Knowledge Graph Embedding with Hierarchical Relation Structure

Zhao Zhang[1], Fuzhen Zhuang[1], Meng Qu[2], Fen Lin[3] and Qing He[1]

[1]Institute of Computing Technology, Chinese Academy of Sciences, China

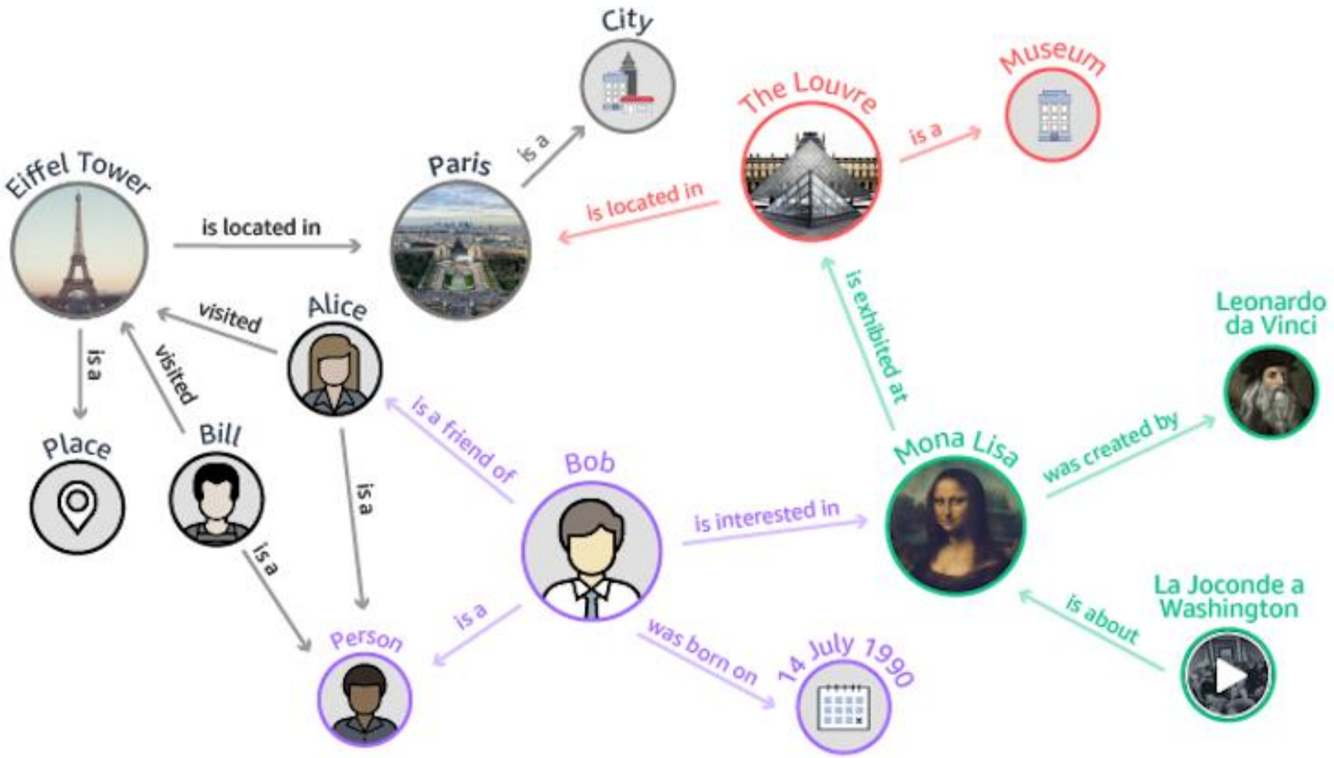
[2]Rutgers Business School, Rutgers University, USA

[3]WeChat Search Application Department, Tencent, China

Outline

- Introduction to Knowledge Graph Embedding
- Related Work and Preliminaries
- Methodology
- Experimental Results and Conclusion

Knowledge Graph (KG)



Knowledge graphs are multi-relational directed graphs composed of entities as nodes and relations as edges.

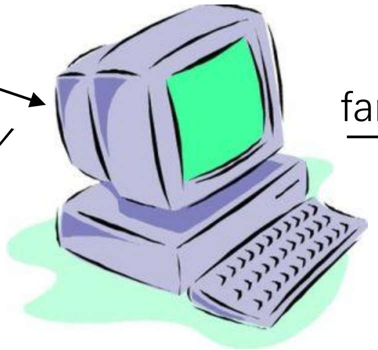
Knowledge Graph Embedding

- Knowledge graphs are comprised of knowledge triples in the form of (h, r, t), e.g. (Beijing, capitalOf, China).

- (Beijing, capitalOf, China)

easy to understand by

hard to manipulate by



familiar with

Embeddings

Applications of KG embeddings

中国科学院计算技术研究所

- Knowledge graph completion
  - (Donald Trump, bornIn, New York) & (New York, cityOf, USA) → (Donald Trump, nationality, USA)
- Knowledge triple classification
  - (Donald Trump, nationality, USA) → True
  - (Donald Trump, nationality, Canada) → False
- Serve as a fundamental step for many other tasks
  - Question Answering
  - Recommender Systems

Motivation

- Semantically similar relations are often observed in large-scales KGs.
  - producerOfAFilm and directorOfAFilm

- There are relations that have multiple semantic meanings and can be split into several sub-relations.
  - partOf
    - (New York, partOf, USA), location-related
    - (monitor, partOf, television), composition-related

Outline

- Introduction to Knowledge Graph Embedding
- Related Work and Preliminaries
- Methodology
- Experimental Results and Conclusion

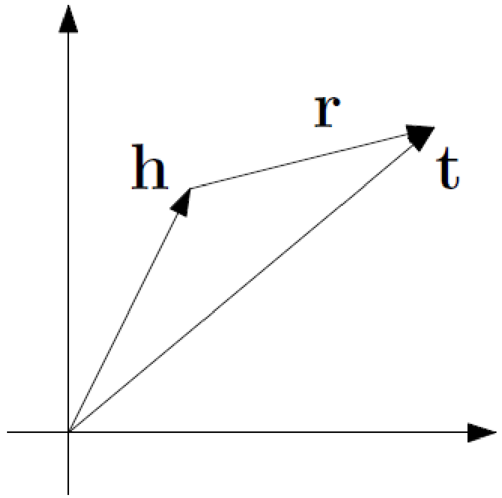
Related work

- Translation-based methods
  - TransE, Bordes et al, 2013
  - TransH, Wang et al, 2014
  - TransR/CTransR, Lin et al, 2015

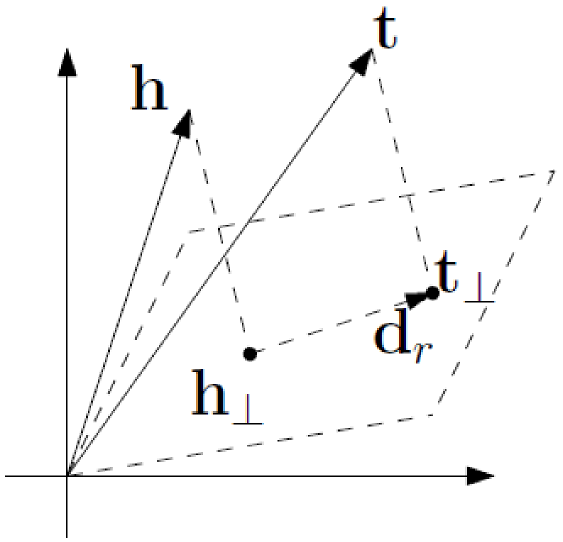
- Tensor Factorization based methods
  - DistMult, Yang et al, 2015
  - ComplEx, Trouillon et al, 2016
  - ANALOGY, Liu et al, 2017

Preliminaries

TransE, $\mathbf{h}+\mathbf{r}=\mathbf{t}$

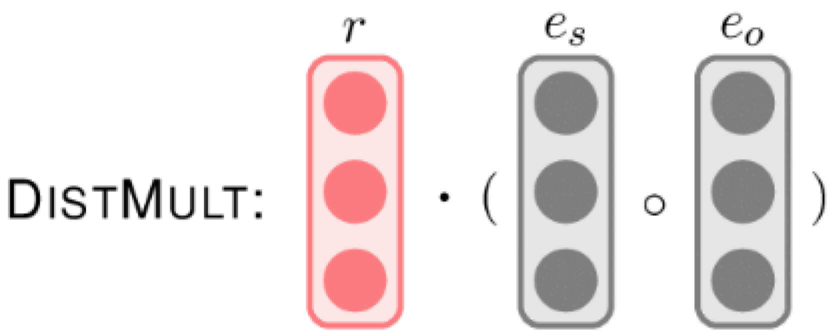


TransH, $\mathbf{h}_{\perp}+\mathbf{r}=\mathbf{t}_{\perp}$



DistMult

$score(h,r,t)=\mathbf{h}^{\mathrm{T}}\mathbf{M}_{r}\mathbf{t}$

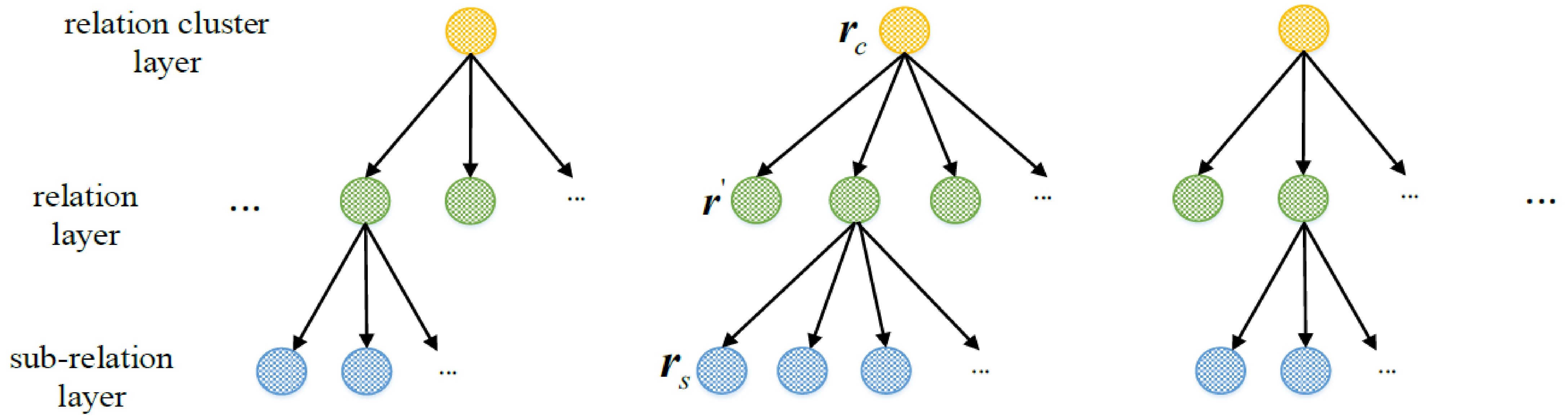


Outline

- Introduction to Knowledge Graph Embedding
- Related Work and Preliminaries
- Methodology
- Experimental Results and Conclusion

Hierarchical Relation Structure

- Semantically similar relations make up **relation clusters**.
- Relations that have multiple semantic meanings are split into several **sub-relations**.



$$\mathbf{r} = \mathbf{r}_c + \mathbf{r}' + \mathbf{r}_s$$

Hierarchical Relation Structure

- Relation clusters and sub-relations are obtained based on the results of TransE

- relation clusters
  - We run k-means on all the relation embeddings in the results of TransE

- sub-relations
  - For each relation and its (h, r, t) triples, we run k-means on all the $\mathbf{h} - \mathbf{t}$ in the results of TransE

Model

中国科学院计算技术研究所

- Relation Embedding, for each r in (h, r, t), $\mathbf{r} = \mathbf{r}_c + \mathbf{r}' + \mathbf{r}_s$
- Score function, TransE $f(h,r,t) = \| \mathbf{h} + \mathbf{r} - \mathbf{t} \|_{L_n}$

  TransE-HRS $f(h,r,t) = \| \mathbf{h} + \mathbf{r}_c + \mathbf{r}' + \mathbf{r}_s - \mathbf{t} \|_{L_n}$

- Loss Function $L_{Total} = L_{Orig} + L_{HRS}$
- HRS Loss $L_{HRS} = \lambda_1 \sum_{\mathbf{r}_c \in C} \| \mathbf{r}_c \|_2^2 + \lambda_2 \sum_{\mathbf{r}' \in C} \| \mathbf{r}' \|_2^2 + \lambda_3 \sum_{\mathbf{r}_s \in S} \| \mathbf{r}_s \|_2^2$

Model Variants

中国科学院计算技术研究所

- Top-Middle Model

$$\mathbf{r} = \mathbf{r}_c + \mathbf{r}'$$

$$L_{HRS} = \lambda_1 \sum_{\mathbf{r}_c \in C} \| \mathbf{r}_c \|_2^2 + \lambda_2 \sum_{\mathbf{r}' \in C} \| \mathbf{r}' \|_2^2$$

- Middle Bottom Model

$$\mathbf{r} = \mathbf{r}' + \mathbf{r}_s$$

$$L_{HRS} = \lambda_2 \sum_{\mathbf{r}' \in C} \| \mathbf{r}' \|_2^2 + \lambda_3 \sum_{\mathbf{r}_s \in S} \| \mathbf{r}_s \|_2^2$$

Outline

中国科学院计算技术研究所

- Introduction to Knowledge Graph Embedding
- Related Work and Preliminaries
- Methodology
- Experimental Results and Conclusion

Dataset

中国科学院计算技术研究所

- We adopt five datasets to conduct experiments.

Table 1: Statistics of the Five Datasets.

Dataset	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	#triples in Train/Valid/Test
FB15k	14,951	1,345	483,142 / 50,000 / 59,071
FB15k-237	14,541	237	272,115 / 17,535 / 20,466
FB13	75,043	13	316,232 / 5,908 / 23,733
WN18	40,943	18	141,442 / 5,000 / 5,000
WN11	38,696	11	112,581 / 2,609 / 10,544

Baselines

- TransE (Bordes et al., 2013)
- TransH (Wang et al., 2014)
- DistMult (Yang et al., 2015)
- CTransR (Lin et al., 2015)
- TransD (Ji et al., 2015):
- TransG (Xiao et al., 2016):

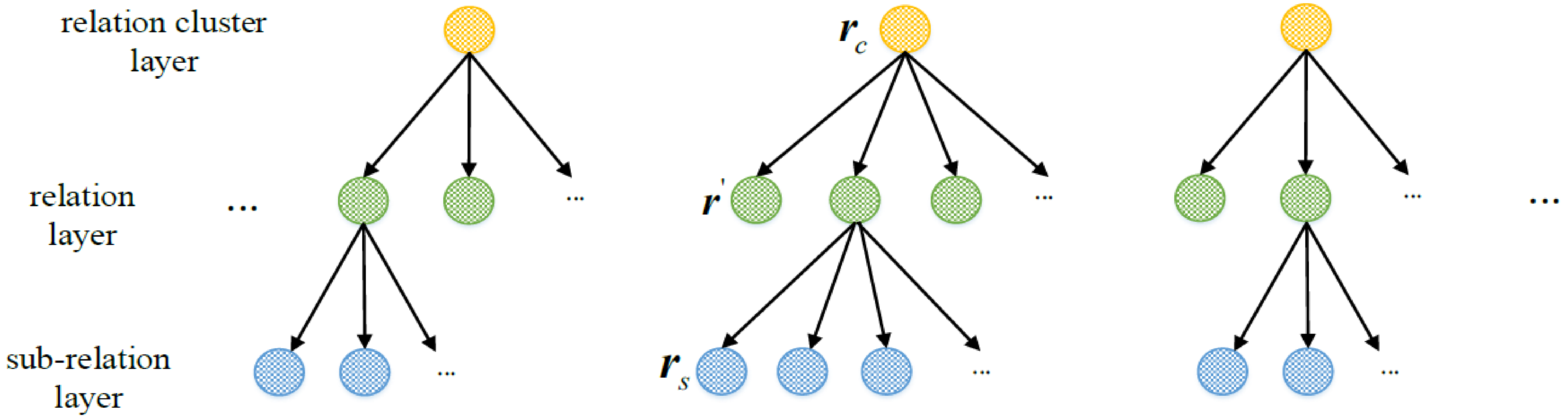
Link Prediction

➢Predict ? in (h, r, ?) or (?, r, t)

	FB15k					FB15k-237					WN18				
	MR	MRR	H10	H3	H1	MR	MRR	H10	H3	H1	MR	MRR	H10	H3	H1
CTransR	81	0.408	0.740	0.573	0.314	279	0.298	0.469	0.301	0.198	228	0.816	0.923	0.842	0.316
TransD	90	0.658	0.781	0.586	0.324	256	0.286	0.453	0.291	0.179	215	0.823	0.928	0.851	0.336
TransG	101	0.672	0.802	0.591	0.322	309	0.304	0.471	0.298	0.182	466	0.830	0.936	0.876	0.764
TransE	91	0.404	0.688	0.493	0.251	375	0.207	0.377	0.227	0.125	**387**	0.408	0.925	0.725	0.067
TransE-top-middle	61	0.463	0.730	0.556	0.315	286	0.258	0.440	0.286	0.170	609	0.402	0.919	0.710	0.058
TransE-middle-bottom	51	0.493	0.738	0.582	0.355	232	0.310	0.486	0.332	0.202	474	**0.496**	**0.945**	**0.890**	**0.112**
TransE-HRS	**49**	**0.510**	**0.767**	**0.610**	**0.361**	**230**	**0.311**	**0.487**	**0.353**	**0.215**	477	0.490	0.943	0.883	0.106
TransH	63	0.394	0.713	0.519	0.210	311	0.211	0.386	0.224	0.132	388	0.437	0.919	0.832	0.039
TransH-top-middle	65	0.477	0.737	0.561	0.308	275	0.272	0.461	0.291	0.185	411	0.416	0.890	0.813	0.034
TransH-middle-bottom	50	0.469	0.742	0.583	0.343	271	0.269	0.466	0.286	0.191	**283**	**0.491**	**0.942**	**0.880**	**0.113**
TransH-HRS	**47**	**0.509**	**0.783**	**0.639**	**0.390**	**243**	**0.309**	**0.491**	**0.346**	**0.216**	296	0.482	0.940	0.861	0.097
DistMult	95	0.642	0.813	0.726	0.523	251	0.244	0.423	0.261	0.159	261	0.806	0.931	0.904	0.713
DistMult-top-middle	85	0.677	0.830	0.746	0.589	243	0.286	0.461	0.291	0.192	246	0.769	0.903	0.853	0.681
DistMult-middle-bottom	83	0.682	0.828	0.758	0.606	246	0.291	0.475	0.306	0.199	226	**0.912**	**0.947**	**0.913**	**0.879**
DistMult-HRS	**72**	**0.739**	**0.846**	**0.799**	**0.661**	**232**	**0.315**	**0.496**	**0.350**	**0.241**	**206**	0.891	0.932	0.901	0.736

Hierarchical Relation Structure

- Semantically similar relations make up **relation clusters**.
- Relations that have multiple semantic meanings are split into several **sub-relations**.



$$\mathbf{r} = \mathbf{r}_c + \mathbf{r}' + \mathbf{r}_s$$

Relation Clusters

中国科学院计算技术研究所

- Examples of Relation Clusters in FB15k

	relations
1	/olympics/olympic_athlete/medals_won./olympics/olympic_medal_honor/country /olympics/olympic_athlete/country./olympics/olympic_athlete_affiliation/country
2	/sports/sports_team/roster./basketball/basketball_roster_position/player, /basketball/basketball_team/roster./sports/sports_team_roster/player /basketball/basketball_team/roster./basketball/basketball_roster_position/player
3	/computer/software/developer, /computer/operating_system/developer, /cvg/computer_videogame/developer

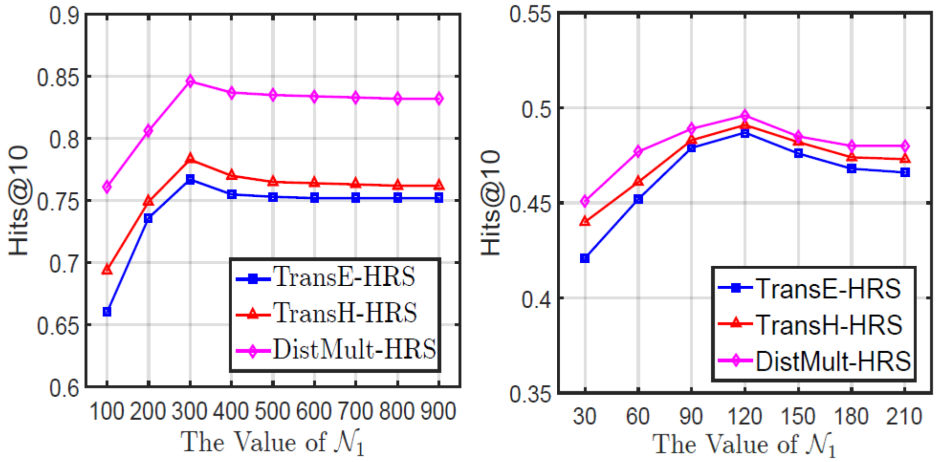
Sub-relations

- Examples of Sub-relations for Relation '/music/artist/genre' in FB15k

	(head, tail)
1	(Steve Stills, Rock Music), (Velvet Underground, Rock Music), (Benjamin Chase Harper, Rock Music), ...
2	(Justin Beiber, Pop Music), (Natalie Maria Cole, Pop Music), (Peter Thorkelson, Pop Music), ...
3	(Billy Preston, R & B), (Earth Wind Fire, Funk Rap), (Alvin Joiner, Hip-hop), ...

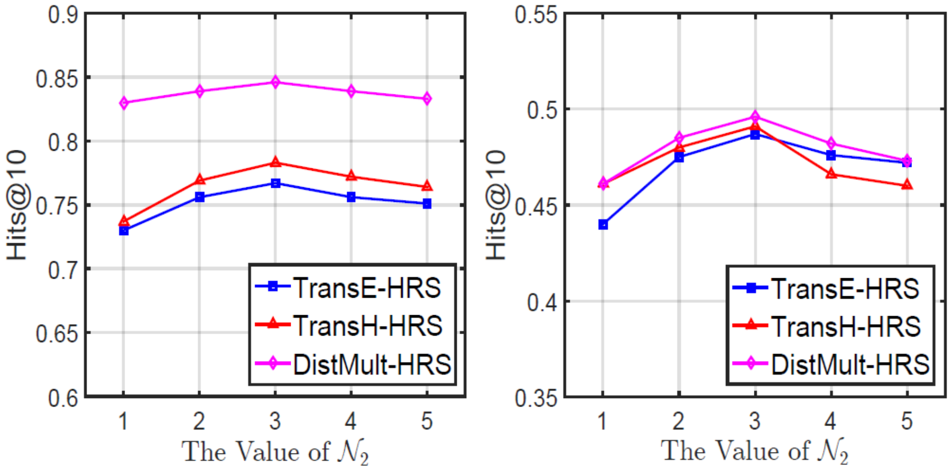
Parameter Study

- The Change of Hits@10 with the Value of N_1 Increasing.



(a) FB15k (b) FB15k-237

- The Change of Hits@10 with the Value of N_2 Increasing.

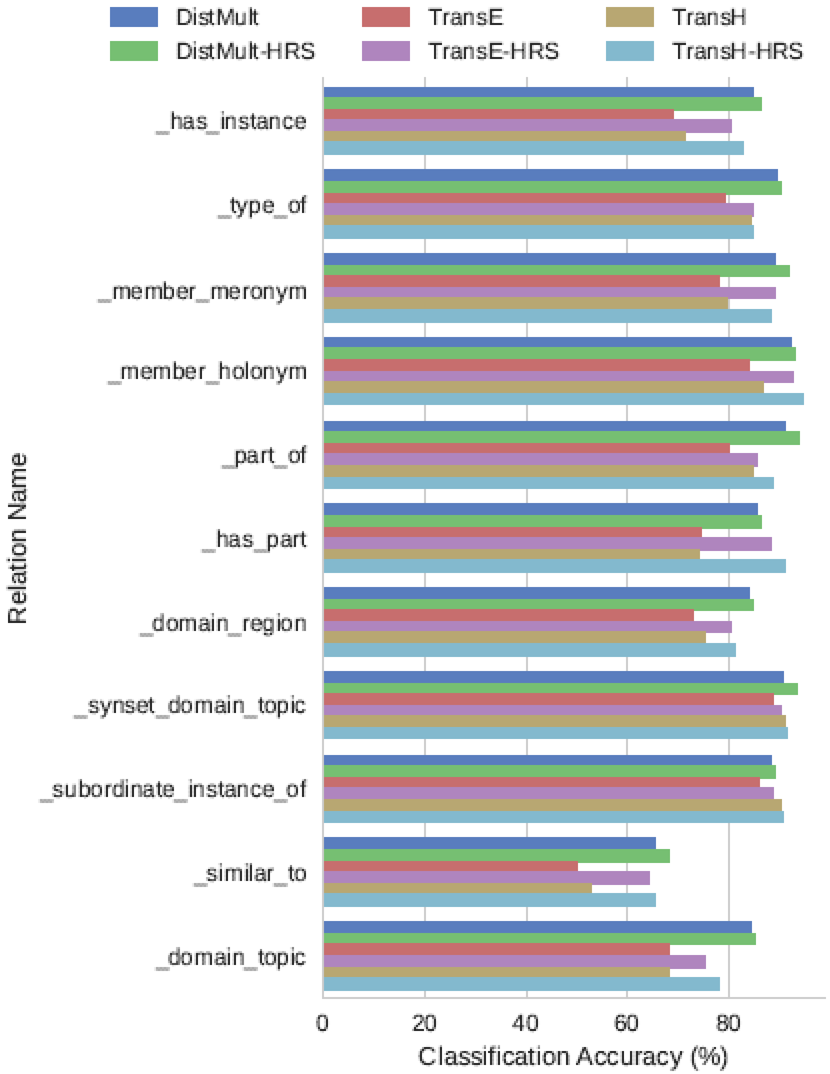


(a) FB15k (b) FB15k-237

Triple Classification

- Predict the label (True or False) given (h, r, t).

Model	WN11	FB13	FB15k	Avg
CTransR	85.7	-	84.4	-
TransD	86.4	89.1	88.2	87.9
TransG	87.4	87.3	88.5	87.7
TransE	$\boxed{75.9}$	81.5	78.7	78.7
TransH	78.8	83.3	81.1	81.1
DistMult	87.1	86.2	86.3	86.5
TransE-HRS	$\boxed{86.8}$	88.4	87.6	87.6
TransH-HRS	**87.6**	88.9	**88.7**	**88.4**
DistMult-HRS	**88.9**	89.0	**89.1**	**89.0**



Conclusion and Future Work

中国科学院计算技术研究所

➢Conclusions

- Leveraging the three-layer hierarchical relation structure is simple but effective for knowledge graph embedding models.
- The technique of utilizing the hierarchical relation structure can be applied to many existing knowledge graph embedding models.

➢Future Work

- Utilize the embeddings of the three layers in a more sophisticated way instead of sum them together.
- Determine the number of relation clusters and sub-relations automatically instead of manually.

Thanks!

Contact info:
zhangzhao2017@ict.ac.cn
zhuangfuzhen@ict.ac.cn
heqing@ict.ac.cn