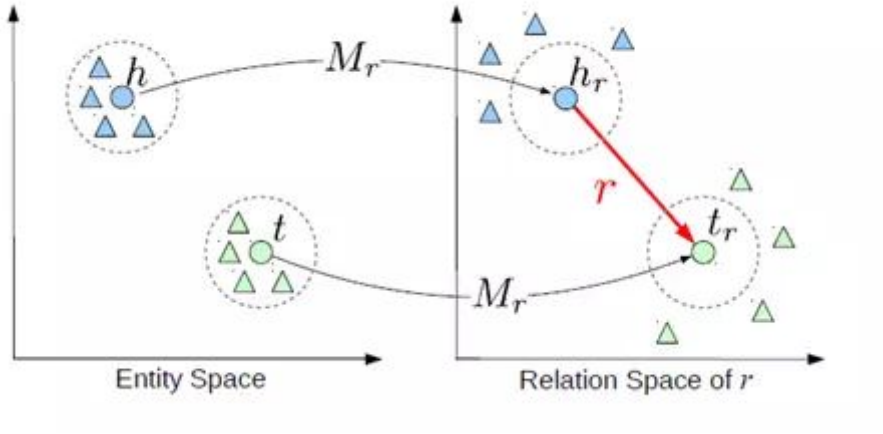
预训练模型学习情况周报 13

姚凯

本周学习

transR

transE和transH都假设实体和关系是在同一语义空间的向量。但每一个实体可以有很多方面，不同的关系关注的是实体的不同方面。TransR模型对不同的关系r建立各自的空间，在计算时先将实体映射到关系空间进行计算。在关系空间做向量的叠加。



$$h_r = hw_r$$

$$t_r = tw_r$$

再在关系空间中训练$h_r + r \approx t_r$（同transE）

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

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

RotatE

能够同时建模对称性/反对称性，互逆性，传递性

Definition 1. *A relation* $\mathtt{r}$ *is* ***symmetric (antisymmetric)*** *if* $\forall \mathtt{x}, \mathtt{y}$

$$\mathtt{r}(\mathtt{x},\mathtt{y}) \Rightarrow \mathtt{r}(\mathtt{y},\mathtt{x}) \quad (\ \mathtt{r}(\mathtt{x},\mathtt{y}) \Rightarrow \neg\mathtt{r}(\mathtt{y},\mathtt{x})\ )$$

A clause with such form is a ***symmetry (antisymmetry)*** *pattern.*

Definition 2. *Relation* $\mathtt{r}_1$ *is* ***inverse*** *to relation* $\mathtt{r}_2$ *if* $\forall \mathtt{x}, \mathtt{y}$

$$\mathtt{r}_2(\mathtt{x},\mathtt{y}) \Rightarrow \mathtt{r}_1(\mathtt{y},\mathtt{x})$$

A clause with such form is a ***inversion*** *pattern.*

Definition 3. *Relation* $\mathtt{r}_1$ *is* ***composed*** *of relation* $\mathtt{r}_2$ *and relation* $\mathtt{r}_3$ *if* $\forall \mathtt{x}, \mathtt{y}, \mathtt{z}$

$$\mathtt{r}_2(\mathtt{x},\mathtt{y}) \wedge \mathtt{r}_3(\mathtt{y},\mathtt{z}) \Rightarrow \mathtt{r}_1(\mathtt{x},\mathtt{z})$$

A clause with such form is a ***composition*** *pattern.*

rotatE模型将实体和关系映射到复数向量空间，将每个关系定义为从head实体到tail实体间的旋转

对三元关系（h, r, t），期望$t = h \circ r$，模长$|r_i| = 1$。其中○表示hadamard积（向量对应元素相乘）

transE不能建模对称性情况

Model	Score Function	Symmetry	Antisymmetry	Inversion	Composition
SE	$-\|W_{r,1}\mathbf{h} - W_{r,2}\mathbf{t}\|$	✗	✗	✗	✗
TransE	$-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$	✗	✓	✓	✓
TransX	$-\|g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t})\|$	✓	✓	✗	✗
DistMult	$\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$	✓	✗	✗	✗
ComplEx	$\mathrm{Re}(\langle \mathbf{h}, \mathbf{r}, \bar{\mathbf{t}} \rangle)$	✓	✓	✓	✗
RotatE	$-\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|$	✓	✓	✓	✓

Table 2: The pattern modeling and inference abilities of several models.

损失函数同样采用负采样

$$L = -log\sigma(\gamma - d_r(h,t)) - \sum_{i=1}^{n} \frac{1}{k} log\sigma(d_r(h'_i, t'_i) - \gamma)$$

$$其中, d_r(h,t) = ||h \circ t - r||$$

实践

TransR

transR在训练时用的transE模型，每个关系r的转换矩阵w_r均为直接对关系r进行embedding得到，而非通过训练得到

```
(myconda) root@Z922Dv:/OpenKE# python my_transR.py
Input Files Path : ./benchmarks/WN18RR/
The toolkit is importing datasets.
The total of relations is 11.
The total of entities is 40943.
The total of train triples is 86835.
Input Files Path : ./benchmarks/WN18RR/
The total of test triples is 3134.
The total of valid triples is 3034.
Finish initializing...
Epoch 0 | loss: 152.253739: 100%|██████████| 1/1 [00:00<00:00,  2.46it/s]
Finish initializing...
Epoch 942 | loss: 0.021703:  94%|
| 942/1000 [41:41<02:33,
Epoch 942 | loss: 0.021703:  94%|
| 943/1000 [41:41<02:30,
Epoch 999 | loss: 0.020034: 100%|| 1000/1000 [44:12<00:00,  2.65s/it]
100%| 3134/3134 [00:09<00:00, 317.13it/s]
no type constraint results:
metric:                 MRR             MR              hit@10          hit@3           hit@1
l(raw):                 0.138375        6386.518066     0.410657        0.237715        0.000000
r(raw):                 0.161296        3928.531982     0.462668        0.271857        0.006701
averaged(raw):          0.149836        5157.524902     0.436662        0.254786        0.003350

l(filter):              0.195995        6363.101562     0.433950        0.374920        0.004467
r(filter):              0.212555        3923.186768     0.473516        0.395660        0.009572
averaged(filter):       0.204275        5143.144043     0.453733        0.385290        0.007020
0.453733
0.45373326539993286
```

TransR链接预测的均值为45.4%

RotatE

实验采用自对抗负采样，不再均匀采样，根据当前嵌入模型对负三元组进行加权采样，加权的概率公式类似sigmoid函数

$$p(h'_j, r, t'_j | \{(h_i, r_i, t_i)\}) = \frac{\exp \alpha f_r(\mathbf{h}'_j, \mathbf{t}'_j)}{\sum_i \exp \alpha f_r(\mathbf{h}'_i, \mathbf{t}'_i)}$$

负采样的错误样本的score值越大（即负样本错的越离谱），权重越大

$$L = -\log \sigma(\gamma - d_r(\mathbf{h}, \mathbf{t})) - \sum_{i=1}^{n} p(h'_i, r, t'_i) \log \sigma(d_r(\mathbf{h}'_i, \mathbf{t}'_i) - \gamma)$$

```
(myconda) root@Z922Dv:/OpenKE# python my_rotatE.py
Input Files Path : ./benchmarks/WN18RR/
The toolkit is importing datasets.
The total of relations is 11.
The total of entities is 40943.
The total of train triples is 86835.
Input Files Path : ./benchmarks/WN18RR/
The total of test triples is 3134.
The total of valid triples is 3034.
Finish initializing...
Epoch 1999 | loss: 19.179349: 100%|████████████████████████████████████████| 2000/2000 [1:46:20<00:00,  3.19s/it]
100%|████████████████████████████████████████| 3134/3134 [00:11<00:00, 280.31it/s]
no type constraint results:
metric:                 MRR             MR              hit@10          hit@3           hit@1
l(raw):                 0.256455        4989.099121     0.461391        0.330249        0.141353
r(raw):                 0.332827        2313.251465     0.571474        0.425654        0.199426
averaged(raw):          0.294641        3651.175293     0.516433        0.377952        0.170389

l(filter):              0.451149        4965.660645     0.528079        0.465858        0.409700
r(filter):              0.491952        2307.915039     0.593172        0.514678        0.439694
averaged(filter):       0.471551        3636.787842     0.560625        0.490268        0.424697
0.560625
0.560625433921814
```

本模型训练采用自对抗负采样，RotatR链接预测的均值为56.1%。模型能建模的情况最多，链接预测的准确率也最高

参考：

https://blog.csdn.net/weixin_40449300/article/details/88771302

https://zhuanlan.zhihu.com/p/158950085