

日期: /

Heterogeneous Graphs and Knowledge Graph Embedding

一个异质图(Heterogeneous graph)由以下集合定义:

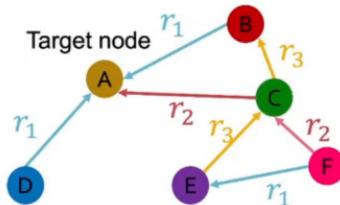
$$G = (V, E, R, T)$$

- 带有节点类型的节点 $v_i \in V$
- 带有关系类型的边 $(v_i, r, v_j) \in E$
- 节点类型 $T(v_i)$
- 边类型 $r \in R$

[Relational GCN]

用于处理异质图的GCN。

对于同质图, GCN表示为: $h^{(l+1)}_v = \sigma \left(\sum_{u \in N(v)} w^{(l)}_{uv} \frac{h_u^{(l)}}{|N(v)|} \right)$



对于异质图, 处理方式是以不同的权重处理不同边

$$h^{(l+1)}_v = \sigma \left(\sum_{r \in R(v)} \sum_{u \in N(v)} \frac{1}{C_{v,r}} W_r^{(l)} h_u^{(l)} + W_o^{(l)} h_v^{(l)} \right), C_{v,r} = |N(v)|$$

日期: /

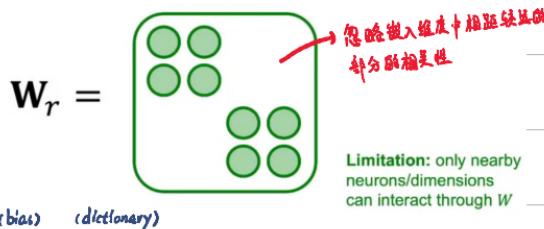
对于 L 层 RGCN, 每种关系的表示需要 L 个不同矩阵;

$W_r^{(1)}, W_r^{(2)} \dots W_r^{(L)}$, 每个矩阵的大小为 $d^{(1)} \times d^{(1)}$

这意味着 RGCN 的参数量会是巨大的, 极可能出现过拟合

$W_r^{(L)}$ 有 2 种规范化手段:

1) 使用块对角矩阵



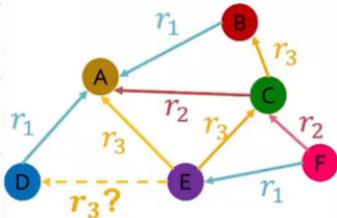
2) 使用基或字典学习

将各关系的权重矩阵表示为基的线性组合

$$W_r = \sum_{b=1}^B \underbrace{a_{rb} \cdot V_b}_{\text{矩阵 } V_b \text{ 是关系 } r \text{ 中的权重}} \text{ bias matrix}$$

异构图数据集的分割也是不同的, 在 link prediction 的任务中, 我们对 r 种关系分别进行直推式的划分, 再将其合并 (否则罕见类型的边可能不能被均匀分配)

日期: /



Link Prediction 任务的一般过程是:

1. 计算正边 (E, r_3, D) 得分
2. 计算负边 $\{(E, r_3, V) | V \in \{B, F\}\}$, since (E, r_3, A)
3. 训练时以 CE 为损失函数进行 GD, 预测时
执行排名 (Rank) 并计算 metrics (命中率, 平均倒数排名...)