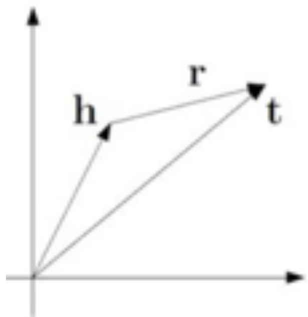
TransE：



Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\quad\quad \boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\quad\quad \mathbf{e} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

$\quad \mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

$\quad S_{batch} \leftarrow$sample$(S, b)$ // sample a minibatch of size b

$\quad T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

$\quad$ **for** $(h, \ell, t) \in S_{batch}$ **do**

$\quad\quad (h', \ell, t') \leftarrow$sample$(S'_{(h,\ell,t)})$ // sample a corrupted triplet

$\quad\quad T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

$\quad$ **end for**

$\quad$ Update embeddings w.r.t. $\displaystyle\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$

end loop

TransH:

hyperplane

TransE在自反，一对多，多对一，多对多的方面处理得不好，如果要满足上述方面就要构建很复杂的模型，不能用transE这样embedding这样简单高效的方法，TransH就是在TransE上优化上述方面。TransH在模型容量和效率之间权衡。

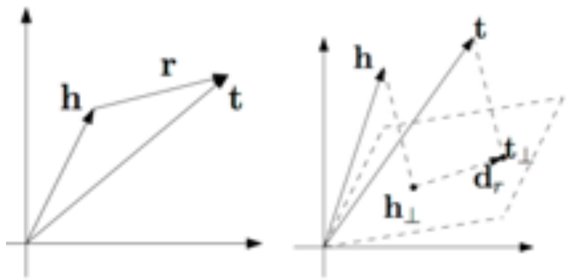
- If $(h, r, t) \in \Delta$ and $(t, r, h) \in \Delta$, i.e., r is a reflexive map, then $\mathbf{r} = \mathbf{0}$ and $\mathbf{h} = \mathbf{t}$.
- If $\forall i \in \{0, \ldots, m\}, (h_i, r, t) \in \Delta$, i.e., r is a many-to-one map, then $\mathbf{h}_0 = \ldots = \mathbf{h}_m$. Similarly, if $\forall i, (h, r, t_i) \in \Delta$, i.e., r is a one-to-many map, then $\mathbf{t}_0 = \ldots = \mathbf{t}_m$.

关系特征两个向量：平移向量wr和范数向量dr

$\|\mathbf{w}_r\|_2 = 1$, it is easy to get

$$\mathbf{h}_\perp = \mathbf{h} - \mathbf{w}_r^\top \mathbf{h} \mathbf{w}_r, \quad \mathbf{t}_\perp = \mathbf{t} - \mathbf{w}_r^\top \mathbf{t} \mathbf{w}_r.$$

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



(a) TransE　　(b) TransH

约束：

$$\forall e \in E, \|\mathbf{e}\|_2 \le 1, //\text{scale} \tag{1}$$
$$\forall r \in R, |\mathbf{w}_r^\top \mathbf{d}_r| / \|\mathbf{d}_r\|_2 \le \epsilon, //\text{orthogonal} \tag{2}$$
$$\forall r \in R, \|\mathbf{w}_r\|_2 = 1, //\text{unit normal vector} \tag{3}$$

使实体在涉及不同关系时具有分布式表示。
利用1对多，多对1来减少错误的负标签。

TransR：

TransE和TransH都假设实体和关系的嵌入在同一空间Rk中。但是，实体可能有多个方面，各种关系关注实体的不同方面。因此，直观的是，某些实体是相似的，因此在实体空间中彼此接近，但在某些特定方面具有相对差异，因此在相应的关系空间中彼此相距甚远。

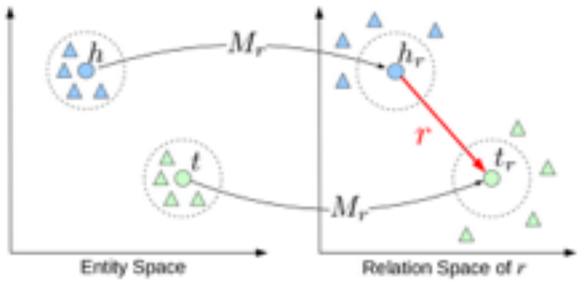


Figure 1: Simple illustration of TransR.

$\mathbf{h}, \mathbf{t} \in \mathbb{R}^k$　　$\mathbf{r} \in \mathbb{R}^d$　k/= d.

$\mathbf{h}_r = \mathbf{h}\mathbf{M}_r, \quad \mathbf{t}_r = \mathbf{t}\mathbf{M}_r.$　　Mr ∈ Rk×d

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

we have $\|\mathbf{h}\|_2 \le 1, \|\mathbf{r}\|_2 \le 1, \|\mathbf{t}\|_2 \le 1, \|\mathbf{h}\mathbf{M_r}\|_2 \le 1, \|\mathbf{t}\mathbf{M_r}\|_2 \le 1.$

CTransR:

为每个关系学习一个独特的向量，这可能代表不适合在这个关系下的所有实体对，因为这些关系通常是相当多样的。

先对所有实体的关系聚类(h – t)，然后再计算：

$$f_r(h, t) = \|\mathbf{h}_{r,c} + \mathbf{r}_c - \mathbf{t}_{r,c}\|_2^2 + \alpha \|\mathbf{r}_c - \mathbf{r}\|_2^2,$$

这样不会相差太远

TransD:

用两个向量表示关系或者实体：一个是实体（关系）的含义；另一个是实体嵌入到关系空间的投影向量。构造实体到关系的动态映射。

CTransR中，所有类型的实体共享相同的映射向量/矩阵。但是，不同类型的实体具有不同的属性和功能，不足以让它们共享相关的相同变换参数。

没有矩阵计算。

TransD比TransR计算更简单。

三元组(h, r, t)：向量 h, hp , r, rp , t, tp

$$\mathbf{M}_{rh} = \mathbf{r}_p \mathbf{h}_p^\top + \mathbf{I}^{m\times n}$$
$$\mathbf{M}_{rt} = \mathbf{r}_p \mathbf{t}_p^\top + \mathbf{I}^{m\times n}$$
$$\mathbf{h}_\perp = \mathbf{M}_{rh}\mathbf{h}, \quad \mathbf{t}_\perp = \mathbf{M}_{rt}\mathbf{t}$$
$$f_r(\mathbf{h},\mathbf{t}) = -\|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2^2$$

$$L = \sum_{\xi\in\Delta}\sum_{\xi'\in\Delta'}[\gamma + f_r(\xi') - f_r(\xi)]_+$$

负样例：Δ′ = {(hl, rk, tk) | hl/= hk ∧yk = 1}∪{(hk,rk,tl) | tl/= tk ∧yk = 1}