不同类型的节点和边缘保持专用的表示方法。为了处理**动态**异构图，我们在 HGT 中引入了**相对时间编码技术**，该技术能够捕捉任意持续时间的动态结构依赖性。为了处理网络规模的图数据，我们设计了**异构小批量图采样算法**--HGSampling--以实现高效和可扩展的训练。在包含 1.79 亿个节点和 20 亿条边的开放学术图谱（Open Academic Graph）上进行的大量实验表明，所提出的 HGT 模型在各种下游任务上的表现始终优于所有最先进的 GNN 基线 9%-21%。

1 知识准备： 异构图转换器的整体架构



- 给定一个以 t 为目标节点、 s_1 和 s_2 为源节点的采样异构子图，HGT 模型将其边 $e_1 = (s_1, t)$ 和 $e_2 = (s_2, t)$ 及其相应的元关系 $\langle \tau(s_1), \phi(e_1), \tau(t) \rangle$ 和 $\langle \tau(s_2), \phi(e_2), \tau(t) \rangle$ 作为输入，学习每个节点的上下文文化表示 $H(L)$ ，并将其用于下游任务。颜色解码节点类型。HGT 包括三个部分：(1) 元关系感知异构相互注意力；(2) 源节点的异构信息传递；(3) 目标特定的异构信息聚合。

图片来源：<https://arxiv.org/pdf/2006.09963v3.pdf> Heterogeneous Graph Transformer

将异构图转化为同构图

- 异构图为管理不同类型的节点和边及其相关特征提供了一个清晰的接口。这在以下情况下尤其有用:
 - 不同类型的节点和边的特征具有不同的数据类型或大小。
 - 用户希望对不同类型的节点和边应用不同的操作。
- 如果上述情况不适用，并且用户不希望在建模中区分节点和边的类型，则DGL允许使用 `dgl.DGLGraph.to_homogeneous()` API将异构图转换为同构图。
- 具体行为如下:
 - 用从0开始的连续整数重新标记所有类型的节点和边。
 - 对所有的节点和边合并用户指定的特征。

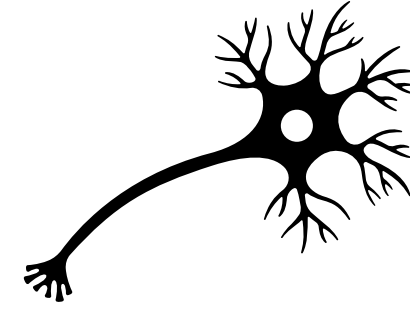
2 案例演示

- 案例来源：
 - <https://paperswithcode.com/paper/heterogeneous-graph-transformer>
- 代码
 - <https://github.com/dmlc/dgl/tree/master/examples/pytorch/hgt>
 - 参考 https://docs.dgl.ai/en/1.0.x/guide_cn/graph-heterogeneous.html#guide-cn-graph-heterogeneous

学习目标

- 来源: <https://github.com/dmlc/dgl/tree/master/examples/pytorch/hgt>
- 来自Readme: 本玩具实验基于 DGL 的官方[教程](<https://docs.dgl.ai/en/0.4.x/generated/dgl.heterograph.html>)。由于 ACM 数据集没有输入特征, 我们只是为每个节点随机分配特征。这一过程可以简单地用任何准备好的特征来代替。
- 理解图异构转化器HGT的原理, 如何预测节点属性。

环境准备



1. 安装Anaconda <https://www.anaconda.com/download/>
2. 在Anaconda建立环境dgl, 根据dgl文档(<https://www.dgl.ai/pages/start.html>)确定**Python版本**
3. 根据代码, 使用pip install -r requirements.txt 安装依赖库并验证库的存在pip list
 - o torch
4. 安装dgl <https://www.dgl.ai/pages/start.html>

例如: conda install -c dgteam dgl

代码结构

文件名	作用	方法
model.py	模型文件	class HGTLayer(nn.Module) class HGT(nn.Module) class HeteroRGCNLayer(nn.Module) class HeteroRGCN(nn.Module)
train_acm.py	训练文件	get_n_params,train
README.md	说明文件	

2.1 代码运行流程

- 问题：预测节点属性
- 训练流程：
 1. 装载数据集
 2. 初始化模型
 3. 训练模型
 4. 评估模型

AchieveFun

2.2 代码演示

- <https://github.com/dmlc/dgl/tree/master/examples/pytorch/hgt>
- `python train_acm.py`
- ACM 数据集包含发表在 KDD、SIGMOD、SIGCOMM、MobiCOMM 和 VLDB 上的论文，分为三类（数据库、无线通信和数据挖掘）。

3 参考资料

- 案例来源：
 - <https://paperswithcode.com/paper/heterogeneous-graph-transformer>
- 代码
 - <https://github.com/dmlc/dgl/tree/master/examples/pytorch/hgt>
 - 参考 https://docs.dgl.ai/en/1.0.x/guide_cn/graph-heterogeneous.html#guide-cn-graph-heterogeneous
- 论文： <https://arxiv.org/pdf/2003.01332v1.pdf> 中文参考 <https://blog.csdn.net/TgqDT3gGaMdkHasLZv/article/details/120755172>

4 回顾【动手做】图神经网络模型

- GraphSAGE: 链接预测
- GCN图卷积网络: 分类
- GAT图注意力网络: 分类
- VGAE 图变分自编码模型: 链接预测
- DiffPool: Hierarchical Graph Representation Learning with Differentiable Pooling 利用可微池进行层次图表示学习
- HGT: Heterogeneous Graph Transformer 异构图转换器
- MVGRL: Contrastive Multi-View Representation Learning on Graphs 图上的对比多视图表征学习
- GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training 图对比学习模型 自监督学习
- Text Generation from Knowledge Graphs with Graph Transformers 利用图转换器从知识图谱生成文本
- GPT-GNN: Generative Pre-Training of Graph Neural Networks 生成式图网络预训练框架

总结

0 工具准备

- Python和Anaconda
- DGL-Deep Graph Library

1 知识准备: 异构图转换器的整体架构

2 案例演示

- 代码运行流程
- 代码结构
- 代码演示

3 参考资料

4 回顾【动手做】图神经网络模型