

Multilingual Knowledge Graph Embeddings for Cross-lingual Knowledge Alignment

Muhao Chen¹, Yingtao Tian², Mohan Yang¹, and Carlo Zaniolo¹
University of California, Los Angeles¹
Stony Brook University²

Outline

- Background
- *MTransE*—A multilingual knowledge graph embedding model
- Evaluation
- Open Challenges and Future Work

Knowledge Graphs

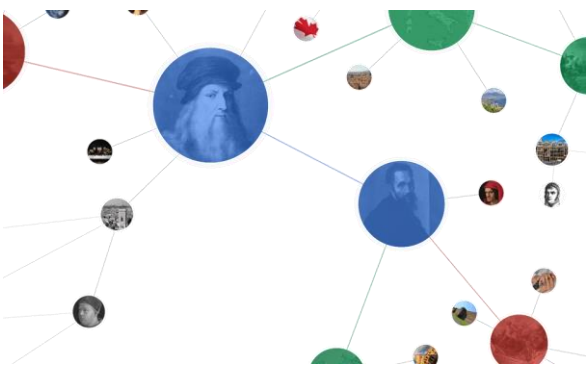
- Symbolic representation of entities and relations

(California, capital city, Sacramento)

Monolingual knowledge:
triples (relation facts of
entities)

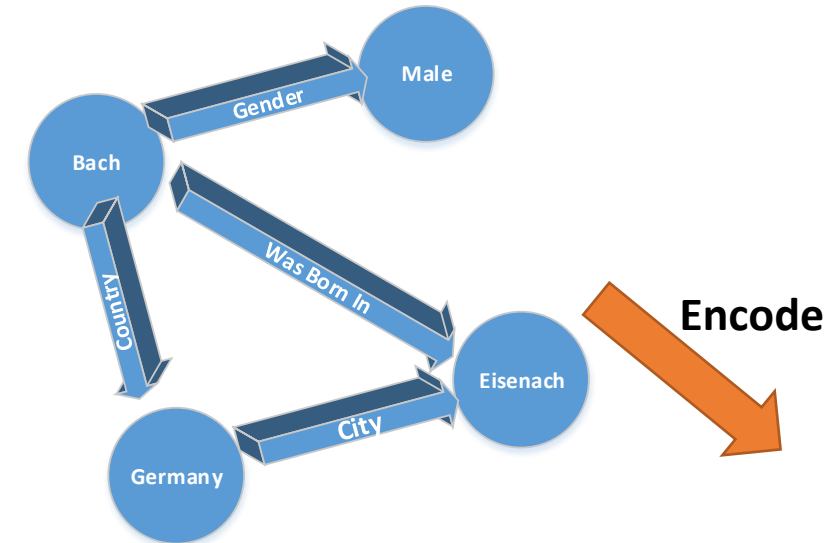
(カリフォルニア, 首都, サクラメント)

Cross-lingual knowledge:
alignment of monolingual
knowledge across
languages



Knowledge Graph Embeddings

- Encode entities as vectors



Knowledge Graph

Paris (0.036, -0.12, ..., 0.323)
 capital (0.102, 0.671, ..., -0.101)
 France (0.138, 0.551, ..., 0.222)
 ...

Embeddings

Semantic similarity of entities

Capture

Enable

Applications

Relational inferences as vector algebra

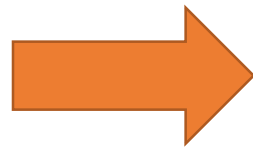
- *France – Paris \approx capital*
- *US – USD \approx currency*
- *Bach – German \approx nationality*
- ...

- KG Completion
- Relation extraction from text
- Question answering

Current KG Embedding Approaches

- Focused on embedding **monolingual** triples (h, r, t)

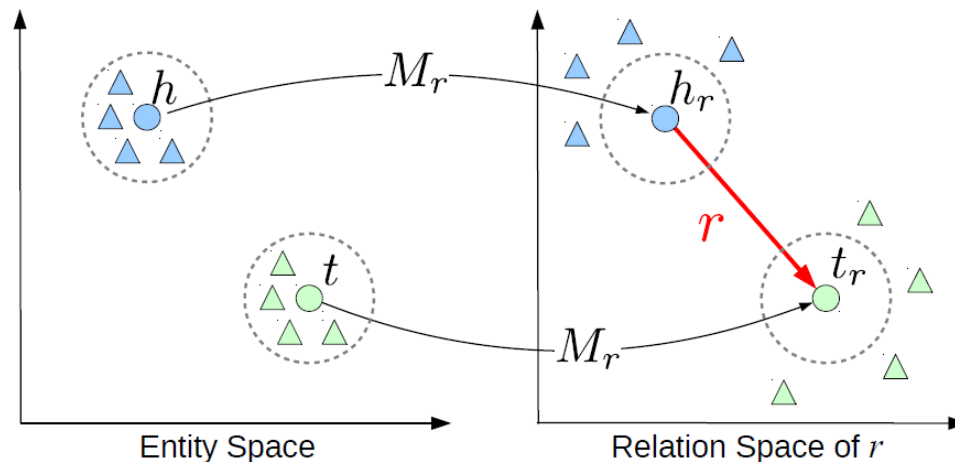
TransE: $h+r \approx t$



Later approaches

- TransH [Wang et al. 2014]
- TransR [Lin et al. 2015]
- TransD [Ji et al. 2015]
- HoIE [Nickle et al. 2016]
- ComplEx [Trouillon et al. 2016]
- ...

Embedding of **monolingual** knowledge seems to be well-addressed.



What about **cross-lingual** knowledge?

Emerging challenge

- Existing works do not characterize cross-lingual knowledge
 - Entity inter-lingual links (ILLs): (ambulance --- Krankenwagen)
 - Triple-wise alignment (TWA): ((State of California, capital city, Sacramento) --- (カリフォルニア, 首都, サクラメント))
 - Many KGs store such knowledge

Why important?

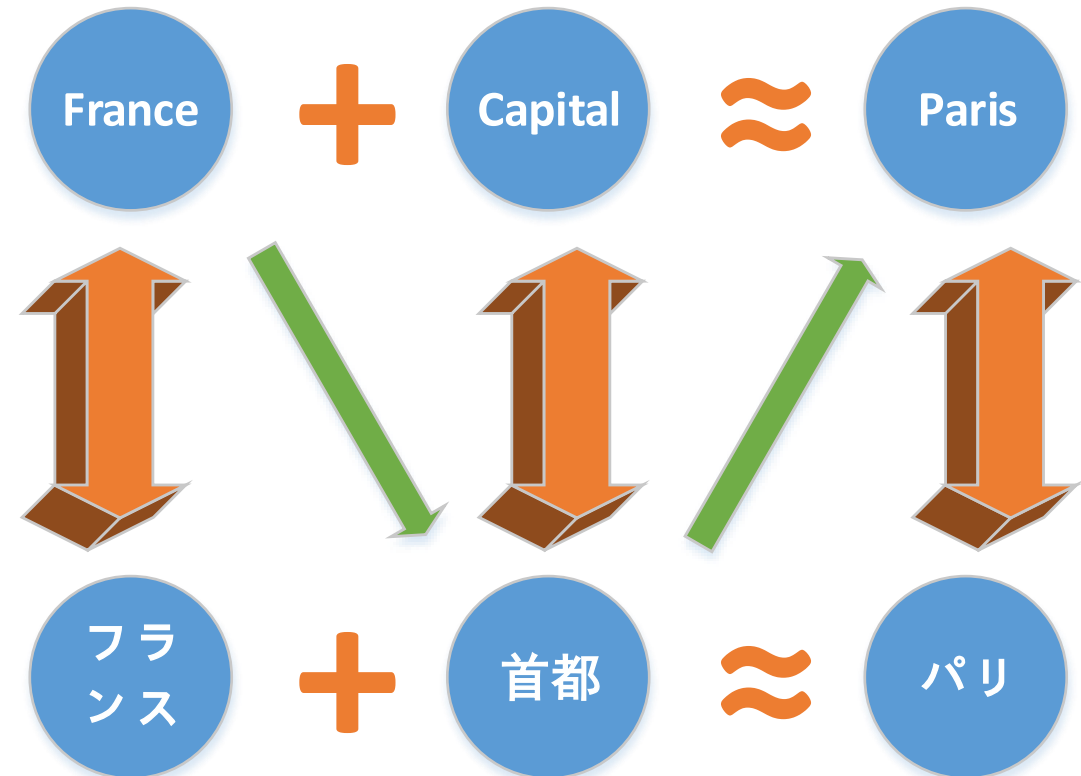
- Enables multilingual semantic representations
- Benefits cross-lingual NLP
 - Knowledge alignment
 - Machine translation
 - Cross-lingual Q&A
 - ...

Difficult to characterize:

- **Fewer samples:** Cross-lingual knowledge currently accounts for a small portion of each KB
- **Larger domains:** Cross-lingual knowledge applies on the entire spaces of involved languages
- **Incoherence:** Language-specific versions of KG are usually incoherent
- **Heterogeneity:** Applies to both entities and monolingual relations with inconsistent vocabularies

What does MTransE use and enable?

- Corpora: (partially-aligned) multilingual KGs
- Enabling: inferable embeddings of multilingual semantics
- Can be applied to:
 - Knowledge alignment
 - Cross-lingual Q&A
 - Multilingual chat-bots
 - ...



MTransE Model Components

- *Knowledge model*

$$S_K = \sum_{L \in \{L_i, L_j\}} \sum_{T \in G_L} \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$$

- *Alignment model*

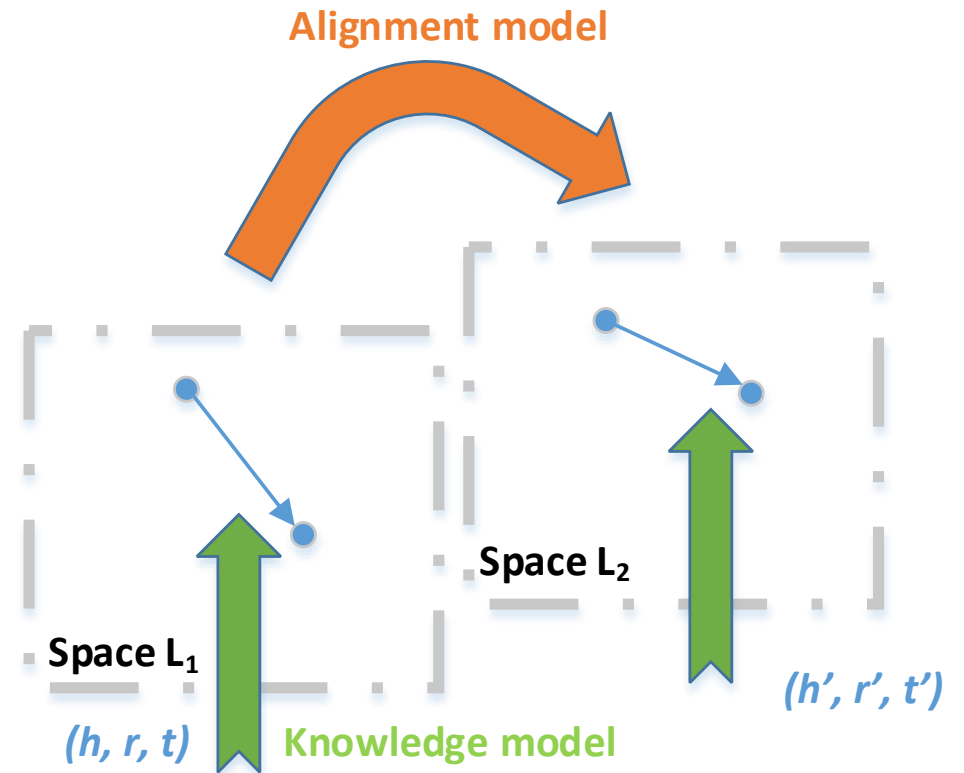
$$S_A = \sum_{(T, T') \in \delta(L_i, L_j)} S_a(T, T')$$

$(T, T') \in \delta(L_i, L_j)$

All aligned triples

- *Objective of learning*

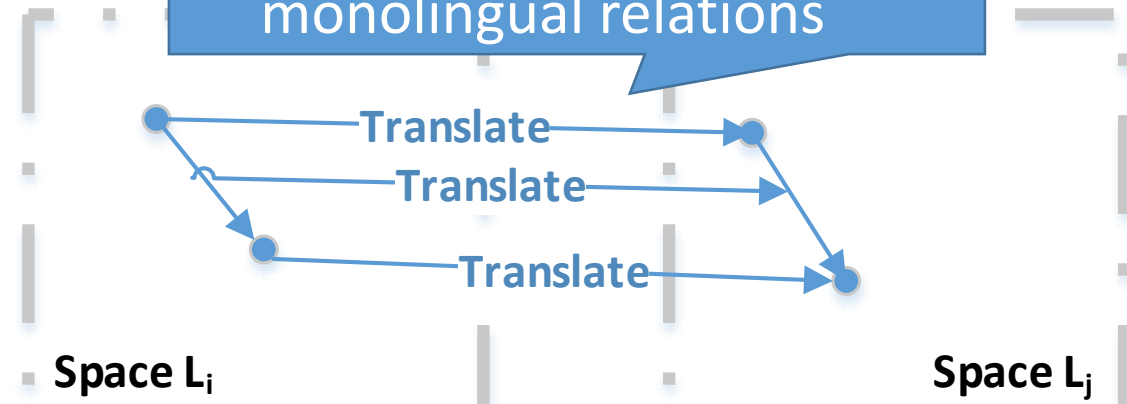
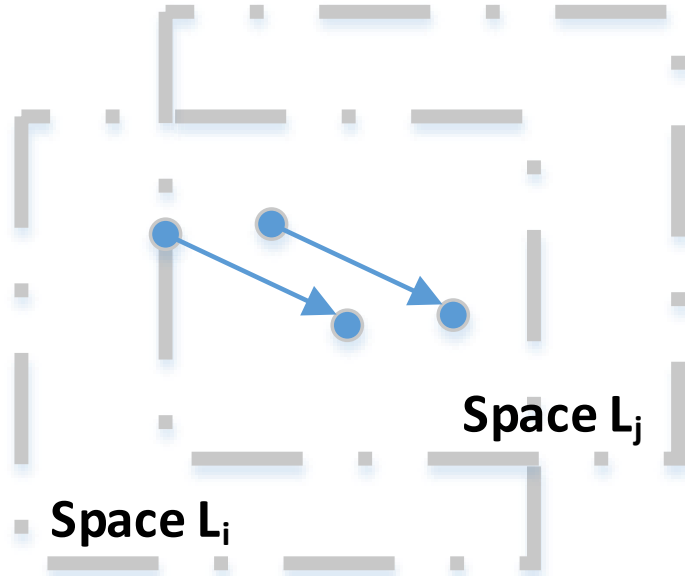
– Minimizing $J(\theta) = S_K + \alpha S_A$



Different alignment techniques

Translation vectors

- Encoding cross-lingual transitions just like monolingual relations

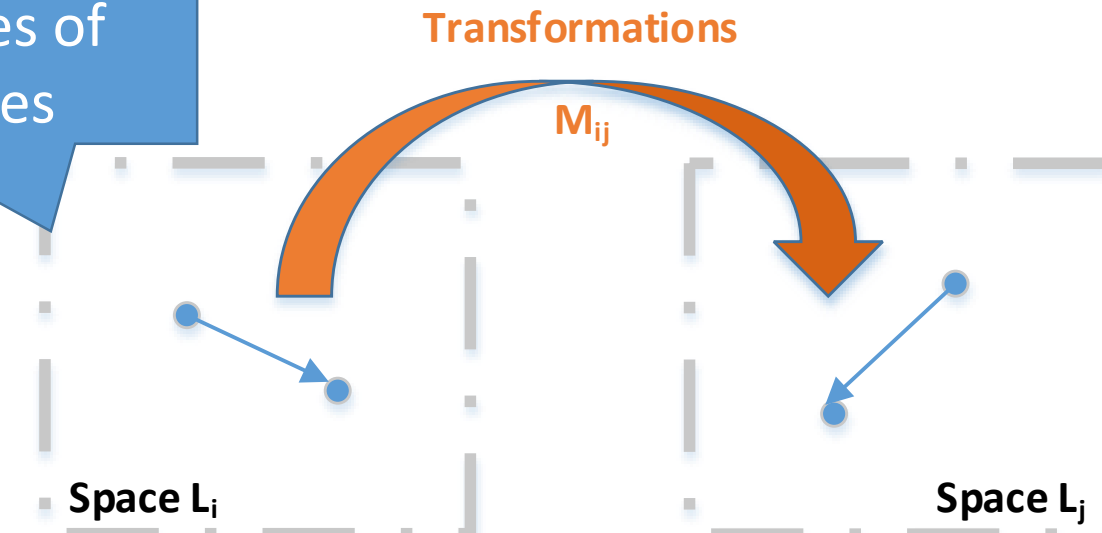


Linear Transformations

- Transformations across embedding spaces of different languages

Axis calibration

- Cross-lingual counterparts have close embeddings



Alignment Scores and Five Model Variants

- Var_i combines the i^{th} alignment model with the knowledge model

Variant	Alignment Score	Remark	
Var_1	$S_{a_1} = \ \mathbf{h} - \mathbf{h}'\ + \ \mathbf{t} - \mathbf{t}'\ $		} Axis Calibration
Var_2	$S_{a_2} = \ \mathbf{h} - \mathbf{h}'\ + \ \mathbf{r} - \mathbf{r}'\ + \ \mathbf{t} - \mathbf{t}'\ $		
Var_3	$S_{a_3} = \ \mathbf{h} + \mathbf{v}_{ij}^e - \mathbf{h}'\ + \ \mathbf{r} + \mathbf{v}_{ij}^r - \mathbf{r}'\ + \ \mathbf{t} + \mathbf{v}_{ij}^e - \mathbf{t}'\ $	$\mathbf{v}_{ij}^e = -\mathbf{v}_{ji}^e, \mathbf{v}_{ij}^r = -\mathbf{v}_{ji}^r$	→ Translation Vector
Var_4	$S_{a_4} = \ \mathbf{M}_{ij}^e \mathbf{h} - \mathbf{h}'\ + \ \mathbf{M}_{ij}^e \mathbf{t} - \mathbf{t}'\ $	$\mathbf{M}_{ij}^e \in \mathbb{R}^{k \times k}, \mathbf{M}_{ij}^r \in \mathbb{R}^{k \times k}$	} Linear Transforms
Var_5	$S_{a_5} = \ \mathbf{M}_{ij}^e \mathbf{h} - \mathbf{h}'\ + \ \mathbf{M}_{ij}^r \mathbf{r} - \mathbf{r}'\ + \ \mathbf{M}_{ij}^e \mathbf{t} - \mathbf{t}'\ $		

Experimental Evaluation

- Cross-lingual knowledge alignment tasks
 - Entity Matching
 - Triple-wise Alignment (TWA) Verification
- Monolingual relation extraction task
- Trilingual data sets
 - Wiki-based (WK3I-15k, WK3I-120k)
 - ConceptNet-based (CN3I)
- Baselines
 - LM [Mikolov et al. 2013] + Knowledge model
 - CCA [Faruqui et al. 2014] + Knowledge mode
 - OT [Xing et al. 2015] + Knowledge models

Table 4.1: Statistics of the WK3I data sets.

Data set	#En triples	#Fr triples	#De triples	#Aligned triples
WK3I-15k	203,502	170,605	145,616	En-Fr:16,470 En-De:37,170
WK3I-120k	1,376,011	767,750	391,108	En-Fr:124,433 En-De:69,413
CN3I	47,696	18,624	25,560	En-Fr:3,668 En-De:8,588

Table 4.2: Number of extra entity inter-lingual links (ILLs).

Data Set	En-Fr	Fr-En	En-De	De-En
WK3I-15k	3,733	3,815	1,840	1,610
WK3I-120k	42,413	41,513	7,567	5,921
CN3I	2,154	2,146	3,485	3,813

These three data sets are available at <https://github.com/muhaochen/MTransE>

Entity Matching

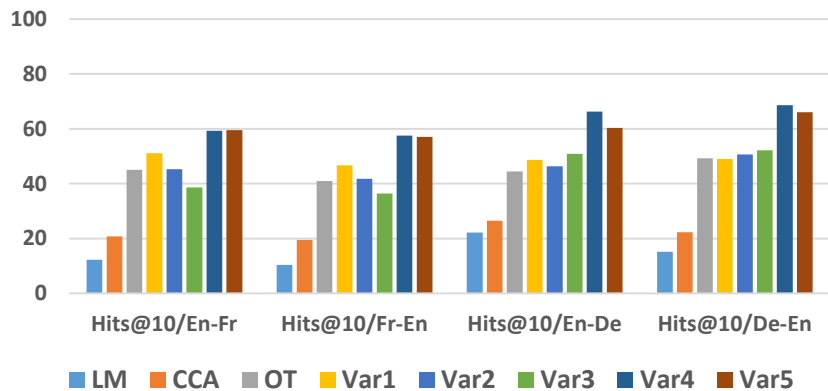
What is the German entity for the English entity “Regulation of Property”?

- Evaluation protocol
 - For each (e, e') , rank e' in the neighborhood of $\tau(e)$
- Training sets
 - Pairs of language-specific graphs and corresponding alignment sets
- Test data
 - Entity Inter-lingual links $\{(e, e')\}$ (Unidirectional)

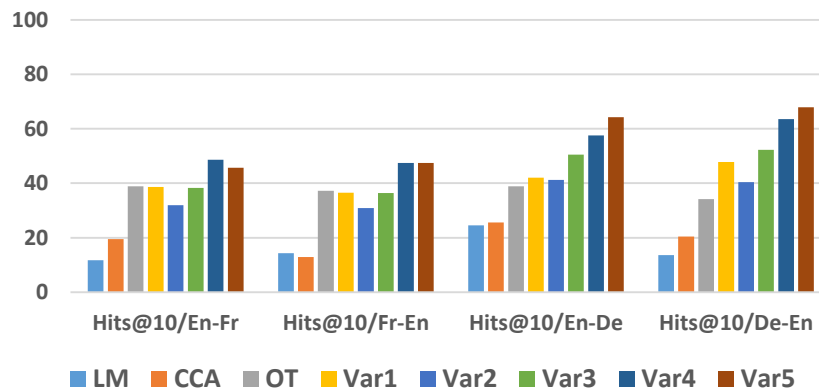
Entity Matching



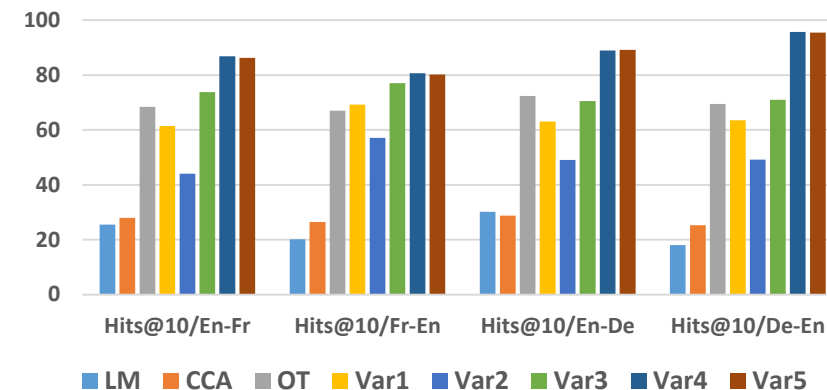
Hits@10 on WK3I-15k



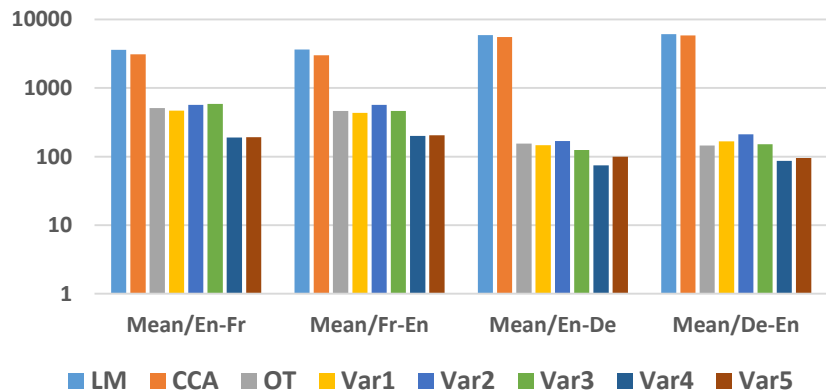
Hits@10 on WK3I-120k



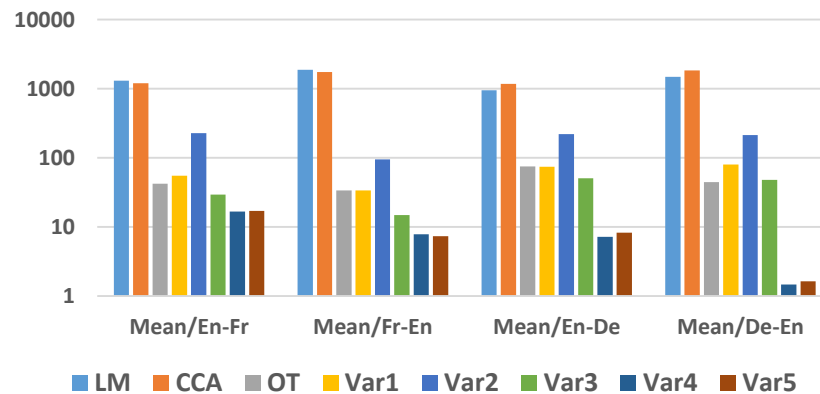
Hits@10 on CN3I



Mean on WK3I-15k



Mean on CN3I

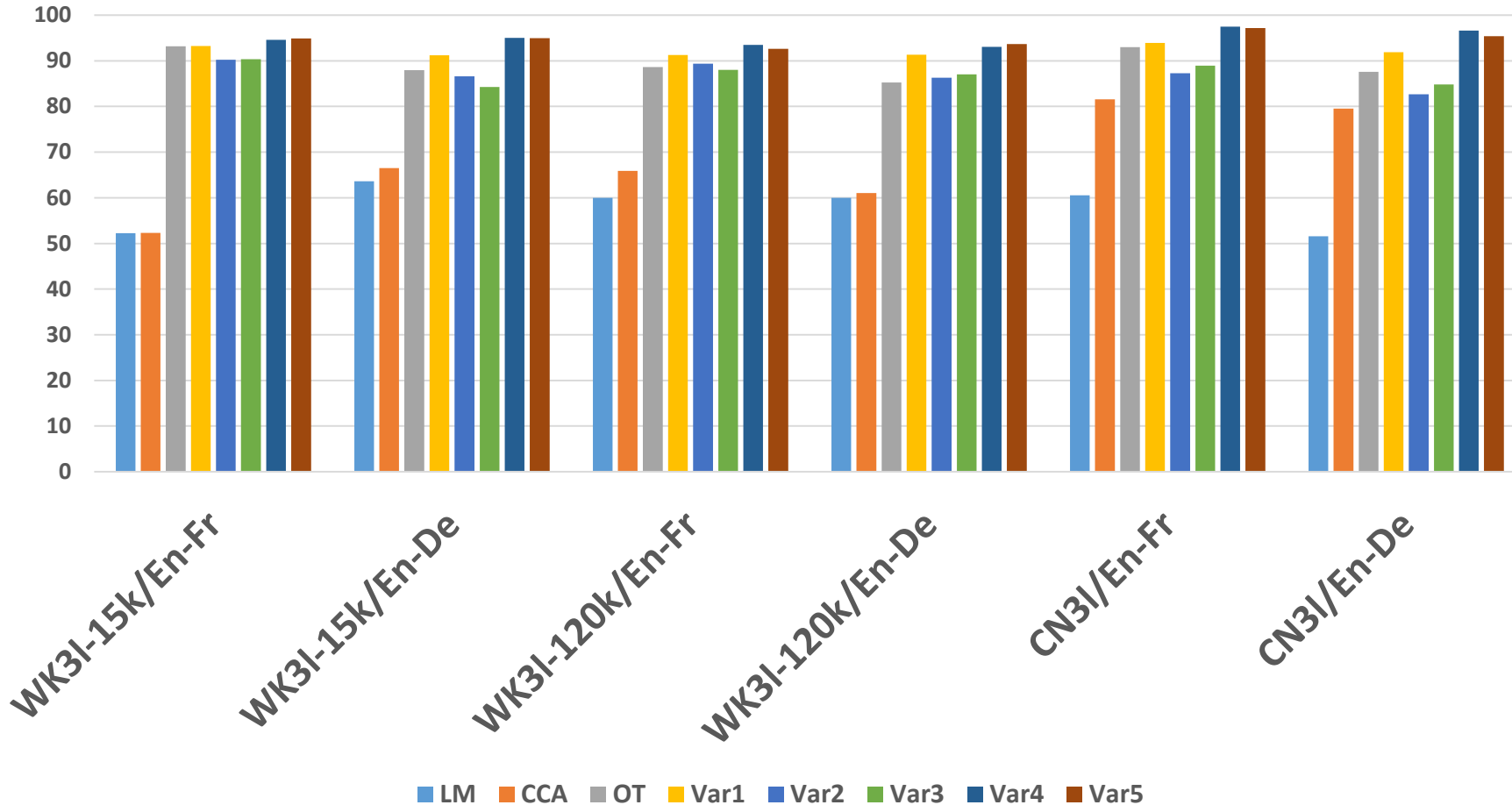


$\text{Var}_4 \approx \text{Var}_5 > \text{Var}_1 \approx \text{Var}_3 \approx \text{OT} > \text{Var}_2 \gg \text{CCA} > \text{LM}$

Axis Calibration	$\text{Var}_1, \text{Var}_2$
Trans. Vectors	Var_3
Linear Transforms	$\text{Var}_4, \text{Var}_5$

Triple-wise Alignment Verification

Accuracy of TWA Verification



$Var_4 \approx Var_5 > Var_1 > Var_2 > Var_3 \approx OT$
 $\gg CCA > LM$

We receive similar evaluation conclusions in all settings.

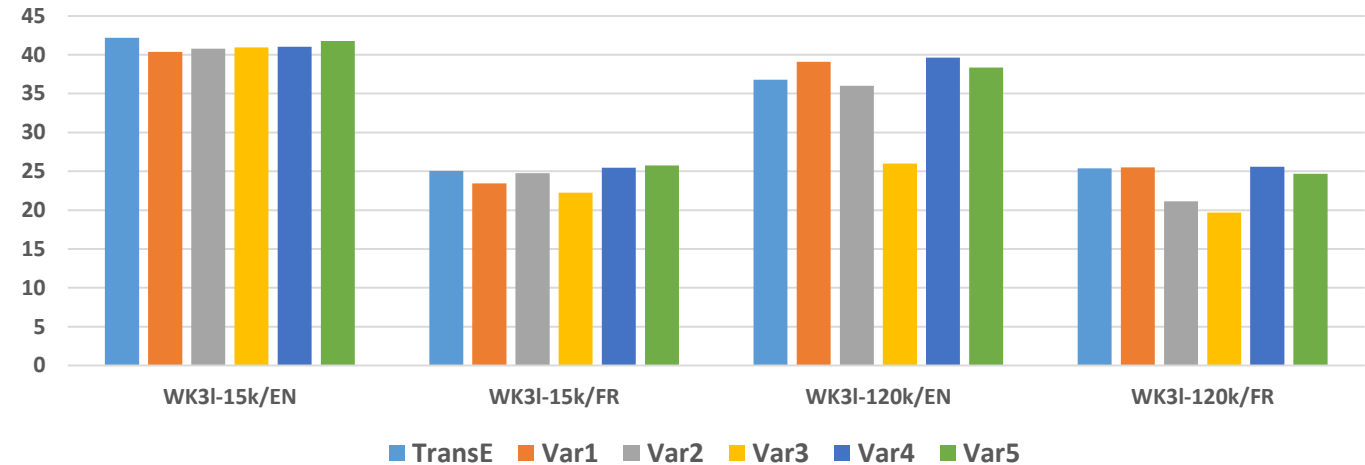
Axis Calibration	Var_1, Var_2
Trans. Vectors	Var_3
Linear Transforms	Var_4, Var_5

Monolingual Relation Extraction (English, French)

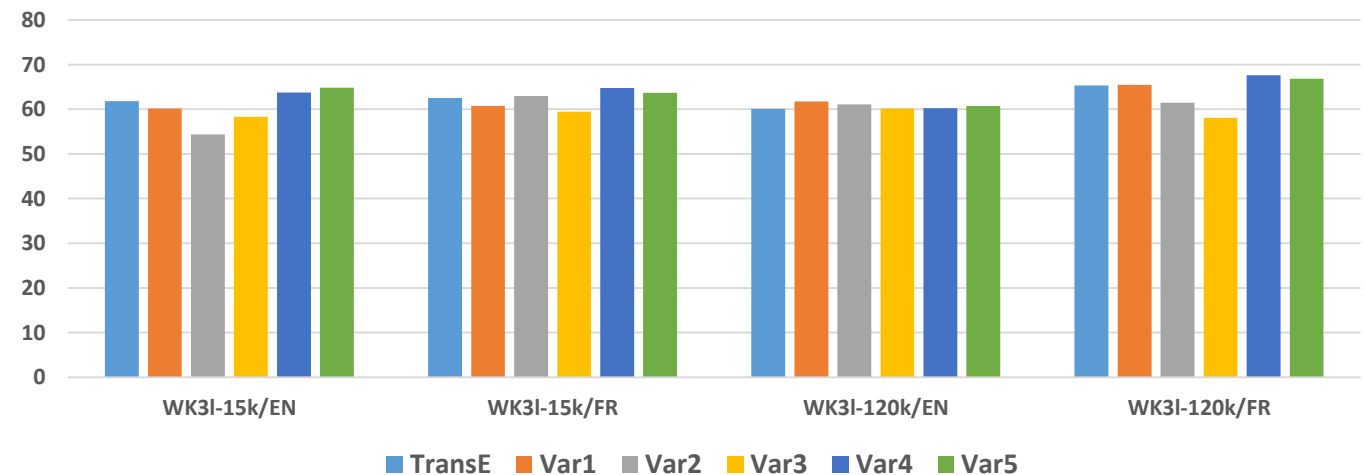
- Train/Test
 - Train Sets: 90% triples and intersecting alignment sets
 - Test Sets: 10% triples
- MTransE preserves well the monolingual relations

Axis Calibration	Var ₁ , Var ₂
Trans. Vectors	Var ₃
Linear Transforms	Var ₄ , Var ₅

Predicting Missing Tails (Hits@10)



Predicting Missing Relations (Hits@10)



Applications based on MTransE

- Multilingual Q&A
- Cross-lingual relation prediction
- Improving monolingual KG completion using multilingual correlation
- Knowledge alignment across knowledge bases

Examples of Cross-lingual Question Answering

Query	Target	Candidates (in ascending order of rank)
(Adam Lambert, genre, ? <i>t</i>)	French	<i>musique indépendante</i> , musique alternative , ode, glam rock
	German	popmusik , dance-pop , no wave, <i>soul</i>
(Ronaldinho, position, ? <i>t</i>)	French	milieu offensif , attaquant , <i>quarterback</i> , <i>latéral gauche</i>
	German	stürmer , <i>linker flügel</i> , angriffsspieler , <i>rechter flgel</i>
(Italy, ? <i>r</i> , Rome)	French	capitale , plus grande ville , chef-lieu , garnison
	German	hauptstadt , hauptort , verwaltungssitz, stadion
(Barack Obama, ? <i>r</i> , George Bush)	French	<i>ministre-prèsident</i> , prèdècesseur , <i>premier ministre</i> , <i>prèsidant du conseil</i>
	German	vorgänger , vorgängerin , besetzung, lied
(? <i>h</i> , instrument, guitar)	French	Brant Bjork , Chris Garneau , <i>David Draiman</i> , Ian Mackaye
	German	Phil Manzanera , <i>Styles P.</i> , <i>Tina Charles</i> , Luke Bryan

Bold-faced ones are correct answers, *italic* ones are close answers.

Improve the embedding model

- Other forms of knowledge models and alignment models
 - Neural knowledge models such as HoIE and ComplEx
 - Other alignment models such as affine transformations
 - Alignment models which consider disambiguation
- Encoding more information from multilingual KGs
 - Entity domains, class templates, entity descriptions, etc
 - Cross-lingual disambiguation
- Jointly embedding with other forms of corpora such as multilingual documents

References

- [Bordes et al., 2013] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In NIPS, pages 2787–2795, 2013.
- [Nickel et al., 2016] Maximilian Nickel, Lorenzo Rosasco, Tomaso Poggio, et al. Holographic embeddings of knowledge graphs. In AAI, 2016.
- [Saxe et al., 2014] Andrew M Saxe, James L McClelland, and Surya Ganguli. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. ICLR, 2014.
- [Wang et al., 2014] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In AAI, 2014.
- [Lin et al., 2015] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In AAI, 2015.
- [Ji et al., 2015] Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix. In ACL, pages 687–696, 2015.
- [Mikolov et al., 2013] Tomas Mikolov, Quoc V Le, and Ilya Sutskever. Exploiting similarities among languages for machine translation. arXiv, 2013.
- [Faruqui and Dyer, 2014] Manaal Faruqui and Chris Dyer. Improving vector space word representations using multilingual correlation. EACL, 2014.
- [Xing et al., 2015] Chao Xing, Dong Wang, Chao Liu, and Yiye Lin. Normalized word embedding and orthogonal transform for bilingual word translation. In NAACL HLT, pages 1006–1011, 2015.

Thank You