Improving Knowledge Graph Embedding Using Simple Constraints

Boyang Ding, **Quan Wang**, Bin Wang, Li Guo

Institute of Information Engineering, Chinese Academy of Sciences

School of Cyber Security, University of Chinese Academy of Sciences

Code and data available at https://github.com/iieir-km/ComplEx-NNE_AER

ACL-18: 15-20 July 2018, Melbourne, Australia

Outline

1 Intro

2 Approach

3 Experiments

4 Summary

1 Intro

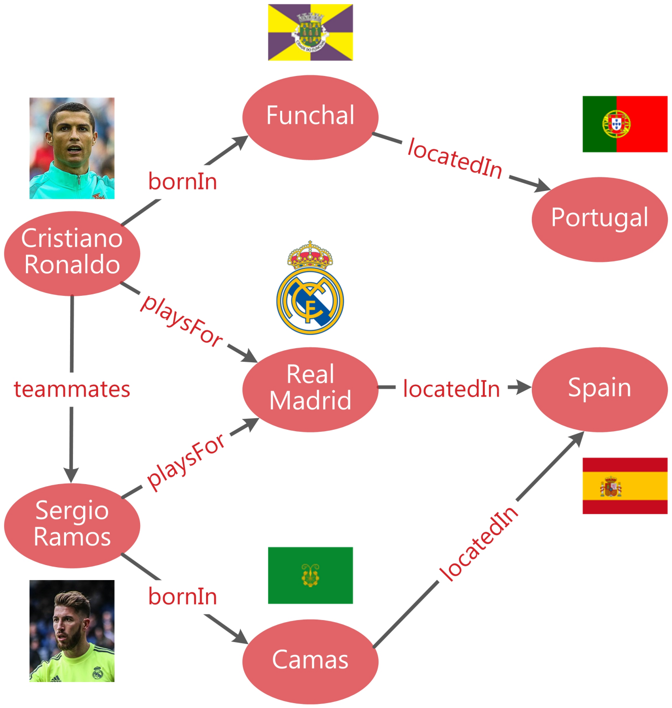
2 Approach

3 Experiments

4 Summary

Knowledge graph

- A directed graph composed of entities (nodes) and relations (edges)



(Cristiano Ronaldo, bornIn, Funchal)

(Cristiano Ronaldo, playsFor, Real Madrid)

(Cristiano Ronaldo, teammates, Sergio Ramos)

(Sergio Ramos, bornIn, Camas)

(Sergio Ramos, playsFor, Real Madrid)

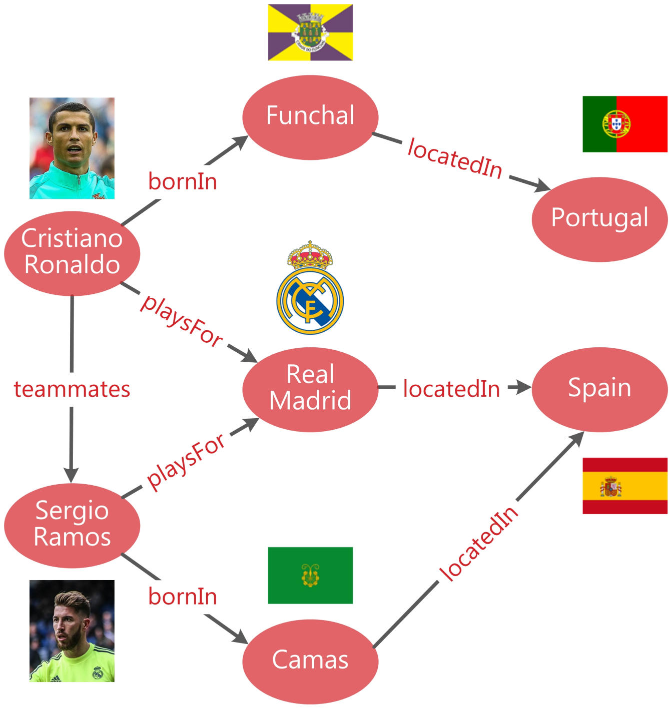
(Funchal, locatedIn, Portugal)

(Real Madrid, locatedIn, Spain)

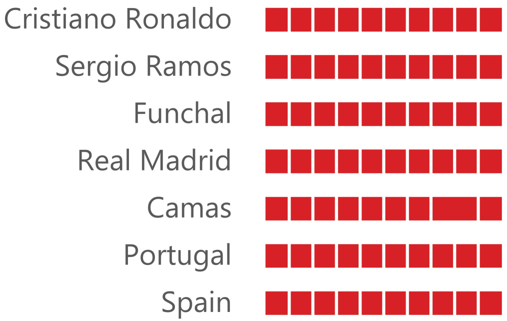
(Camas, locatedIn, Spain)

Knowledge graph embedding

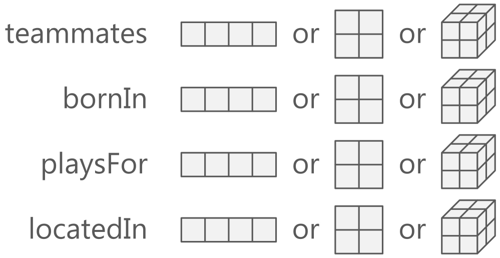
- Learn to represent entities and relations in continuous vector spaces



Entities as points in vector spaces (vectors)

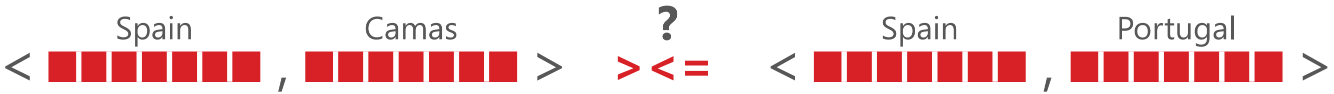


Relations as operations between entities (vectors/matrices/tensors)

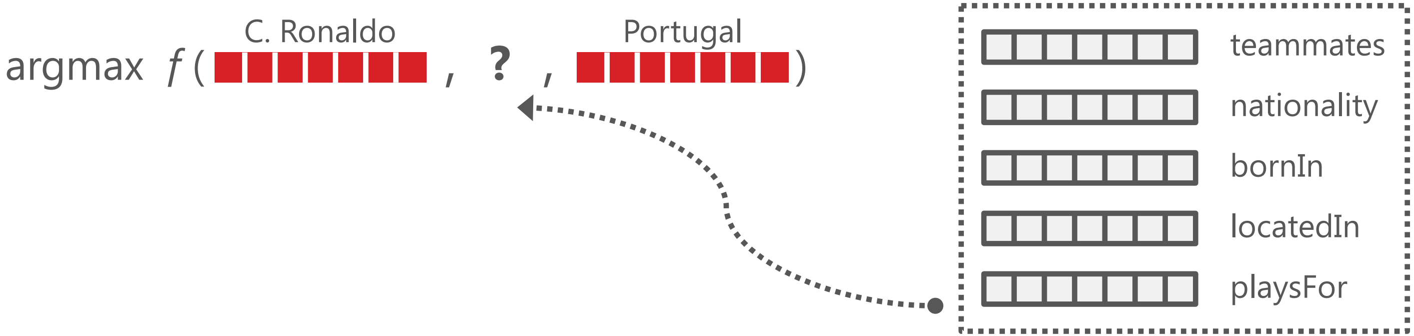


Knowledge graph embedding (cont.)

- Easy computation and inference on knowledge graphs
  - Is **Spain** more similar to **Camas** (a municipality located in Spain) or **Portugal** (both Portugal and Spain are European countries)?



  - What is the relationship between **Cristiano Ronaldo** and **Portugal**?



Previous approaches

- Early works
  - Simple models developed over RDF triples, e.g., TransE, RESCAL, DistMult, ComplEx, ect
- Recent trends
  - Designing more complicated triple scoring models
    Usually with higher computational complexity
  - Incorporating extra information beyond RDF triples
    Not always applicable to all knowledge graphs

This work

- Using simple constraints to improve knowledge graph embedding
  - Non-negativity constraints on entity representations
  - Approximate entailment constraints on relation representations
- Benefits
  - More predictive embeddings
  - More interpretable embeddings
  - Low computational complexity

1 Intro

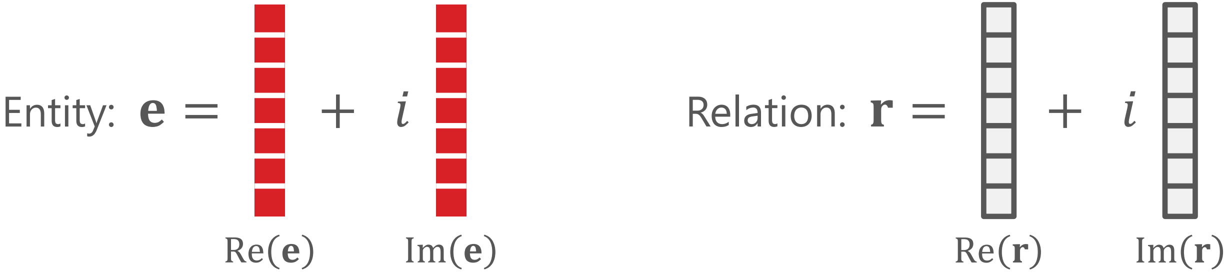
2 Approach

3 Experiments

4 Summary

Basic embedding model: ComplEx

- Entity and relation representations: complex-valued vectors



- Triple scoring function: multi-linear dot product

$$\phi(e_i, r_k, e_j) \triangleq \mathrm{Re}(\langle \mathbf{e}_i, \mathbf{r}_k, \bar{\mathbf{e}}_j \rangle)$$
$$\triangleq \mathrm{Re}\left(\sum_{\ell} [\mathbf{e}_i]_\ell [\mathbf{r}_k]_\ell [\bar{\mathbf{e}}_j]_\ell\right)$$

  - Triples with higher scores are more likely to be true

Non-negativity of entity representations

- Intuition
  - Uneconomical to store negative properties of an entity/concept

Positive properties of cats
- Cats are mammals
- Cats eat fishes
- Cats have four legs

√

Negative properties of cats
- Cats are not vehicles
- Cats do not have wheels
- Cats are not used for communication

X

- Non-negativity constraints

$$\boxed{0 \leq \mathrm{Re}(\mathbf{e}), \mathrm{Im}(\mathbf{e}) \leq \mathbf{1}, \quad \forall e \in \mathcal{E}}$$

non-negativity

⇓

sparsity & interpretability

Approximate entailment for relations

- Approximate entailment
  - $r_p \xrightarrow{\lambda} r_q$: relation r_p approximately entails relation r_q with confidence level λ
  - $\texttt{bornIn} \xrightarrow{0.8} \texttt{nationality}$: a person born in a country is very likely, but not necessarily, to have a nationality of that country
  - Can be derived automatically by modern rule mining systems

Approximate entailment for relations (cont.)

- Approximate entailment constraints
  - Strict entailment $r_p \to r_q$ ($\lambda = +\infty$)

$$\phi(e_i, r_p, e_j) \leq \phi(e_i, r_q, e_j), \quad \forall e_i, e_j \in \mathcal{E} \qquad (*)$$

  - A sufficient condition for $(*)$

$$\mathrm{Re}(\mathbf{r}_p) \leq \mathrm{Re}(\mathbf{r}_q), \;\; \mathrm{Im}(\mathbf{r}_p) = \mathrm{Im}(\mathbf{r}_q) \qquad (**)$$

  - Introducing confidence λ and allowing slackness in $(**)$

A higher confidence level shows less tolerance for violating the constraints

$$\boxed{\begin{aligned} \lambda\big(\mathrm{Re}(\mathbf{r}_p) - \mathrm{Re}(\mathbf{r}_q)\big) &\leq \alpha \\ \lambda\big(\mathrm{Im}(\mathbf{r}_p) - \mathrm{Im}(\mathbf{r}_q)\big)^2 &\leq \beta \end{aligned}}$$

- **Avoid grounding**
- **Handle uncertainty**

Overall model

- Basic embedding model of ComplEx + non-negativity constraints + approximate entailment constraints

$$\min_{\Theta,\{\alpha,\beta\}} \sum_{\mathcal{D}^+\cup\mathcal{D}^-} \log\left(1+\exp(-y_{ijk}\phi(e_i,r_k,e_j))\right)$$

logistic loss for ComplEx

$$+\mu\sum\nolimits_{\mathcal{T}} \mathbf{1}^\top(\boldsymbol{\alpha}+\boldsymbol{\beta})+\eta\|\Theta\|_2^2,$$

$$\text{s.t.}\ \ \lambda\left(\mathrm{Re}(\mathbf{r}_p)-\mathrm{Re}(\mathbf{r}_q)\right)\leq\boldsymbol{\alpha},$$

$$\lambda\left(\mathrm{Im}(\mathbf{r}_p)-\mathrm{Im}(\mathbf{r}_q)\right)^2\leq\boldsymbol{\beta},$$

approximate entailment constraints on relation representations

$$\boldsymbol{\alpha},\boldsymbol{\beta}\geq\mathbf{0},\quad \forall r_p\xrightarrow{\lambda}r_q\in\mathcal{T},$$

$$\mathbf{0}\leq\mathrm{Re}(\mathbf{e}),\mathrm{Im}(\mathbf{e})\leq\mathbf{1},\quad \forall e\in\mathcal{E}.$$

non-negativity constraints on entity representations

Complexity analysis

- Space complexity: $\mathcal{O}(nd + md)$ → the same as that of ComplEx
  - n is the number of entities
  - m is the number of relations
  - d is the dimensionality of the embedding space

the same as that of ComplEx

- Time complexity per iteration: $\mathcal{O}(sd + \bar{n}d + td) \sim \mathcal{O}(sd)$
  - s is the average number of triples in a mini-batch
  - $\bar{n}$ is the average number of entities in a mini-batch
  - t is the total number of approximate entailments

1 Intro

2 Approach

3 Experiments

4 Summary

Experimental setups

- Datasets
  - WN18: subset of WordNet
  - FB15k: subset of Freebase
  - DB100k: subset of DBpedia
  - Training/validation/test split
- Approximate entailment
  - Automatically extracted by AMIE+ with confidence level higher than 0.8

Dataset	# Ent	# Rel	# Train/Valid/Test			# Cons
WN18	40,943	18	141,442	5,000	5,000	17
FB15K	14,951	1,345	483,142	50,000	59,071	535
DB100K	99,604	470	597,572	50,000	50,000	56

$\text{hypernym}^{-1} \xrightarrow{1.00} \text{hyponym}$
$\text{synset\_domain\_topic\_of}^{-1} \xrightarrow{0.99} \text{member\_of\_domain\_topic}$
$\text{instance\_hypernym}^{-1} \xrightarrow{0.98} \text{instance\_hyponym}$
$\text{/people/place\_of\_birth}^{-1} \xrightarrow{1.00} \text{/location/people\_born\_here}$
$\text{/film/directed\_by}^{-1} \xrightarrow{0.98} \text{/director/film}$
$\text{/country/admin\_divisions} \xrightarrow{0.91} \text{/country/1st\_level\_divisions}$
$\text{owner} \xrightarrow{0.95} \text{owning\_company}$
$\text{child}^{-1} \xrightarrow{0.92} \text{parent}$
$\text{distributing\_company} \xrightarrow{0.92} \text{distributing\_label}$

Experimental setups (cont.)

- Link prediction
  - To complete a triple (e_i, r_k, e_j) with e_i or e_j missing
- Baselines
  - Simple embedding models based on RDF triples
  - Other extensions of ComplEx incorporating logic rules
  - Recently developed neural network architectures
- Our approaches
  - ComplEx-NNE: only with non-negativity constraints
  - ComplEx-NNE+AER: also with approximate entailment constraints

Link prediction results

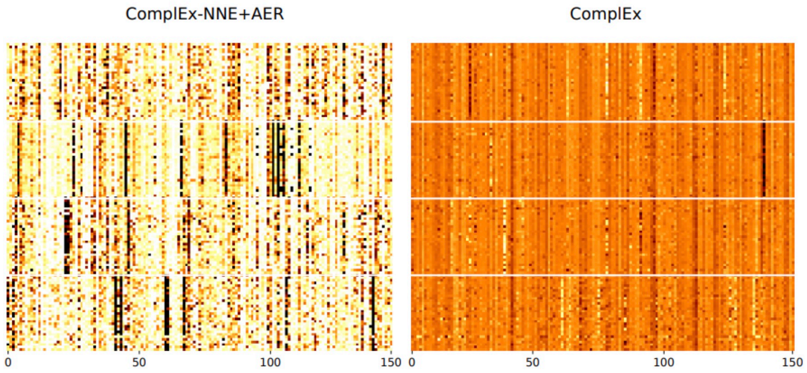
	WN18			FB15K			DB100K		
	MRR	HITS@1	HITS@3	MRR	HITS@1	HITS@3	MRR	HITS@1	HITS@3
TransE(2013)	0.454	0.089	0.823	0.380	0.231	0.472	0.111	0.016	0.164
DistMult(2015)	0.822	0.728	0.914	0.654	0.546	0.733	0.233	0.115	0.301
HolE(2016)	0.938	0.930	0.945	0.524	0.402	0.613	0.260	0.182	0.309
ComplEx(2016)	0.941	0.936	0.945	0.692	0.599	0.759	0.242	0.126	0.312
ANALOGY(2017)	0.942	0.939	0.944	0.725	0.646	0.785	0.252	0.143	0.323
RUGE(2018)	—	—	—	0.768	0.703	0.815	0.246	0.129	0.325
ComplEx^R(2017)	0.940	—	0.943	—	—	—	0.253	0.167	0.294
R-GCN(2017)	0.814	0.686	0.928	0.651	0.541	0.736	—	—	—
R-GCN+(2017)	0.819	0.697	0.929	0.696	0.601	0.760	—	—	—
ConvE(2018)	0.942	0.935	**0.947**	0.745	0.670	0.801	—	—	—
Single DistMult(2017)	0.797	—	—	0.798	—	—	—	—	—
ComplEx-NNE	0.941	0.937	0.944	0.727	0.659	0.772	0.298	0.229	0.330
ComplEx-NNE+AER	**0.943**	**0.940**	0.945	**0.803**	**0.761**	**0.831**	**0.306**	**0.244**	**0.334**

Simple embedding models   Incorporating logic rules   Neural network architectures

ComplEx-NNE+AER can beat very strong baselines just by introducing the simple constraints

Analysis on entity representations

- Visualization of entity representations
  - Pick 4 types reptile/wine region /species/programming language, and randomly select 30 entities from each type
  - Visualize the representations of these entities learned by ComplEx and ComplEx-NNE+AER

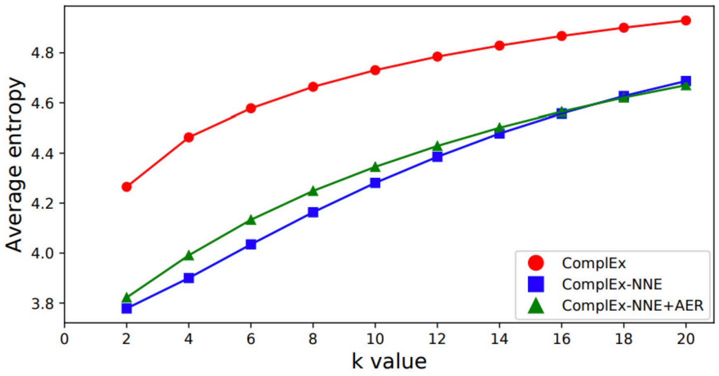


Compact and interpretable entity representations

- Each entity is represented by only a relatively small number of "active" dimensions
- Entities with the same type tend to activate the same set of dimensions

Analysis on entity representations (cont.)

- Semantic purity of latent dimensions
  - For each latent dimension, pick top K percent of entities with the highest activation values on this dimension
  - Calculate the entropy of the type distribution of these entities

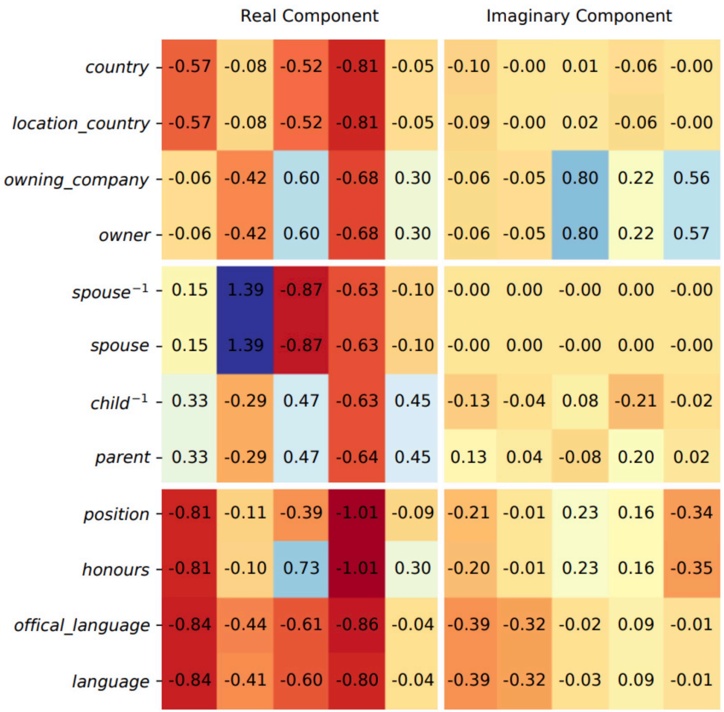


Latent dimensions with higher semantic purity

- A lower entropy means entities along this dimension tend to have the same type (higher semantic purity)

Analysis on relation representations

- Visualization of relation representations



Encode logical regularities quite well

Equivalence $r_p \leftrightarrow r_q$

$$\mathrm{Re}(\mathbf{r}_p) = \mathrm{Re}(\mathbf{r}_q)$$
$$\mathrm{Im}(\mathbf{r}_p) = \mathrm{Im}(\mathbf{r}_q)$$

Inversion $r_p \leftrightarrow r_q^{-1}$

$$\mathrm{Re}(\mathbf{r}_p) = \mathrm{Re}(\mathbf{r}_q)$$
$$\mathrm{Im}(\mathbf{r}_p) = -\mathrm{Im}(\mathbf{r}_q)$$

Ordinary entailment

$$\mathrm{Re}(\mathbf{r}_p) \leq \mathrm{Re}(\mathbf{r}_q)$$
$$\mathrm{Im}(\mathbf{r}_p) = \mathrm{Im}(\mathbf{r}_q)$$

1 Intro

2 Approach

3 Experiments

4 Summary

This work

- Using simple constraints to improve knowledge graph embedding
  - Non-negativity constraints on entity representations
  - Approximate entailment constraints on relation representations
- Experimental results
  - Effective
  - Efficient
  - Interpretable embeddings

Code and data available at https://github.com/iieir-km/ComplEx-NNE_AER

Thank you!

Q&A

wangquan@iie.ac.cn