Multiple Run Ensemble Learning with Low-Dimensional KGEs

ESR 10, Cosimo Gregucci

25 November 2022

Motivation

- link prediction - SOTA obtained at high dimensions
  → incresed training costs & risk of overfitting

- Instead of training a large sized embedding model, the paper proposes to execute *multiple run of a low-dimension embedding model, and combine them*
  → **combination of low-dimensional KGEs outperforms the corrisponding high-dimensional one**

Setting

- $k \times d_l = 1 \times d_h$

k = number of multiple runs
d_l = dimension of each low-dimensional KGE
d_h = dimension of the high-dimensional KGE

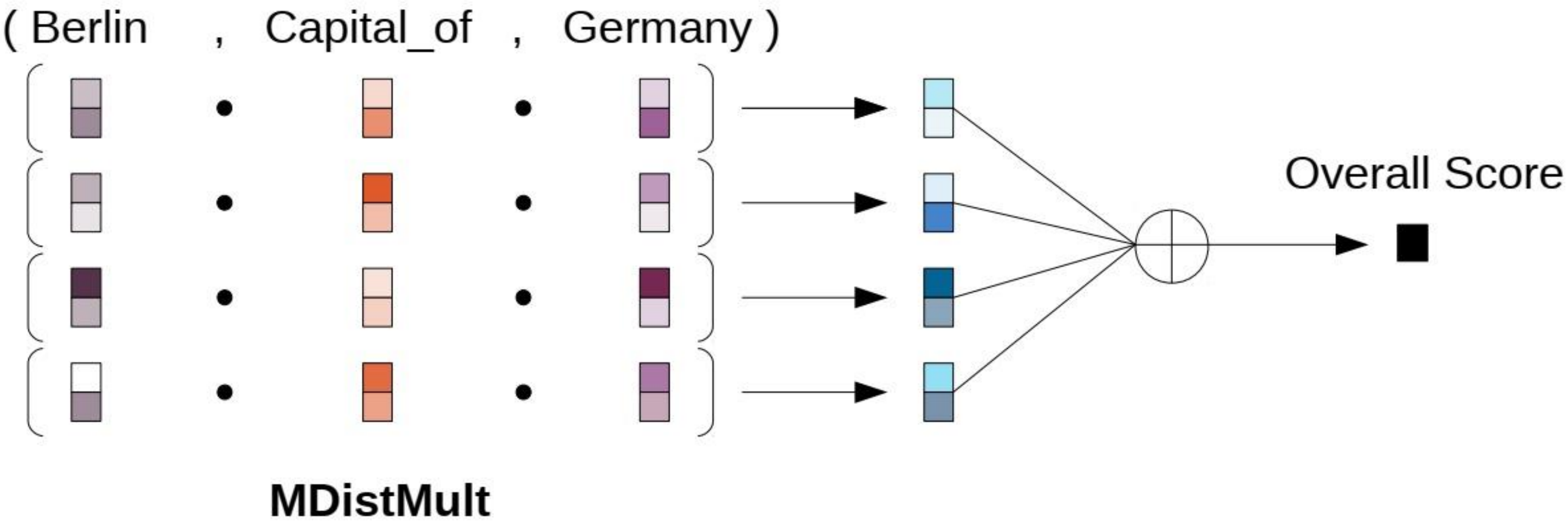
Approach – KGE model *M*

1. k copies of *M* are generated.
   M_j → j-th copy of M
   d_l dimensional embedding → $(\mathbf{h}_j, \mathbf{r}_j, \mathbf{t}_j)$
2. Each M_j is trained separately
3. Score function:

$$f(h,r,t) = \frac{1}{k}\sum_{j=1}^{k} f_M(\boldsymbol{h}_j, \boldsymbol{r}_j, \boldsymbol{t}_j)$$

where $f_M(\boldsymbol{h}_j, \boldsymbol{r}_j, \boldsymbol{t}_j)$ → score of a triple (h,r,t) computed by the j-th copy of the model

Example – *MDistMult*



Generalization capabilities

TransE:

- Symmetric relation (eg. Similarto)

1. h + r = t → t + r ≠ h → f(h,r,t) ≠ 0, f(t,r,h)=0
2. t + r = h → h + r ≠ t → f(t,r,h) ≠ 0, f(h,r,t) =0

By randomly initialize each run of the ensemble, **MTransE** might be able to model symmetrical pattern

Example with different relational patterns of FB15K

Relation	TransE (d=200×1)	TransE (d=1200×1)	MTransE (d=200×6)
1-1	0.642	0.663(↑**0.021**)	0.660(↑**0.018**)
1-n	0.739	0.748(↑**0.009**)	0.780(↑**0.041**)
n-1	0.639	0.650(↑**0.011**)	0.678(↑**0.039**)
n-n	0.706	0.709(↑**0.003**)	0.739(↑**0.033**)
symmetric	0.358	0.360(↑**0.002**)	0.411(↑**0.053**)

Experiments

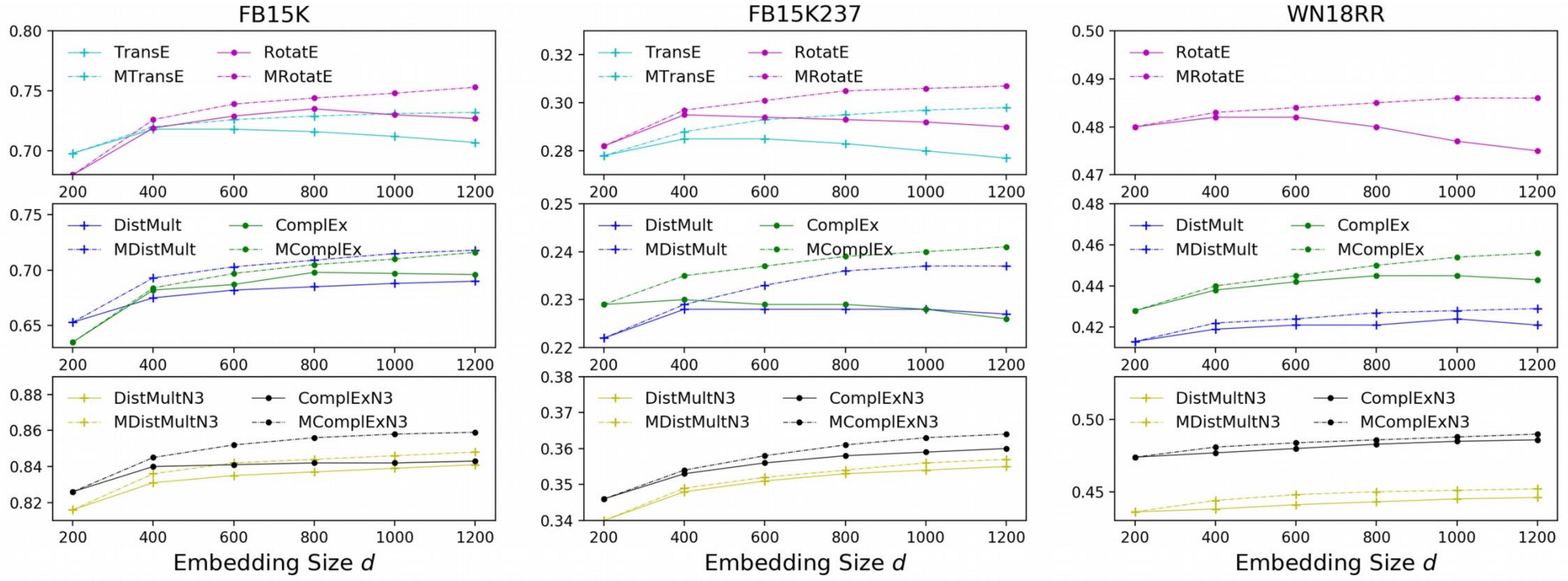
- Datasets statistics

Dataset	$\|\mathcal{E}\|$	$\|\mathcal{R}\|$	#Train	#Valid	#Test
FB15K	14951	1345	483142	50000	59071
FB15K237	14541	237	272115	17535	20466
WN18RR	40943	11	86835	3034	3134

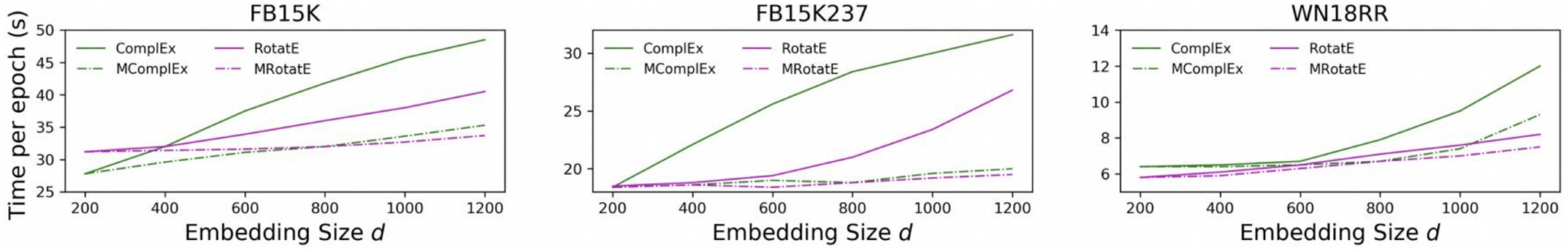
- Metrics:
  - MRR: Mean Reciprocal Rank
  - Hits@N: Proportion of correct entities ranked in the top N

Experiments

MRR:



Training time per epoch:



Experiments - Freebase

	FB15K				FB15K-237			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE ($d = 1200$)	0.704	0.604	0.781	0.862	0.277	0.186	0.303	0.464
MTransE ($d = 6 * 200$)	**0.732**	**0.640**	**0.802**	**0.876**	**0.298**	**0.202**	**0.329**	**0.491**
DitMult ($d = 1200$)	0.688	0.573	0.781	0.855	0.227	0.142	0.249	0.390
MDitMult ($d = 6 * 200$)	**0.718**	**0.603**	**0.815**	**0.883**	**0.237**	**0.160**	**0.260**	**0.399**
ComplEx ($d = 1200$)	0.696	0.580	0.791	0.862	0.226	0.139	0.249	0.398
MComplEx ($d = 6 * 200$)	**0.710**	**0.590**	**0.810**	**0.886**	**0.240**	**0.162**	**0.264**	**0.400**
RotatE ($d = 1200$)	0.727	0.630	0.802	0.868	0.290	0.197	0.319	0.478
MRotatE ($d = 6 * 200$)	**0.753**	**0.656**	**0.832**	**0.891**	**0.307**	**0.213**	**0.338**	**0.496**
DitMultN3 ($d = 1200$)	0.836	0.796	0.865	0.909	0.355	0.260	0.390	0.547
MDitMultN3 ($d = 6 * 200$)	**0.848**	**0.813**	**0.869**	**0.910**	**0.357**	**0.263**	**0.392**	**0.548**
ComplExN3 ($d = 1200$)	0.843	0.802	0.871	0.910	0.360	0.265	0.395	0.549
MComplExN3 ($d = 6 * 200$)	**0.859**	**0.829**	**0.877**	**0.911**	**0.364**	**0.268**	**0.400**	**0.555**

Conclusion

- Multiple run of low-dimensional KGE models:
  - outperforms the corresponding high dimensional KGE
  - sligthly increse the generalization capabilities of the model