RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space

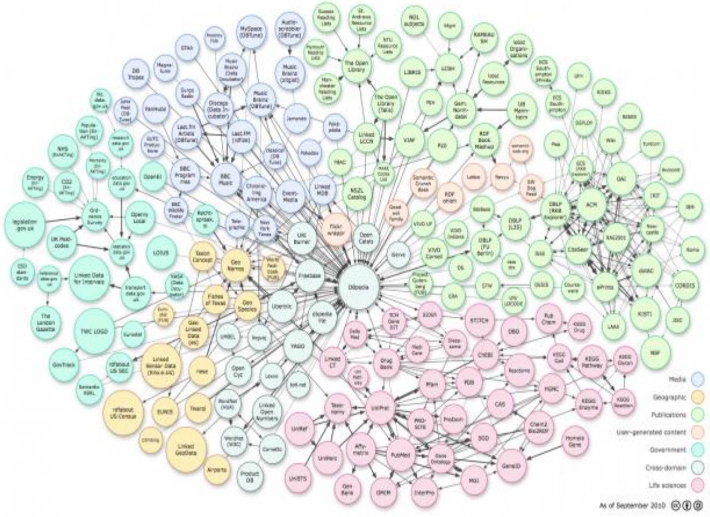
Jian Tang

Mila & HEC Montreal

Zhiqing Sun, Zhihong Deng, Jian-Yun Nie, and **Jian Tang**. “RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space.” In submission to ICLR’19.

Knowledge Graphs

- A set of facts represented as triplets
  - (head entity, relation, tail entity)



- A variety of applications
  - Question answering
  - Search
  - Recommender Systems
  - Natural language understanding
  - …

Google Knowledge Graph

Microsoft SATORI

Browse the Knowledge Base!

NELL: Never-Ending Language Learning

Freebase

DBpedia

yago select knowledge

OpenIE
(Reverb, OLLIE)

Knowledge Graphs are Incomplete

- A fundamental task: **predicting missing links**
- A lot of existing work on knowledge graph embedding for link prediction
- The Key Idea: model and infer the **relation patterns** in knowledge graphs according to observed knowledge facts.
- Example:

Barack_Obama **BornIn** United_States

Barack_Obama **Nationality** American

Parents of Parents are Grandparents

Relation Patterns

- Symmetric/Antisymmetric Relations
  - Symmetric: e.g., Marriage
  - Antisymmetric: e.g., Filiation
- Formally:

***r* is Symmetric:** $r(x, y) \Rightarrow r(y, x)$ if $\forall\ x, y$

***r* is Antisymmetric:** $r(x, y) \Rightarrow \neg r(y, x)$ if $\forall\ x, y$

Relation Patterns

- Inverse Relations
  - Hypernym and hyponym
- Formally:

r_1 **is inverse to relation** r_2**:** $r_2(x, y) \Rightarrow r_1(y, x)$ if $\forall\ x, y$

Relation Patterns

- Composition Relations
  - My mother's husband is my father
- Formally:

r_1 is composed of relation r_2 and relation r_3: $r_2(x, y) \land r_3(y, z) \Rightarrow r_1(x, z)$ if $\forall\ x, y, \mathrm{z}$

Related Work on Knowledge Graph Embedding

- Representing entities as **embeddings**
- Representing relations as **embeddings** or **matrices**

Model	Score Function	
SE (Bordes et al., 2011)	$-\lVert W_{r,1}\mathbf{h} - W_{r,2}\mathbf{t}\rVert$	$\mathbf{h},\mathbf{t} \in \mathbb{R}^k, W_{r,\cdot} \in \mathbb{R}^{k\times k}$
TransE (Bordes et al., 2013)	$-\lVert \mathbf{h} + \mathbf{r} - \mathbf{t}\rVert$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{R}^k$
TransX	$-\lVert g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t})\rVert$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{R}^k$
DistMult (Yang et al., 2014)	$\langle \mathbf{r}, \mathbf{h}, \mathbf{t}\rangle$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{R}^k$
ComplEx (Trouillon et al., 2016)	$\mathrm{Re}(\langle \mathbf{r}, \mathbf{h}, \overline{\mathbf{t}}\rangle)$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{C}^k$
HolE (Nickel et al., 2016)	$\langle \mathbf{r}, \mathbf{h} \otimes \mathbf{t}\rangle$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{R}^k$
ConvE (Dettmers et al., 2017)	$\langle \sigma(\mathrm{vec}(\sigma([\overline{\mathbf{r}}, \overline{\mathbf{h}}] * \Omega))W), \mathbf{t}\rangle$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{R}^k$
RotatE	$-\lVert \mathbf{h} \circ \mathbf{r} - \mathbf{t}\rVert^1$	$\mathbf{h},\mathbf{r},\mathbf{t} \in \mathbb{C}^k, \lvert r_i\rvert = 1$

Abilities in Inferring the Relation Patterns

- None of existing methods are able to model and infer all the three types of relation patterns

Model	Score Function	Symmetry	Antisymmetry	Inversion	Composition
SE	$-\|W_{r,1}\mathbf{h} - W_{r,2}\mathbf{t}\|$	✗	✗	✗	✗
TransE	$-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$	✗	✓	✓	✓
TransX	$-\|g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t})\|$	✓	✓	✗	✗
DistMult	$\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$	✓	✗	✗	✗
ComplEx	$\mathrm{Re}(\langle \mathbf{h}, \mathbf{r}, \overline{\mathbf{t}} \rangle)$	✓	✓	✓	✗
RotatE	$-\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|$	✓	✓	✓	✓

Our Contributions

- A new knowledge graph embedding model RotatE
  - Each relation as a rotation from the source entity to the target entity in the complex vector space
- RotatE is able to model and infer all the three types of relation patterns
- An efficient and effective negative sampling algorithm for optimizing RotatE
- State-of-the-art results on all the benchmarks for link prediction on knowledge graphs

RotatE: Relation as Rotation in Complex Space

- Representing head and tail entities in complex vector space, i.e., $\mathbf{h}, \mathbf{t} \in \mathbb{C}^{\boldsymbol{k}}$
- Define each relation **r** as an element-wise rotation from the head entity $\mathbf{h}$ to the tail entity **t,** i.e.,

$$\mathbf{t} = \mathbf{h} \circ \boldsymbol{r}, \quad \text{where } |r_i|=1$$

- $\circ$ is the element-wise product. More specifically, we have $\mathrm{t_i} = \mathrm{h_i r_i}$, and

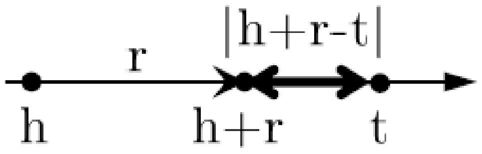
$$\mathrm{r_i} = e^{i\theta_{r,i}},$$

- where $\theta_{r,i}$ is the phase angle of **r** in the i-th dimension.

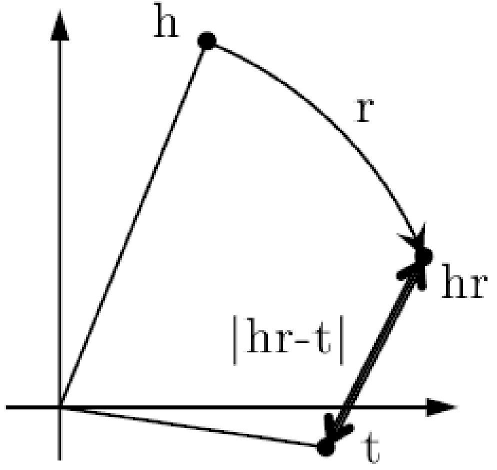
Geometric Interpretation

- Define the distance function of RotatE as

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



(a) TransE models $\mathbf{r}$ as translation in real line.



(b) RotatE models $\mathbf{r}$ as rotation in complex plane.

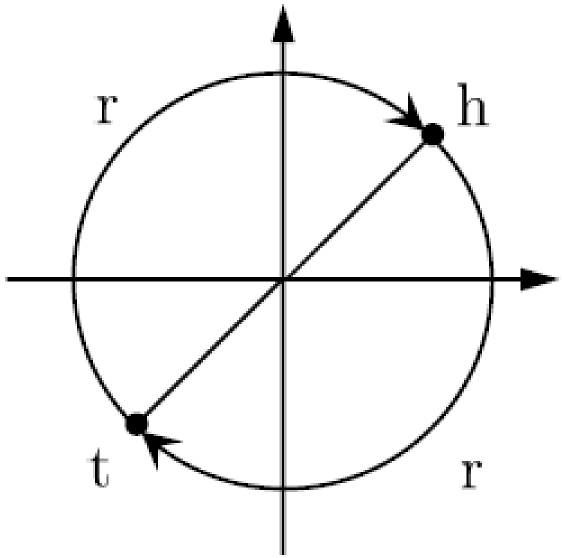
Modeling the Relation Patterns with RotatE

- A relation **r** is **symmetric** if and only if $r_i = \pm 1$, i.e.,

$$\theta_{r,i} = 0 \; or \; \pi$$

- An example on the space of $\mathbb{C}$

$r_i = -1$ or $\theta_{r,i} = \pi$



Modeling the Relation Patterns with RotatE

- A relation r is **antisymmetric** if and only if $\mathbf{r} \circ \mathbf{r} \neq \mathbf{1}$

- Two relations r_1 and r_2 are **inverse** if and only if $\mathbf{r}_2 = \overline{\mathbf{r}_1}$, i.e.,

$$\theta_{2,i} = -\theta_{1,i}$$

- A relation $\boldsymbol{r_3} = e^{i\boldsymbol{\theta_3}}$ is a **composition** of two relations $\boldsymbol{r_1} = e^{i\boldsymbol{\theta_1}}$ and $\boldsymbol{r_2} = e^{i\boldsymbol{\theta_2}}$ if only if $\boldsymbol{r_3} = \boldsymbol{r_1} \circ \boldsymbol{r_2}$, i.e.,

$$\boldsymbol{\theta_3} = \boldsymbol{\theta_1} + \boldsymbol{\theta_2}$$

Optimization

- Negative sampling loss

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

- γ is a fixed margin, σ is the sigmoid function, and $(\boldsymbol{h}_i', \boldsymbol{r}, \boldsymbol{t}_i')$ is the i-th negative triplet.

Self-adversarial Negative Sampling

- Traditionally, the negative samples are drawn in an uniform way
  - Inefficient as training goes on since many samples are obviously false
  - Does not provide useful information
- A self-adversarial negative sampling
  - Sample negative triplets according to the current embedding model
  - Starts from easier samples to more and more difficult samples
  - Curriculum Learning

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

- α is the temperature of sampling. $f_r(h'_j, t'_j)$ measures the salience of the triplet

The Final Objective

- Instead of sampling, treating the sampling probabilities as weights.

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

Experiments: Data Sets

- **FB15K**: a subset of Freebase. The main relation types are **symmetry/antisymmetry** and **inversion** patterns.
- **WN18**: a subset of WordNet. The main relation types are **symmetry/antisymmetry** and **inversion** patterns.
- **FB15K-237**: a subset of FB15K, where inversion relations are deleted. The main relation types are **symmetry/antisymmetry** and **composition** patterns.
- **WN18RR**: a subset of WN18, where inversion relations are deleted. The main relation types are **symmetry/antisymmetry** and **composition** patterns.

Dataset	#entity	#relation	#training	#validation	#test
FB15k	14,951	1,345	483,142	50,000	59,071
WN18	40,943	18	141,442	5,000	5,000
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134

Compared Algorithms

- Baseline algorithms
  - TransE (Bordes et al. 2013)
  - DistMult (Yang et al. 2014)
  - ComplEx (Trouillon et al. 2016)
  - ConvE (Dettmers et al. 2017)
- Our algorithms
  - RotatE
  - pRotatE: a variant of our algorithm by constraining the modulus of the entity embeddings to be the same, i.e., $|h_i| = |t_i| = C$

Results on FB15k and WN18

- RotatE performs the best
- pRotatE performs similarly to RotatE

	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	-	-	**.893**	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	**.799**	**.750**	.829	.884	**254**	.947	.942	.950	.957
RotatE	**40**	.797	.746	**.830**	.884	309	**.949**	**.944**	**.952**	**.959**

Results on FB15k-237 and WN18RR

- RotatE performs the best
- RotatE performs significantly better than pRotatE
  - A lot of composition patterns on the two data sets
  - Modulus information are important for modeling the composition patterns

	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	**2923**	.462	.417	.479	.552
RotatE	**177**	**.338**	**.241**	**.375**	**.533**	3340	**.476**	**.428**	**.492**	**.571**

Results on Countries (Bouchard et al. 2015)

- A carefully designed dataset to explicitly test the capabilities for modeling the composition patterns
  - Three subtasks S1, S2, S3
  - From easy to difficult

	Countries (AUC-PR)			
	DistMult	ComplEx	ConvE	RotatE
S1	**1.00 ± 0.00**	0.97 ± 0.02	**1.00 ± 0.00**	**1.00 ± 0.00**
S2	0.72 ± 0.12	0.57 ± 0.10	0.99 ± 0.01	**1.00 ± 0.00**
S3	0.52 ± 0.07	0.43 ± 0.07	0.86 ± 0.05	**0.95 ± 0.00**

Comparing Different Negative Sampling Techniques

- Different negative sampling techniques
  - Uniform sampling
  - KBGAN (Cai&Wang, 2017) : a generator for generating negative samples and a discriminator for training knowledge graph embeddings
  - our Self-adversarial technique

	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	**.949**
self-adversarial	**.298**	**.475**	**.223**	**.510**	**.736**	.947

Table: The performance of different techniques with TransE

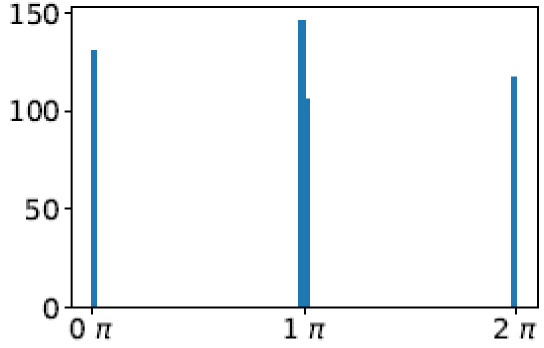
Further Experiments

- To make a fair comparison, we compare with other models, which are also trained with self-adversarial
- Similar results are observed

	FB15k		FB15k-237		Countries (AUC-ROC)		
	MRR	H@10	MRR	H@10	S1	S2	S3
TransE	.735	.871	.332	.531	**1.00 ± 0.00**	**1.00 ± 0.00**	**0.96 ± 0.00**
ComplEx	.780	**.890**	.319	.509	**1.00 ± 0.00**	0.98 ± 0.00	0.88 ± 0.01
RotatE	**.797**	.884	**.338**	**.533**	**1.00 ± 0.00**	**1.00 ± 0.00**	0.95 ± 0.00

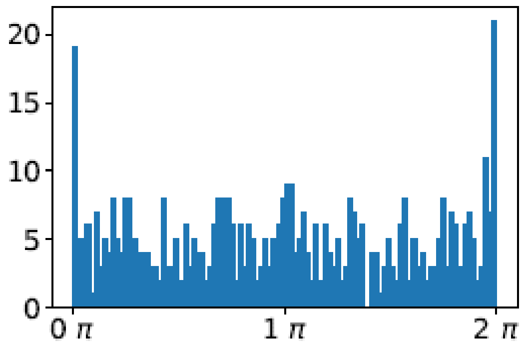
Implicit Relation Pattern Inference: Symmetric

- Ignore the positions in the relation embedding $\boldsymbol{\theta_r}$ and plot the histogram of the phrase angle of each element in the relation embedding, i.e., $\theta_{r,i}$



(a) similar_to

A **symmetric** relation

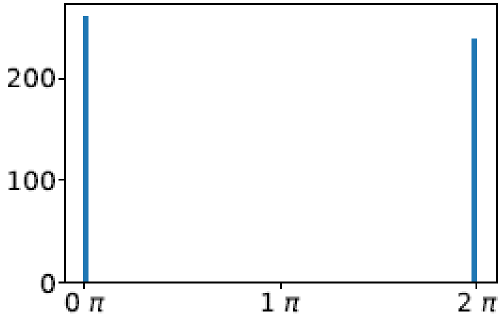


(b) hypernym

A **random** relation

Implicit Relation Pattern Inference: Inversion

- Ignore the positions in the relation embedding $\boldsymbol{\theta_r}$ and plot the histogram of $\theta_{1,i} + \theta_{2,i}$

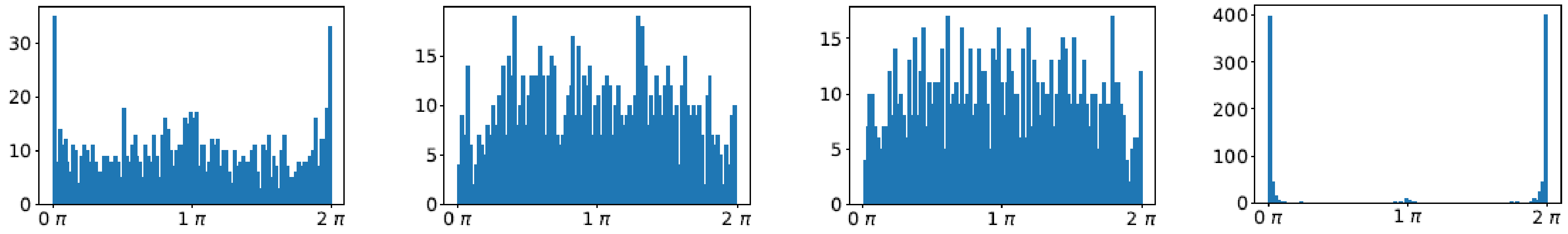


(c) hypernym ∘ hyponym

hypernym is the **inverse** relation of hyponym

Implicit Relation Pattern Inference: Composition

- Ignore the positions in the relation embedding $\boldsymbol{\theta_r}$ and plot the histogram $\theta_{1,i} + \theta_{2,i} - \theta_{3,i}$



(d) $\mathtt{for}_1$ (e) $\mathtt{winner}$ (f) $\mathtt{for}_2$ (g) $\mathtt{for}_2^{-1} \circ \mathtt{winner} \circ \mathtt{for}_1$

for2 is a **composition** of the relation for1 and winner.
for1: award_nominee/award_nominations./award/awrd_nomination/nominated_for
winner: award_category/winners./award/award_honor/award_winner
for2: award_category/nominees./award/award_nomination/nominated_for

Summary

- Modeling relation patterns is critical for knowledge base completion
  - Symmetric/Antisymmetric, Inverse, and composition
- RotatE: define each relation as a **rotation** from the head entity to the tail entity in the complex vector space
  - Capable of modeling and inferring all the three types of relation patterns
- A new self-adversarial negative sampling approach
  - Sampling the negative samples according to current embeddings
  - Curriculum learning
- State-of-the-art results on all existing benchmark data sets

Thanks!

jian.tang@hec.ca