AUTOMATED KNOWLEDGE BASE CONSTRUCTION 2021 | irvine

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

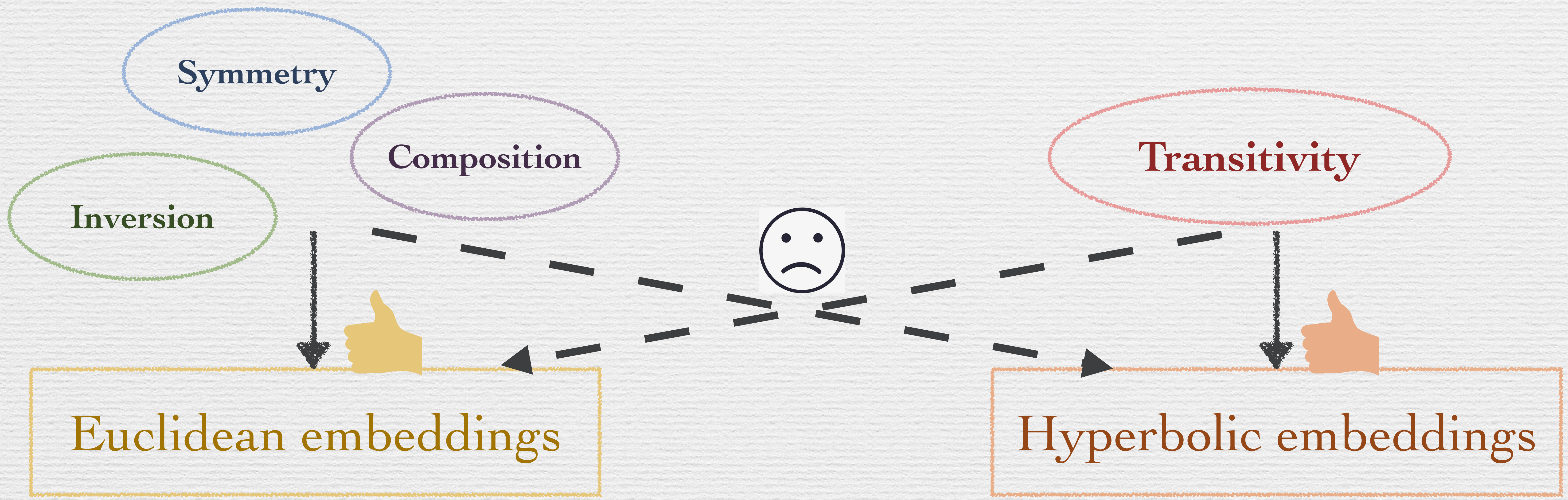
Manifold Alignment across Geometric Spaces for Knowledge Base Representation Learning

Huiru Xiao *HKUST*

Yangqiu Song *HKUST*

Motivation

☆Knowledge bases have multi-relations with **distinctive properties**.

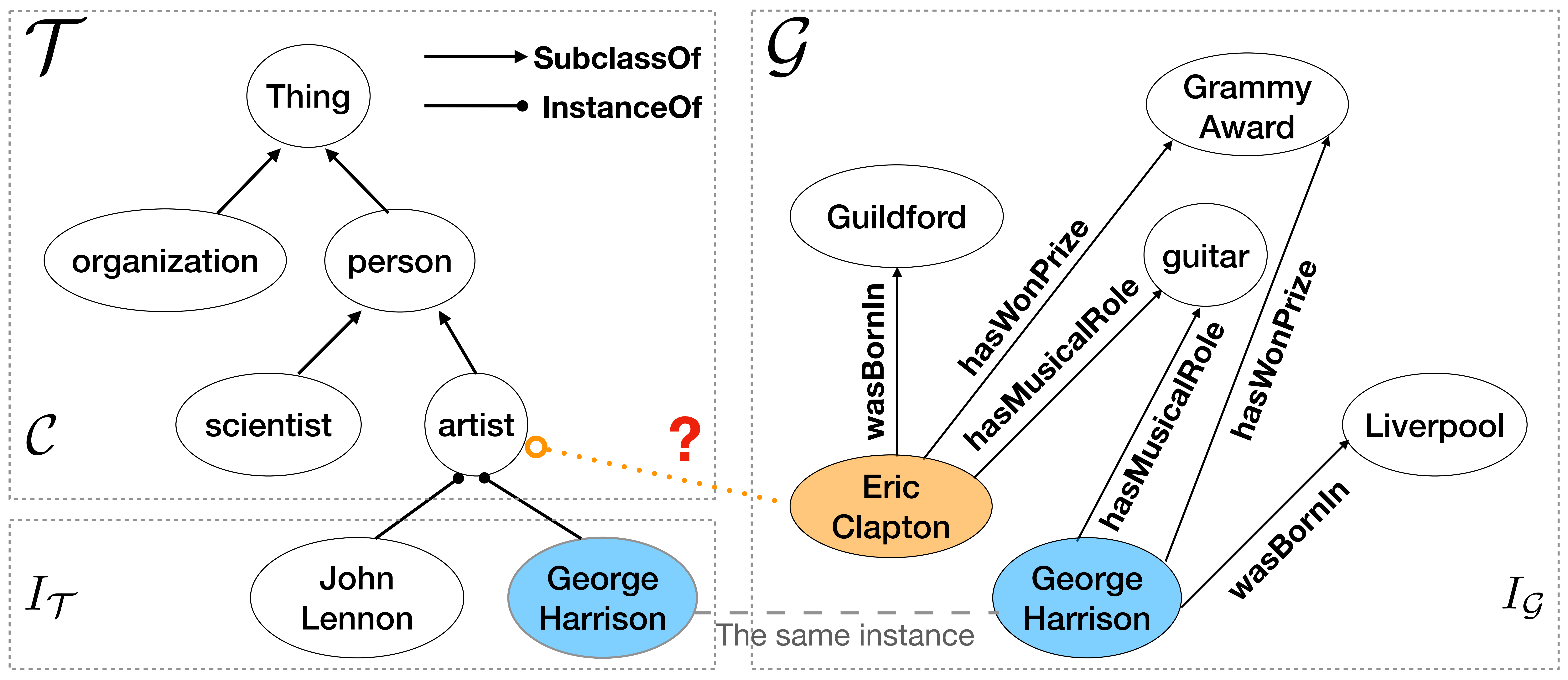


Building a representation learning framework for all relation properties is highly difficult.

☆Propose to learn the knowledge base embeddings **in different geometric spaces** and apply **manifold alignment** to align the shared entities.

Task

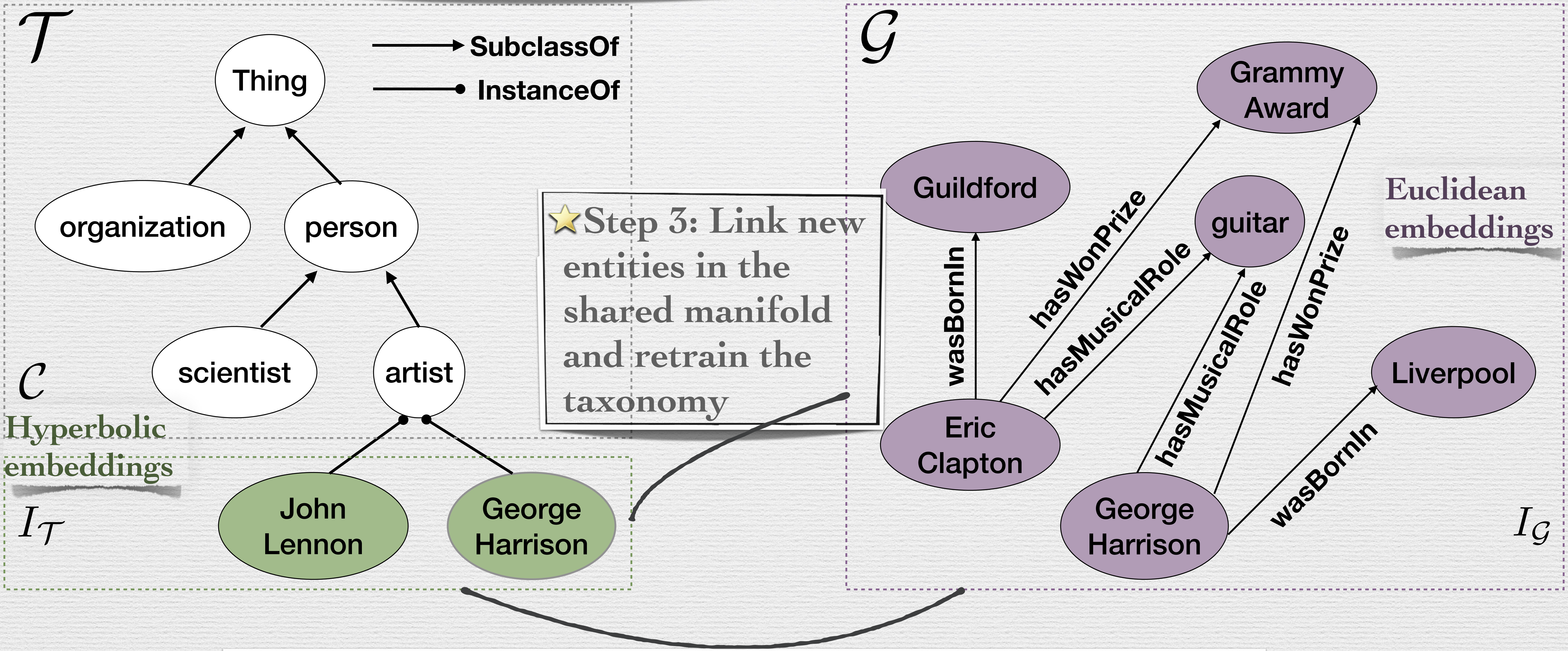
⭐Predict the type for the out-of-taxonomy entity.



Method

☆Step 1.1: Pretrain the taxonomy by hyperbolic embeddings in the hyperboloid model

☆Step 1.2: Pretrain the knowledge graph by TransE in the Euclidean space



☆Step 3: Link new entities in the shared manifold and retrain the taxonomy

☆Step 2: Apply manifold alignment to align the embeddings of taxonomy entities and knowledge graph entities

Main Results

Dimension	5		10		20		100	
	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1	MRR	Hits@1
TransE	‡2.42	‡3.91	‡5.41	‡9.91	‡6.80	‡14.14	‡8.77	‡19.21
ComplEx	‡0.09	‡0.00	‡11.97	‡22.59	‡14.32	‡24.20	‡20.21	‡34.39
RotatE	‡1.76	‡1.39	‡3.50	‡2.94	‡4.23	‡3.37	‡37.68	‡79.95
TransC	‡32.62	‡66.69	‡36.15	‡78.55	‡36.67	‡84.03	‡34.28	‡81.32
HAKE	‡7.85	‡12.62	‡15.58	‡22.70	‡10.16	‡17.29	†37.90	‡70.88
JOIE	‡27.76	‡48.01	‡39.93	‡84.68	‡41.64	‡92.62	‡42.12	‡94.61
MurP	‡42.18	‡98.19	‡42.61	‡98.59	‡42.72	‡99.60	**42.91**	‡99.90
AttH	-	-	†40.60	†92.47	‡42.10	‡97.05	‡42.78	‡98.76
HyperKA	‡16.93	‡34.51	‡21.65	‡41.30	‡19.00	‡42.98	‡24.51	‡61.92
GeoAlign	**42.74**	**99.64**	**42.83**	**99.84**	**42.81**	**99.92**	‡42.82	**99.93**

Table 3: Results of MRR(%) and Hits@1(%) in different em
requires the dimension to be even because of the diagonal
not applicable to 5-dimensional space. The best results ar
results are underlined. The statistically significance metr
< 0.05 or ‡ if p-values < 0.001.

Experimental results on the out-of-taxonomy entity typing task demonstrate that our approach has significantly good performances, especially in low dimensions and on small training rates.

Training rate	0.1	0.2	0.3	0.4
Training/Test entities	1,401/12,605	2,802/11,204	4,202/9,804	5,603/8,403
Training/Test edges	42,764/110,879	54,637/99,006	67,788/85,855	80,190/73,453
MurP	‡82.83	‡86.56	‡88.16	‡88.45
AttH	†84.32	‡86.62	‡87.80	88.40
GeoAlign	**88.50**	**88.78**	**89.09**	**88.90**

Table 4: Results of MAP(%) under different training rates on wikiObjects in 50-dimension. The training rate is used to randomly split the taxonomy entities. Training edges represents the number of training edges in the training taxonomy wikiObjects (the training edges of YAGOfacts is 392,335 all the time). The best results are shown in boldface. The statistically significance metrics are marked with either † if p-values < 0.05 or ‡ if p-values < 0.001.