Summary of related papers on Graph Transformer

Reported by Yidan Liu

gesis
Leibniz-Institut
für Sozialwissenschaften

2022_SIGIR_LogiformerA Two-Branch Graph Transformer Network for Interpretable Logical Reasoning

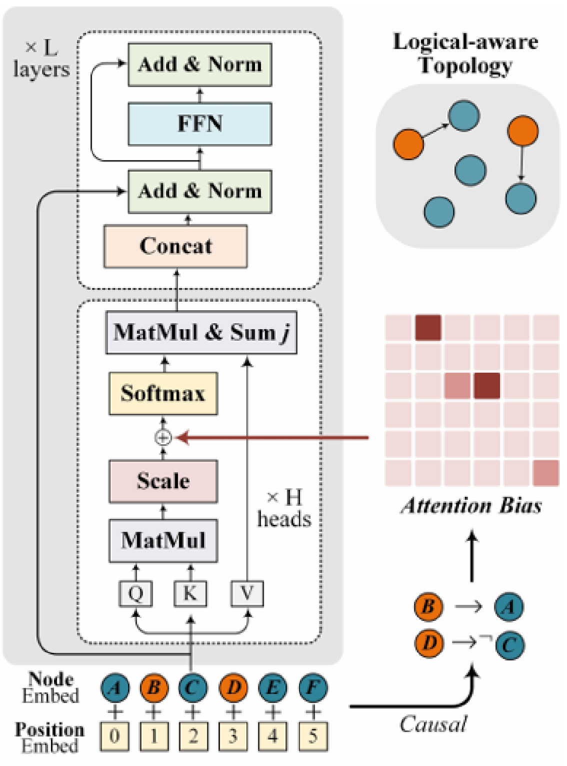


Figure 3: The illustration of logical-aware graph transformer. The inputs are the node sequence as well as the topology and the outputs are omitted.

$$Q = \mathrm{V_i} \cdot \mathrm{W^Q}, \quad K = \mathrm{V_i} \cdot \mathrm{W^K}, \quad V = \mathrm{V_i} \cdot \mathrm{W^V}, \tag{6}$$

$$A = \frac{QK^{\mathrm{T}}}{\sqrt{d_k}}, \quad Att(Q,K,V) = \mathrm{softmax}(A) \cdot V, \tag{7}$$

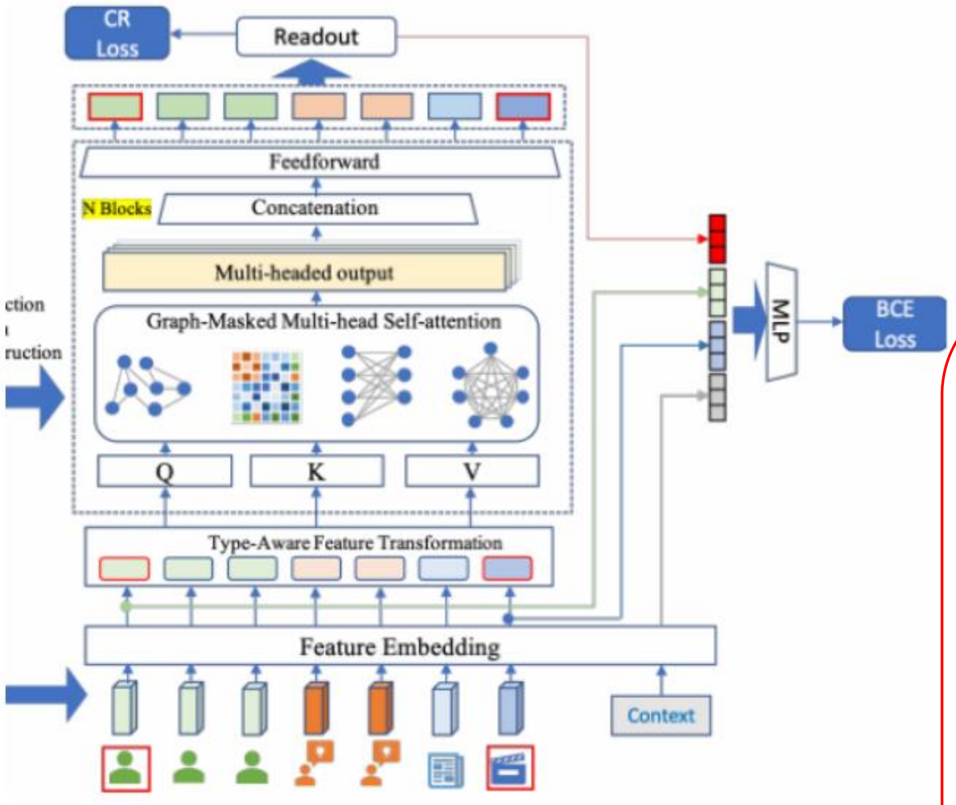
$$A' = \frac{QK^{\mathrm{T}}}{\sqrt{d_k}} + \mathrm{M}_{cas}. \tag{8}$$

$$Att_{MH}(Q,K,V) = [Head_1; ...; Head_H] \cdot \mathrm{W^H}, \tag{9}$$

$$\mathrm{V}_{cas} = \mathrm{V}_{cas}^{(L-1)} + \mathrm{V}_{cas}^{(L)}, \tag{10}$$

$$\mathrm{V}_{occ} \in \mathbb{R}^{K_{occ} \times d}$$

2022_SIGIR_Neighbour Interaction based Click-Through Rate Prediction via Graph-masked Transformer



Heterogeneous Node Feature Transformation layer:

$$\mathbf{x}_i = \mathbf{W}_i \mathbf{f}_i \qquad \text{User or item: } [\mathbf{f}_1, \cdots, \mathbf{f}_k]$$

$$\mathbf{h}_i = \mathrm{Linear}^{t(i)}(\mathbf{x}_i). \quad (1)$$

Graph-masked Multi-head Self-attention:

$$e_{ij} = \frac{(\mathbf{Q}\mathbf{h}_i)^\top (\mathbf{K}\mathbf{h}_j)}{\sqrt{d}},$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}, \quad (2)$$

$$\mathbf{z}_i = \sum_{j=1}^{n} \alpha_{ij} (\mathbf{V}\mathbf{h}_j),$$

$$e_{ij} = f_m\left(\frac{(\mathbf{Q}\mathbf{h}_i)^\top (\mathbf{K}\mathbf{h}_j)}{\sqrt{d}}, \mathbf{M}_{ij}\right), \quad (3)$$

$$f_m(x, \lambda) = \begin{cases} \lambda x & \lambda \neq 0 \\ -\infty & \lambda = 0. \end{cases} \quad (4)$$

$$Z_i = \mathrm{FFN}(\mathbf{W}^O \mathrm{Concat}(\mathbf{z}_i^1, \cdots, \mathbf{z}_i^H)), \quad (5)$$

$$\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{|\mathcal{N}_{uv}|}\}.$$

$$g_{uv} = \mathrm{Readout}(\mathbf{Z}), \quad (6)$$