



# **RotatE: KNOWLEDGE GRAPH EMBEDDING BY RELATIONAL ROTATION IN COMPLEX SPACE**

**2024-07-30**

**presenter : Sooho Moon**

**DMAIS**

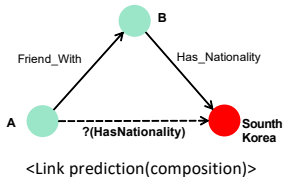
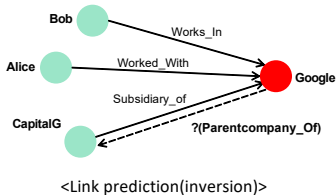
**Zhiqing Sun, Zhi-Hong Deng,  
Jian-Yun Nie, Jian Tang**

**ICLR 2019**

- **Main interest**
- **Problem with previous models**
- **Architecture**
- **Experiments**
- **Conclusion**

- 지식 그래프 : 일반 방향 그래프의 특수한 형태, 노드와(entity), 간선(relation)으로 이루어짐

- **Link prediction** : Entity와 relation 사이의 관계를 knowledge base(KB)에서 추론하는 task
- **Relation Composition** : Link prediction 중 하나인 다수의 관계들을 연결해 새로운 관계를 추론하는 task



# Problem with previous models

- Relation의 기초적 요소인 symmetry, antisymmetry, inversion, composition에 대한 과거 모델들의 학습 범주

Model	Score Function	Symmetry	Antisymmetry	Inversion	Composition
SE	$-\ W_{r,1}h - W_{r,2}t\ $	✗	✗	✗	✗
TransE	$-\ h + r - t\ $	✗	✓	✓	✓
TransX	$-\ g_{r,1}(h) + r - g_{r,2}(t)\ $	✓	✓	✗	✗
DistMult	$\langle h, r, t \rangle$	✓	✗	✗	✗
Complex	$\text{Re}(\langle h, r, t \rangle)$	✓	✓	✓	✗

$\mathcal{g}$  = knowledge graph,  $(h, r, t) \in \mathcal{g}$

- RotatE의 학습 범주

RotatE	$-\ h \circ r - t\ $	✓	✓	✓	✓
--------	----------------------	---	---	---	---

- TorusE(Ebisu & Ichise, 2018) 모델은 임베딩을 compact Lie group에서의 해석으로 접근함
  - 임베딩의 크기를 고정하고 임베딩 벡터들이 torus위에 놓여있게 함
- RotatE는 TorusE에서 더 나아가 복소 공간으로 확장하여 더 다양한 특성들을 학습할 수 있음

$$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k, |r_i| = 1$$

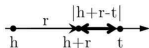
- 오일러 공식을 이용해 복소 평면에서 임베딩 벡터들을 나타냄

$$e^{i\theta} = \cos \theta + i \sin \theta \quad \mathbf{r} = \begin{bmatrix} e^{i\theta_1} \\ e^{i\theta_2} \\ \vdots \\ e^{i\theta_{k-1}} \\ e^{i\theta_k} \end{bmatrix}$$

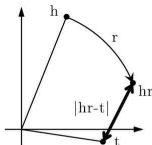
- 주어진 triplet (h,r,t)에 대한 임베딩 벡터들이 Hadamard(element-wise) product에 대한 등식을 만족하기를 바람

$$\mathbf{t} = \mathbf{h} \circ \mathbf{r} \quad \longrightarrow \quad t_i = h_i r_i, \text{ where } h_i, r_i, t_i \in \mathbb{C} \text{ and } |r_i| = 1$$

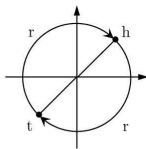
- RotatE가 복소 공간으로 다양한 특성들을 학습할 수 있는 이유 (e.g., 임베딩 차원이 1일때)



(a) TransE models  $r$  as translation in real line.



(b) RotatE models  $r$  as rotation in complex plane.



(c) RotatE: an example of modeling symmetric relations  $r$  with  $r_i = -1$

Figure 1: Illustrations of TransE and RotatE with only 1 dimension of embedding.

- 어떻게 복소 공간에서의 벡터 회전이 symmetry, antisymmetry, inverse, composition을 모두 포착하는가?(참고로 Hadamard 곱은 결합법칙이 성립함)

- **Symmetry** //  $r(x,y), r(y,x)$  가 참일 때

$$y = r \circ x \wedge x = r \circ y \Rightarrow r \circ r = 1$$

- **Antisymmetry** //  $r(x,y), \neg r(y,x)$  가 참일 때

$$y = r \circ x \wedge x \neq r \circ y \Rightarrow r \circ r \neq 1$$



- **Inverse** //  $r_1(x,y), r_2(y,x)$  가 참일 때

$$y = r_1 \circ x \wedge x = r_2 \circ y \Rightarrow r_1 = r_2^{-1}$$

- **Composition** //  $r_1(x,y), r_2(y,z), r_3(x,z)$  가 참일 때

$$z = r_1 \circ x \wedge y = r_2 \circ x \wedge z = r_3 \circ y \Rightarrow r_1 = r_2 \circ r_3$$

- Distance function

$$d_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|$$

- Optimization(corrupted triplet(  $(h'_i, r, t'_i)$  ))

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

$\gamma$  = fixed margin

$\sigma$  = sigmoid function

## ▪ Corrupted triplet negative sampling의 문제점

- 명확히 틀린 negative sample이 모이면 **P**가 필요만큼 작지 않아도 **N**이 너무 작아져 필요만큼 penalize 되지 않을 수 있음(KBGAN(NAAACL-HLT, 2018))

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

## ▪ Negative sample에 가중치를 부여해 해결(self-adversarial negative sampling)

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

- Final optimization function

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

- 비용이 요구되지만 성능 측정은 이후 experiment에서 비교해 본다.

- **Metrics** : **MR**(lower the better), **MRR**(higher the better), **Hits@n**(higher the better)
- **pRotate** : Entity 임베딩의 norm을 상수(C)로 고정시킨 모델이며  
C가 커질수록 RotatE는 TransE로 퇴화할 수 있음

$$\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| = 2C \left\| \sin \frac{\theta_h + \theta_r - \theta_t}{2} \right\|$$

$$\text{let } \theta_h = c\mathbf{h}', \theta_r = c\mathbf{r}', \theta_t = c\mathbf{t}' \text{ and } C = 1/c \quad \longrightarrow \quad \lim_{c \rightarrow 0} \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| = \|\mathbf{h}' + \mathbf{r}' - \mathbf{t}'\|$$

- Symmetry/antisymmetry, inversion datasets(FB15k, WN18)

	FB15k					WN18				
	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE [♥]	-	.463	.297	.578	.749	-	.495	.113	.888	.943
DistMult [♦]	42	.798	-	-	<b>.893</b>	655	.797	-	-	.946
HolE	-	.524	.402	.613	.739	-	.938	.930	.945	.949
ComplEx	-	.692	.599	.759	.840	-	.941	.936	.945	.947
ConvE	51	.657	.558	.723	.831	374	.943	.935	.946	.956
pRotatE	43	<b>.799</b>	<b>.750</b>	.829	.884	<b>254</b>	.947	.942	.950	.957
RotatE	<b>40</b>	.797	.746	<b>.830</b>	.884	309	<b>.949</b>	<b>.944</b>	<b>.952</b>	<b>.959</b>

- Symmetry/antisymmetry, composition datasets(FB15k-237, WN18RR)

	FB15k-237					WN18RR				
	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE [♥]	357	.294	-	-	.465	3384	.226	-	-	.501
DistMult	254	.241	.155	.263	.419	5110	.43	.39	.44	.49
ComplEx	339	.247	.158	.275	.428	5261	.44	.41	.46	.51
ConvE	244	.325	.237	.356	.501	4187	.43	.40	.44	.52
pRotatE	178	.328	.230	.365	.524	<b>2923</b>	.462	.417	.479	.552
RotatE	<b>177</b>	<b>.338</b>	<b>.241</b>	<b>.375</b>	<b>.533</b>	3340	<b>.476</b>	<b>.428</b>	<b>.492</b>	<b>.571</b>

각각 subset of FB15k, WN18

- Experiments on Countries(composition) dataset

	Countries (AUC-PR)			
	DistMult	Complex	ConvE	RotatE
S1	<b><math>1.00 \pm 0.00</math></b>	$0.97 \pm 0.02$	<b><math>1.00 \pm 0.00</math></b>	<b><math>1.00 \pm 0.00</math></b>
S2	$0.72 \pm 0.12$	$0.57 \pm 0.10$	$0.99 \pm 0.01$	<b><math>1.00 \pm 0.00</math></b>
S3	$0.52 \pm 0.07$	$0.43 \pm 0.07$	$0.86 \pm 0.05$	<b><math>0.95 \pm 0.00</math></b>

S# : 길이가 #인 composition 추론에 대한 정확도



## ▪ 다양한 negative sampling technique에 대한 평가

- **uniform** : Random extraction from  $\mathcal{E}$
- **KBGAN** : 본 자료에서 소개한 sampling 기법을 모델화 시킨 방법
- **self-adversarial** : KBGAN을 TransE로 튜닝

	FB15k-237		WN18RR		WN18	
	MRR	H@10	MRR	H@10	MRR	H@10
uniform	.242	.422	.186	.459	.433	.915
KBGAN (Cai & Wang, 2017)	.278	.453	.210	.479	.705	<b>.949</b>
self-adversarial	<b>.298</b>	<b>.475</b>	<b>.223</b>	<b>.510</b>	<b>.736</b>	.947

- 1-to-1, 1-to-N, N-to-1, N-to-N relation에 대한 평가

Relation Category	1-to-1	1-to-N	N-to-1	N-to-N	1-to-1	1-to-N	N-to-1	N-to-N
Tasks	Prediction Head (Hits@10)				Prediction Tail (Hits@10)			
TransE	.437	.657	.182	.472	.437	.197	.667	.500
TransH (bern)	.668	.876	.287	.645	.655	.398	.833	.672
KG2E_KL (bern)	.923	.946	.660	.696	.926	.679	.944	.734
TransE	.894	<b>.972</b>	.567	.880	.879	.671	<b>.964</b>	.910
ComplEx	<b>.939</b>	.969	<b>.692</b>	<b>.893</b>	<b>.938</b>	<b>.823</b>	.952	.910
RotatE	.922	.967	.602	<b>.893</b>	.923	.713	.961	<b>.922</b>
Tasks	Prediction Head (MRR)				Prediction Tail (MRR)			
TransE	.701	.912	.424	.737	.701	.561	.894	.761
ComplEx	.832	.914	<b>.543</b>	.787	.826	<b>.661</b>	.869	.800
RotatE	<b>.878</b>	<b>.934</b>	.465	<b>.803</b>	<b>.872</b>	.611	<b>.909</b>	<b>.832</b>

## ▪ Previous problems

- Link prediction의 기초적인 relation(symmetric/antisymmetric, inversion, composition)을 모두 학습할 수 있는 모델이 없음

## ▪ RotatE

- 복소 공간에서의 회전을 임베딩에 사용해 다양한 표현을 학습 가능
- 새로운 self-adversarial negative sampling 기법을 사용해 모델 기능 향상
- 모델의 크기가 KB의 크기에 linear하므로 더 거대한 KB로 확장 가능함

Thank You!



**Contact:** Sooho Moon (Email: [moonwalk725@cau.ac.kr](mailto:moonwalk725@cau.ac.kr))