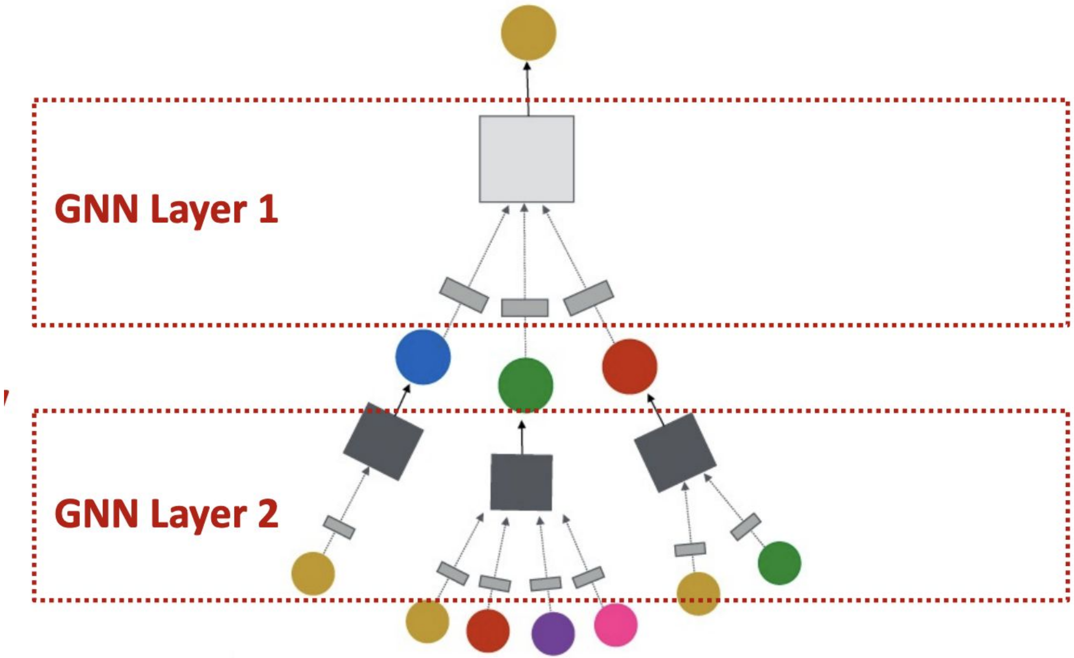
Глубинное обучение в анализе графовых данных

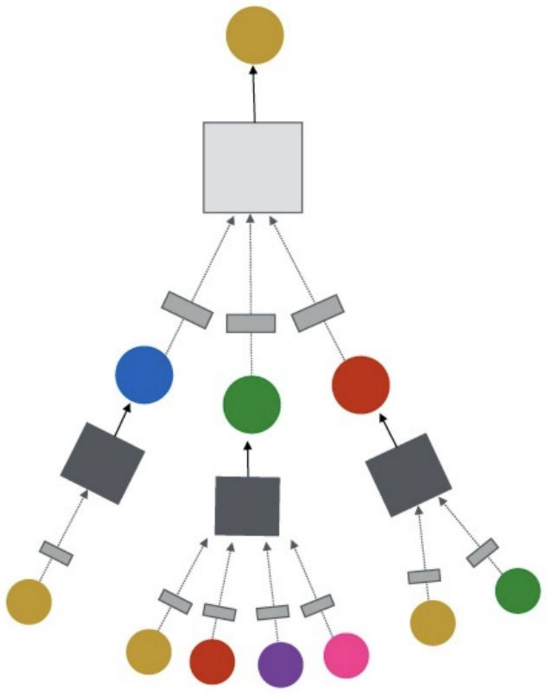
8. Глубинное обучение на графах знаний

в предыдущих сериях…

Связь слоев

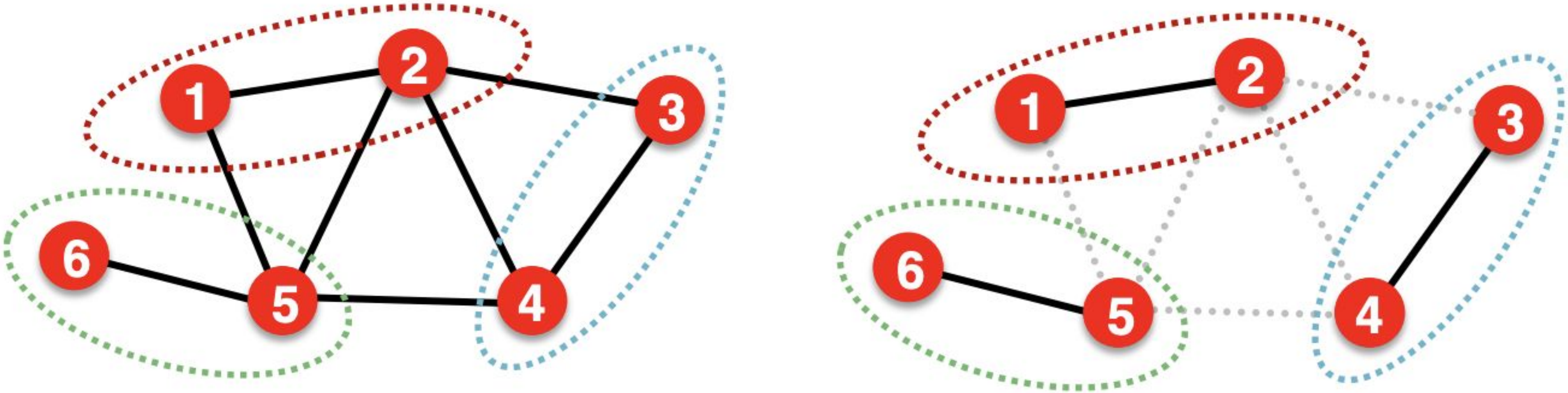


Аугментации



(4) Graph augmentation

Splitting

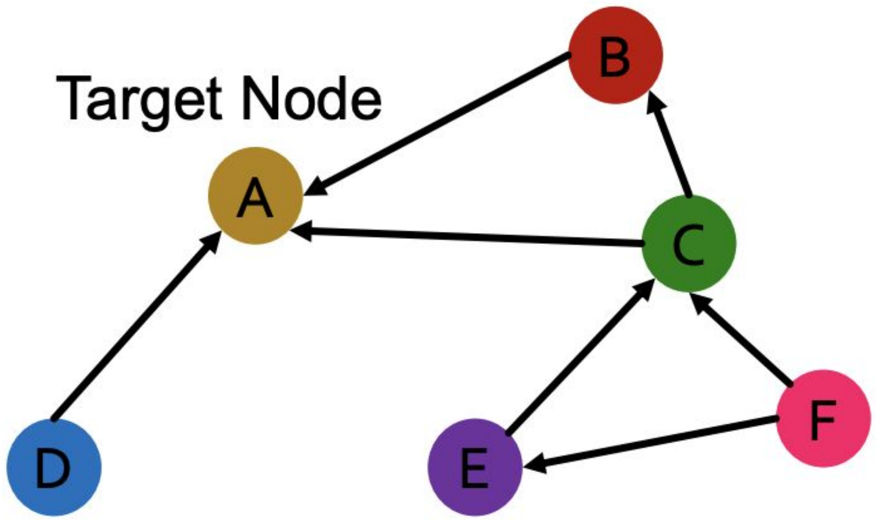


Графы знаний

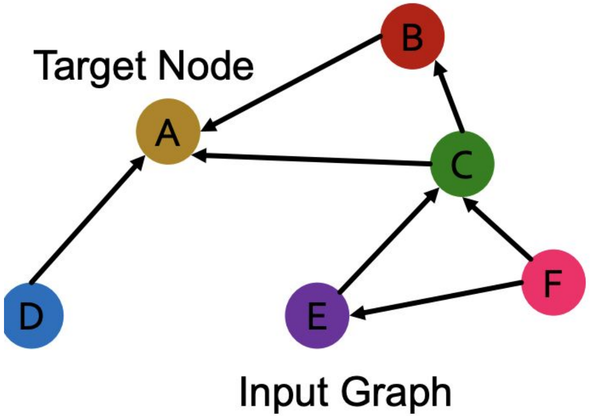
Гетерогенный граф

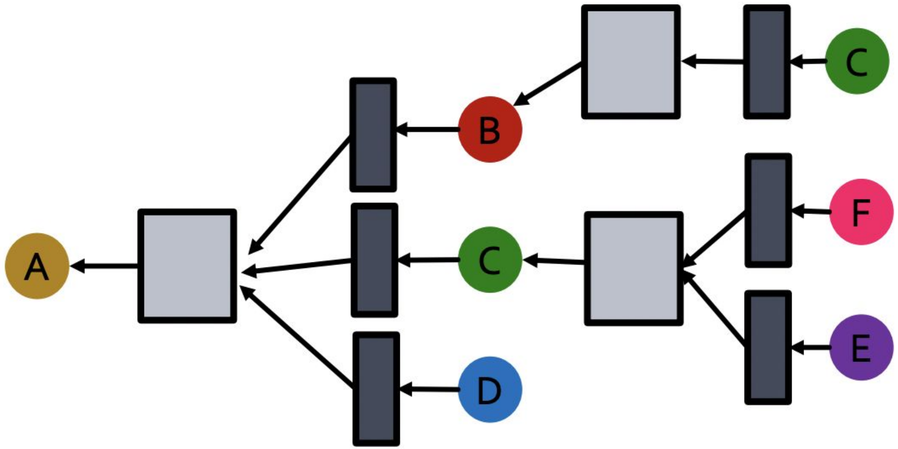
$$\boldsymbol{G = (V, E, R, T)}$$

Направленный граф

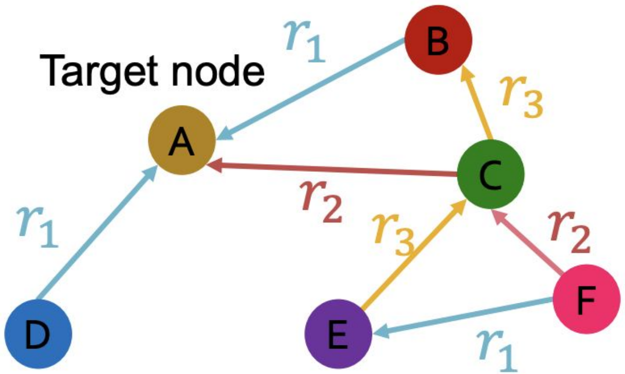


Вычислительный граф



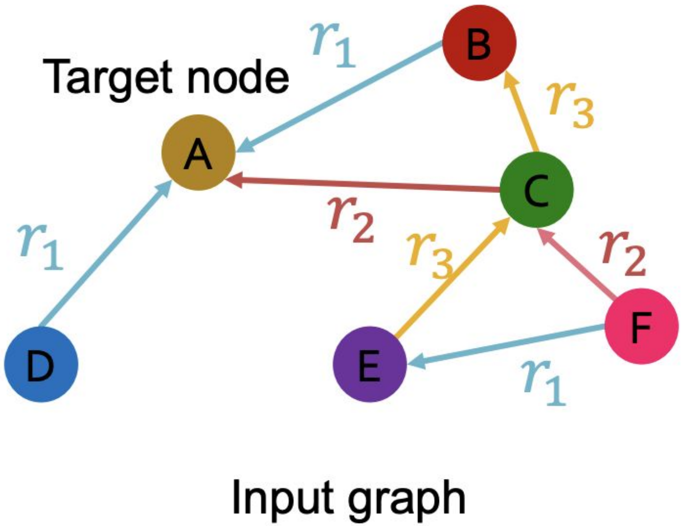


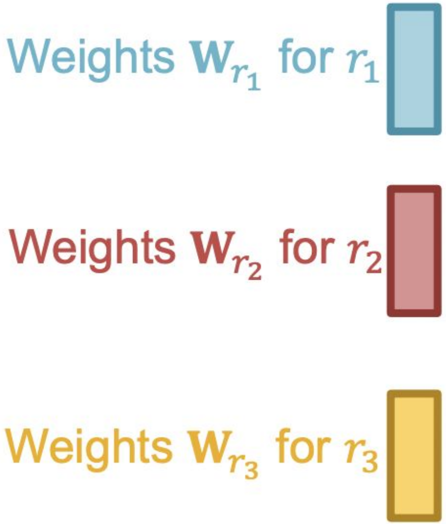
Граф знаний



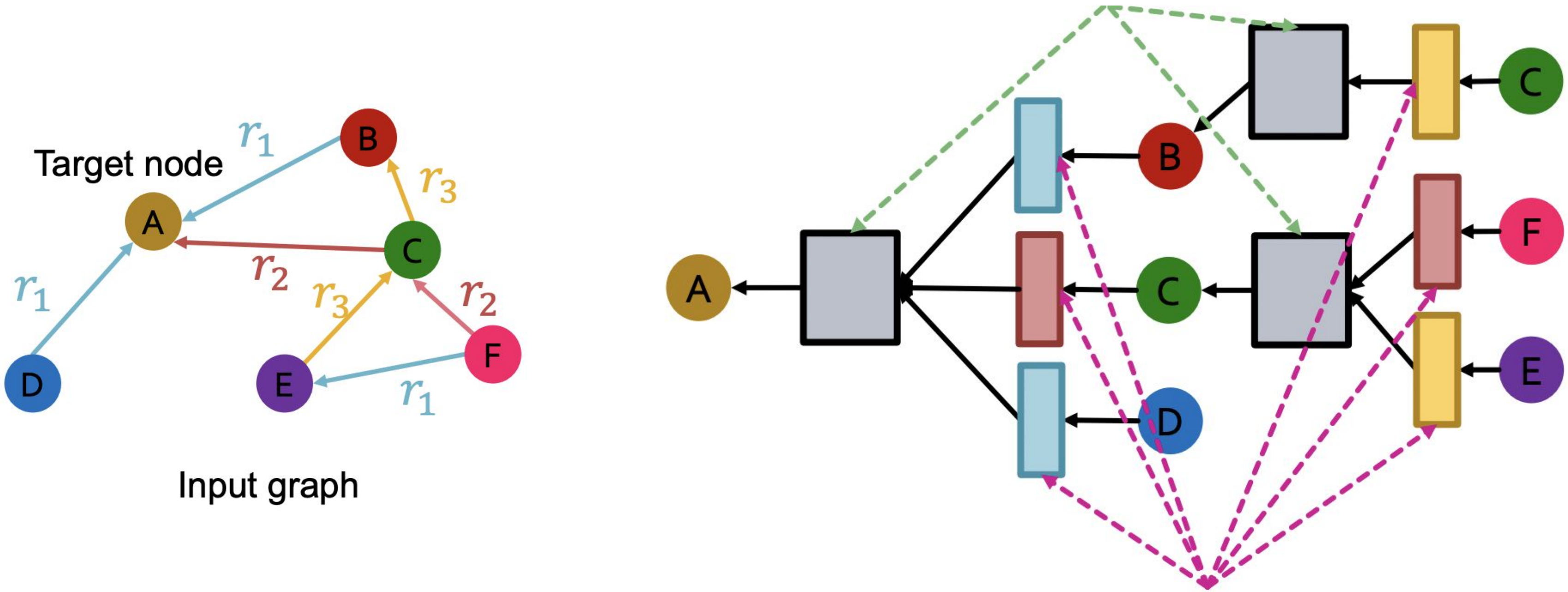
Input graph

Слои сообщений





Вычислительный граф в графе знаний



RGCN

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\sum_{r\in R}\sum_{u\in N_v^r}\frac{1}{c_{v,r}}\mathbf{W}_r^{(l)}\mathbf{h}_u^{(l)} + \mathbf{W}_0^{(l)}\mathbf{h}_v^{(l)}\right)$$

1) Message

$$\mathbf{m}_{u,r}^{(l)} = \frac{1}{c_{v,r}}\mathbf{W}_r^{(l)}\mathbf{h}_u^{(l)}$$

$$\mathbf{m}_v^{(l)} = \mathbf{W}_0^{(l)}\mathbf{h}_v^{(l)}$$

2) Aggregation

$$\mathbf{h}_v^{(l+1)} = \sigma\left(\mathrm{Sum}\left(\left\{\mathbf{m}_{u,r}^{(l)}, u \in N(v)\right\} \cup \left\{\mathbf{m}_v^{(l)}\right\}\right)\right)$$

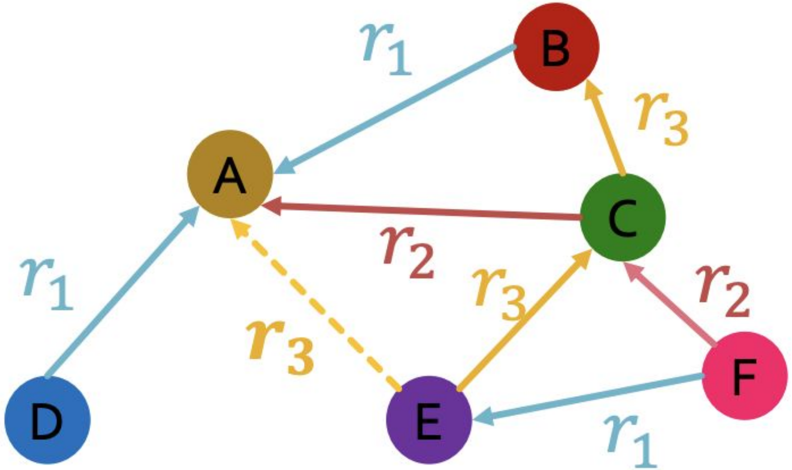
Проблемы с RGCN

Из-за большого числа параметров быстро переобучается

Варианты решения проблемы

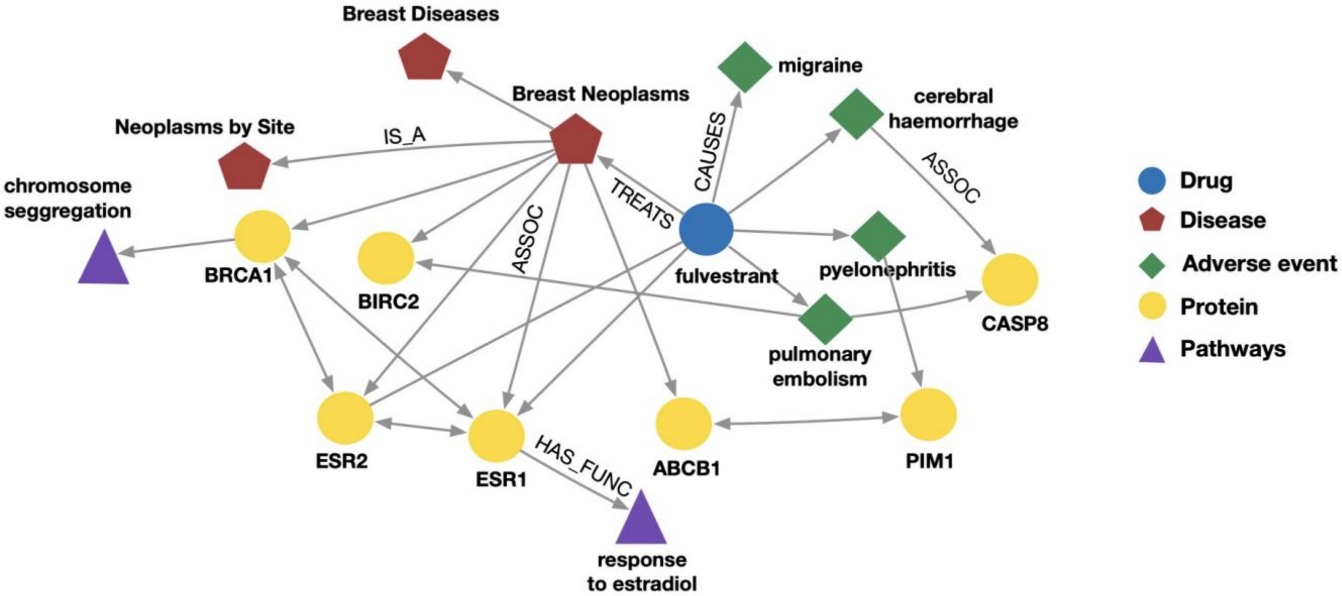
1) диагональная матрица
2) базисы

Пример



$$\ell = -\log \sigma\left(f_{r_3}(h_E, h_A)\right) - \log(1 - \sigma(f_{r_3}(h_E, h_B)))$$

Графы знаний



TransE

$$f_r(h, t) = -||\mathbf{h} + \mathbf{r} - \mathbf{t}||$$

TransE алгоритм обучения

Algorithm 1 Learning TransE

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

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

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

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

Entities and relations are initialized uniformly, and normalized

loop

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

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

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

Negative sampling with triplet that does not appear in the KG

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

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

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

end for

$\boldsymbol{d}$ represents distance (negative of score)

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

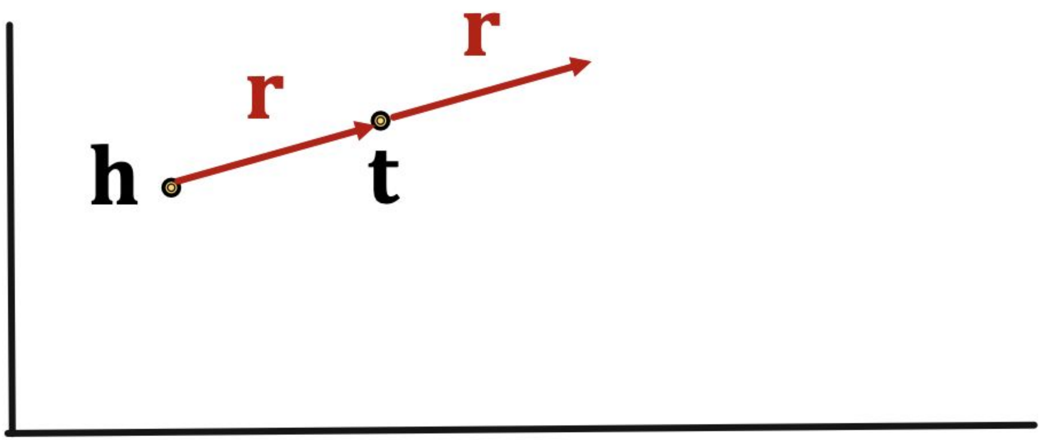
end loop

Contrastive loss: favors lower distance (or higher score) for valid triplets, high distance (or lower score) for corrupted ones

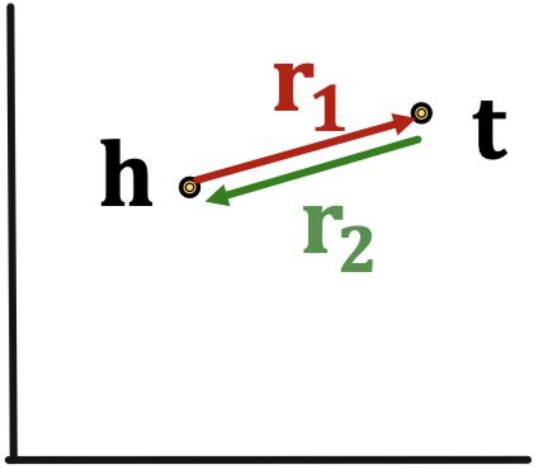
Шаблоны отношений

- (анти)Симметрия : $r(h,t) \Rightarrow r(t,h) \;\; (r(h,t) \Rightarrow \neg r(t,h)) \quad \forall h,t$
- Инверсия : $r_2(h,t) \Rightarrow r_1(t,h)$
- Транзитивность : $r_1(x,y) \wedge r_2(y,z) \Rightarrow r_3(x,z) \quad \forall x,y,z$
- 1 to N : $r(h,t_1), r(h,t_2), \ldots, r(h,t_n)$ are all True

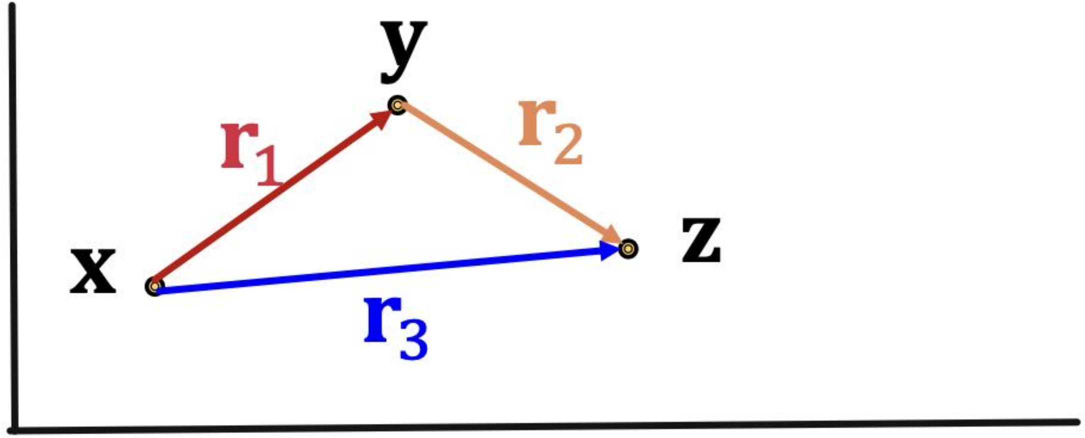
TransE: антисимметричность



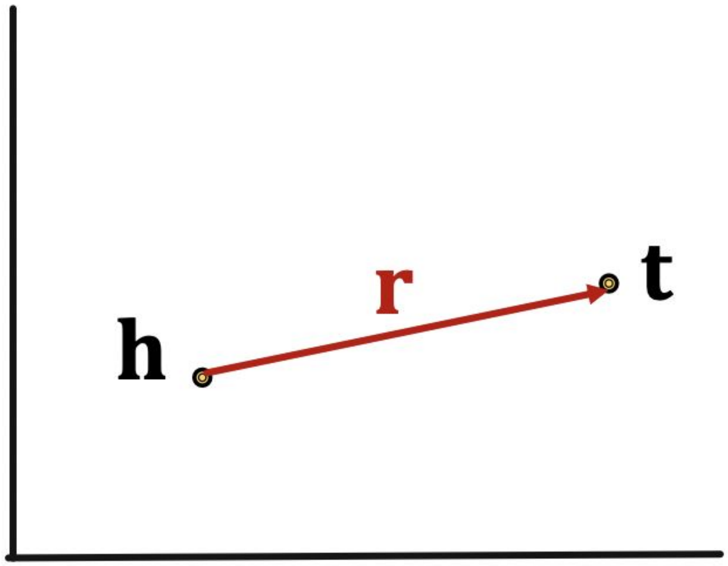
TransE: инверсивность



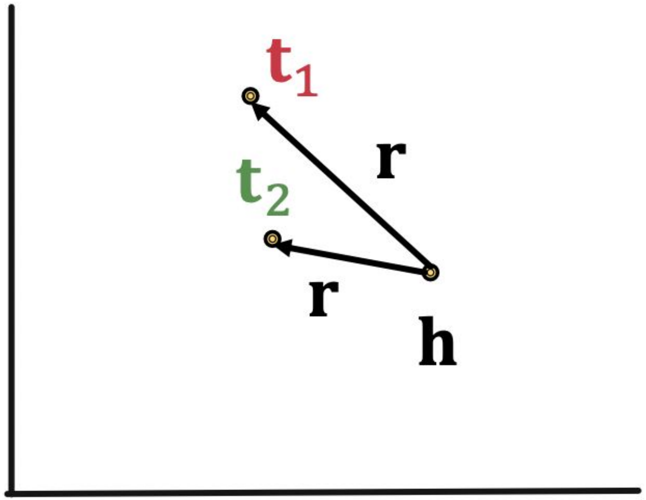
TransE: композитность



TransE: симметричность



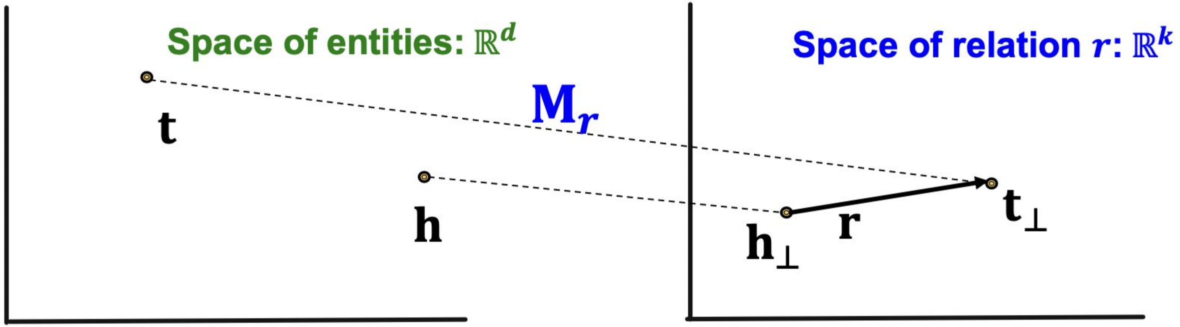
TransE: 1-to-N



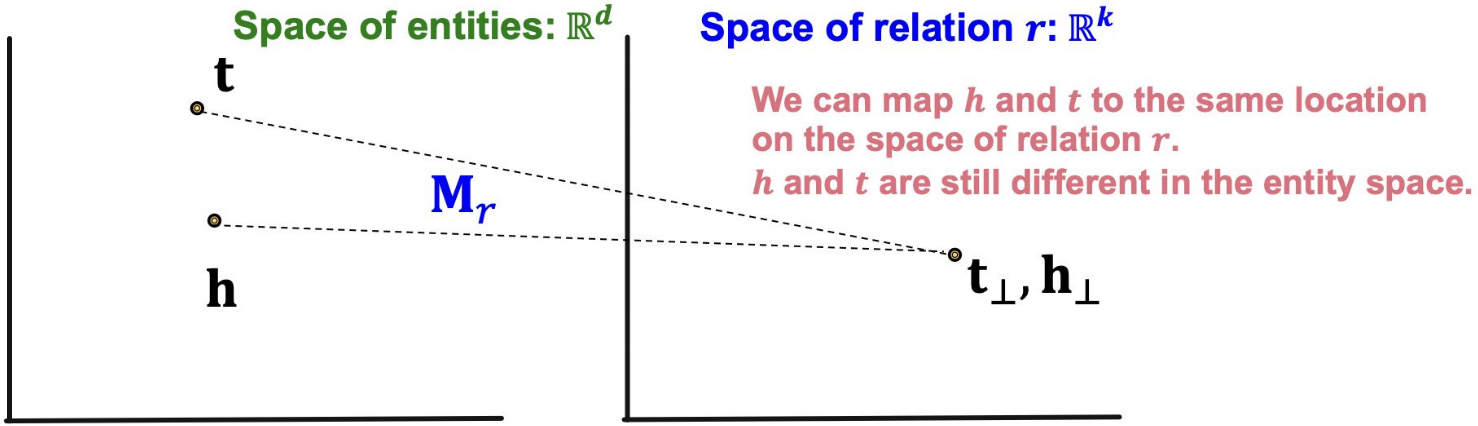
TransR

$$\mathbf{h}_{\perp} = \mathbf{M}_r\mathbf{h}, \ \mathbf{t}_{\perp} = \mathbf{M}_r\mathbf{t}$$

$$f_r(h, t) = -||\mathbf{h}_{\perp} + \mathbf{r} - \mathbf{t}_{\perp}||$$



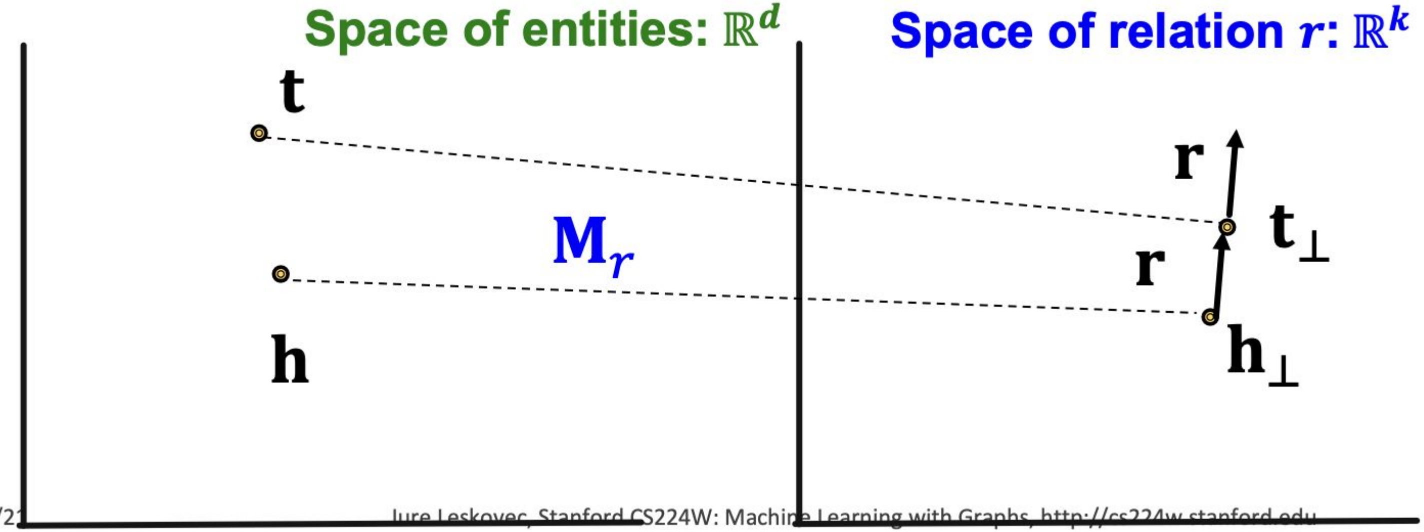
TransR: симметричность



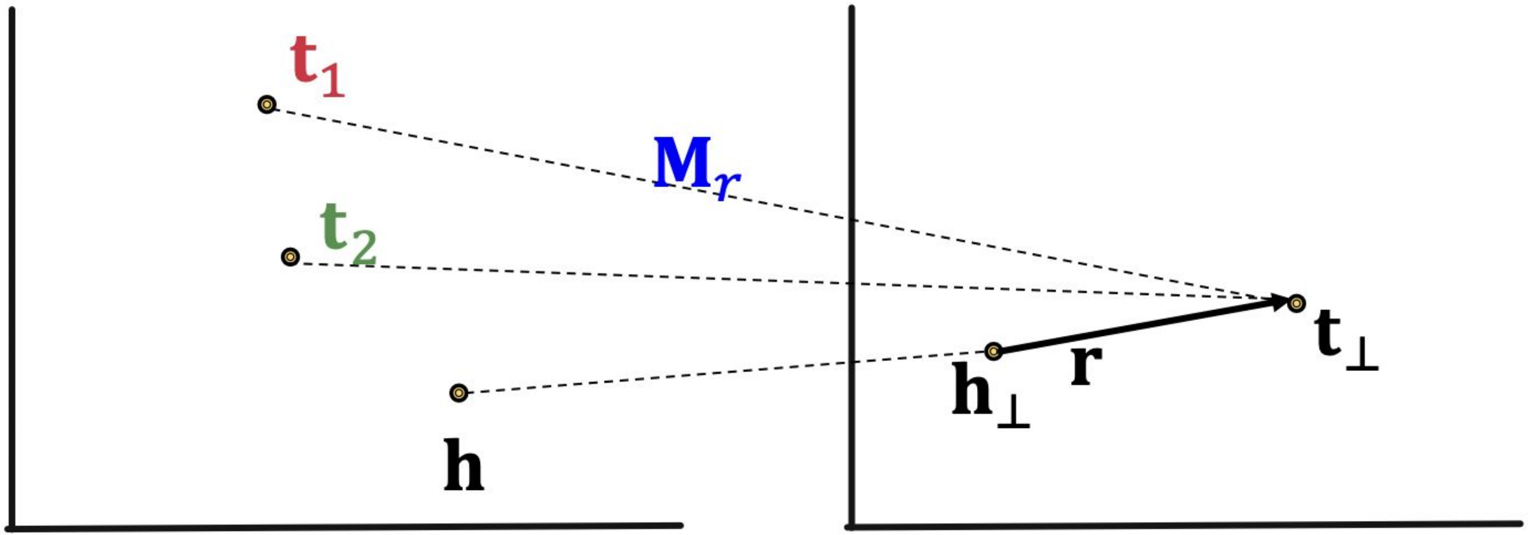
TransR: антисимметричность

$$\mathbf{r} \neq 0, \mathbf{M}_r\mathbf{h} + \mathbf{r} = \mathbf{M}_r\mathbf{t},$$

$$\text{Then } \mathbf{M}_r\mathbf{t} + \mathbf{r} \neq \mathbf{M}_r\mathbf{h} ✓$$

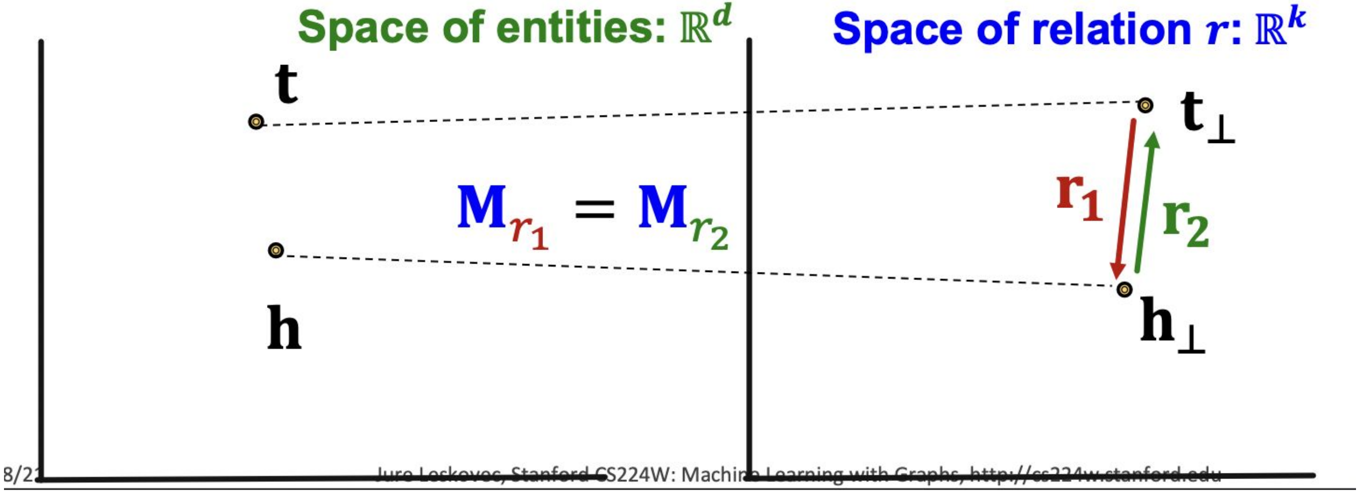


TransR: 1-to-N



$\mathbf{t}_1$
$\mathbf{M}_r$
$\mathbf{t}_2$
$\mathbf{h}$
$\mathbf{h}_\perp$
$\mathbf{r}$
$\mathbf{t}_\perp$

TransR: инверсивность



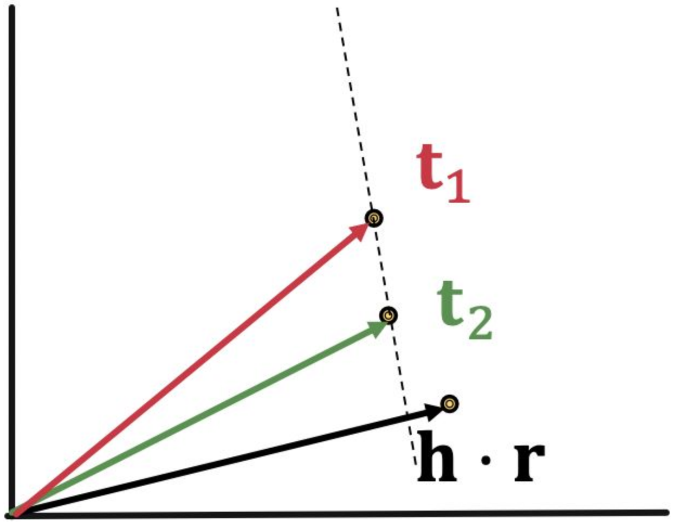
DistMult

$$f_r(h,t) = <\mathbf{h},\mathbf{r},\mathbf{t}> = \sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \mathbf{t}_i$$

$$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{R}^k$$

DistMult: 1-to-N

$$< \mathbf{h}, \mathbf{r}, \mathbf{t}_1 > = < \mathbf{h}, \mathbf{r}, \mathbf{t}_2 >$$



DistMult: симметричность

$$f_r(h,t) = < \mathbf{h}, \mathbf{r}, \mathbf{t} > = \sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \mathbf{t}_i =$$

$$< \mathbf{t}, \mathbf{r}, \mathbf{h} > = f_r(t,h)$$

DistMult: антисимметричность

$$f_r(h,t) = < \mathbf{h}, \mathbf{r}, \mathbf{t} > = < \mathbf{t}, \mathbf{r}, \mathbf{h} > = f_r(t,h)$$

DistMult: инверсивность

$$f_{r_2}(h, t) =< \mathbf{h}, \mathbf{r_2}, \mathbf{t} > = < \mathbf{t}, \mathbf{r_1}, \mathbf{h} >= f_{r_1}(t, h)$$

ComplEx

$$f_r(h, t) = \mathrm{Re}(\sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \bar{\mathbf{t}}_i)$$

ComplEx: симметричность

When $\mathrm{Im}(\mathbf{r}) = 0$, we have

- $f_r(h, t) = \mathrm{Re}(\sum_i \mathbf{h}_i \cdot \mathbf{r}_i \cdot \bar{\mathbf{t}}_i) = \sum_i \mathrm{Re}(\mathbf{r}_i \cdot \mathbf{h}_i \cdot \bar{\mathbf{t}}_i) = \sum_i \mathbf{r}_i \cdot \mathrm{Re}(\mathbf{h}_i \cdot \bar{\mathbf{t}}_i) = \sum_i \mathbf{r}_i \cdot \mathrm{Re}(\bar{\mathbf{h}}_i \cdot \mathbf{t}_i) = \sum_i \mathrm{Re}(\mathbf{r}_i \cdot \bar{\mathbf{h}}_i \cdot \mathbf{t}_i) = f_r(t, h)$