Self-Attention Graph Pooling

ICML 2019

Junhyun Lee, Inyeop Lee, Jaewoo Kang

Introduction

- *Convolutional neural networks in graph*
  - *Convolution*
  - *Downsampling (pooling)*
- *Previous researches*
  - *Only graph topology*
  - *Diffpool : quadratic storage complexity; many parameters*
  - *gPool : no topology*

Introduction

- Self-Attention Graph Pooling
  - hierarchical
  - end-to-end
  - few parameters
  - both node features and graph topology

Related Work

- Topology based pooling
  - Eigendecomposition
- Global pooling : consider graph features
  - Set2Set
  - SortPool
- Hierarchical pooling : capture structural information
  - Diffpool
  - gPool

Method

- The key point of SAGPool is that it uses a GNN to provide self-attention scores.

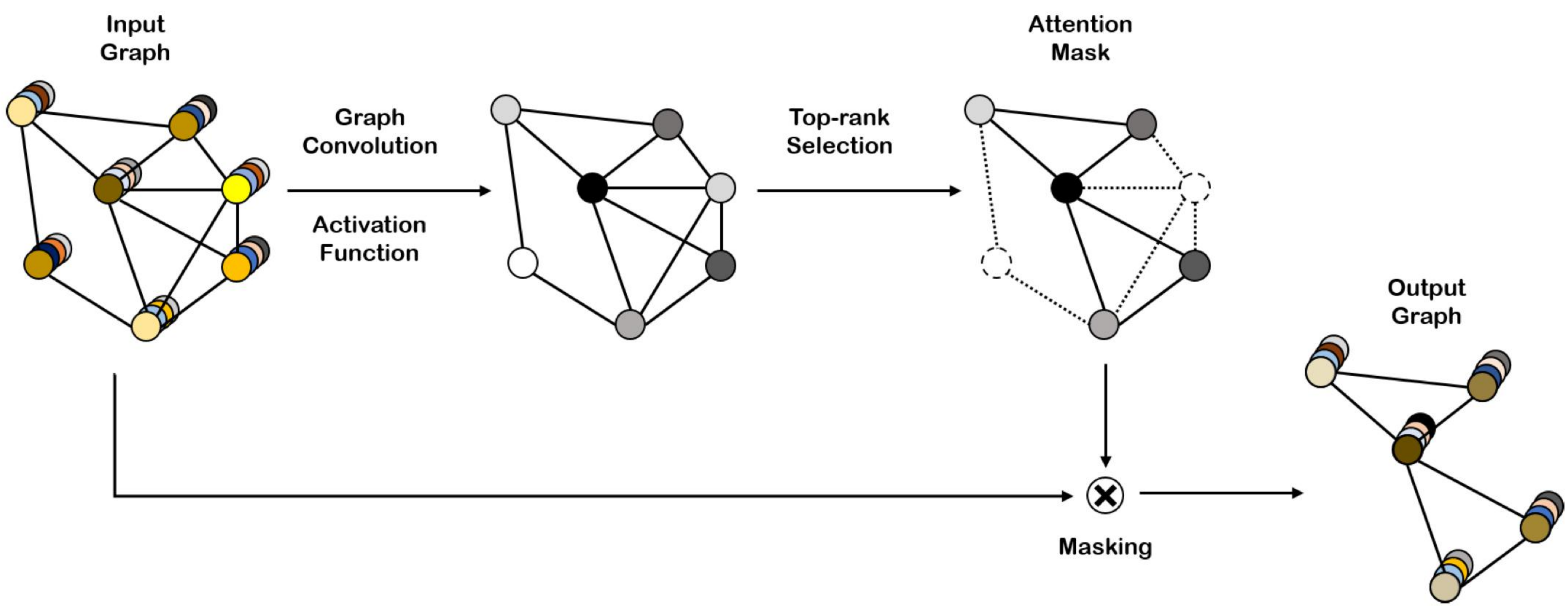

Figure 1. An illustration of the SAGPool layer.

Method

- Self-attention mask
  - the self-attention score $Z \in \mathbb{R}^{N \times 1}$ is calculated as follows.

$$Z = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta_{att}) \quad (3)$$

$\Theta_{att} \in \mathbb{R}^{F \times 1}$ is the only parameters

  - The result is based on both graph features and topology.
  - Node selection $\mathrm{idx} = \text{top-rank}(Z, \lceil kN \rceil), \quad Z_{mask} = Z_{\mathrm{idx}}$
  - Graph pooling $X' = X_{\mathrm{idx},:}, \; X_{out} = X' \odot Z_{mask}, \; A_{out} = A_{\mathrm{idx},\mathrm{idx}}$

Method

- Variation of SAGPool
  - The generalized equatio $Z = \sigma(\mathrm{GNN}(X, A))$
  - Variation 1 $Z = \sigma(\mathrm{GNN}(X, A + A^2))$
  - Variation 2 $Z = \sigma(\mathrm{GNN}_2(\sigma(\mathrm{GNN}_1(X, A)), A))$
  - Variation : $Z = \frac{1}{M}\sum_m \sigma(\mathrm{GNN}_m(X, A))$

Method

- Model Architecture
  - Convolution layer

$$h^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}h^{(l)}\Theta)$$

  - Readout

$$s = \frac{1}{N}\sum_{i=1}^{N} x_i \,||\, \max_{i=1}^{N} x_i$$

  - Global pooling architecture
  - Hierarchical pooling architecture

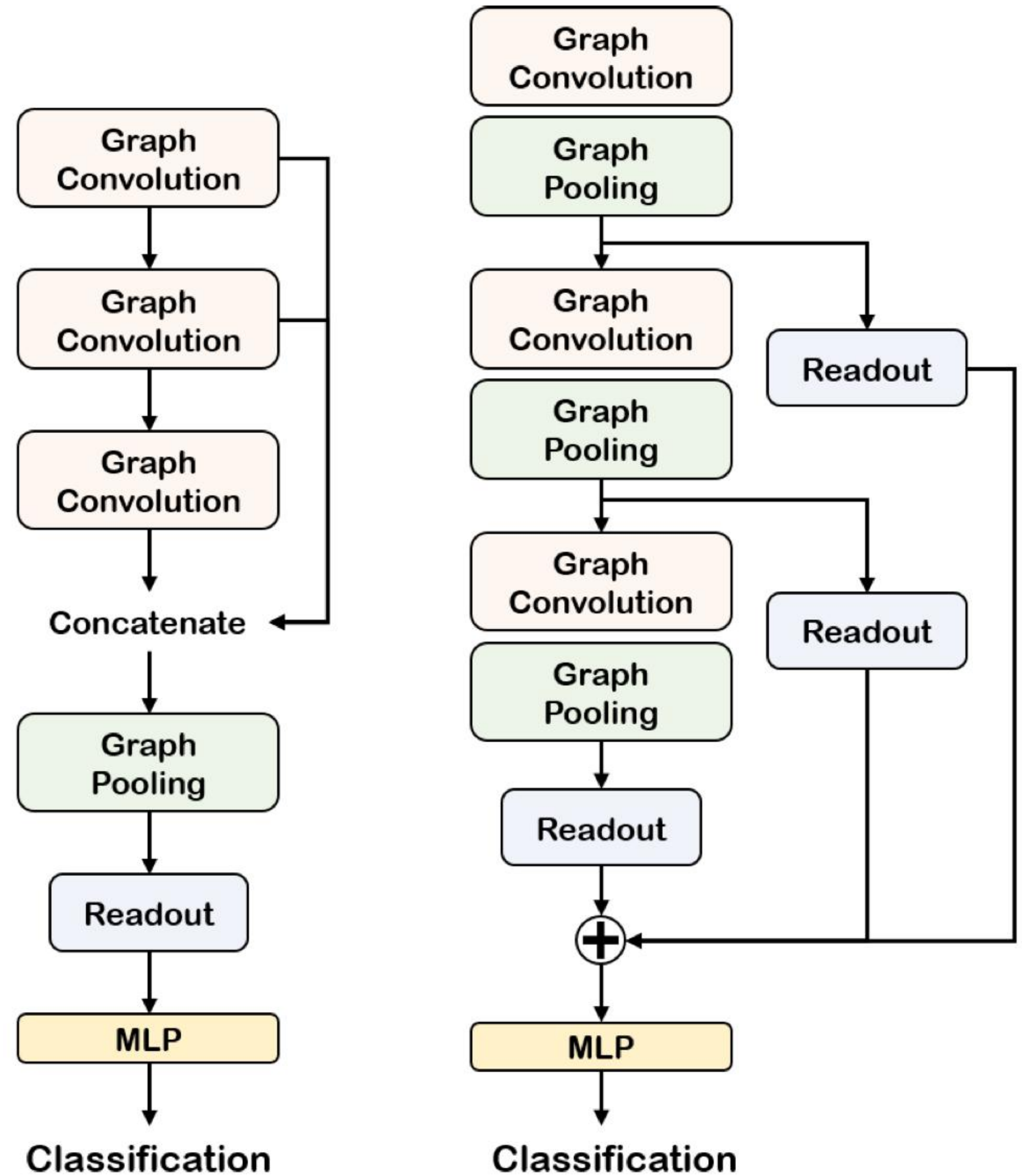

Figure 2. The global pooling architecture (left) and the hierarchical pooling architecture (right). These architectures are applied to all the baselines and SAGPool for a fair comparison. In this paper, the architecture on the left side is referred to as $POOL_g$ and the architecture on the right side is referred to as $POOL_h$ with the $POOL$ method (e.g. SAGPool$_g$, gPool$_h$).

Experiments

Table 4. Experimental results of SAGPool$_h$ variants. We compare ChebConv(K=2) (Defferrard et al., 2016), GCNConv (Kipf & Welling, 2016), SAGEConv (Hamilton et al., 2017), and GATConv(heads=6) (Velikovi et al., 2018). GCNConv is applied to SAGPool$_h$, SAGPool$_{h,\text{augmentation}}$, SAGPool$_{h,\text{serial}}$, and SAGPool$_{h,\text{parallel}}$.

Graph Convolution	**D&D**	**PROTEINS**
SAGPool$_h$	76.45 ± 0.97	71.86 ± 0.97
SAGPool$_{h,\text{Cheb}}$	75.82 ± 0.79	71.98 ± 0.93
SAGPool$_{h,\text{SAGE}}$	76.28 ± 1.06	71.93 ± 0.82
SAGPool$_{h,\text{GAT}}$	75.49 ± 0.93	71.98 ± 1.01
SAGPool$_{h,\text{augmentation}}$	77.07 ± 0.82	71.82 ± 0.81
SAGPool$_{h,\text{serial,2layers}}$	76.68 ± 0.96	72.17 ± 0.87
SAGPool$_{h,\text{parallel},M=2}$	75.79 ± 0.96	72.05 ± 0.43
SAGPool$_{h,\text{parallel},M=4}$	76.77 ± 0.61	71.66 ± 0.98

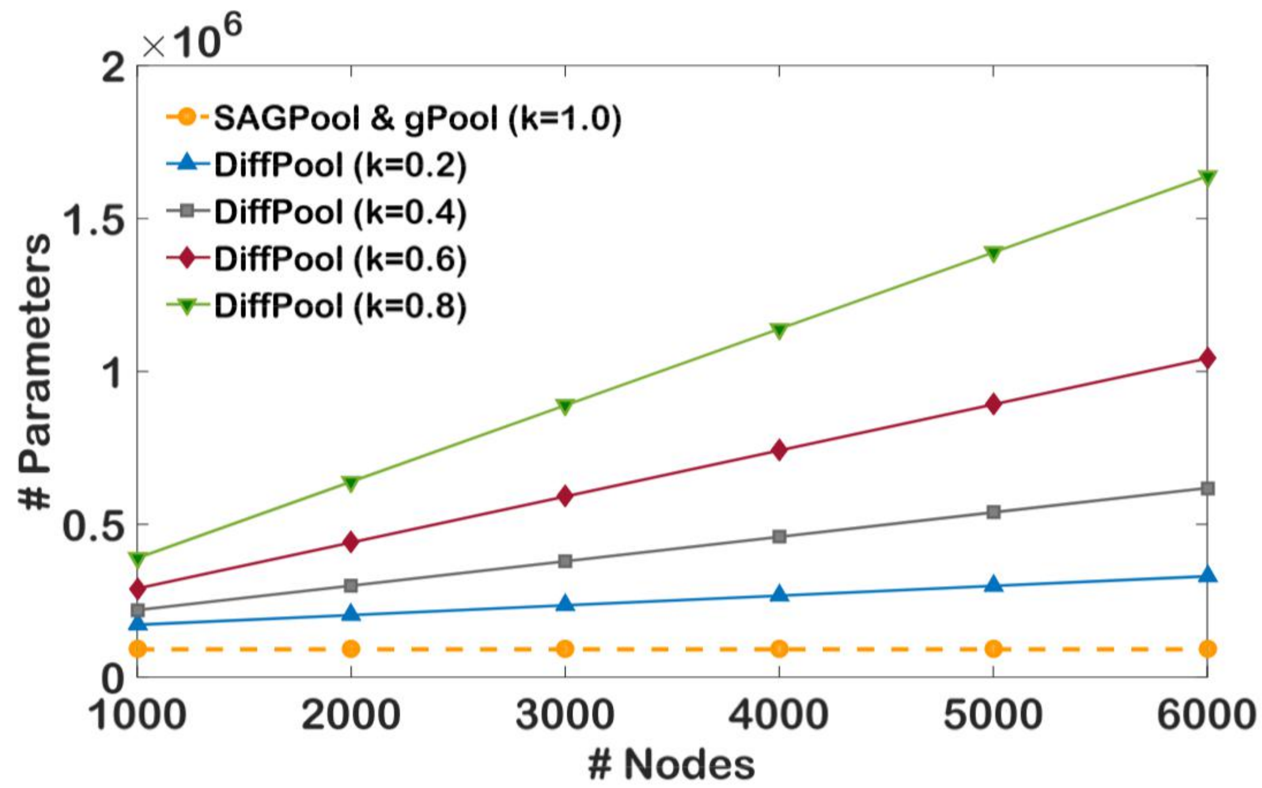

Figure 3. The increase in the number of parameters according to the number of graph nodes. The x-axis label denotes the number of input graph nodes and the y-axis label denotes the number of parameters of the hierarchical pooling models: the number of input node features is 128, the hidden feature size is 128, and the number of classes is 2. Equation (3) is used as a graph convolution of SAGPool. k denotes the pooling ratio and $k = 1.0$ indicates that the entire node is preserved after pooling. gPool and SAGPool have a consistent number of parameters regardless of the input graph size and the pooling ratio.

Experiments

- Analysis
  - 比较global and hierarchical pooling
  - 解释SAGPool方法如何解决gPool方法的缺点
  - 比较SAGPool与DiffPool的效率
  - 分析SAGPool变体
- Limitation
  - Cannot parameterize the pooling ratio k