KNOWLEDGE GRAPH EMBEDDING

INCORPORATING ADDITIONAL INFORMATION

汇报人：黄婷

2017.11.03

目录

DIRECTORY

- KG embedding with facts alone
    1. Translational Distance Models
    2. Semantic Matching Models
- Incorporating additional information

Translational Distance Models

Method	Ent. embedding	Rel. embedding	Scoring function $f_r(h,t)$	Constraints/Regularization
TransE [14]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\Vert\mathbf{h}+\mathbf{r}-\mathbf{t}\Vert_{1/2}$	$\Vert\mathbf{h}\Vert_2=1, \Vert\mathbf{t}\Vert_2=1$
TransH [15]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^d$	$-\Vert(\mathbf{h}-\mathbf{w}_r^\top\mathbf{h}\mathbf{w}_r)+\mathbf{r}-(\mathbf{t}-\mathbf{w}_r^\top\mathbf{t}\mathbf{w}_r)\Vert_2^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1$ $\vert\mathbf{w}_r^\top\mathbf{r}\vert/\Vert\mathbf{r}\Vert_2\le\epsilon, \Vert\mathbf{w}_r\Vert_2=1$
TransR [16]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k, \mathbf{M}_r\in\mathbb{R}^{k\times d}$	$-\Vert\mathbf{M}_r\mathbf{h}+\mathbf{r}-\mathbf{M}_r\mathbf{t}\Vert_2^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert\mathbf{M}_r\mathbf{h}\Vert_2\le 1, \Vert\mathbf{M}_r\mathbf{t}\Vert_2\le 1$
TransD [50]	$\mathbf{h},\mathbf{w}_h\in\mathbb{R}^d$ $\mathbf{t},\mathbf{w}_t\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^k$	$-\Vert(\mathbf{w}_r\mathbf{w}_h^\top+\mathbf{I})\mathbf{h}+\mathbf{r}-(\mathbf{w}_r\mathbf{w}_t^\top+\mathbf{I})\mathbf{t}\Vert_2^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert(\mathbf{w}_r\mathbf{w}_h^\top+\mathbf{I})\mathbf{h}\Vert_2\le 1$ $\Vert(\mathbf{w}_r\mathbf{w}_t^\top+\mathbf{I})\mathbf{t}\Vert_2\le 1$
TranSparse [51]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k, \mathbf{M}_r(\theta_r)\in\mathbb{R}^{k\times d}$ $\mathbf{M}_r^1(\theta_r^1), \mathbf{M}_r^2(\theta_r^2)\in\mathbb{R}^{k\times d}$	$-\Vert\mathbf{M}_r(\theta_r)\mathbf{h}+\mathbf{r}-\mathbf{M}_r(\theta_r)\mathbf{t}\Vert_{1/2}^2$ $-\Vert\mathbf{M}_r^1(\theta_r^1)\mathbf{h}+\mathbf{r}-\mathbf{M}_r^2(\theta_r^2)\mathbf{t}\Vert_{1/2}^2$	$\Vert\mathbf{h}\Vert_2\le 1, \Vert\mathbf{t}\Vert_2\le 1, \Vert\mathbf{r}\Vert_2\le 1$ $\Vert\mathbf{M}_r(\theta_r)\mathbf{h}\Vert_2\le 1, \Vert\mathbf{M}_r(\theta_r)\mathbf{t}\Vert_2\le 1$ $\Vert\mathbf{M}_r^1(\theta_r^1)\mathbf{h}\Vert_2\le 1, \Vert\mathbf{M}_r^2(\theta_r^2)\mathbf{t}\Vert_2\le 1$
TransM [52]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\theta_r\Vert\mathbf{h}+\mathbf{r}-\mathbf{t}\Vert_{1/2}$	$\Vert\mathbf{h}\Vert_2=1, \Vert\mathbf{t}\Vert_2=1$

Semantic Matching Models

Method	Ent. embedding	Rel. embedding	Scoring function $f_r(h,t)$	Constraints/Regularization
RESCAL [13]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{h}^\top\mathbf{M}_r\mathbf{t}$	$\\|\mathbf{h}\\|_2\le 1, \\|\mathbf{t}\\|_2\le 1, \\|\mathbf{M}_r\\|_F\le 1$ $\mathbf{M}_r=\sum_i \pi_r^i \mathbf{u}_i\mathbf{v}_i^\top$ (required in [17])
TATEC [64]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d, \mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{h}^\top\mathbf{M}_r\mathbf{t}+\mathbf{h}^\top\mathbf{r}+\mathbf{t}^\top\mathbf{r}+\mathbf{h}^\top\mathbf{D}\mathbf{t}$	$\\|\mathbf{h}\\|_2\le 1, \\|\mathbf{t}\\|_2\le 1, \\|\mathbf{r}\\|_2\le 1$ $\\|\mathbf{M}_r\\|_F\le 1$
DistMult [65]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathbf{h}^\top\mathrm{diag}(\mathbf{r})\mathbf{t}$	$\\|\mathbf{h}\\|_2=1, \\|\mathbf{t}\\|_2=1, \\|\mathbf{r}\\|_2\le 1$
HolE [62]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathbf{r}^\top(\mathbf{h}\star\mathbf{t})$	$\\|\mathbf{h}\\|_2\le 1, \\|\mathbf{t}\\|_2\le 1, \\|\mathbf{r}\\|_2\le 1$
ComplEx [66]	$\mathbf{h},\mathbf{t}\in\mathbb{C}^d$	$\mathbf{r}\in\mathbb{C}^d$	$\mathrm{Re}(\mathbf{h}^\top\mathrm{diag}(\mathbf{r})\bar{\mathbf{t}})$	$\\|\mathbf{h}\\|_2\le 1, \\|\mathbf{t}\\|_2\le 1, \\|\mathbf{r}\\|_2\le 1$
ANALOGY [68]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{h}^\top\mathbf{M}_r\mathbf{t}$	$\\|\mathbf{h}\\|_2\le 1, \\|\mathbf{t}\\|_2\le 1, \\|\mathbf{M}_r\\|_F\le 1$ $\mathbf{M}_r\mathbf{M}_r^\top=\mathbf{M}_r^\top\mathbf{M}_r$ $\mathbf{M}_r\mathbf{M}_{r'}=\mathbf{M}_{r'}\mathbf{M}_r$
SME [18]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$(\mathbf{M}_u^1\mathbf{h}+\mathbf{M}_u^2\mathbf{r}+\mathbf{b}_u)^\top(\mathbf{M}_v^1\mathbf{t}+\mathbf{M}_v^2\mathbf{r}+\mathbf{b}_v)$ $((\mathbf{M}_u^1\mathbf{h})\circ(\mathbf{M}_u^2\mathbf{r})+\mathbf{b}_u)^\top((\mathbf{M}_v^1\mathbf{t})\circ(\mathbf{M}_v^2\mathbf{r})+\mathbf{b}_v)$	$\\|\mathbf{h}\\|_2=1, \\|\mathbf{t}\\|_2=1$
NTN [19]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{b}_r\in\mathbb{R}^k, \underline{\mathbf{M}}_r\in\mathbb{R}^{d\times d\times k}$ $\mathbf{M}_r^1,\mathbf{M}_r^2\in\mathbb{R}^{k\times d}$	$\mathbf{r}^\top\tanh(\mathbf{h}^\top\underline{\mathbf{M}}_r\mathbf{t}+\mathbf{M}_r^1\mathbf{h}+\mathbf{M}_r^2\mathbf{t}+\mathbf{b}_r)$	$\\|\mathbf{h}\\|_2\le 1, \\|\mathbf{t}\\|_2\le 1, \\|\mathbf{r}\\|_2\le 1$ $\\|\mathbf{b}_r\\|_2\le 1, \\|\underline{\mathbf{M}}_r^{[:,:,i]}\\|_F\le 1$ $\\|\mathbf{M}_r^1\\|_F\le 1, \\|\mathbf{M}_r^2\\|_F\le 1$

Model Comparison

- First, models which represent entities and relations as vectors are more efficient.
- Second, models which represent relations as matrices or tensors usually have higher complexity in both space and time.
- Finally, models based on neural network architectures generally have higher complexity in time.

01

Entity Types

Entity Types

- *AlfredHitchcock* has the type of *Person*
- *Psycho* the type of *CreativeWork*
- (*Psycho*, IsA, *CreativeWork*)

Entity Types ----- semantically smooth embedding （SSE）

Definition： Require entities of the same type to stay close to each other in the embedding space

- Laplacian eigenmaps : requires an entity to lie close to every other entity in the same category

$$\mathcal{R}_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\|\mathbf{e}_i - \mathbf{e}_j\|_2^2 w_{ij}^1,$$

- locally linear embedding: The latter represents an entity as a linear combination of its nearest neighbors, i.e., entities within the same category.

$$\mathcal{R}_2 = \sum_{i=1}^{n}\|\mathbf{e}_i - \sum_{e_j \in \mathbb{N}_{e_i}} w_{ij}^2 \mathbf{e}_j\|_2^2,$$

Entity Types ----- type-embodied knowledge representation learning (TKRL)

Definition: TKRL is a translational distance model with type-specific entity projections. Given a fact (h, r, t), it first projects h and t with type-specific projection matrices, and then models r as a translation between the two projected entities.

Scoring Function: $f_r(h,t) = -\|\mathbf{M}_{rh}\mathbf{h} + \mathbf{r} - \mathbf{M}_{rt}\mathbf{t}\|_1,$

$$\mathbf{M}_{rh} = \frac{\sum_{i=1}^{n_h} \alpha_i \mathbf{M}_{c_i}}{\sum_{i=1}^{n_h} \alpha_i}, \quad \alpha_i = \begin{cases} 1, & c_i \in \mathbb{C}_{r_h}, \\ 0, & c_i \notin \mathbb{C}_{r_h}, \end{cases}$$

$$\begin{aligned} \text{addition:} \quad & \mathbf{M}_{c_i} = \beta_1 \mathbf{M}_{c_i^{(1)}} + \cdots + \beta_\ell \mathbf{M}_{c_i^{(\ell)}}; \\ \text{multiplication:} \quad & \mathbf{M}_{c_i} = \mathbf{M}_{c_i^{(1)}} \circ \cdots \circ \mathbf{M}_{c_i^{(\ell)}}. \end{aligned}$$

02

Relation Paths

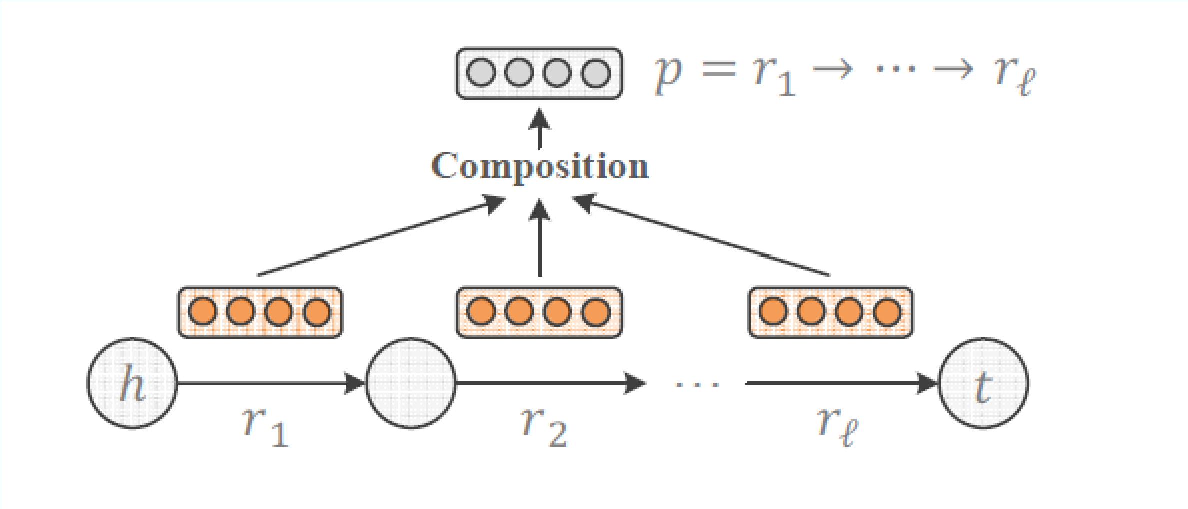
Relation Paths

Definition: A relation path is typically defined as a sequence of relations $r1 \rightarrow \cdots \rightarrow r\ell$ through which two entities can be connected on the graph.

e.g.: *AlfredHitchcock* → *BornIn* → （*Leytonstone*） → *LocatedIn* → *England*

Relation Paths ----- path-based TransE (PTransE)

Definition: Given a path $p = r1 \longrightarrow \cdots \longrightarrow r\ell$ linking two entities h and t, as well as the vector representations $r1, \cdots, r\ell$ of the constituent relations



$$\begin{aligned} \text{addition:} \quad & \mathbf{p} = \mathbf{r}_1 + \cdots + \mathbf{r}_\ell; \\ \text{multiplication:} \quad & \mathbf{p} = \mathbf{r}_1 \circ \cdots \circ \mathbf{r}_\ell; \\ \text{RNN:} \quad & \mathbf{c}_i = f\left(\mathbf{W}[\mathbf{c}_{i-1}; \mathbf{r}_i]\right). \end{aligned}$$

Relation Paths ----- path-based TransE (PTransE)

Definition: The path *p* is then required to be consistent with a direct relation *r* between the two entities, i.e., ||p− r|| tends to be small if(*h*, *r*, *t*) holds.

$$\mathcal{L}_{\text{path}} = \frac{1}{Z} \sum_{p \in \mathbb{P}(h,t)} R(p|h,t) \cdot \ell(p,r),$$

$$\ell(p,r) = \sum_{r'} \max(0, \gamma + \|\mathbf{p} - \mathbf{r}\|_1 - \|\mathbf{p} - \mathbf{r}'\|_1),$$

Relation Paths ----- extensions of both the TransE model and the RESCAL model

The score of (*h*, *p*, *t*) :

$$f_p(h,t) = -\|\mathbf{h} + (\mathbf{r}_1 + \cdots + \mathbf{r}_\ell) - \mathbf{t}\|_1,$$

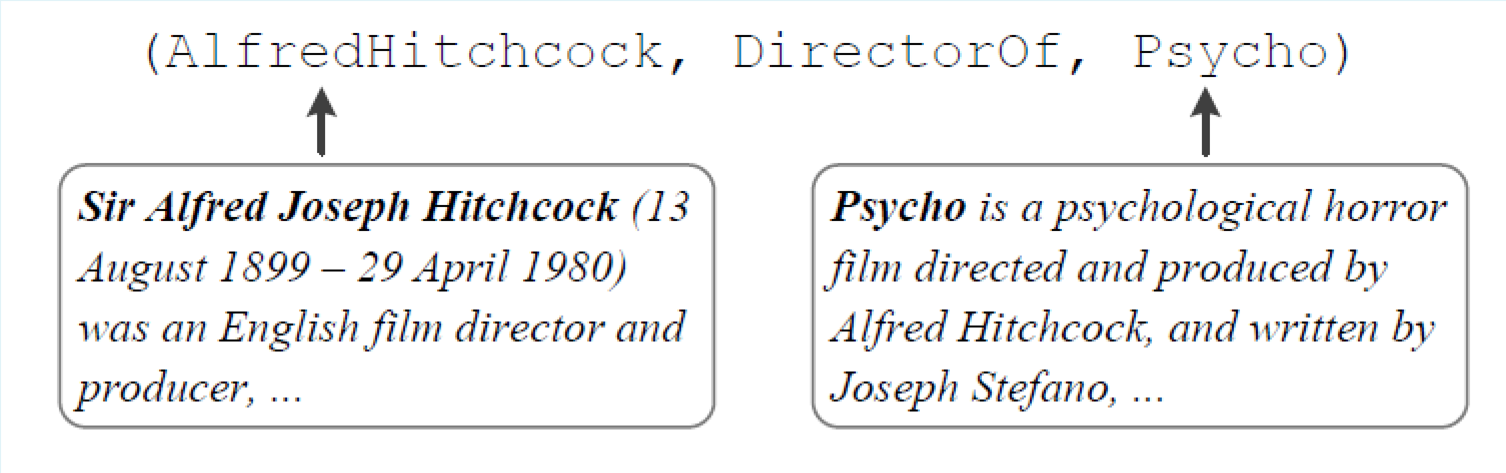
$$f_p(h,t) = \mathbf{h}^\top(\mathbf{M}_1 \circ \cdots \circ \mathbf{M}_\ell)\mathbf{t}.$$

03

Textual Descriptions

Textual Descriptions

Definition: Actually, in most KGs there are concise descriptions for entities which contain rich semantic information about them.



Textual Descriptions ----- joint model

Definition: The key idea is to align the given KG with an auxiliary text corpus, and then jointly conduct KG embedding and word embedding.

knowledge model, text model, and alignment model

$$\mathcal{L} = \mathcal{L}_K + \mathcal{L}_T + \mathcal{L}_A.$$

Textual Descriptions ----- joint model

- **Knowledge Model:**

score function: $z(\mathbf{h},\mathbf{r},\mathbf{t}) = b - \frac{1}{2}\|\mathbf{h}+\mathbf{r}-\mathbf{t}\|^2$

conditional probability: $\Pr(h|r,t) = \frac{\exp\{z(\mathbf{h},\mathbf{r},\mathbf{t})\}}{\sum_{\tilde{h}\in\mathcal{I}}\exp\{z(\tilde{\mathbf{h}},\mathbf{r},\mathbf{t})\}}$

conditional likelihoods:

$$\mathcal{L}_f(h,r,t) = \log\Pr(h|r,t) + \log\Pr(t|h,r) + \log\Pr(r|h,t) \quad (2)$$

Textual Descriptions ----- joint model

- **Text Model:**

Relational Concurrence Assumption. *If two words w and v concur in some context, e.g., a window of text, then there is a relation r_{wv} between the two words. That is, we can state the triplet of (w, r_{wv}, v) is a fact.*

$$z(\mathbf{w}, \mathbf{r}_{wv}, \mathbf{v}) \triangleq z(\mathbf{w}', \mathbf{v}) = b - \frac{1}{2}\|\mathbf{w}' - \mathbf{v}\|^2 \quad (4)$$

and

$$\Pr(w|r_{wv}, v) \triangleq \Pr(w|v) = \frac{\exp\{z(\mathbf{w}', \mathbf{v})\}}{\sum_{\tilde{w}\in\mathcal{V}} \exp\{z(\tilde{\mathbf{w}}', \mathbf{v})\}} \quad (5)$$

$$\mathcal{L}_T = \sum_{(w,v)\in\mathcal{C}} n_{wv} \log \Pr(w|v).$$

Textual Descriptions ----- joint model

- **Alignment Model:** align the two entity embedding space and word embedding space into the same one
- for most Wikipedia (English) pages, there is an unique corresponding entity e_v in Freebase

$$(w, v) \rightarrow (w, e_v)$$

$$\mathcal{L}_{AA} = \sum_{(w,v)\in\mathcal{C}, v\in\mathcal{A}} \log \Pr(w|e_v)$$

Textual Descriptions ----- description-embodied knowledge representation learning (DKRL)

Definition： The aim is to extend TransE so as to further handle entity descriptions.

$$f_r(h,t) = -\|\mathbf{h}_s + \mathbf{r} - \mathbf{t}_s\|_1 - \|\mathbf{h}_d + \mathbf{r} - \mathbf{t}_d\|_1 - \|\mathbf{h}_s + \mathbf{r} - \mathbf{t}_d\|_1 - \|\mathbf{h}_d + \mathbf{r} - \mathbf{t}_s\|_1,$$

Textual Descriptions ----- text-enhanced KG embedding model (TEKE)

Definition: Given a KG and a text corpus, TEKE first annotates entities in the corpus and constructs a co-occurrence network composed of entities and words. Then, for each entity *e*, TEKE defines its textual context *n(e)* as its neighbors in the co-occurrence network, i.e., words co-occurring frequently with the entity in the text corpus.

$$\hat{\mathbf{h}} = \mathbf{A}\mathbf{n}(h) + \mathbf{h},$$
$$\hat{\mathbf{t}} = \mathbf{A}\mathbf{n}(t) + \mathbf{t},$$
$$\hat{\mathbf{r}} = \mathbf{B}\mathbf{n}(h, t) + \mathbf{r}.$$

04

Logical Rules

Logical Rules

Definition： e.g., HasWife(x, y) → HasSpouse(x, y) stating that any two entities linked by the relation HasWife should also be linked by the relation HasSpouse.

Logical Rules ----- A joint model which embeds KG facts and logical rules simultaneously

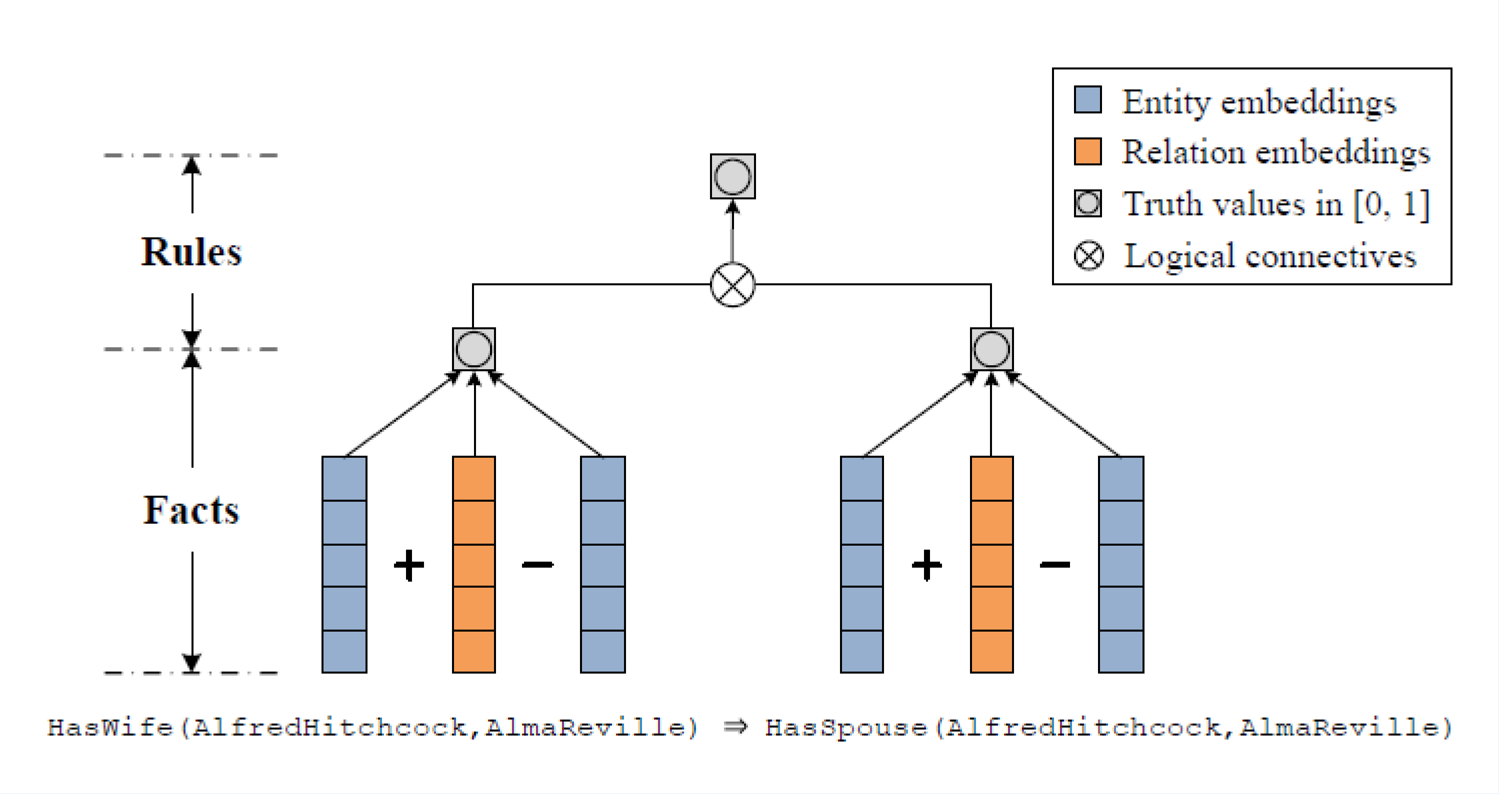
- A fact *(h, r, t)* is taken as a ground atom, with its truth value defined as：

$$I(h, r, t) = 1 - \frac{1}{3\sqrt{d}}\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_1,$$

- Logical rules are first instantiated into ground rules, e.g., the universally quantified rule ∀x, y : HasWife(x, y) ⇒ HasSpouse(x, y) can be grounded into HasWife(AlfredHitchcock, AlmaReville) ⇒ HasSpouse(AlfredHitchcock, AlmaReville).

Logical Rules ----- A joint model which embeds KG facts and logical rules simultaneously

- Ground rules are then interpreted as complex formula constructed by combining ground atoms:



谢谢聆听