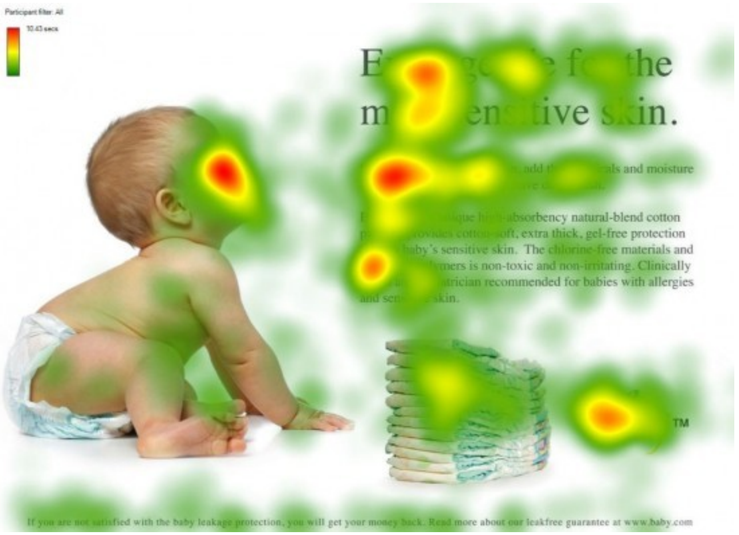
Transformer and GNN

Chengpei 2023-04-10

Transformer and GNN

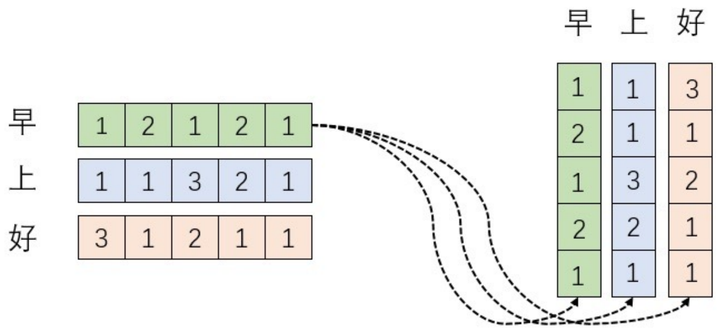
- self-attention
- Transformer
- Graph transformer

attention

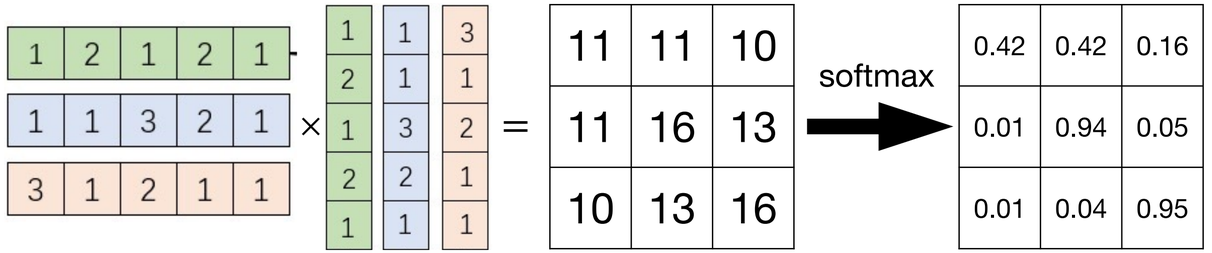


注意力机制：让模型学习不同区域的重要性，忽略无关信息，关注重要信息

simple self-attention



$X \in R^{N \times D}$

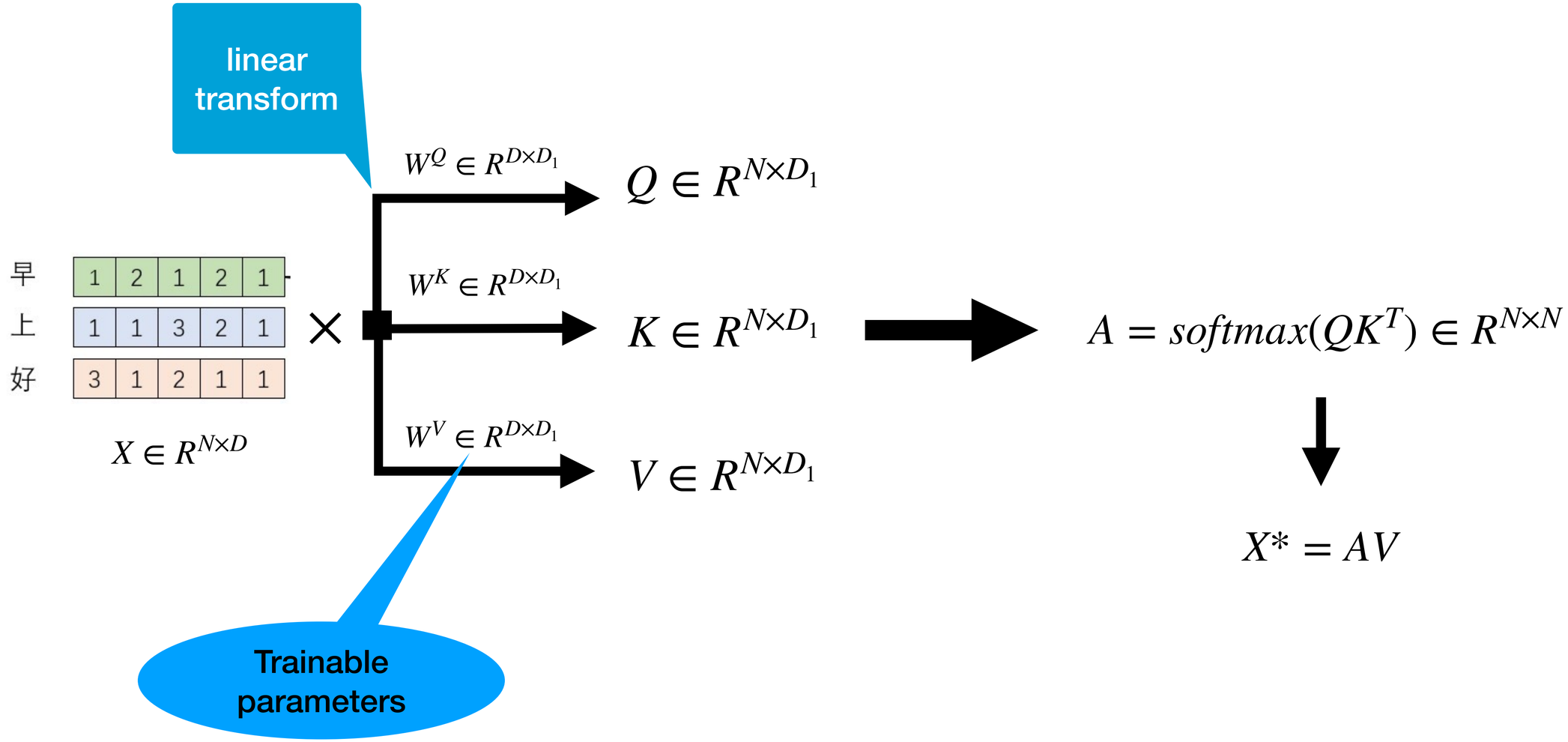


$$A = softmax(XX^T) \in R^{N \times N}$$

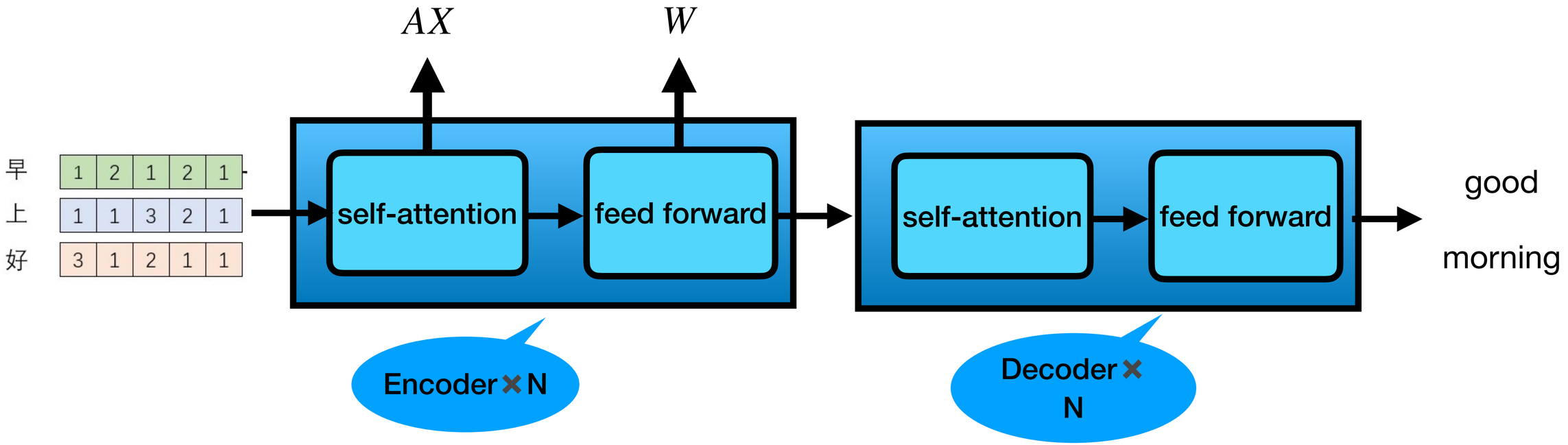
$$X^* = AX$$

No trainable parameters

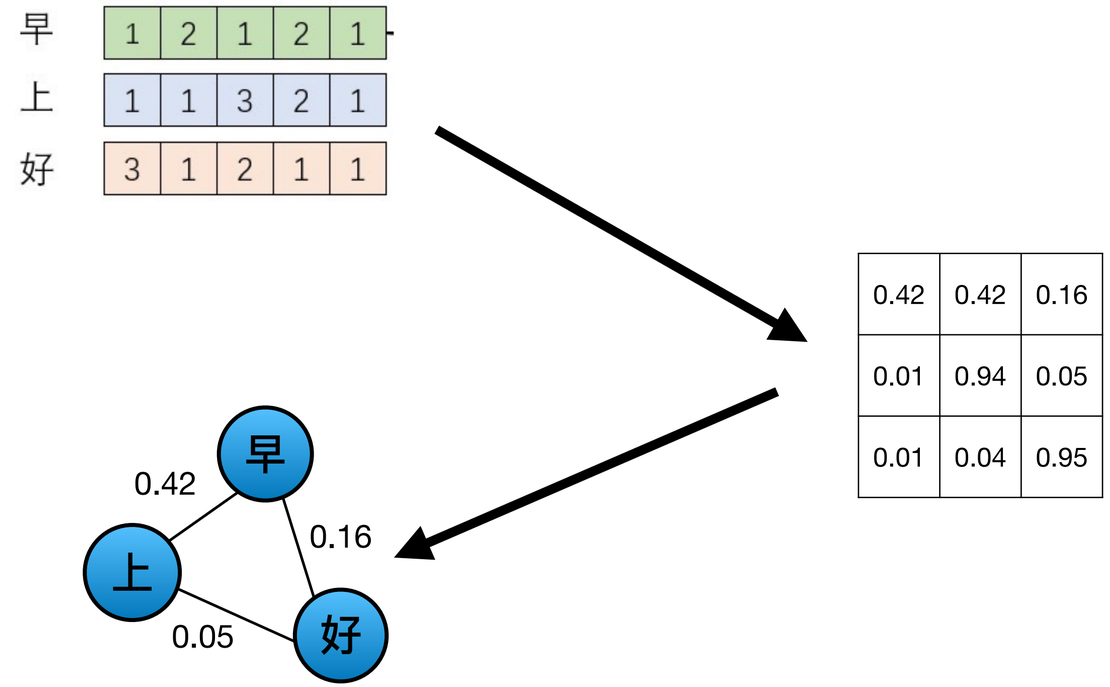
trainable self-attention



Transformer

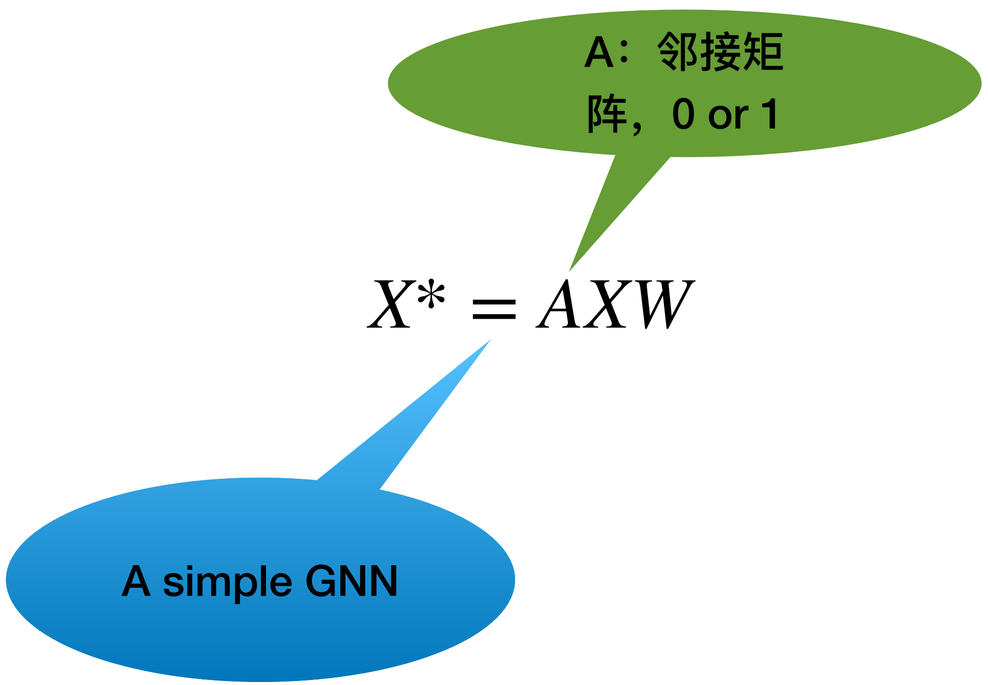


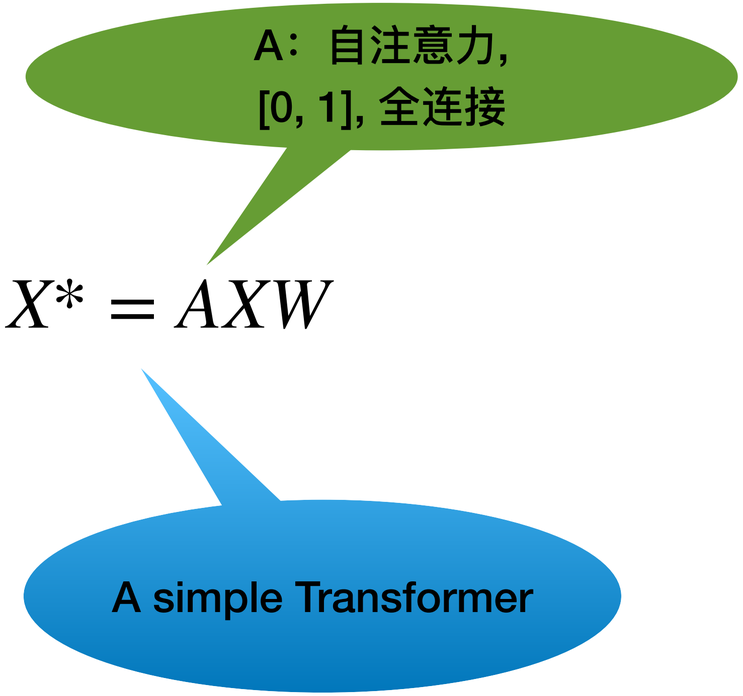
attention matrix is adjacency matrix



如果将一个句子看成一个Graph，每个词是一个node，注意力矩阵A实际上是整个图的带权邻接矩阵，并且整个图是全连接的

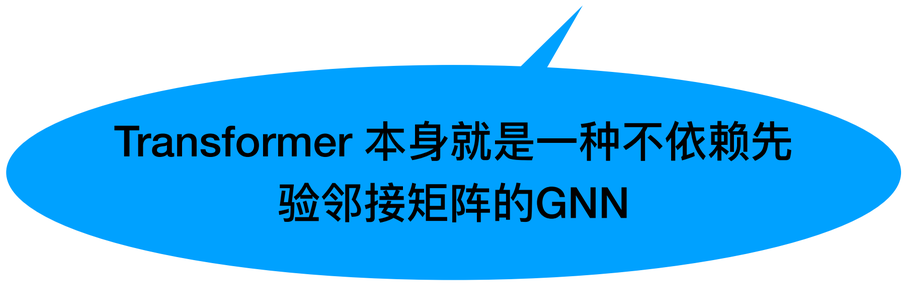
attention matrix is adjacency matrix



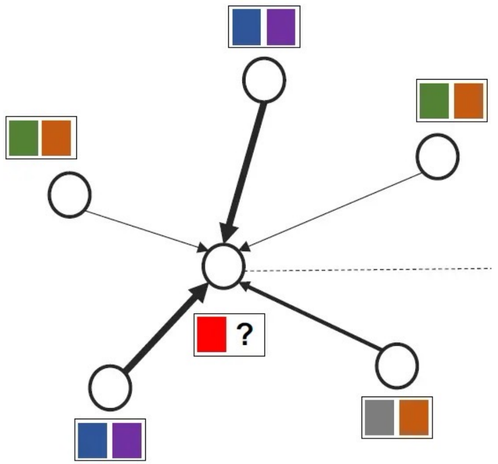


Transformer is a special GNN

- Transformer不把图的邻接矩阵（先验信息）作为模型输入，而是通过注意力机制自己学习‘邻接矩阵’，仅输入节点表示，因此每个节点会通过自注意力（学习到的邻接矩阵）聚合来自其它所有节点的信息。
- GNN（MPNN）同时接受节点特征和邻接矩阵矩阵（先验信息）作为模型输入，每个节点在学习过程中只聚合邻居节点的信息（使用了邻接矩阵先验）。



Transformer is equivalent to GAT



GAT: 使用自注意力机制，为节点的每个邻居节点学习一个权重，在聚合邻居节点特征的时候，乘上权重

当GAT处理的图是全连接图的时候，GAT就是Transformer