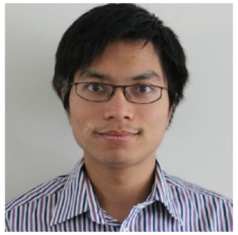
A Convolutional Neural Network-based Model for Knowledge Base Completion

Dat Quoc Nguyen

Joint work with:

Dai Quoc Nguyen, Tu Dinh Nguyen and Dinh Phung



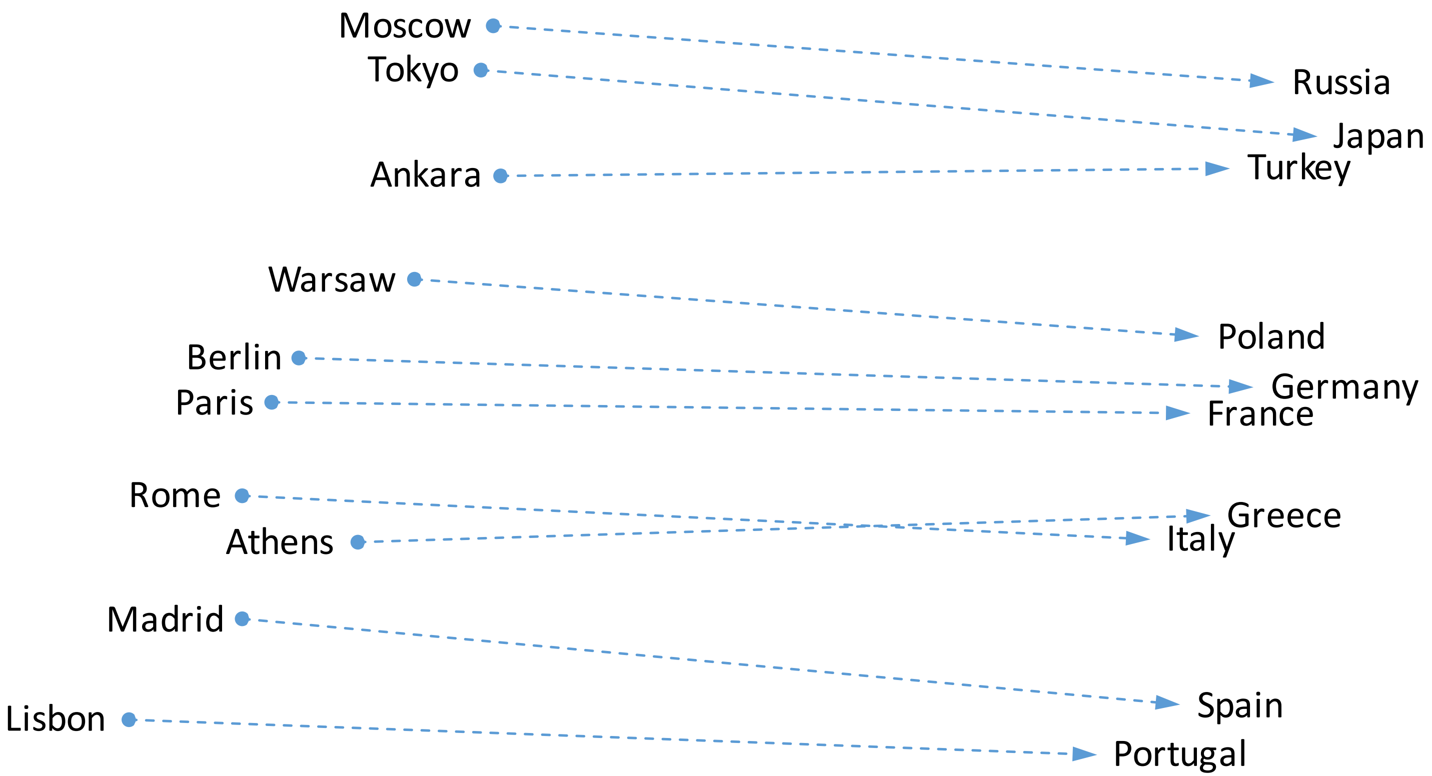
April 16, 2018

Introduction

- Word vectors learned from a large corpus can model relational similarities or linguistic regularities between pairs of words as translations in the projected vector space:

$$\boldsymbol{v}_{king} - \boldsymbol{v}_{man} \approx \boldsymbol{v}_{queen} - \boldsymbol{v}_{woman}$$

A “royal” relationship between “king” and “man”, and between “queen” and “woman”



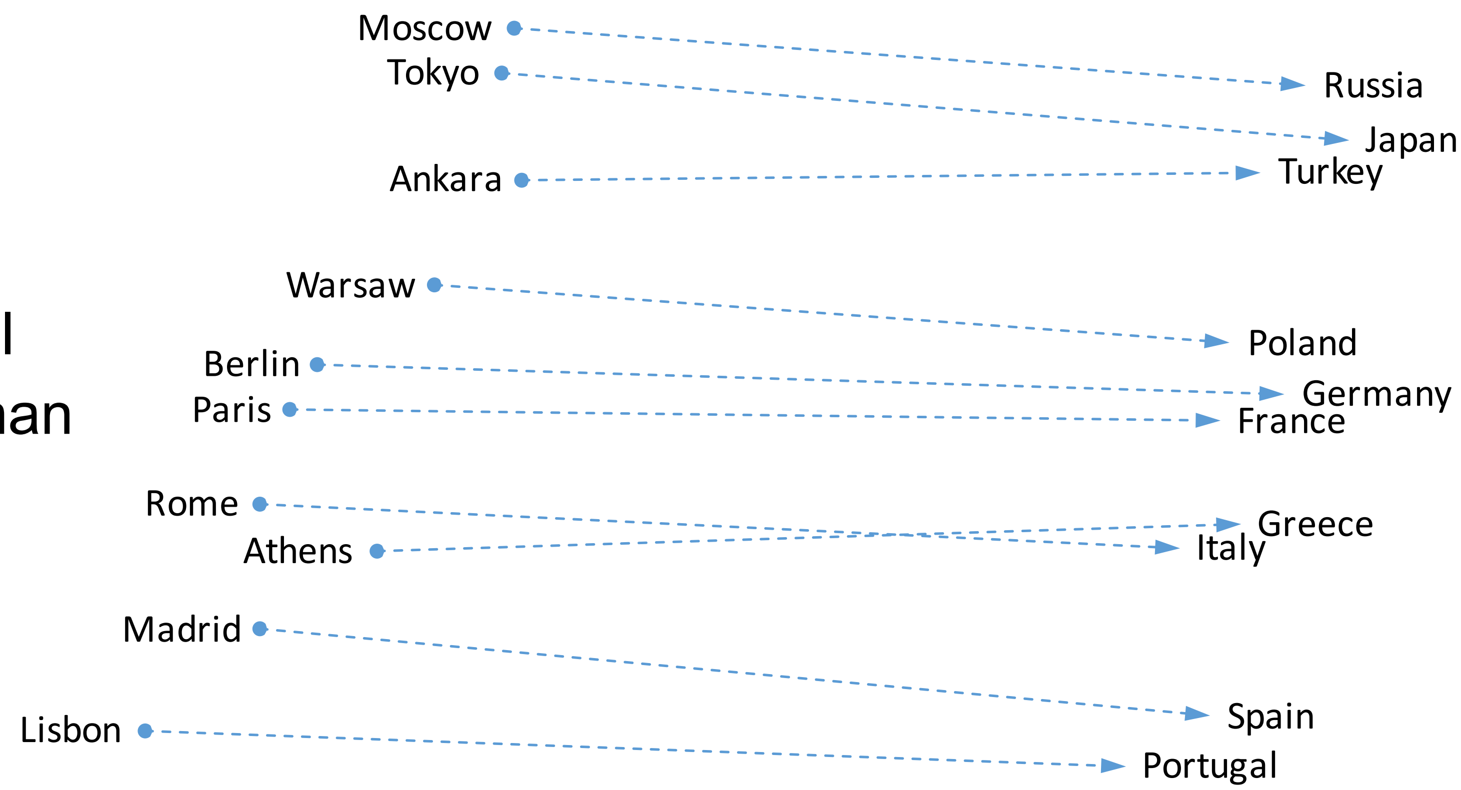
$$\boldsymbol{v}_{Japan} - \boldsymbol{v}_{Tokyo} \approx \boldsymbol{v}_{Germany} - \boldsymbol{v}_{Berlin}$$

$$\boldsymbol{v}_{Germany} - \boldsymbol{v}_{Berlin} \approx \boldsymbol{v}_{Italy} - \boldsymbol{v}_{Rome}$$

$$\boldsymbol{v}_{Italy} - \boldsymbol{v}_{Rome} \approx \boldsymbol{v}_{Portugal} - \boldsymbol{v}_{Lisbon}$$

Introduction

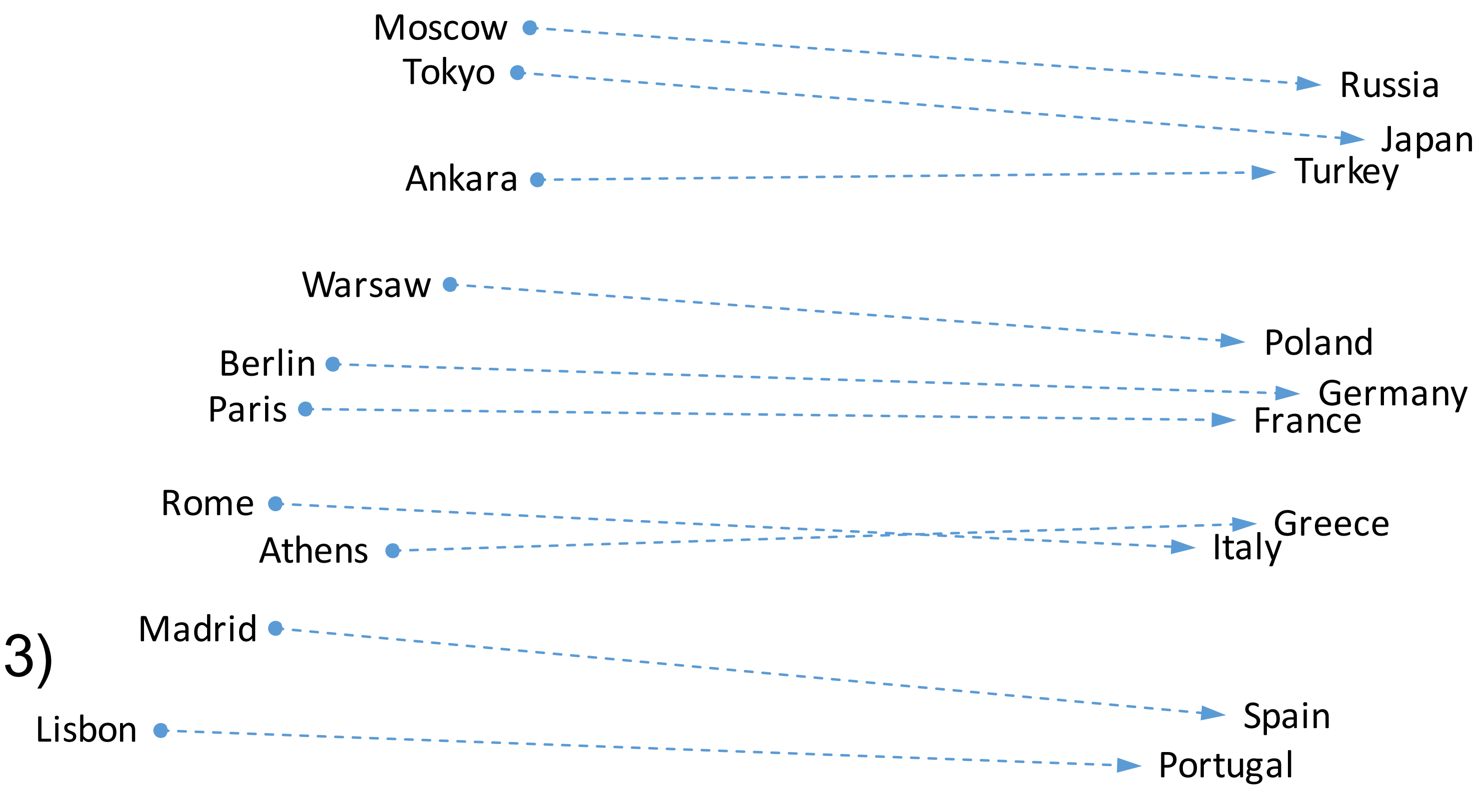
- Let consider the country and capital pairs to be pairs of entities rather than word types
    - We represent country and capital entities by low-dimensional and dense vectors
    - The relational similarity between word pairs is presumably to capture a **"is capital of"** relationship between country and capital entities
    - We also represent this relationship by a translation vector in the entity vector space



$$\boldsymbol{v}_{Tokyo} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Japan} \approx \mathbf{0}$$
$$\boldsymbol{v}_{Berlin} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Germany} \approx \mathbf{0}$$
$$\boldsymbol{v}_{Rome} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Italy} \approx \mathbf{0}$$
$$\boldsymbol{v}_{Lisbon} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Portugal} \approx \mathbf{0}$$

Introduction

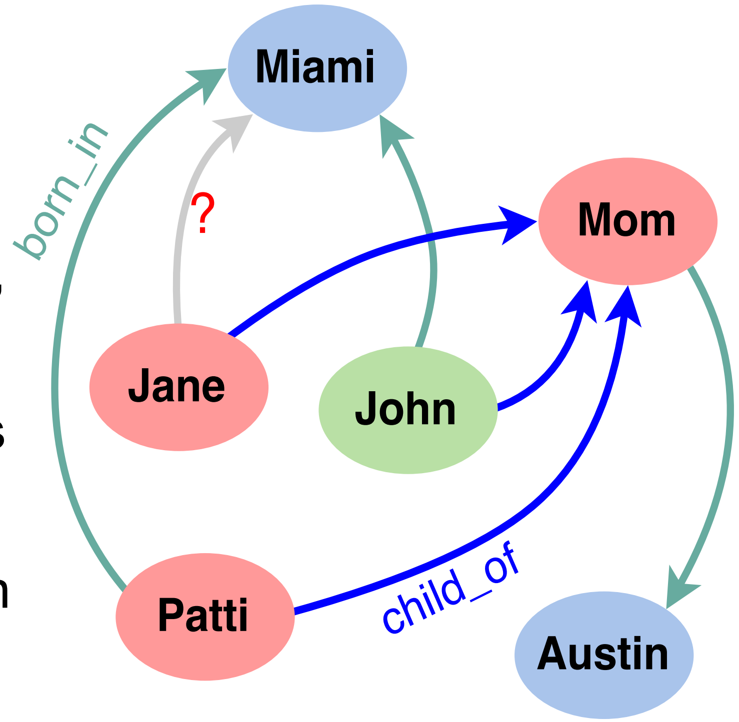
- This intuition inspired the TransE model—a well-known embedding model for **KB completion** or link prediction in KBs (Bordes et al., 2013)



$$\begin{aligned}
v_{Tokyo} + v_{is_capital_of} - v_{Japan} &\approx \mathbf{0} \\
v_{Berlin} + v_{is_capital_of} - v_{Germany} &\approx \mathbf{0} \\
v_{Rome} + v_{is_capital_of} - v_{Italy} &\approx \mathbf{0} \\
v_{Lisbon} + v_{is_capital_of} - v_{Portugal} &\approx \mathbf{0}
\end{aligned}$$

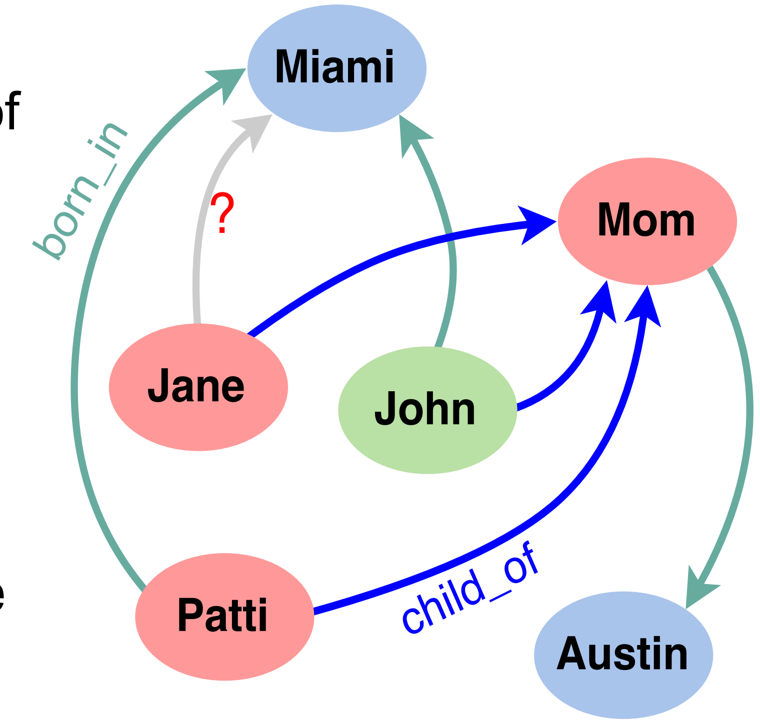
Introduction

- KBs are collections of real-world triples, where each triple or fact (h, r, t) represents some relation **r** between a head entity **h** and a tail entity **t**
  - Entities are real-world things or objects such as persons, places, organizations, music tracks or movies
  - Each relation type defines a certain relationship between entities
  - E.g., the relation type "*child of*" relates person entities with each other, while the relation type "*born in*" relates person entities with place entities
- KBs thus are useful resources for many NLP tasks



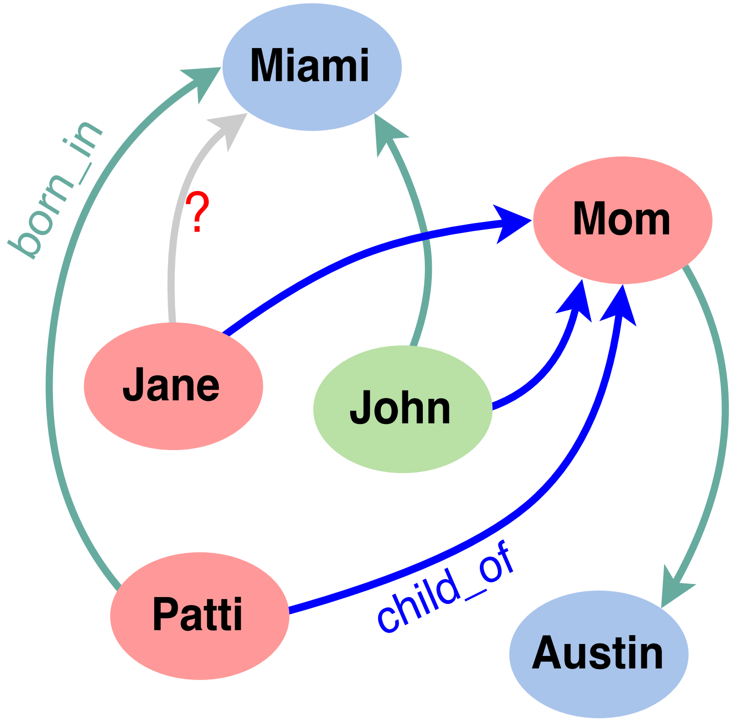
Introduction

- Issue: KBs are far from complete
  - In English DBpedia 2014, 60% of person entities miss a place of birth and 58% of the scientists do not have a fact about what they are known for (Krompaß et al., 2015)
  - In Freebase, 71% of 3 million person entities miss a place of birth, 75% do not have a nationality while 94% have no facts about their parents (West et al., 2014)
- So, a question answering application based on an incomplete KB would not provide a correct answer given a correctly interpreted question
  - It would be impossible to answer the question "where was Jane born ?"



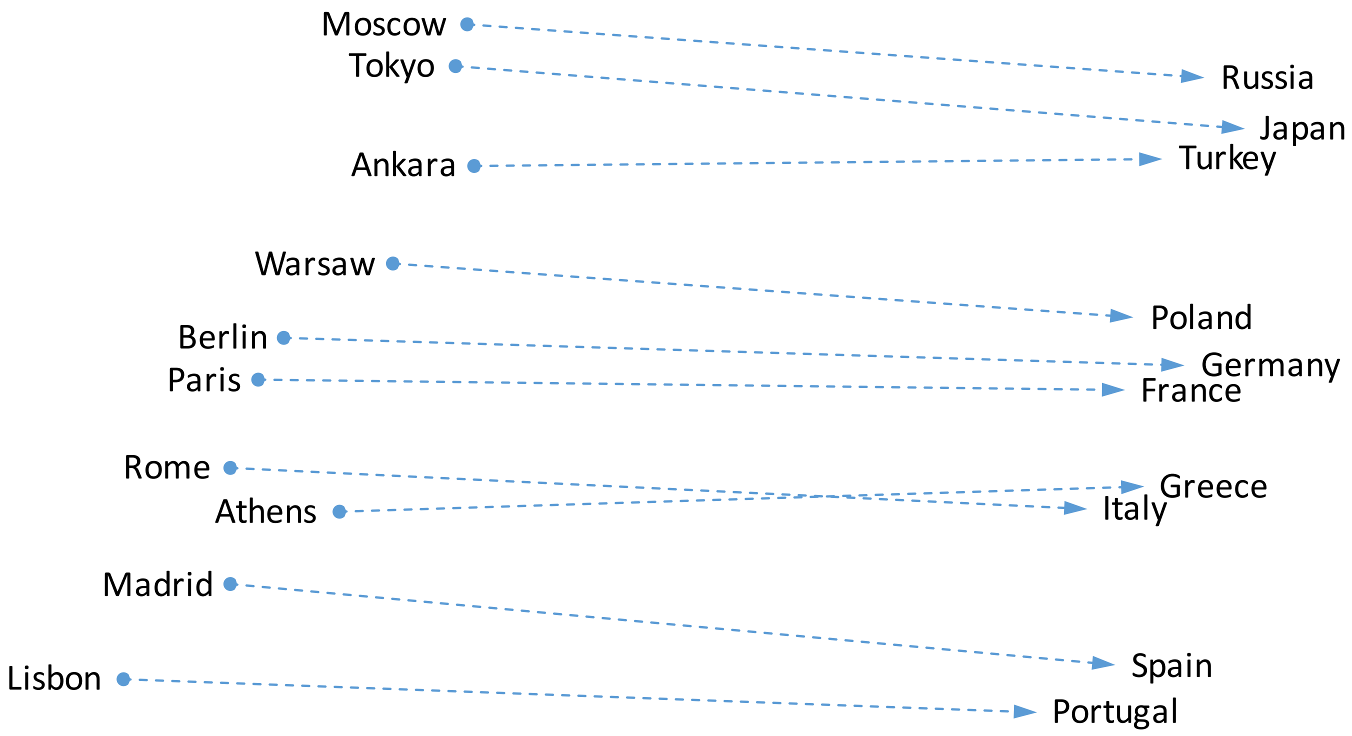
Introduction

- KB completion or Link prediction:
  - Predict whether a relationship/triple not in the KB is likely to be true, i.e., to add new triples by leveraging existing triples in the KB
  - E.g., Predict the missing tail entity in the incomplete triple (Jane, born in, ?) or predict whether the triple (Jane, born in, Miami) is correct or not



Embedding models for KB completion

- Embedding models for KB completion have been proven to give state-of-the-art link prediction performances
  - Entities are represented by latent feature vectors
  - Relation types are represented by latent feature vectors, matrices or third-order tensors



$$\boldsymbol{v}_{Tokyo} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Japan} \approx \mathbf{0}$$

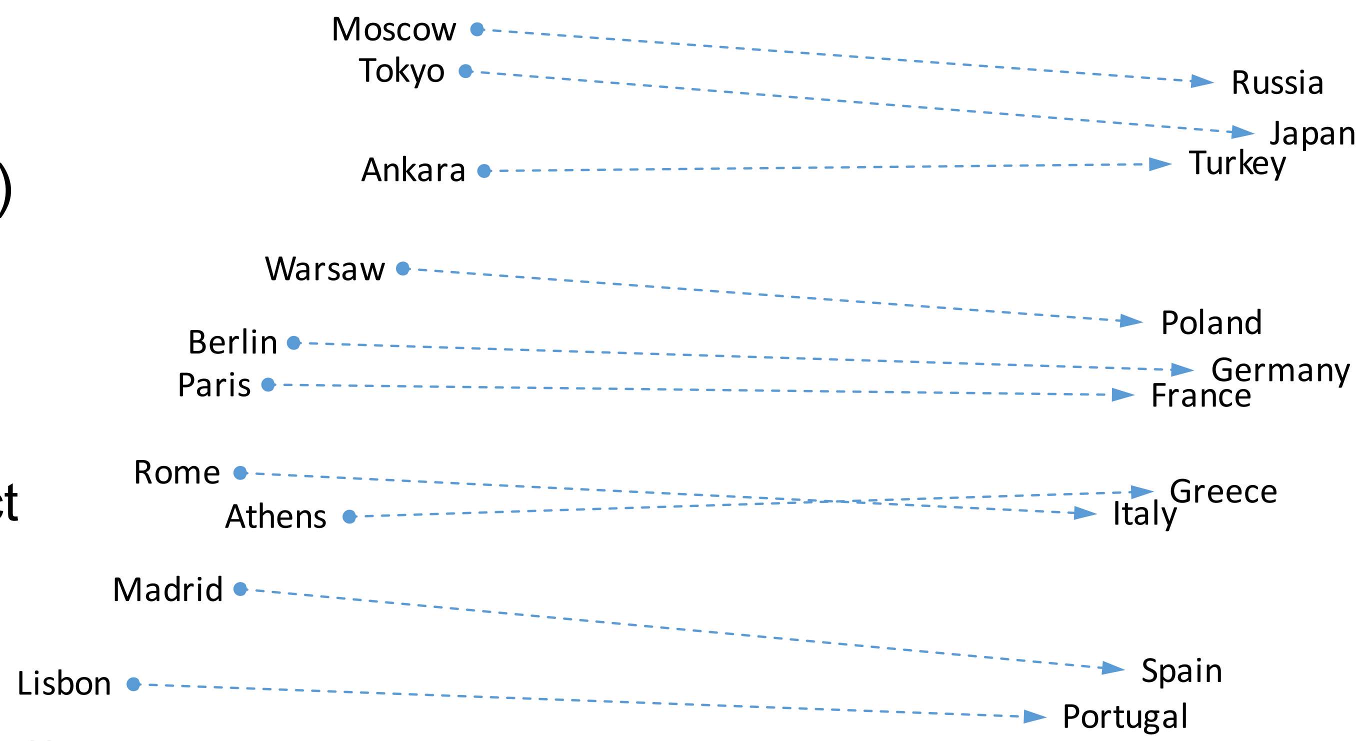
$$\boldsymbol{v}_{Berlin} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Germany} \approx \mathbf{0}$$

$$\boldsymbol{v}_{Rome} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Italy} \approx \mathbf{0}$$

$$\boldsymbol{v}_{Lisbon} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Portugal} \approx \mathbf{0}$$

Embedding models for KB completion

- For each triple (h, r, t), the embedding models define a score function f(h; r; t) of its implausibility:
  - Choose f such that the score f(h, r, t) of a correct triple (h, r, t) is <u>smaller</u> than the score f(h', r', t') of an incorrect triple f(h', r', t')
- TransE:



$$f_{\text{TransE}}(h, r, t) = \|\boldsymbol{v}_h + \boldsymbol{v}_r - \boldsymbol{v}_t\|_{\ell_{1/2}}$$

$$\|\boldsymbol{v}_{Tokyo} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Japan}\|_{\ell_{1/2}} \quad < \quad \|\boldsymbol{v}_{Tokyo} + \boldsymbol{v}_{is\_capital\_of} - \boldsymbol{v}_{Portugal}\|_{\ell_{1/2}}$$

Embedding models for KB completion

Model	Score function $f(h, r, t)$
Unstructured	$\|\boldsymbol{v}_h - \boldsymbol{v}_t\|_{\ell_{1/2}}$
SE	$\|\mathbf{W}_{r,1}\boldsymbol{v}_h - \mathbf{W}_{r,2}\boldsymbol{v}_t\|_{\ell_{1/2}}$; $\mathbf{W}_{r,1}, \mathbf{W}_{r,2} \in \mathbb{R}^{k\times k}$
TransE	$\|\boldsymbol{v}_h + \boldsymbol{v}_r - \boldsymbol{v}_t\|_{\ell_{1/2}}$; $\boldsymbol{v}_r \in \mathbb{R}^k$
STransE	$\|\mathbf{W}_{r,1}\boldsymbol{v}_h + \boldsymbol{v}_r - \mathbf{W}_{r,2}\boldsymbol{v}_t\|_{\ell_{1/2}}$; $\mathbf{W}_{r,1}, \mathbf{W}_{r,2} \in \mathbb{R}^{k\times k}$; $\boldsymbol{v}_r \in \mathbb{R}^k$
DISTMULT	$\boldsymbol{v}_h^\top \mathbf{W}_r \boldsymbol{v}_t$; $\mathbf{W}_r$ is a diagonal matrix $\in \mathbb{R}^{k\times k}$
Bilinear-COMP	$\boldsymbol{v}_h^\top \mathbf{W}_{r_1}\mathbf{W}_{r_2}...\mathbf{W}_{r_m}\boldsymbol{v}_t$; $\mathbf{W}_{r_1}, \mathbf{W}_{r_2}, ..., \mathbf{W}_{r_m} \in \mathbb{R}^{k\times k}$
TransE-COMP	$\|\boldsymbol{v}_h + \boldsymbol{v}_{r_1} + \boldsymbol{v}_{r_2} + ... + \boldsymbol{v}_{r_m} - \boldsymbol{v}_t\|_{\ell_{1/2}}$; $\boldsymbol{v}_{r_1}, \boldsymbol{v}_{r_2}, ..., \boldsymbol{v}_{r_m} \in \mathbb{R}^k$
ConvE	$\boldsymbol{v}_t^\top g\left(\mathsf{vec}\left(g\left(\mathsf{concat}(\overline{\boldsymbol{v}}_h, \overline{\boldsymbol{v}}_r) * \boldsymbol{\Omega}\right)\right)\boldsymbol{W}\right)$; g denotes a non-linear function
Our **ConvKB**	$\mathrm{w}^\top \mathsf{concat}\left(g\left([\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t] * \boldsymbol{\Omega}\right)\right)$; $*$ denotes a convolution operator

Embedding models for KB completion

- A common objective function is the following margin-based function:

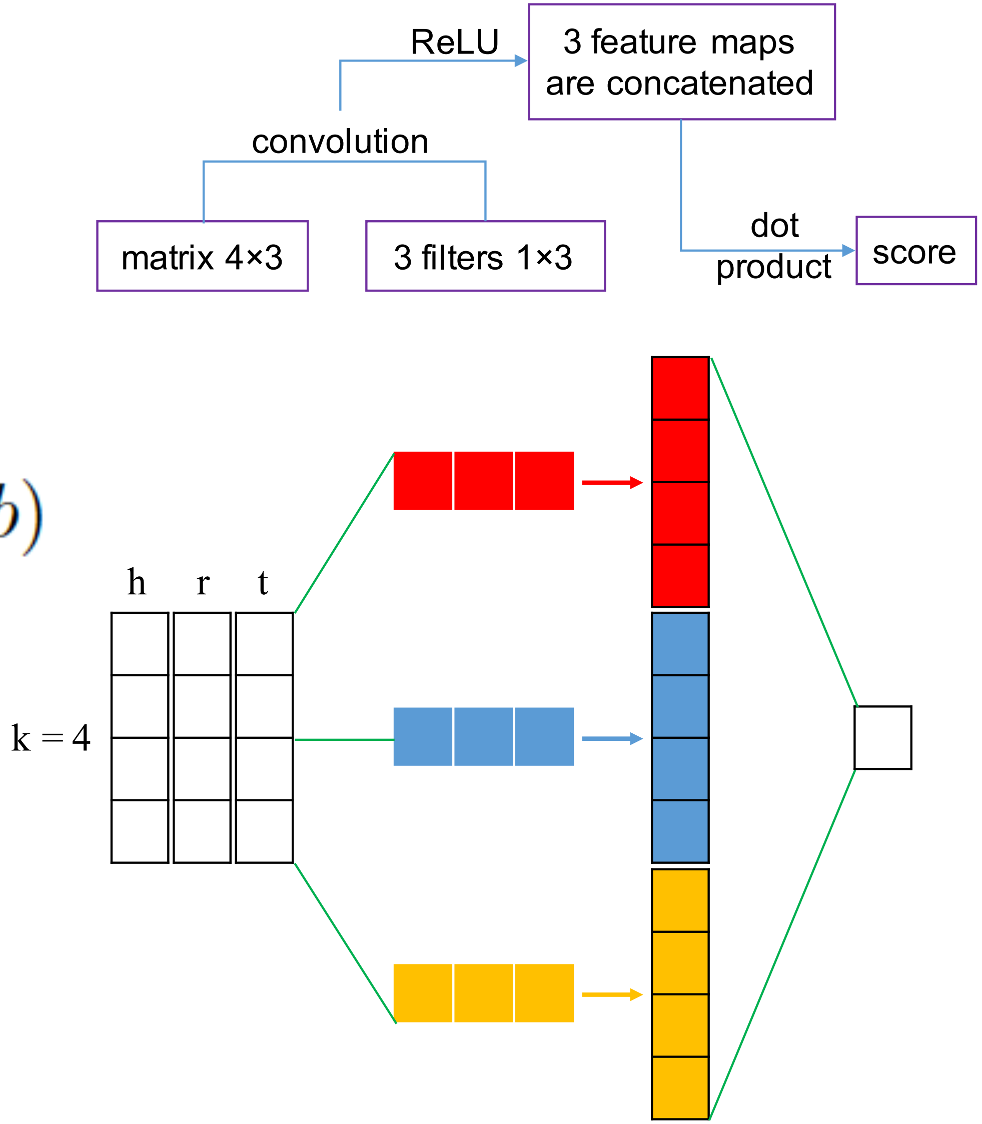
$$\mathcal{L} = \sum_{\substack{(h,r,t)\in\mathcal{G} \\ (h',r,t')\in\mathcal{G}'_{(h,r,t)}}} max(0, \gamma + f(h,r,t) - f(h',r,t'))$$

  - γ is the margin hyper-parameter
  - $\mathcal{G}'_{(h,r,t)}$ is the set of incorrect triples generated by corrupting the correct triple $(h, r, t) \in \mathcal{G}$

A CNN-based model for KB completion

- Each embedding triple $(\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t)$ are viewed as a 3-column matrix $\boldsymbol{A} = [\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t]$
- Each filter $\boldsymbol{\omega} \in \mathbb{R}^{1\times3}$ is repeatedly operated over every row of $\boldsymbol{A}$ to generate a feature map $\boldsymbol{v} = [v_1, v_2, ..., v_k]$ with $v_i = g\left(\boldsymbol{\omega} \cdot \boldsymbol{A}_{i,:} + b\right)$
- Let $\boldsymbol{\Omega}$ and $*$ denote the set of filters and a convolution operator, respectively
- Our ConvKB formally defines a score function as

$$f(h, r, t) = \mathsf{concat}\left(g\left([\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t] * \boldsymbol{\Omega}\right)\right) \cdot \mathbf{w}$$

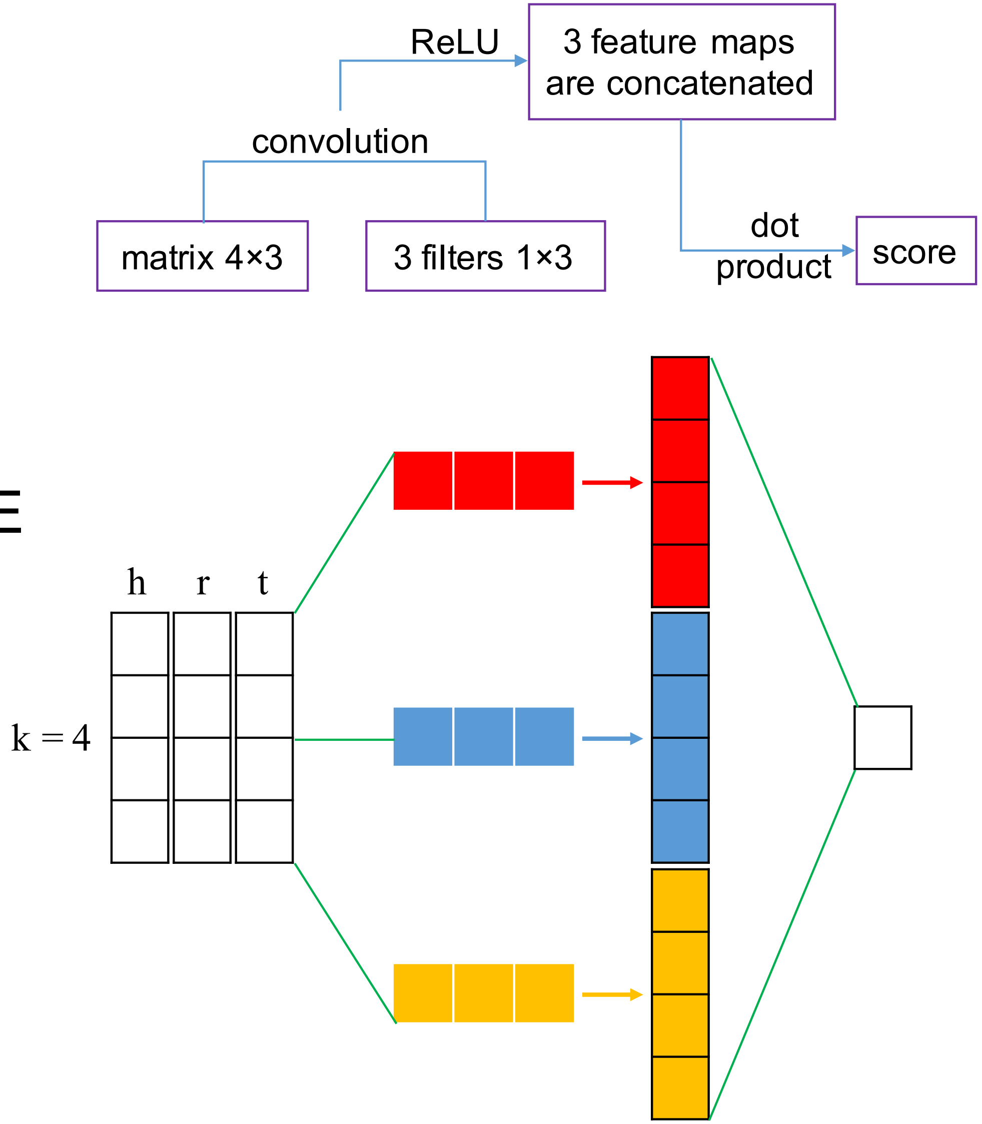


A CNN-based model for KB completion

- ConvKB formally defines a score function as

$$f(h,r,t) = \text{concat}\left(g\left([\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t] * \boldsymbol{\Omega}\right)\right) \cdot \mathbf{w}$$

- If we only use one filter with g be the vector norm and fix $\boldsymbol{\omega} = [1, 1, -1]$ and $\mathbf{w} = \mathbf{1}$ during training, then ConvKB reduces to the plain TransE
  - ConvKB model can be viewed as an extension of TransE



KB completion experiments

Dataset	$\|\mathcal{E}\|$	$\|\mathcal{R}\|$	#Triples in train/valid/test		
FB15k-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134
FB13	75,043	13	316,232	5,908	23,733
WN11	38,696	11	112,581	2,609	10,544

On both FB13 and WN11, each validation and test set also contains the same number of incorrect triples as the number of correct triples

- Link prediction task:
    - Predict h given (?, r, t) or predict t given (h, r, ?) where ? denotes the missing element
    - Corrupt each correct test triple (h, r, t) by replacing either h or t by each of the possible entities
    - Rank these candidates by their implausibility value
    - Metrics: mean rank (MR), mean reciprocal rank (MRR), and Hits@10 (i.e., the proportion of the valid test triples ranking in top 10 predictions)

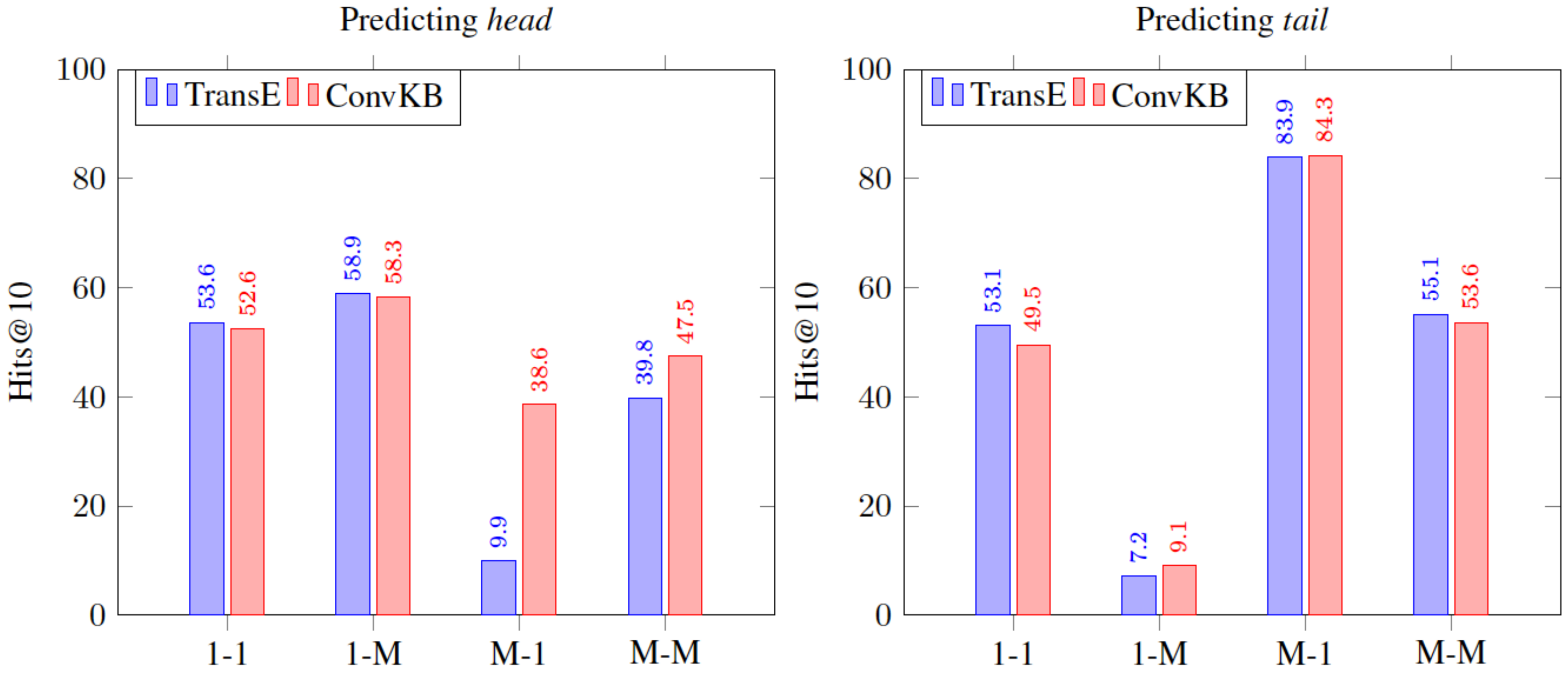
KB completion experiments

- Link prediction results:

Method	WN18RR			FB15k-237		
	MR	MRR	H@10	MR	MRR	H@10
IRN (Shen et al., 2017)	–	–	–	211	–	46.4
KBGAN (Cai and Wang, 2018)	–	0.213	48.1	–	0.278	45.8
DISTMULT (Yang et al., 2015) [⋆]	5110	0.43	49	254	0.241	41.9
ComplEx (Trouillon et al., 2016) [⋆]	5261	0.44	51	339	0.247	42.8
ConvE (Dettmers et al., 2018)	5277	**0.46**	48	246	0.316	49.1
TransE (Bordes et al., 2013) (our results)	3384	0.226	50.1	347	0.294	46.5
Our ConvKB model	**2554**	0.248	**52.5**	257	**0.396**	**51.7**
KB_{LRN} (García-Durán and Niepert, 2017)	–	–	–	**209**	0.309	49.3
R-GCN+ (Schlichtkrull et al., 2017)	–	–	–	–	0.249	41.7
Neural LP (Yang et al., 2017)	–	–	–	–	0.240	36.2
Node+LinkFeat (Toutanova and Chen, 2015)	–	–	–	–	0.293	46.2

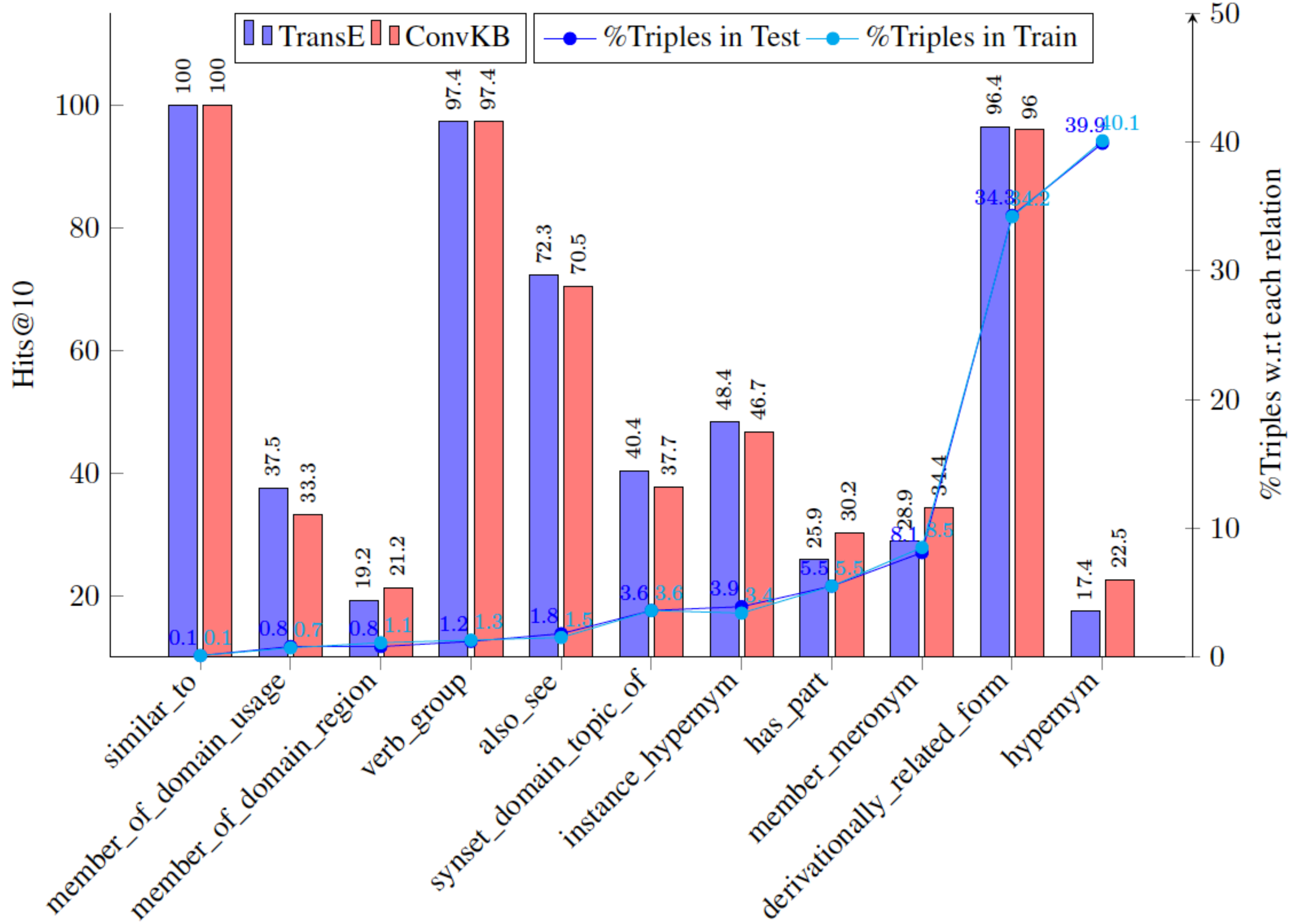
KB completion experiments

- Hits@10 on FB15k-237:



KB completion experiments

- Hits@10 on WN18RR:



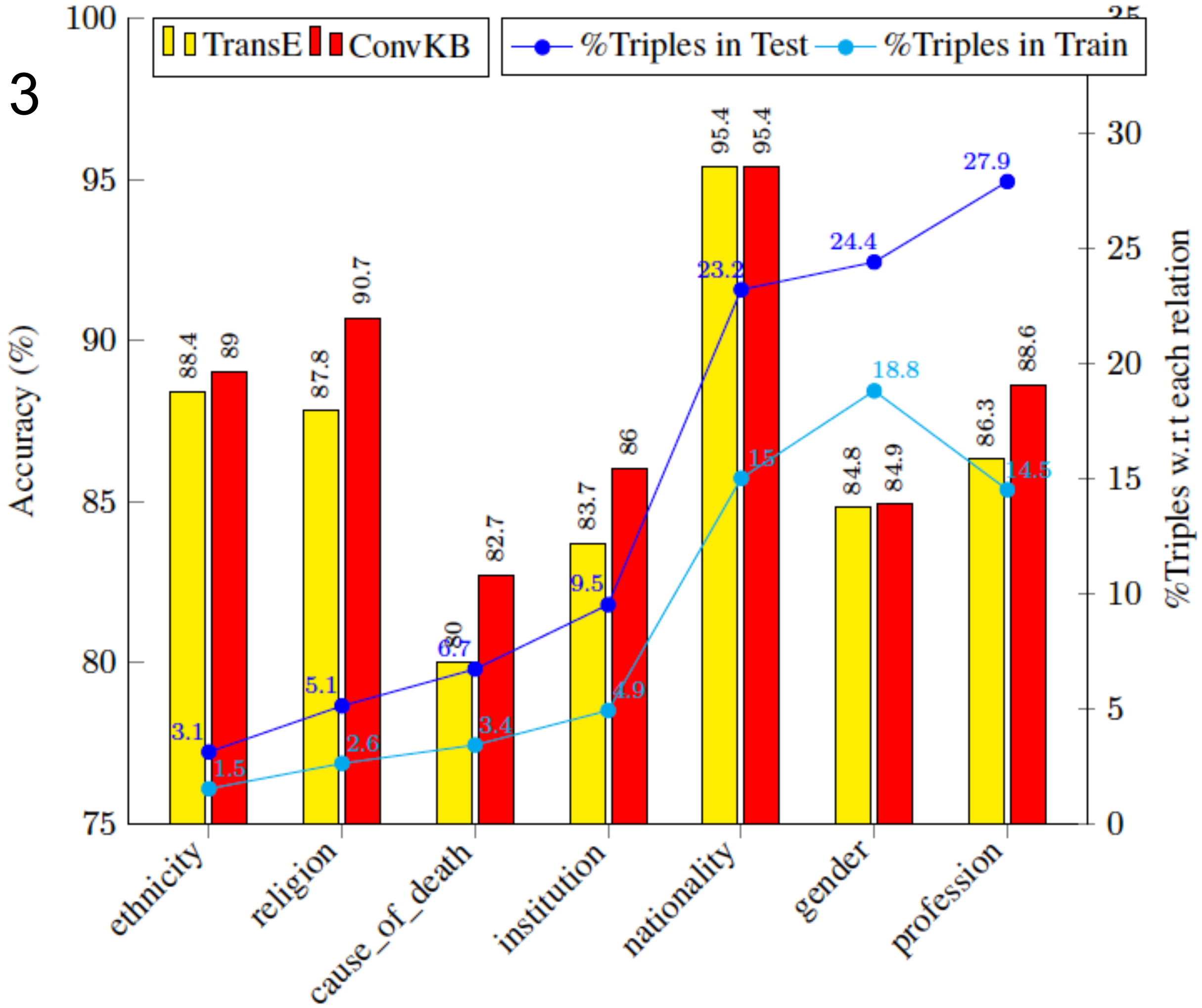
KB completion experiments

- <u>Triple classification task</u>:
  - Predict whether a triple (h, r, t) is correct or not
  - Set a relation-specific threshold θ_r for each relation type r
  - For an unseen test triple (h, r, t), if f(h, r, t) is smaller than θ_r then the triple will be classified as correct, otherwise incorrect
  - Relation-specific thresholds are determined by maximizing the micro-averaged accuracy on the validation set

Method	**WN11**	**FB13**	Avg.
NTN [41]	70.6	87.2	78.9
TransH [53]	78.8	83.3	81.1
TransR [27]	85.9	82.5	84.2
TransD [19]	86.4	**89.1**	<u>87.8</u>
TransR-FT [12]	86.6	82.9	84.8
TranSparse-S [20]	86.4	88.2	87.3
TranSparse-US [20]	86.8	87.5	87.2
ManifoldE [56]	<u>87.5</u>	87.2	87.4
TransG [57]	87.4	87.3	87.4
lppTransD [64]	86.2	88.6	87.4
TransE [5] (our results)	86.5	87.5	87.0
Our ConvKB model	**87.6**	<u>88.8</u>	**88.2**
TransE-NMM [33]	86.8	88.6	87.7
TEKE_H [52]	84.8	84.2	84.5
Bilinear-COMP [16]	77.6	86.1	81.9
TransE-COMP [16]	80.3	87.6	84.0

KB completion experiments

- Triple classification results on FB13



Conclusion

- We have proposed an embedding model ConvKB for knowledge base completion
- ConvKB outperforms previous state-of-the-art models:
  - On two benchmark link prediction datasets WN18RR and FB15k-237
  - On two benchmark triple classification datasets WN11 and FB13
- Code: https://github.com/daiquocnguyen/ConvKB
- References:

  - Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen and Dinh Phung. 2018. A Novel Embedding Model for Knowledge Base Completion Based on Convolutional Neural Network. In *Proceedings of NAACL-HLT 2018*, to appear. https://arxiv.org/abs/1712.02121
  - Dat Quoc Nguyen. 2017. An overview of embedding models of entities and relationships for knowledge base completion. https://arxiv.org/abs/1703.08098