



图注意力网络 (GAT)

作者: Calvin

QQ: 179209347

Mail: 179209347@qq.com

介绍

笔记简介:

- 面向对象: 深度学习初学者
- 依赖课程: **线性代数, 统计概率**, 优化理论, 图论, 离散数学, 微积分, 信息论

知乎专栏:

<https://zhuanlan.zhihu.com/p/693738275>

Github & Gitee 地址:

https://github.com/mymagicpower/AIAS/tree/main/deep_learning

https://gitee.com/mymagicpower/AIAS/tree/main/deep_learning

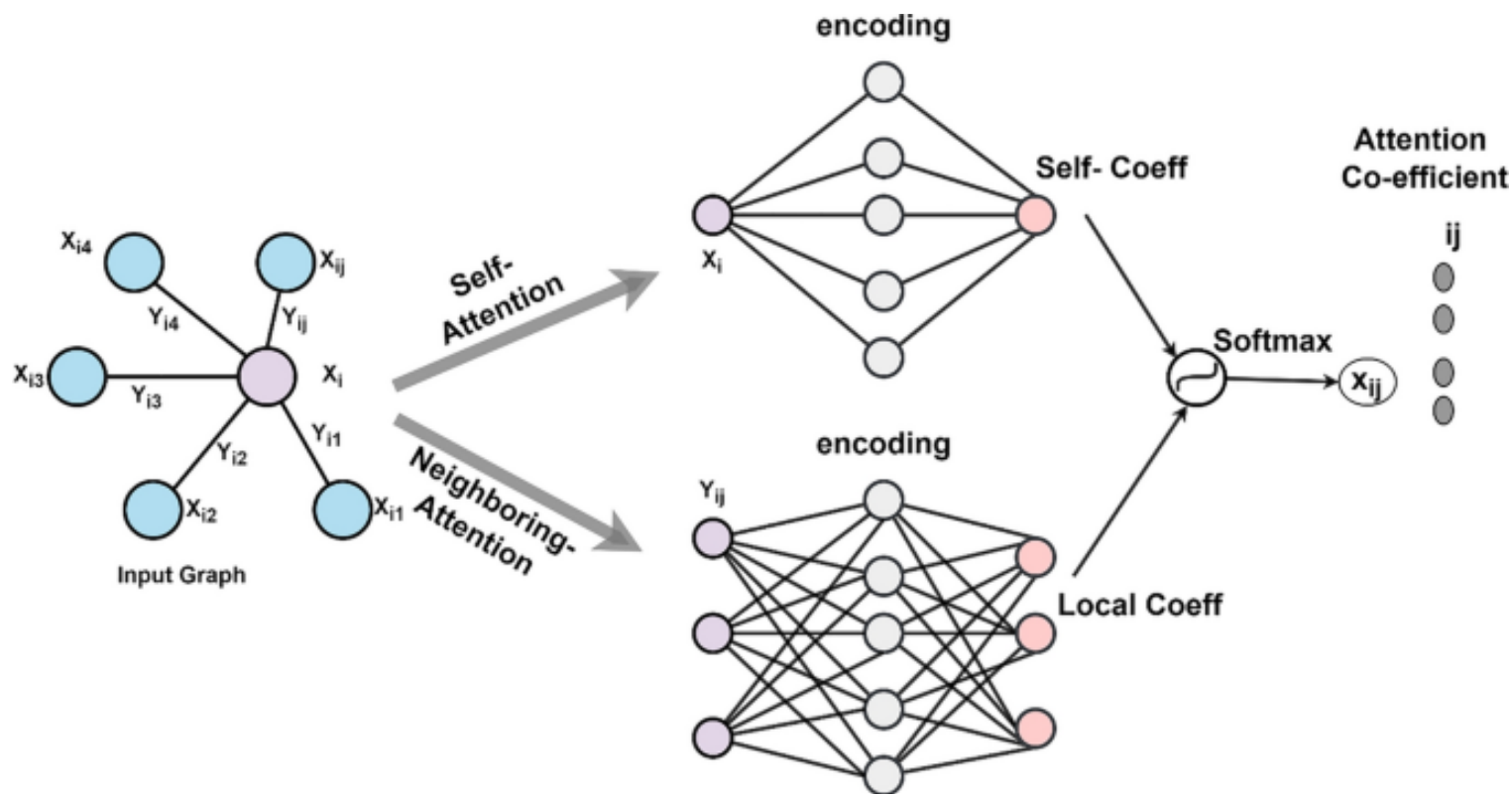
* 版权声明:

- 仅限用于个人学习
- 禁止用于任何商业用途

图注意力网络 (GAT - Graph Attention Network)

图注意力网络 (Graph Attention Network, 简称GAT) 利用注意力机制来学习节点之间的关系, 从而在图数据上实现节点分类、节点级别的特征学习等任务。

GAT的核心思想是在每个节点上计算注意力系数, 以确定节点与其邻居节点之间的重要性。这种注意力机制使得模型能够对不同节点之间的关系赋予不同的权重, 从而更好地捕捉图数据中的局部结构和全局信息。



GAT – 输入与输出

图注意力网络的输入：

在图注意力网络中，其节点的特征表示和普通的图神经网络中的节点的特征表示是类似的，都是采用embedding的方式对节点的特征表示进行向量化。对于图注意力神经网络而言，其初始的输入也是各个节点的特征组合。用公式表示就是：

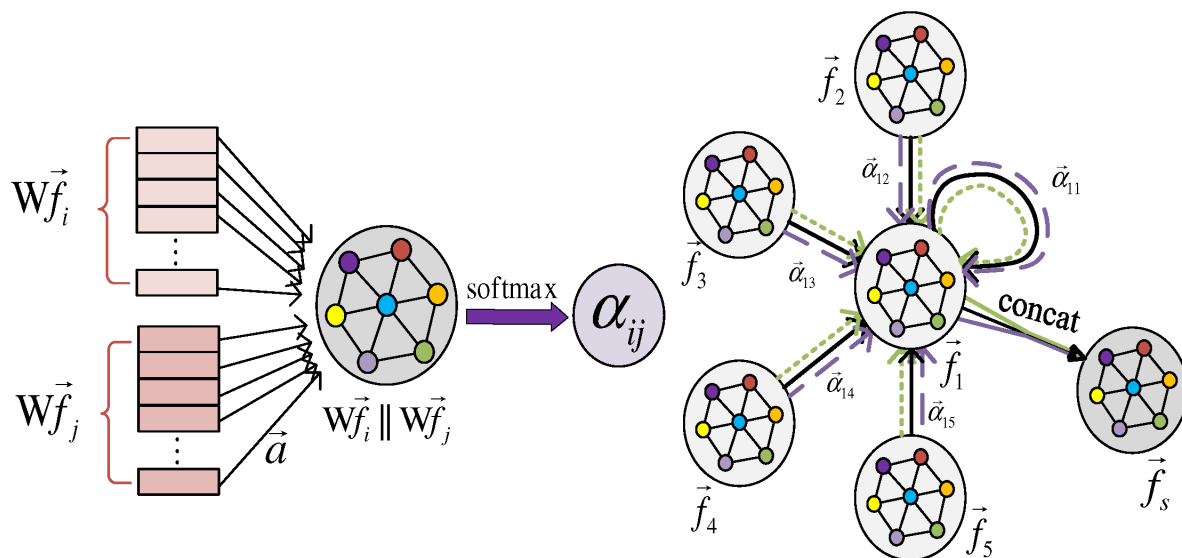
$$h = h_1, h_2, h_3, \dots, h_n, h_i \in \mathbb{R}^F$$

其中n表示图中节点的数量。F表示的是每一个节点的特征表示的数量。

图注意力网络的输出：

经过神经网络内部的运算，特神经网络对于下一层输出仍然是一系列的特征，用公式表达是：

$$h' = h'_1, h'_2, \dots, h'_n h'_i \in \mathbb{R}^{F'}$$



GAT - 注意力机制

注意力机制机制通过在模型训练过程中动态调整模型对数据/特征不同部分的权重提升模型的学习效果。



GAT 在图神经网络中引入了 Attention，希望关注邻居节点集合中比较重要的部分，淡化其余部分。

GAT - 注意力机制

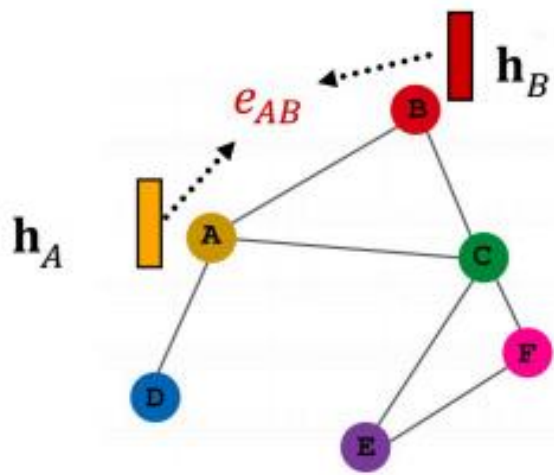
首先, 介绍下如何学习节点对 (i, j) 之间的注意力值 e_{ij} 。很明显, 为了计算 e_{ij} , 注意力网络 a 需要同时考虑两个节点的影响, 如下式:

$$e_{ij} = a(W\vec{h}_i W\vec{h}_j)$$

其中, \vec{h}_i , \vec{h}_j 分别是节点 i 和 j 的表示, W 是一个投影矩阵。

注意力网络可以有很多的设计方式, 这里作者将节点 i 和 j 的表示进行了拼接, 再映射为一个标量。需要注意, 这里拼接导致 $e_{ij} \neq e_{ji}$, 也就是说注意力值 e_{ij} 是非对称的。

$$e_{ij} = \text{LeakyReLU}(\vec{a}^T [\mathbf{W}\vec{h}_i \parallel \mathbf{W}\vec{h}_j])$$



GAT - 注意力机制

除了拼接操作以外，聚合邻居信息时，需要对每个节点的所有邻居的注意力进行归一化。归一化之后的注意力权重 α_{ij} 才是真正的聚合系数。

$$\alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}$$

这里的归一化进一步导致了注意力权重的非对称性。因为在归一化的过程中，每个节点的归一化对象并不一样。 e_{ij} 是针对节点 i 的所有邻居进行归一化，而 e_{ji} 是针对节点 j 的所有邻居进行归一化。综上，在求解注意力权重 α_{ij} 的过程中，分子和分母都是非对称的，进而也是非对称的。完整的图注意力公式如下：

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i || W\vec{h}_j]))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\vec{a}^T [W\vec{h}_i || W\vec{h}_k]))}$$

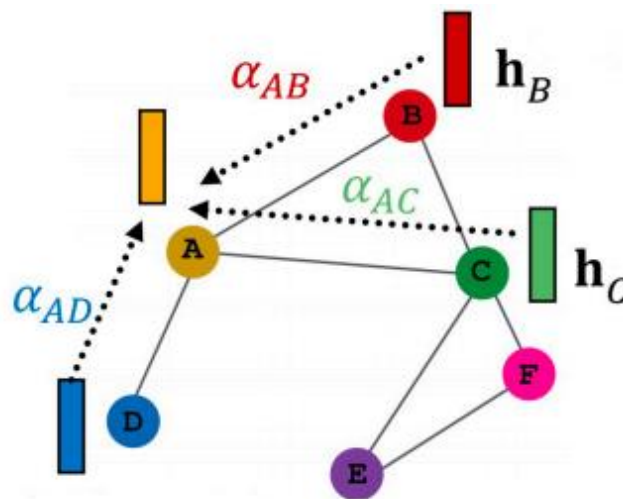
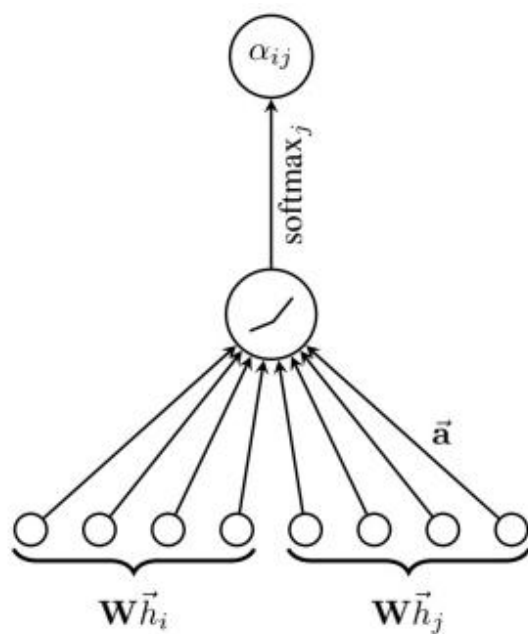
其中 $||$ 表示拼接操作。

GAT - 注意力机制

网络的输出特征可以通过下式得到:

$$\vec{h}'_i = \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij} W \vec{h}_j \right)$$

σ 为一个非线性激活函数。



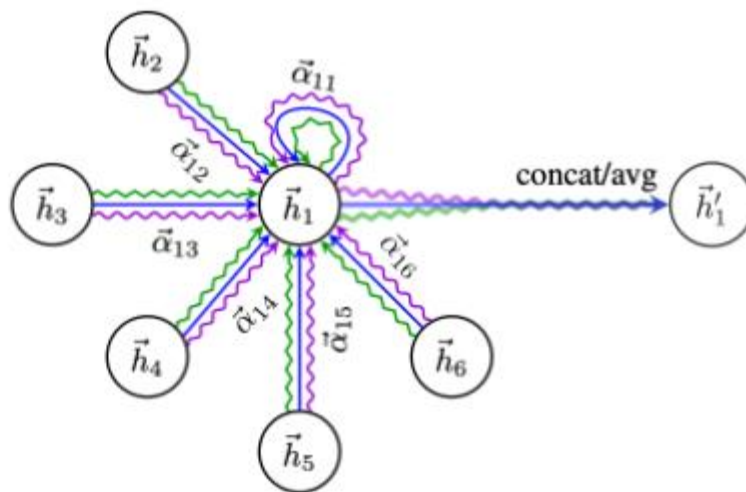
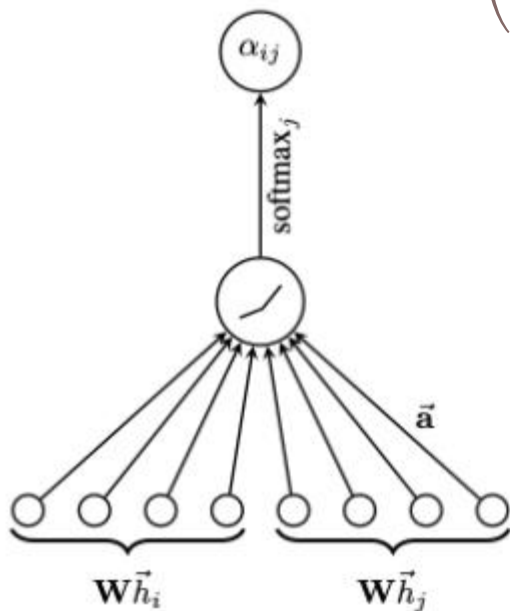
GAT – 多头注意力机制

为了提高模型的拟合能力，GAT 还引入了多头注意力机制 (multi-head attention)，即同时使用多个独立的注意力机制来计算 attention 因子，并将它们的结果合并（拼接或求和）：

$$\vec{h}'_i = \prod_{k=1}^K \sigma \left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right)$$

对于网络的最终层，用求均值的方式：

$$\vec{h}'_i = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k W^k \vec{h}_j \right)$$





Thank

You