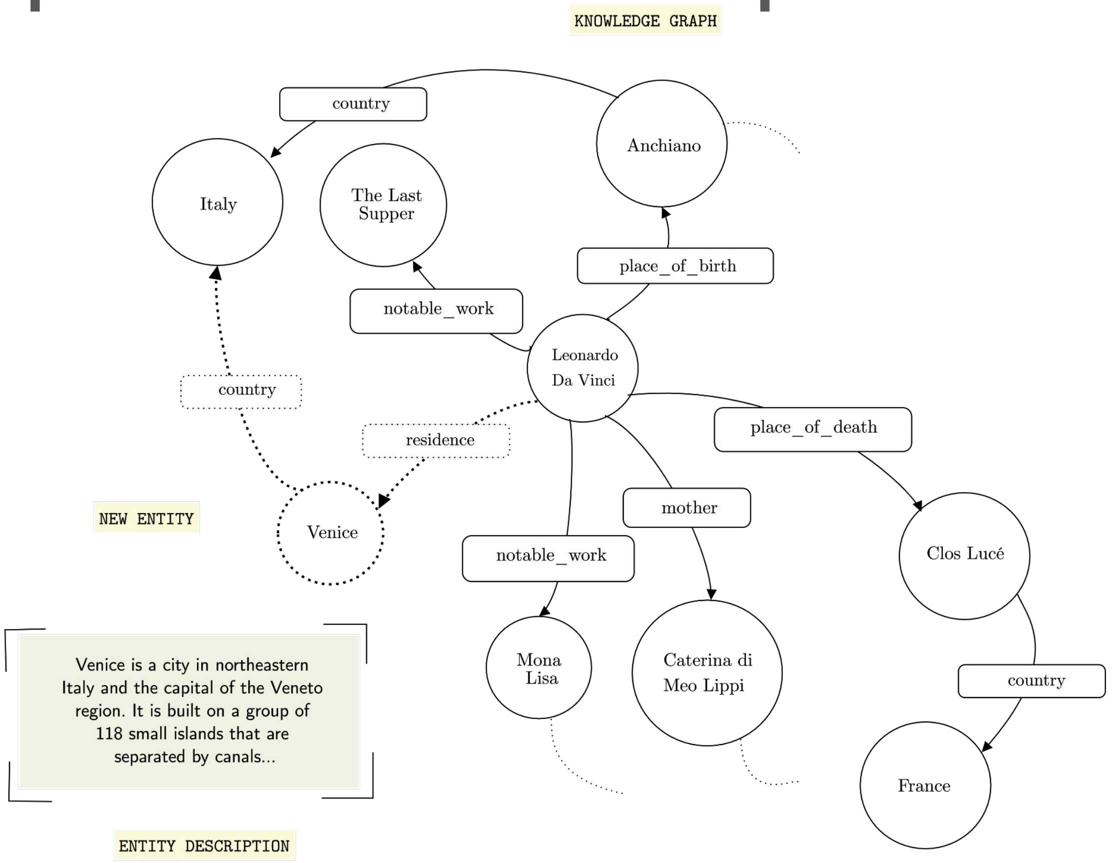
A Joint Training Framework for Open-World Knowledge Graph Embeddings

Karthik Venkat Ramanan, Beethika Tripathi,
Mitesh M. Khapra, Balaraman Ravindran

Robert Bosch Centre for Data Science and AI
Indian Institute of Technology, Madras

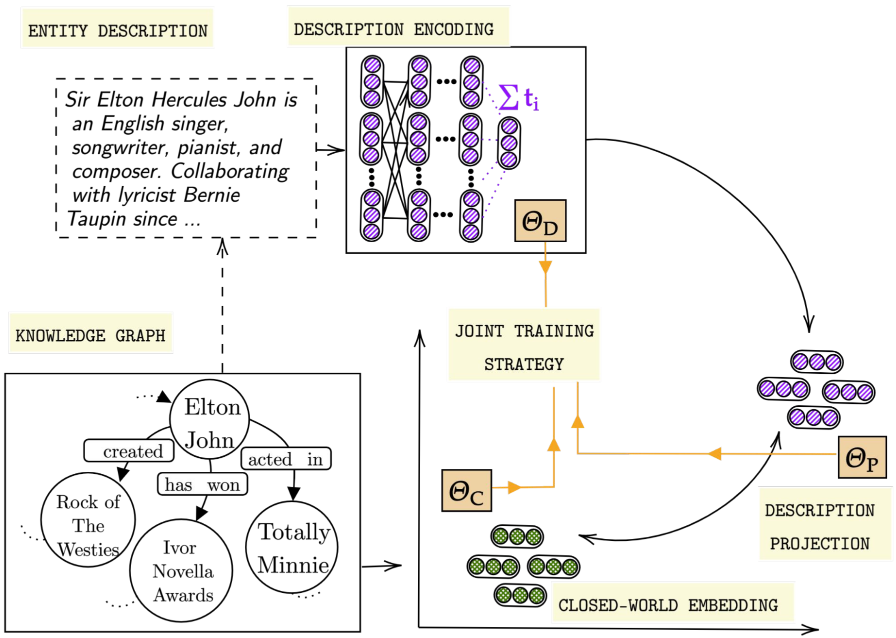
Open-World KG Completion



Essential Properties

- Open-world Embedding Generation
- Efficient Ranking
- Sequence-Size Aware
- Joint Training

FOlK - Framework for Open-World KG embeddings



FOlK(s)

We train our description embedding module, structural embedding module and projection module jointly using a combination of losses

$$\mathscr{L} = \sum_{(h,r,t)\in\mathcal{T}^\circ} log(1 + exp(-I_{h,r,t} Re(\langle \boldsymbol{h}, \boldsymbol{r}, \overline{\boldsymbol{t}} \rangle)))$$

$$\mathscr{L}_{proj} = \sum_{e\in\mathcal{E}} ||\boldsymbol{\phi_P}(\boldsymbol{\phi_D}(\boldsymbol{e})) - \boldsymbol{e}||_2$$

$$\mathscr{L}_X(h,r,t) = \mathscr{L}(\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{\phi_P}(\boldsymbol{\phi_D}(\boldsymbol{t}))) + \mathscr{L}(\boldsymbol{\phi_P}(\boldsymbol{\phi_D}(\boldsymbol{h})), \boldsymbol{r}, \boldsymbol{t})$$

$$\mathscr{L}_{\texttt{FOlK}} = \overbrace{\mathscr{L} + \lambda \mathscr{L}_{reg}}^{\text{closed-world completion}} + \overbrace{\alpha \mathscr{L}_{proj} + \beta \mathscr{L}_X}^{\text{open-world alignment}}$$

ROBERT BOSCH CENTRE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IIT MADRAS

FOlK(l)

Significant performance benefit on using contextual embedding models with long descriptions. We train the three modules phase-wise.

Algorithm 1: Algorithm for FOlK(l)

Input: Triplets: $\mathcal{T}_{train}$, $\mathcal{T}_{valid}$, $\mathcal{T}_{valid}^{open}$
Output: $\boldsymbol{\Theta_S}, \boldsymbol{\Theta_{D_l}} \boldsymbol{\Theta_P}$
Initialise $\boldsymbol{\Theta_{D_l}}$ on RoBERTa's pre-training tasks
Initialise $\boldsymbol{\Theta_S}$ and $\boldsymbol{\Theta_P}$
while *MRR no longer improves on* $\mathcal{T}_{valid}$ **do**
  Train $\boldsymbol{\Theta_S}$ by optimising $\mathcal{L} + \lambda \mathcal{L}_{reg}$;
end
$i \leftarrow 0$
while *MRR no longer improves on* $\mathcal{T}_{valid}^{open}$ **do**
  Phase 1: Freeze $\boldsymbol{\Theta_{D_l}}$ and $\boldsymbol{\Theta_P}$
  if $i \neq 0$ **then**
    Train $\boldsymbol{\Theta_S}$ by optimising $\mathcal{L} + \lambda \mathcal{L}_{reg} + \alpha \mathcal{L}_{proj} + \beta \mathcal{L}_X$; `// Until MRR no longer improves on` $\mathcal{T}_{valid}$
  end
  Phase 2: Freeze Θ_S
  Train $\boldsymbol{\Theta_{D_l}}$ and $\boldsymbol{\Theta_P}$ by optimising $\alpha \mathcal{L}_{proj} + \beta \mathcal{L}_X$; `// //Until MRR no longer improves on` $\mathcal{T}_{valid}^{open}$
  $i \leftarrow i + 1$;
end

ROBERT BOSCH CENTRE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IIT MADRAS

Results - Open-World KG Completion

We outperform existing open-world KG embedding models across the board. Average MRR improvement of 35%

	YAGO3-10-Open				WN18RR-Open				FB15k-237-OWE(L)			
Model	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
JointE	5.1	1.8	4.5	11.0	8.2	4.5	8.2	16.0	10.3	5.1	10.9	20.0
DKRL-CNN	2.6	1.5	2.2	4.1	2.5	1.1	2.4	5.1	19.9	13.9	21.7	32.1
DKRL-CBOW	2.7	1.6	2.4	4.2	2.4	1.0	2.3	4.9	20.7	14.5	22.6	33.4
ConMask	17.3	10.3	18.9	31.3	23.3	10.3	22.7	38.4	21.1	14.0	23.4	34.6
OWE	21.6	14.9	23.3	34.3	21.7	17.3	23.4	29.4	32.4	25.1	35.6	46.0
FOlK(l). iter. 1	25.7	19.0	27.5	38.9	35.6	30.9	37.9	45.5	42.4	33.6	45.7	57.2
FOlK(l). iter. 2	**26.5**	**19.5**	**28.0**	**40.0**	**40.3**	**32.2**	**40.8**	**50.0**	**43.6**	**34.8**	**47.6**	**59.8**

ROBERT BOSCH CENTRE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IIT MADRAS

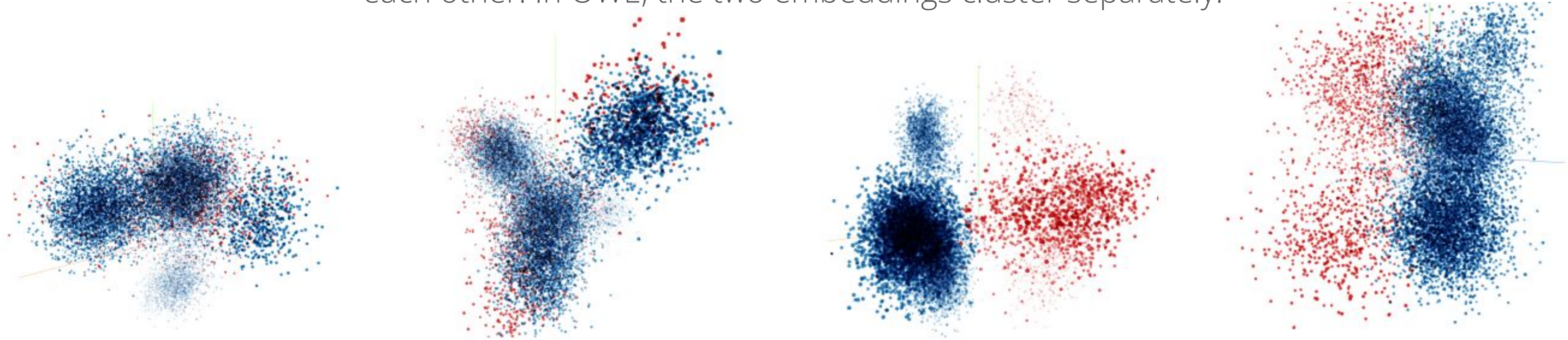
Open-World KG Completion - Short Descriptions

FB15k-237-OWE has an average description length of 4.9. We obtain an MRR improvement of 11%

Model	FB15k-237-OWE			
	MRR	H@1	H@3	H@10
JointE	6.7	2.5	7.0	14.2
DKRL-CNN	19.0	13.0	21.2	31.0
DKRL-CBOW	19.3	13.1	21.5	31.9
ConMask	9.1	3.7	9.5	20.5
OWE	35.2	27.8	38.6	49.1
FOlK(l) iter. 1	38.8	29.9	42.6	54.5
FOlK(l) iter. 2	**39.1**	**32.1**	42.5	52.1
FOlK(s)	**39.1**	30.3	**43.0**	**56.1**

Geometric properties

Visualization of embeddings using PCA. The closed-world and projected embeddings are differentiated by red and blue dots respectively. FOIK embeddings are indistinguishable from each other. In OWE, the two embeddings cluster separately.



(a) FO1K(s) on FB15k-237-OWE

(b) FO1K(l) on FB15k-237-OWE(L)

(c) OWE on FB15k-237-OWE

(d) OWE on FB15k-237-OWE(L)

Geometric properties

Eigenvalue Similarity is a measure of the geometric similarity between the two spaces. Hubness is the minimum % of induced embeddings that are the nearest neighbours to at least N% of structural embeddings. We use ↑ to indicate that a higher value is better and vice-versa

	FB15k-237-OWE		FB15k-237-OWE(L)	
	Eig. Sim.↓	Hub. 10%↑	Eig. Sim.↓	Hub. 10%↑
OWE (Offline)	998	0.2	34981	0.3
`FOlK` (Joint)	**15.3**	**1.3**	**1.5**	**0.8**

Thank You!

RBC DSAI
ROBERT BOSCH CENTRE FOR DATA SCIENCE AND ARTIFICIAL INTELLIGENCE IIT MADRAS