Simple and Deep Graph Convolutional Networks

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, Yaliang Li

ICML2020

Problem

- Graph Convolutional Networks are stuck in a shallow archtecture due to the over-smoothing issue.
- This paper proposes some modifications to the convolution of the GCNs, enabling a deep architecture.

Variants of GCNs

- Vanilla GCN:

$$\mathbf{H}^{(\ell+1)} = \sigma\left(\tilde{\mathbf{P}}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)}\right)$$

- APPNP:

$$\boldsymbol{H}^{(\ell+1)} = (1-\alpha)\tilde{\boldsymbol{P}}\boldsymbol{H}^{(\ell)} + \alpha\boldsymbol{H}^{(0)}$$

- JKNet:

$$\text{Aggregate}(\left[\mathbf{H}^{(1)}, \ldots, \mathbf{H}^{(K)}\right])$$

- DropEdge:

$$\mathbf{H}^{(\ell+1)} = \sigma\left(\tilde{\mathbf{P}}_{\text{drop}}\mathbf{H}^{(\ell)}\mathbf{W}^{(\ell)}\right)$$

Deep GCNs

- Initial residual connection:

$$(1-\alpha_\ell)\tilde{\mathbf{P}}\mathbf{H}^{(\ell)}+\alpha_\ell\mathbf{H}^{(0)}$$

- Identity mapping:

$$(1-\beta_\ell)\mathbf{I}_n+\beta_\ell\mathbf{W}^{(\ell)}$$

- Proprogation rule of the deep GCNs:

$$\mathbf{H}^{(\ell+1)}=\sigma\left(\left((1-\alpha_\ell)\tilde{\mathbf{P}}\mathbf{H}^{(\ell)}+\alpha_\ell\mathbf{H}^{(0)}\right)\left((1-\beta_\ell)\mathbf{I}_n+\beta_\ell\mathbf{W}^{(\ell)}\right)\right)$$

- α_ℓ is recommended to set to 0.1 or 0.2
- $\beta_\ell = \log(\frac{\lambda}{\ell}+1) \approx \frac{\lambda}{\ell}$ where λ is a hyper-parameter (they set to 0.5)

Results

Table 2. Summary of classification accuracy (%) results on Cora, Citeseer, and Pubmed. The number in parentheses corresponds to the number of layers of the model.

Method	Cora	Citeseer	Pubmed
GCN	81.5	71.1	79.0
GAT	83.1	70.8	78.5
APPNP	83.3	71.8	80.1
JKNet	81.1 (4)	69.8 (16)	78.1 (32)
JKNet(Drop)	83.3 (4)	72.6 (16)	79.2 (32)
Incep(Drop)	83.5 (64)	72.7 (4)	79.5 (4)
GCNII	**85.5** $\pm$ **0.5** (64)	**73.4** $\pm$ **0.6** (32)	80.2 $\pm$ 0.4 (16)
GCNII*	85.3 $\pm$ 0.2 (64)	73.2 $\pm$ 0.8 (32)	**80.3** $\pm$ **0.4** (16)

Results

Table 3. Summary of classification accuracy (%) results with various depths.

Dataset	Method	Layers 2	4	8	16	32	64
Cora	GCN	**81.1**	80.4	69.5	64.9	60.3	28.7
	GCN(Drop)	**82.8**	82.0	75.8	75.7	62.5	49.5
	JKNet	-	80.2	80.7	80.2	**81.1**	71.5
	JKNet(Drop)	-	**83.3**	82.6	83.0	82.5	83.2
	Incep	-	77.6	76.5	81.7	**81.7**	80.0
	Incep(Drop)	-	82.9	82.5	83.1	83.1	**83.5**
	GCNII	82.2	82.6	84.2	84.6	85.4	**85.5**
	GCNII*	80.2	82.3	82.8	83.5	84.9	**85.3**
Citeseer	GCN	**70.8**	67.6	30.2	18.3	25.0	20.0
	GCN(Drop)	**72.3**	70.6	61.4	57.2	41.6	34.4
	JKNet	-	68.7	67.7	**69.8**	68.2	63.4
	JKNet(Drop)	-	72.6	71.8	**72.6**	70.8	72.2
	Incep	-	69.3	68.4	**70.2**	68.0	67.5
	Incep(Drop)	-	**72.7**	71.4	72.5	72.6	71.0
	GCNII	68.2	68.9	70.6	72.9	**73.4**	73.4
	GCNII*	66.1	67.9	70.6	72.0	**73.2**	73.1
Pubmed	GCN	**79.0**	76.5	61.2	40.9	22.4	35.3
	GCN(Drop)	**79.6**	79.4	78.1	78.5	77.0	61.5
	JKNet	-	78.0	**78.1**	72.6	72.4	74.5
	JKNet(Drop)	-	78.7	78.7	79.1	**79.2**	78.9
	Incep	-	77.7	**77.9**	74.9	OOM	OOM
	Incep(Drop)	-	**79.5**	78.6	79.0	OOM	OOM
	GCNII	78.2	78.8	79.3	**80.2**	79.8	79.7
	GCNII*	77.7	78.2	78.8	**80.3**	79.8	80.1