

1. GAT 的注意力聚合机制

给定无向图 $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, 节点特征矩阵 $\mathbf{H} \in \mathbb{R}^{N \times F}$ 。单层 Graph Attention Network (GAT) 将邻居信息按照可学习的注意力权重 α_{ij} 聚合到中心节点 i , 具体步骤如下:

1. 线性映射

$$\mathbf{h}_i = \mathbf{W}\mathbf{H}_i, \quad \mathbf{h}_j = \mathbf{W}\mathbf{H}_j, \quad \mathbf{W} \in \mathbb{R}^{F' \times F}.$$

2. 相关性打分

$$e_{ij} = \text{LeakyReLU}(\mathbf{a}^\top [\mathbf{h}_i \parallel \mathbf{h}_j]), \quad \mathbf{a} \in \mathbb{R}^{2F'}.$$

3. Soft-max 归一化

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik})}, \quad \sum_{j \in \mathcal{N}(i)} \alpha_{ij} = 1.$$

4. 加权聚合并激活

$$\mathbf{H}'_i = \sigma\left(\sum_{j \in \mathcal{N}(i)} \alpha_{ij} \mathbf{h}_j\right), \quad \sigma = \text{ELU 或 ReLU}.$$

2. GAT 相对 GCN 的优势及适用场景

维度	GCN	GAT 优势
邻居权重	固定 (度归一化)	可学习, 细粒度区分邻居贡献
谱依赖	需全图拉普拉斯	仅局部邻接, 无需谱分解
归纳能力	同一图上训练推断	支持在未见过的新图推断
可解释性	弱	注意力权重可视化解释模型决策
过平滑风险	高层数时显著	多头差异化缓解过平滑
典型优势任务	同质静态图	异构图、邻里贡献差异明显, 或需归纳推断、解释性的任务 (推荐、知识图谱、PPI 等)

3. 多头注意力在 GAT 中的实现

设有 K 个独立注意力头，第 k 头的参数为 $\mathbf{W}^{(k)}, \mathbf{a}^{(k)}$ 。每头独立完成步骤 1–4，得到输出 $\mathbf{h}_i'^{(k)}$ 。

- 隐藏层融合（拼接）

$$\mathbf{H}_i' = \parallel_{k=1}^K \mathbf{h}_i'^{(k)}.$$

- 输出层融合（平均或求和）

$$\mathbf{H}_i' = \frac{1}{K} \sum_{k=1}^K \mathbf{h}_i'^{(k)}.$$

多头机制提供以下益处：

1. **降低方差**：独立头相当于模型集成，提高稳定性。
2. **子空间表达**：各头可关注不同语义或结构模式。
3. **容量提升**：拼接增加特征维度，提高表示能力。
4. **缓解过平滑**：头间差异化聚合抑制信息过度混合。

4. GAT 在 Cora 数据集上的关键实验指标

项目	设定 / 结果
数据集	Cora: 2 708 节点, 5 429 边, 7 类
网络结构	隐藏层: 8 头 × 8 维; 输出层: 1 头 × 7 维
正则化	Dropout 0.6 (输入与注意力权重)
优化器	Adam, 学习率 0.005, 权衰减 5×10^{-4}
提前停止	验证损失 100 epoch 无下降即停止
节点分类准确率	83.0 ± 0.7 % (原论文结果)

5关键实验指标

该脚本启动后首先解析命令行超参数，自动决定 CPU 或 GPU，然后用 `torch_geometric.datasets.Planetoid` 下载并加载 Cora 图数据。核心亮点在于模型通过字符串“`--model + --impl`”动态导入：只要输入 `gat`、`gcn` 或 `gin` 即可切换网络结构，`no_pyg` 代表完全用纯 PyTorch 实现的算子，`pyg` 则调用 `torch-geometric` 的高效封装，两套实现接口完全一致，便于横向对比。训练阶段采用 Adam 优化器，按验证集准确率实时保存最佳权重并记录三条准确率曲线；训练结束后自动输出最优指标、绘制并保存曲线图，以及序列化的 checkpoint。整条流程无需手动改代码即可在 GAT、GCN、GIN 之间来回切换，并直观比较纯 torch 与 PyG 版本的性能差异。

6 实验结果

Model	Implementation	Val Acc (%)	Test Acc (%)
GCN	PyG	79.60	81.20
GAT	PyG	80.40	82.10
GIN	PyG	60.80	61.10
GCN	no_PyG	80.20	81.20
GAT	no_PyG	79.20	80.80
GIN	no_PyG	49.60	47.90

训练过程见tutorial.ipynb。