

第八章 知识图谱

戴洪良

计算机科学与技术学院/人工智能学院

hongldai@nuaa.edu.cn

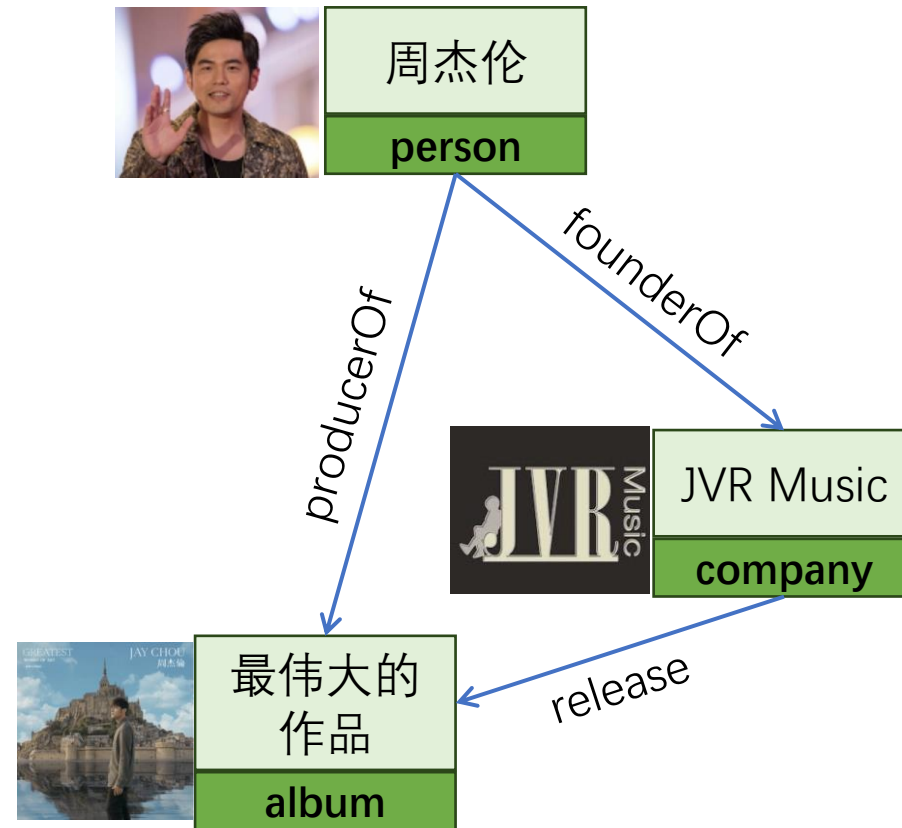


知识图谱嵌入

- 知识图谱嵌入 (Knowledge Graph Embedding, KGE) 也被称为知识表示学习 (Knowledge Representation Learning, KRL)
- 目标：将知识图谱中的实体和关系表示为稠密低维的实值向量
 - 相比传统的one-hot encode表示，可捕捉丰富的语义信息，方便应用于各类神经网络模型

知识图谱嵌入设定

- 知识图谱包含实体和关系
 - 节点代表实体
 - 边代表关系
- 知识图谱嵌入一般将知识图谱视为三元组集合：
 - (头实体, 关系, 尾实体)
 - (head, relation, tail)



知识图谱嵌入的应用

- Link Prediction (链接预测)
 - 预测未知的正确关系, 可用于对知识图谱补全 (Knowledge Graph Completion)
 - 常被用于评估知识图谱嵌入方法的效果
- 问答系统
 - 为知识库问答 (Knowledge Base Question Answering, KBQA) 提供支持
- 推荐系统
 - 利用知识图谱中的知识提高推荐质量
 - 如: 用有关电影的知识更好地为用户推荐电影
- 知识图谱对齐
 - 利用向量表示对齐不同知识图谱中的实体

知识图谱嵌入方法

- 基于Translation的方法
 - TransE、TransH、TransR、TransD等
- 其他方法
 - 基于bilinear function、神经网络模型、利用其他信息等

知识图谱嵌入方法

- 基于Translation的方法
 - TransE、TransH、TransR、TransD等
- 其他方法
 - 基于bilinear function、神经网络模型、利用其他信息等

知识图谱嵌入

- TransE (Bordes et al., 2013)模型

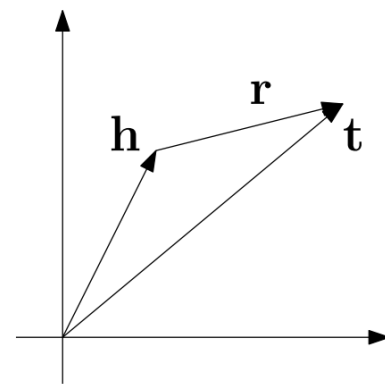
- 对每条知识 (head, relation, tail) 中的relation看作从head到tail的平移操作。
- h, t, r 分别为head实体、tail实体和关系relation的向量表示，希望有：

$$\mathbf{h} + \mathbf{r} \approx \mathbf{t}$$

- 打分函数（对三元组的合理性（plausibility）打分）：

$$f(h, r, t) = -d(\mathbf{h} + \mathbf{r}, \mathbf{t})$$

其中 $d(\mathbf{x}_1, \mathbf{x}_2)$ 表示向量 \mathbf{x}_1 和 \mathbf{x}_2 的距离



TransE

• 模型训练

Algorithm 1 Learning TransE

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

```
1: initialize  $\ell \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$  for each  $\ell \in L$ 
2:            $\ell \leftarrow \ell / \|\ell\|$  for each  $\ell \in L$ 
3:            $\mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$  for each entity  $e \in E$ 
4: loop
5:    $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$  for each entity  $e \in E$ 
6:    $S_{batch} \leftarrow \text{sample}(S, b)$  // sample a minibatch of size  $b$ 
7:    $T_{batch} \leftarrow \emptyset$  // initialize the set of pairs of triplets
8:   for  $(h, \ell, t) \in S_{batch}$  do
9:      $(h', \ell, t') \leftarrow \text{sample}(S'_{(h, \ell, t)})$  // sample a corrupted triplet
10:     $T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$ 
11:   end for
12:   Update embeddings w.r.t.  $\sum_{((h, \ell, t), (h', \ell, t')) \in T_{batch}} \nabla [\gamma + d(\mathbf{h} + \ell, \mathbf{t}) - d(\mathbf{h}' + \ell, \mathbf{t}')]_+$ 
13: end loop
```

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

注：这里用 ℓ 表示关系

其中 $S'_{(h, \ell, t)} = \{(h', \ell, t) | h' \in E\} \cup \{(h, \ell, t') | t' \in E\}$

TransE

- 训练后的理想效果：

- 打分函数能做出高质量预测

如果 (h, r, t) 是正确三元组， (h, r, t') 是错误三元组，则

$$f(h, r, t) > f(h, r, t')$$

- 且为实体和关系学习到的向量能捕捉它们的语义信息

- 如：法国+首都=巴黎；相似实体的向量表示是相似的

使用Link Prediction评估嵌入效果

- Link Prediction（链接预测）常被用于知识图谱嵌入的效果评估
- 具体操作为，对测试集中的真实三元组 (h, r, t)
 - 使用打分函数根据尾实体 t 和关系 r 预测头实体 h ，或
 - 使用打分函数根据头实体 h 和关系 r 预测尾实体 t

如根据 t 和 r 预测 h ，则可对知识图谱中每个实体 e 用打分函数 f 计算分数: $f(e, r, t)$

之后对所有实体对应分数排序，可得到正确实体的打分排名，即rank

评测指标**Mean Rank**: 测试样例中正确实体的rank的平均

评测指标**Hits@10**: 测试样例中正确实体排前10的比率

TransE

• TransE的实验效果

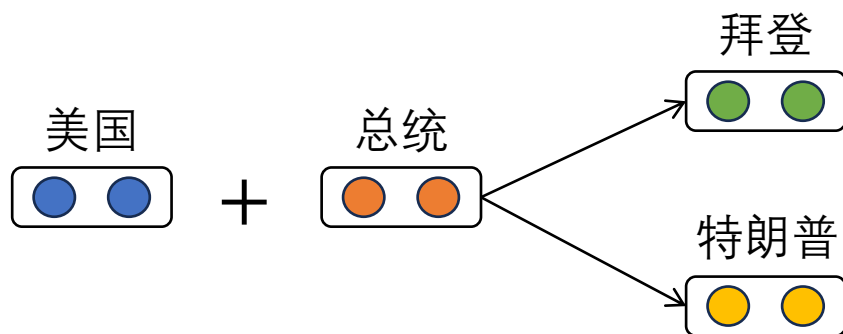
DATASET	WN				FB15K				FB1M	
METRIC	MEAN RANK		HITS@10 (%)		MEAN RANK		HITS@10 (%)		MEAN RANK	HITS@10 (%)
<i>Eval. setting</i>	<i>Raw</i>	<i>Filt.</i>	<i>Raw</i>	<i>Filt.</i>	<i>Raw</i>	<i>Filt.</i>	<i>Raw</i>	<i>Filt.</i>	<i>Raw</i>	<i>Raw</i>
Unstructured [2]	315	304	35.3	38.2	1,074	979	4.5	6.3	15,139	2.9
RESCAL [11]	1,180	1,163	37.2	52.8	828	683	28.4	44.1	-	-
SE [3]	1,011	985	68.5	80.5	273	162	28.8	39.8	22,044	17.5
SME(LINEAR) [2]	545	533	65.1	74.1	274	154	30.7	40.8	-	-
SME(BILINEAR) [2]	526	509	54.7	61.3	284	158	31.3	41.3	-	-
LFM [6]	469	456	71.4	81.6	283	164	26.0	33.1	-	-
TransE	263	251	75.4	89.2	243	125	34.9	47.1	14,615	34.0

WN: 基于WordNet构建的数据集，以synsets为实体，共约4万实体

FB15K和FB1M: 基于Freebase构建的数据集

TransE

- TransE存在的问题：不能很好地处理复杂关系
- 如：1-N, N-1, N-N关系



TransE会使拜登和特朗普的向量变得相同



TransH

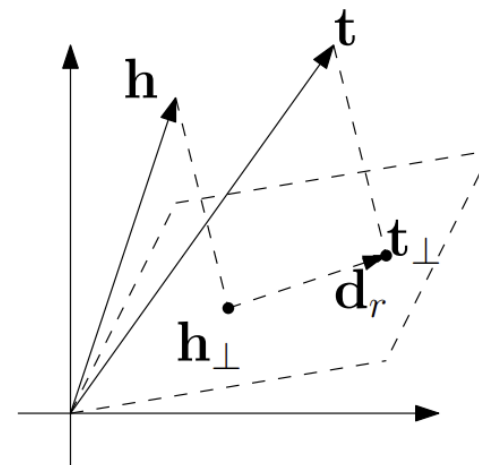
- TransH将每种关系建模为一个超平面，打分时将三元组中的头实体和尾实体分别映射到该超平面中

设超平面参数向量（法向量）为 \mathbf{w}_r ，且 $\|\mathbf{w}_r\|_2 = 1$ ，则

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

- 打分函数：

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



TransH

- TransH的效果

Dataset	WN18				FB15k			
Metric	MEAN		HITS@10		MEAN		HITS@10	
	Raw	Filt.	Raw	Filt.	Raw	Filt.	Raw	Filt.
Unstructured (Bordes et al. 2012)	315	304	35.3	38.2	1,074	979	4.5	6.3
RESCAL (Nickel, Tresp, and Kriegel 2011)	1,180	1,163	37.2	52.8	828	683	28.4	44.1
SE (Bordes et al. 2011)	1,011	985	68.5	80.5	273	162	28.8	39.8
SME (Linear) (Bordes et al. 2012)	545	533	65.1	74.1	274	154	30.7	40.8
SME (Bilinear) (Bordes et al. 2012)	526	509	54.7	61.3	284	158	31.3	41.3
LFM (Jenatton et al. 2012)	469	456	71.4	81.6	283	164	26.0	33.1
TransE (Bordes et al. 2013b)	263	251	75.4	89.2	243	125	34.9	47.1
TransH (unif.)	318	303	75.4	86.7	211	84	42.5	58.5
TransH (bern.)	400.8	388	73.0	82.3	212	87	45.7	64.4

TransR

- TransE和TransH都将实体和关系嵌入到同一个向量空间中
- TransR认为实体和关系是完全不同的对象，不应该对它们用同一个向量空间
- TransR将实体和关系分别用不同的空间建模，即实体空间（entity space）和关系空间（relation spaces）

TransR

- TransR将实体和关系分别用不同的空间建模，即实体空间（entity space）和关系空间（relation spaces）

对每个实体依然用一个向量表示： $\mathbf{h} \in R^k, \mathbf{t} \in R^k$

为每个关系 r 用一个向量表示： $\mathbf{r} \in R^d$

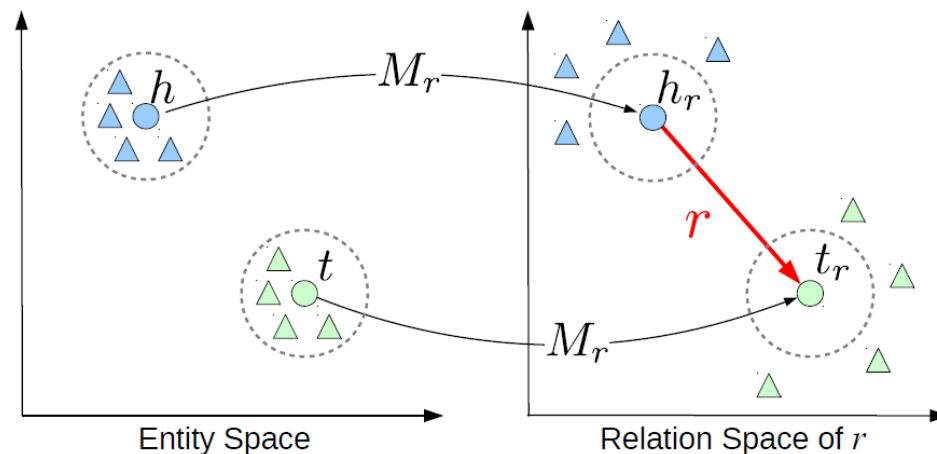
为每个关系 r 设置一个映射矩阵 $M_r \in R^{k \times d}$

将实体向量映射到关系空间：

$$\mathbf{h}_r = \mathbf{h}M_r, \quad \mathbf{t}_r = \mathbf{t}M_r$$

打分函数：

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



TransR

- TransR的效果

Data Sets	WN18				FB15K			
Metric	Mean Rank		Hits@10 (%)		Mean Rank		Hits@10 (%)	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
Unstructured (Bordes et al. 2012)	315	304	35.3	38.2	1,074	979	4.5	6.3
RESCAL (Nickel, Tresp, and Kriegel 2011)	1,180	1,163	37.2	52.8	828	683	28.4	44.1
SE (Bordes et al. 2011)	1,011	985	68.5	80.5	273	162	28.8	39.8
SME (linear) (Bordes et al. 2012)	545	533	65.1	74.1	274	154	30.7	40.8
SME (bilinear) (Bordes et al. 2012)	526	509	54.7	61.3	284	158	31.3	41.3
LFM (Jenatton et al. 2012)	469	456	71.4	81.6	283	164	26.0	33.1
TransE (Bordes et al. 2013)	263	251	75.4	89.2	243	125	34.9	47.1
TransH (unif) (Wang et al. 2014)	318	303	75.4	86.7	211	84	42.5	58.5
TransH (bern) (Wang et al. 2014)	401	388	73.0	82.3	212	87	45.7	64.4
TransR (unif)	232	219	78.3	91.7	226	78	43.8	65.5
TransR (bern)	238	225	79.8	92.0	198	77	48.2	68.7
CTransR (unif)	243	230	78.9	92.3	233	82	44	66.3
CTransR (bern)	231	218	79.4	92.3	199	75	48.4	70.2

TransD

- 在TransH和TransR中，不同的种类的实体共享相同的映射向量或矩阵， 但一个关系的头尾实体的种类和属性可能有较大差别
 - 如 (Biden, nationality, US)
- TransR引入了空间映射， 但是模型参数急剧增加
- TransD将每个对象（实体、关系）嵌入为两个向量： 语义向量、映射向量

TransD

- TransD将每个对象（实体、关系）嵌入为两个向量：语义向量、映射向量

给定三元组 (h, r, t) ，相关的向量表示为： $\mathbf{h}, \mathbf{h}_p, \mathbf{r}, \mathbf{r}_p, \mathbf{t}, \mathbf{t}_p$

其中 $\mathbf{h}_p, \mathbf{r}_p, \mathbf{t}_p$ 为映射向量

基于它们，可得到映射矩阵：

$$\mathbf{M}_{rh} = \mathbf{r}_p \mathbf{h}_p^\top + \mathbf{I}^{m \times n}$$

$$\mathbf{M}_{rt} = \mathbf{r}_p \mathbf{t}_p^\top + \mathbf{I}^{m \times n}$$

打分函数：

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

$$\mathbf{h}_\perp = \mathbf{M}_{rh} \mathbf{h}, \quad \mathbf{t}_\perp = \mathbf{M}_{rt} \mathbf{t}$$

TransD

- TransD的效果

Data sets	WN18				FB15K			
Metric	Mean Rank		Hits@10		Mean Rank		Hits@10	
	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt
Unstructured (Bordes et al. 2012)	315	304	35.3	38.2	1,074	979	4.5	6.3
RESCAL (Nickle, Tresp, and Kriegel 2011)	1,180	1,163	37.2	52.8	828	683	28.4	44.1
SE (Bordes et al. 2011)	1,011	985	68.5	80.5	273	162	28.8	39.8
SME (linear) (Bordes et al.2012)	545	533	65.1	74.1	274	154	30.7	40.8
SME (Bilinear) (Bordes et al. 2012)	526	509	54.7	61.3	284	158	31.3	41.3
LFM (Jenatton et al. 2012)	469	456	71.4	81.6	283	164	26.0	33.1
TransE (Bordes et al. 2013)	263	251	75.4	89.2	243	125	34.9	47.1
TransH (unif) (Wang et al. 2014)	318	303	75.4	86.7	211	84	42.5	58.5
TransH (bern) (Wang et al. 2014)	401	388	73.0	82.3	212	87	45.7	64.4
TransR (unif) (Lin et al. 2015)	232	219	78.3	91.7	226	78	43.8	65.5
TransR (bern) (Lin et al. 2015)	238	225	79.8	92.0	198	77	48.2	68.7
CTransR (unif) (Lin et al. 2015)	243	230	78.9	92.3	233	82	44.0	66.3
CTransR (bern) (Lin et al. 2015)	231	218	79.4	92.3	199	75	48.4	70.2
TransD (unif)	242	229	79.2	92.5	211	67	49.4	74.2
TransD (bern)	224	212	79.6	92.2	194	91	53.4	77.3

其他

- TranSparse
- TransM
- ManiFoldE
- TransF
- TransA
- ...

知识图谱嵌入方法

- 基于Translation的方法
 - TransE、TransH、TransR、TransD等
- 其他方法
 - 基于bilinear function、神经网络模型、利用其他信息等

MLP

- 将头实体、关系、尾实体的向量表示拼接后使用全连接的神经网络模型，打分函数：

$$f(h, r, t) = \mathbf{w}^T \tanh(M[\mathbf{h}, \mathbf{r}, \mathbf{t}])$$

- (Xin et al., 2014)发现使用该种方法学到的实体和向量的表示也捕捉到了语义特征
 - 如, *children* 关系的向量与 *parent*, *spouse*, *birth-place* 相近

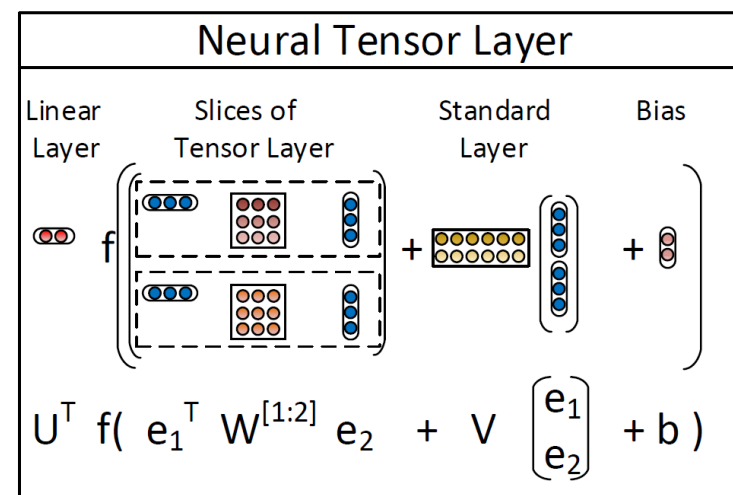
NTN (Neural Tensor Networks)

- NTN用张量神经网络 (Neural Tensor Networks) 对头尾实体间存在某种关系的合理性 (plausibility) 建模, 其使用的打分函数为:

$$g(e_1, R, e_2) = u_R^T f \left(e_1^T W_R^{[1:k]} e_2 + V_R \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + b_R \right)$$

其中 $W_R^{[1:k]}$ 为张量, $e_1^T W_R^{[1:k]} e_2$ 结果为 k 维向量

f 为非线性函数



DistMult

- 使用一个简单的bilinear function作为打分函数:

$$f(h, r, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t}$$

其中 M_r 为关系 r 对应的参数矩阵, 且为对角矩阵。

将 M_r 设置为对角矩阵减少了参数数量, 加快了计算速度。

DistMult

- 使用一个简单的bilinear function作为打分函数:

$$f(h, r, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t}$$

打分函数有什么不足之处?

- (Yang et al., 2015)发现, 使用对角矩阵比完整矩阵效果反而还可能变得更好:



	FB15k		FB15k-401		WN	
	MRR	HITS@10	MRR	HITS@10	MRR	HITS@10
NTN	0.25	41.4	0.24	40.5	0.53	66.1
Bilinear+Linear	0.30	49.0	0.30	49.4	0.87	91.6
TransE (DISTADD)	0.32	53.9	0.32	54.7	0.38	90.9
Bilinear	0.31	51.9	0.32	52.2	0.89	92.8
Bilinear-diag (DISTMULT)	0.35	57.7	0.36	58.5	0.83	94.2

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

ComplEx

- DistMult不能把非对称的关系建模好

$$f(h, r, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t}$$

- ComplEx将向量表示扩展到复数空间，对这一问题进行了处理

Complex Embeddings for Simple Link Prediction

Model	WN18					FB15K				
	MRR		Hits at			MRR		Hits at		
	Filter	Raw	1	3	10	Filter	Raw	1	3	10
CP	0.075	0.058	0.049	0.080	0.125	0.326	0.152	0.219	0.376	0.532
TransE	0.454	0.335	0.089	0.823	0.934	0.380	0.221	0.231	0.472	0.641
DistMult	0.822	0.532	0.728	0.914	0.936	0.654	0.242	0.546	0.733	0.824
HolE*	0.938	0.616	0.93	0.945	0.949	0.524	0.232	0.402	0.613	0.739
ComplEx	0.941	0.587	0.936	0.945	0.947	0.692	0.242	0.599	0.759	0.840

融合文本描述的方法

- DKRL
- 每个实体有两个向量表示
 - 基于结构的向量表示 \mathbf{e}_s
 - 基于文本描述的向量表示 \mathbf{e}_d

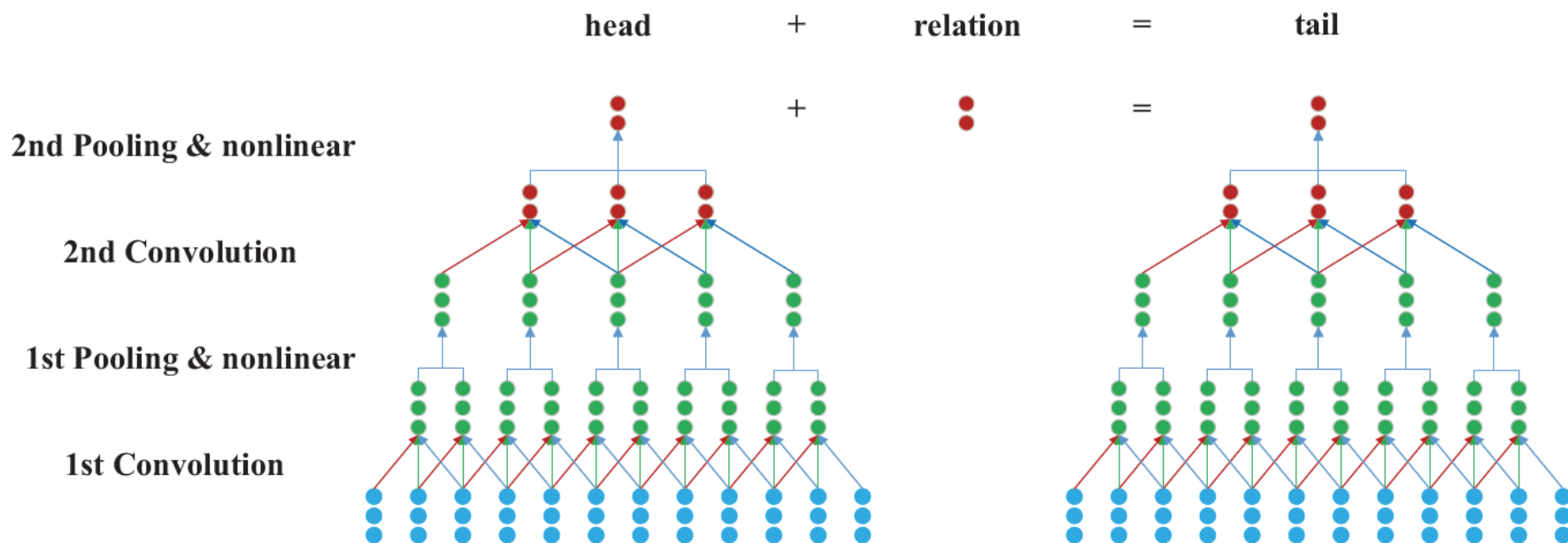
打分函数：

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

目的：让两种向量都能表现实体间关系，并让它们相互促进

融合文本描述的方法

- DKRL采用了CNN获取文本描述的向量表示



融合文本描述的方法

- DKRL实验效果

正常设定下并未比TransE好很多

Table 2: Evaluation results on entity prediction

Metric	Mean Rank		Hits@ 10(%)	
	Raw	Filter	Raw	Filter
TransE	210	119	48.5	66.1
DKRL(CNN)	200	113	44.3	57.6
DKRL(CNN)+TransE	181	91	49.6	67.4

Table 3: Evaluation results on relation prediction

Metric	Mean Rank		Hits@ 1(%)	
	Raw	Filter	Raw	Filter
TransE	2.91	2.53	69.5	90.2
DKRL(CNN)	2.91	2.55	69.8	89.0
DKRL(CNN)+TransE	2.41	2.03	69.8	90.8

但该方法可处理所涉及实体还未在知识图谱内时情况

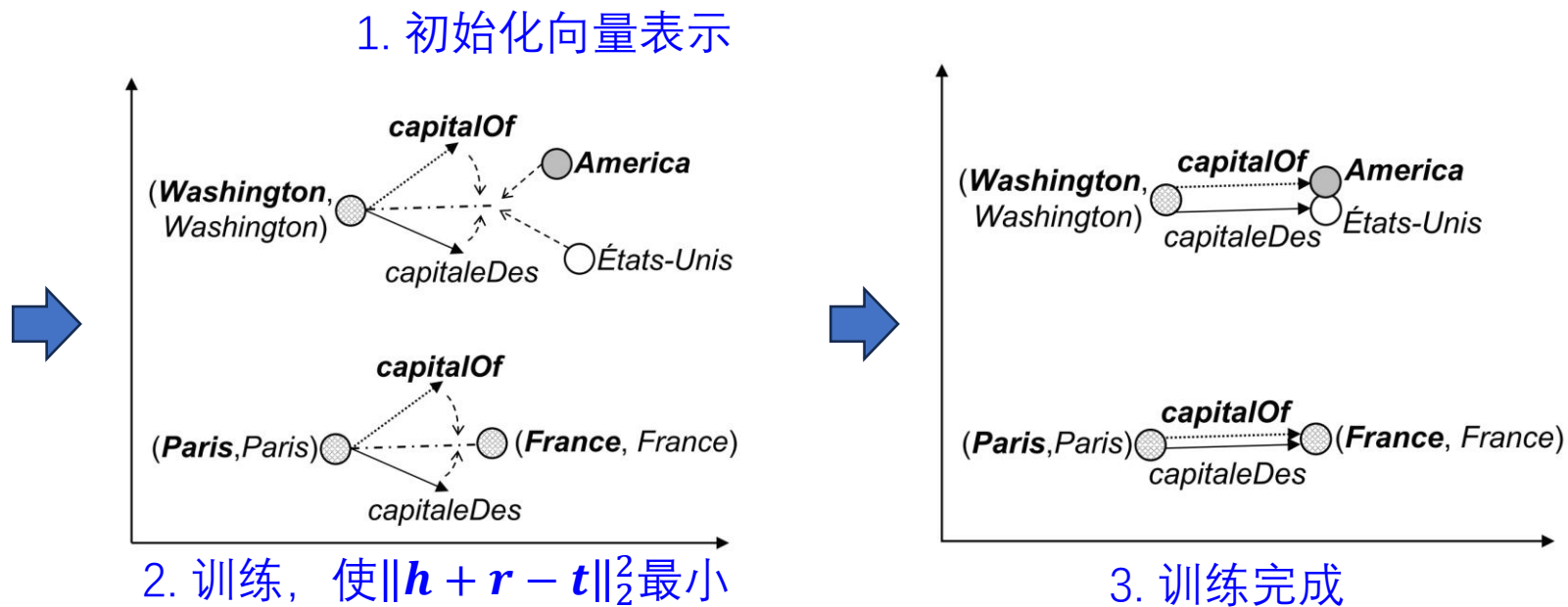
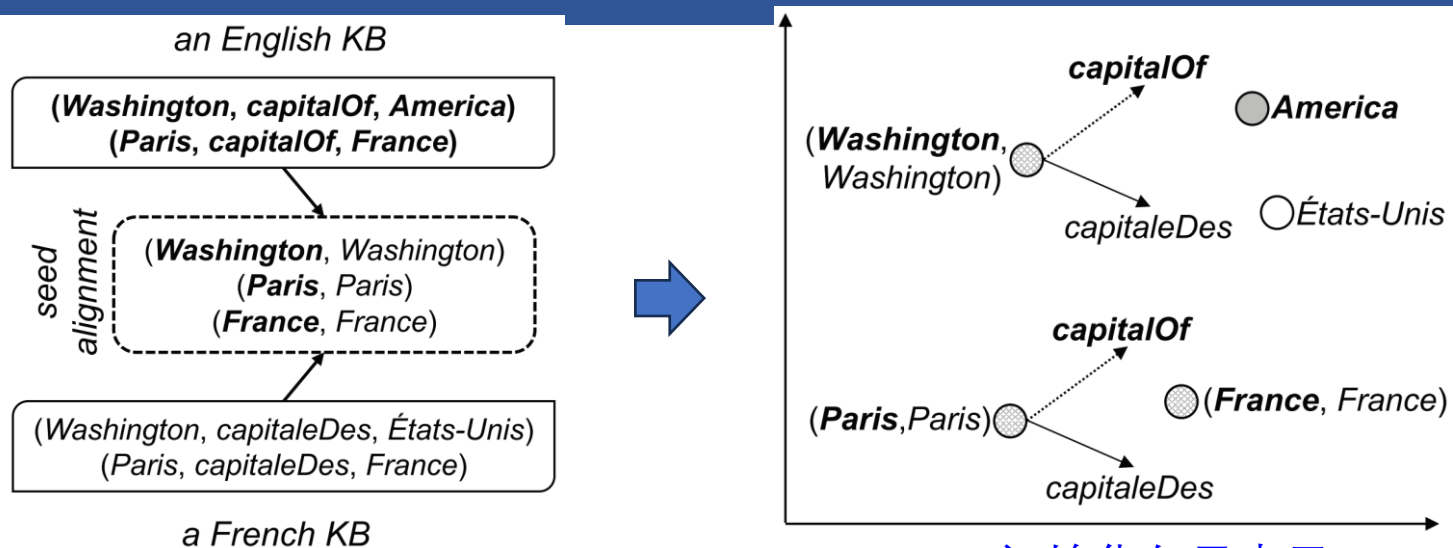
Table 5: Evaluation results on entity prediction in zero-shot scenario

Metric	$d - e$	$e - d$	$d - d$	Total
Partial-CNN	26.8	20.8	69.5	24.8
CNN	31.2	26.1	72.5	29.5

知识图谱嵌入应用案例

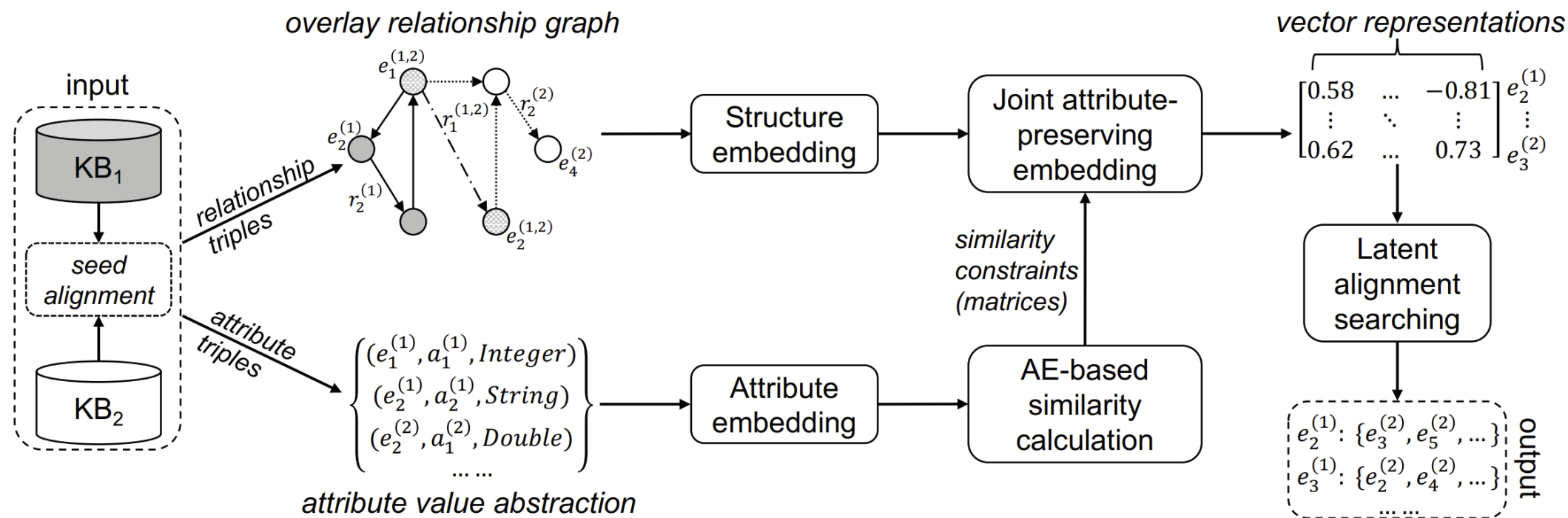
- 基于知识图谱嵌入的实体对齐
 - 从两个不同的知识图谱中识别出表示同一现实对象的实体
- 思路：
 - 同时为两个知识图谱中的实体和关系学习向量表示，并让已知相同的实体共用一个向量表示
 - 最终位于两个知识图谱的相同实体的向量表示会较相似

基于知识图谱嵌入的实体对齐



基于知识图谱嵌入的实体对齐

- (Sun et al., 2017)的方法, 加入了对实体属性的额外考虑



使用基于实体属性得到的实体间相似度辅助嵌入向量学习

END
