

预训练模型学习情况周报 11

姚凯

一、 本周学习：

知识图谱实体及关系表示小结

1.One-hot Representation

缺点：相近词汇相似度为 0

2.Distributional Representation(分布表示)

独热码表示维度太高，分布式表示用上下文在低维空间表示一个词。上下文相似的词，其语义也相同。

优点：解决 nlp 数据稀疏问题；解决跨领域，跨对象问题；多任务学习的统一表示

3.TransE

目的：将实体与实体间的关系表示成向量

对于三元组<h,r,t>，h 和 t 为实体，通过翻译模型，输入两个向量得到三元组缺的另一个向量，并满足 $t \approx h + l$

损失函数构造：

$$\mathcal{L} = \sum_{(h,\ell,t) \in S} \sum_{(h',\ell,t') \in S_{(h,\ell,t)}} [\gamma + d(\mathbf{h} + \ell, \mathbf{t}) - d(\mathbf{h}' + \ell, \mathbf{t}')]_+$$

前一项为正例，后一项为负例。反例通过对正例三元组(h,r,t)中的 h 或 t 随机替换获得。

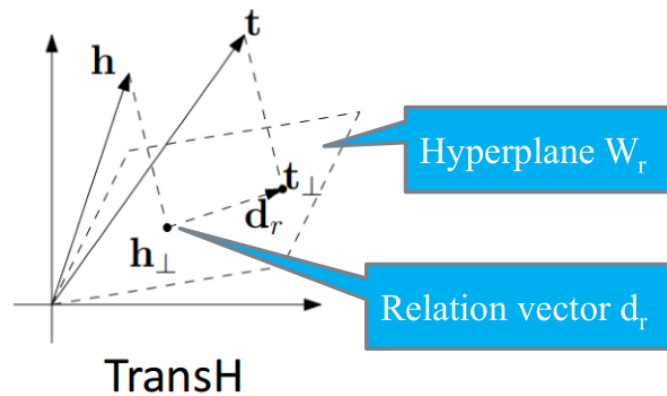
式子中的 d 为向量间的 L1 或 L2 距离

不能处理对称关系，一对多关系。可以处理组合关系，

4.TransH

改进点：对于 TranE, TransH 解决一对多，多对多的问题

为每一个关系 r ，建模一个超平面 w_r 。度量 h 和 t 时，先投影，在投影的超平面上求距离。损失函数类似 TransE



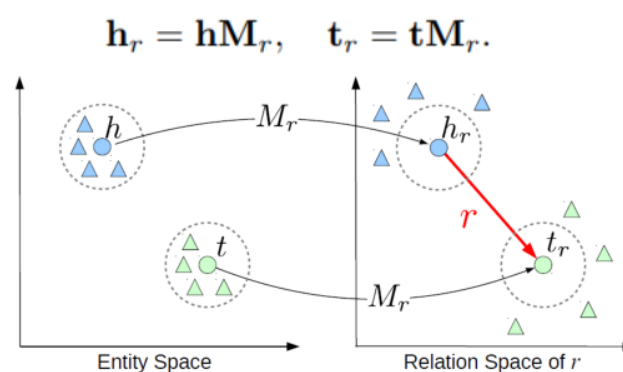
$$\mathcal{L} = \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'_{(h,r,t)}} [f_r(h, t) + \gamma - f_{r'}(h', t')]_+,$$

$$g_r(h) = h - w_r^{\top} h w_r \quad g_r(t) = t - w_r^{\top} t w_r$$

$f_r(h, t)$ 表示 $g_r(h)$ 与 $g_r(t)$ 的 L2 距离

5. TransR

改进点：之前关系向量与实体向量都在同一向量空间，相似实体在实体空间中的嵌入也相似。关系关注的是不同实体某一特定方面，故两个实体的 embedding 在关系空间中应该有很大差异。



损失函数类似 TransH:

$$\mathcal{L} = \sum_{(h,r,t) \in \Delta} \sum_{(h',r',t') \in \Delta'_{(h,r,t)}} [f_r(\mathbf{h}, \mathbf{t}) + \gamma - f_{r'}(\mathbf{h}', \mathbf{t}')]_+,$$

其中 $f_{\mathbf{r}}(h, t) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_2^2$.

二、 下周学习

进行实践，训练满足条件的 TransE, TransH 模型