

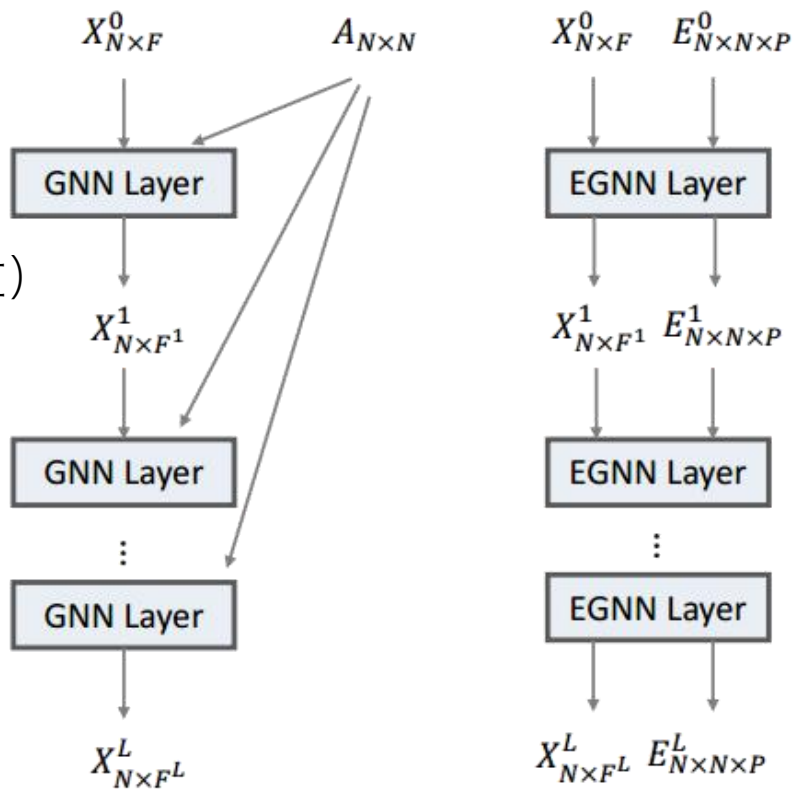
Exploiting Edge Features for Graph Neural Networks

基于边特征的图神经网络

CVPR 2019

GAT和GCN的局限性

- 没有充分使用边的特征
 - GAT只用到了连接与否
 - GCN只用到一维边的特征（权重）
- 原始邻居矩阵可能存在噪声



EGNN的创新点

- 新的注意力机制：
 - 利用多维的边的特征
- 新的网络结构：
 - 跨神经网络层的基于注意的边的自适应能力
- 去噪改进：
 - 双随机边归一化
- 有向图处理：
 - 对有向边编码多维边的特征

双随机归一化

$$\tilde{E}_{ijp} = \frac{\hat{E}_{ijp}}{\sum_{k=1}^N \hat{E}_{ikp}}$$

$$E_{ijp} = \sum_{k=1}^N \frac{\tilde{E}_{ikp} \tilde{E}_{jkp}}{\sum_{v=1}^N \tilde{E}_{vkp}}$$

EGNN(A)

$$X^l = \sigma \left[\bigparallel_{p=1}^P \left(\alpha_{\cdot,p}^l(X^{l-1}, E_{\cdot,p}^{l-1}) g^l(X^{l-1}) \right) \right]$$

$$g^l(X^{l-1}) = X^{l-1} W^l,$$

$$\hat{\alpha}_{ijp}^l = f^l(X_{i\cdot}^{l-1}, X_{j\cdot}^{l-1}) E_{ijp}^{l-1},$$

$$\alpha_{\cdot,p}^l = \text{DS}(\hat{\alpha}_{\cdot,p}^l),$$

$$f^l(X_{i\cdot}^{l-1}, X_{j\cdot}^{l-1}) = \exp \left\{ L \left(a^T [X_{i\cdot}^{l-1} W^l \| X_{j\cdot}^{l-1} W^l] \right) \right\}$$

EGNN(C)

$$X^l = \sigma \left[\bigparallel_{p=1}^P \left(E_{\cdot,p} X^{l-1} W^l \right) \right]$$

有向图边特征

$$\begin{bmatrix} \hat{E}_{ijp} & \hat{E}_{jip} & \hat{E}_{ijp} + \hat{E}_{jip} \end{bmatrix}$$

- X 是一个 $N \times F$ 的矩阵，表示全图的节点的特征
- X_{ij} 表示第 i 个节点的第 j 个特征
- X_i 表示第 i 个节点的 F 维的特征
- E 是一个 $N \times N \times P$ 的tensor，表示图上边的特征
- E_{ij} 表示连接节点 i 和节点 j 的边的 P 维的特征向量
- E_{ijp} 表示 E_{ij} 中的第 p 个通道
- $E_{ij} = 0$ 表示节点 i 和节点 j 之间没有边连接
- \mathcal{N}_i 表示节点 i 的邻居集合

引文网络（分类）

对于节点分类问题，沿着最后一个维度，每个节点embedding vector XiL都应用一个softmax operator

Table 1: Summary of citation network datasets

	Cora	Citeseer	Pubmed
# Nodes	2708	3327	19717
# Edges	5429	4732	44338
# Node Features	1433	3703	500
# Classes	7	6	3

Dataset	Cora		CiteSeer		Pubmed	
Splitting	Sparse	Dense	Sparse	Dense	Sparse	Dense
GCN	72.9 ± 0.8%	72.0 ± 1.2%	69.2 ± 0.7%	75.3 ± 0.4%	83.3 ± 0.4%	83.4 ± 0.2%
GAT	75.5 ± 1.1%	79.0 ± 1.0%	69.5 ± 0.5%	74.9 ± 0.5%	83.4 ± 0.1%	83.4 ± 0.2%
EGNN(C)[W]	82.7 ± 0.6%	87.6 ± 0.6%	69.3 ± 0.6%	76.0 ± 0.5%	84.5 ± 0.2%	84.3 ± 0.4%
EGNN(A)[W]	82.7 ± 0.6%	86.6 ± 0.6%	69.4 ± 0.5%	74.9 ± 0.8%	83.1 ± 0.2%	82.7 ± 0.2%
EGNN(C)[D]	81.8 ± 0.5%	85.1 ± 0.5%	70.6 ± 0.3%	75.0 ± 0.3%	84.3 ± 0.1%	84.1 ± 0.1%
EGNN(C)[DW]	83.2 ± 0.3%	87.4 ± 0.4%	70.3 ± 0.3%	75.4 ± 0.5%	84.1 ± 0.1%	84.1 ± 0.1%
EGNN(C)[M]	80.2 ± 0.4%	86.1 ± 0.5%	69.4 ± 0.3%	76.8 ± 0.4%	86.2 ± 0.2%	86.7 ± 0.1%
EGNN(C)[MW]	82.3 ± 0.4%	87.2 ± 0.4%	69.4 ± 0.3%	77.1 ± 0.4%	86.2 ± 0.1%	86.4 ± 0.3%
EGNN(C)[DM]	83.0 ± 0.3%	88.8 ± 0.3%	69.5 ± 0.3%	76.7 ± 0.4%	86.0 ± 0.1%	86.0 ± 0.1%
EGNN(C)[DMW]	83.4 ± 0.3%	88.5 ± 0.4%	69.5 ± 0.3%	76.6 ± 0.4%	85.8 ± 0.1%	85.6 ± 0.2%
EGNN(A)[A]	76.0 ± 1.0%	79.1 ± 1.0%	69.5 ± 0.4%	74.6 ± 0.3%	83.4 ± 0.1%	83.6 ± 0.2%
EGNN(A)[AW]	82.6 ± 0.6%	86.3 ± 0.9%	69.4 ± 0.4%	74.9 ± 0.4%	83.7 ± 0.2%	82.8 ± 0.3%
EGNN(A)[D]	80.1 ± 1.0%	85.4 ± 0.5%	70.1 ± 0.4%	74.7 ± 0.4%	84.3 ± 0.2%	84.2 ± 0.1%
EGNN(A)[DW]	82.7 ± 0.4%	87.2 ± 0.5%	69.5 ± 0.3%	74.5 ± 0.5%	83.9 ± 0.2%	83.3 ± 0.2%
EGNN(A)[M]	81.7 ± 0.4%	87.9 ± 0.4%	69.4 ± 0.3%	75.7 ± 0.3%	85.5 ± 0.1%	86.0 ± 0.1%
EGNN(A)[MW]	82.8 ± 0.3%	87.0 ± 0.6%	69.1 ± 0.3%	76.3 ± 0.5%	85.2 ± 0.2%	85.3 ± 0.3%
EGNN(A)[ADM]	82.5 ± 0.3%	88.4 ± 0.3%	69.4 ± 0.4%	76.5 ± 0.3%	85.7 ± 0.1%	86.7 ± 0.1%
EGNN(A)[ADMW]	83.1 ± 0.4%	88.4 ± 0.3%	69.3 ± 0.3%	76.3 ± 0.5%	85.6 ± 0.2%	85.7 ± 0.2%

分子分析（全图预测：分类、回归）

对于整个图的预测（分类或者回归），在XL的第一个维度应用一个池化层pooling layer使得特征矩阵简化为一个整个图的单个vector embedding。然后一个全连接层应用于该向量上，全连接层的输出可以被用来做回归的预测，或者是分类。

Dataset	Tox21 (AUC)		Lipo (RMSE)		Freesolv (RMSE)	
	Validation	Test	Validation	Test	Validation	Test
RF	0.78 ± 0.01	0.75 ± 0.03	0.87 ± 0.02	0.86 ± 0.04	1.98 ± 0.07	1.62 ± 0.14
Weave	0.79 ± 0.02	0.80 ± 0.02	0.88 ± 0.06	0.89 ± 0.04	1.35 ± 0.22	1.37 ± 0.14
EGNN(C)	0.82 ± 0.01	0.82 ± 0.01	0.80 ± 0.02	0.75 ± 0.01	1.07 ± 0.08	1.09 ± 0.08
EGNN(A)	0.82 ± 0.01	0.81 ± 0.01	0.79 ± 0.02	0.75 ± 0.01	1.09 ± 0.12	1.01 ± 0.12