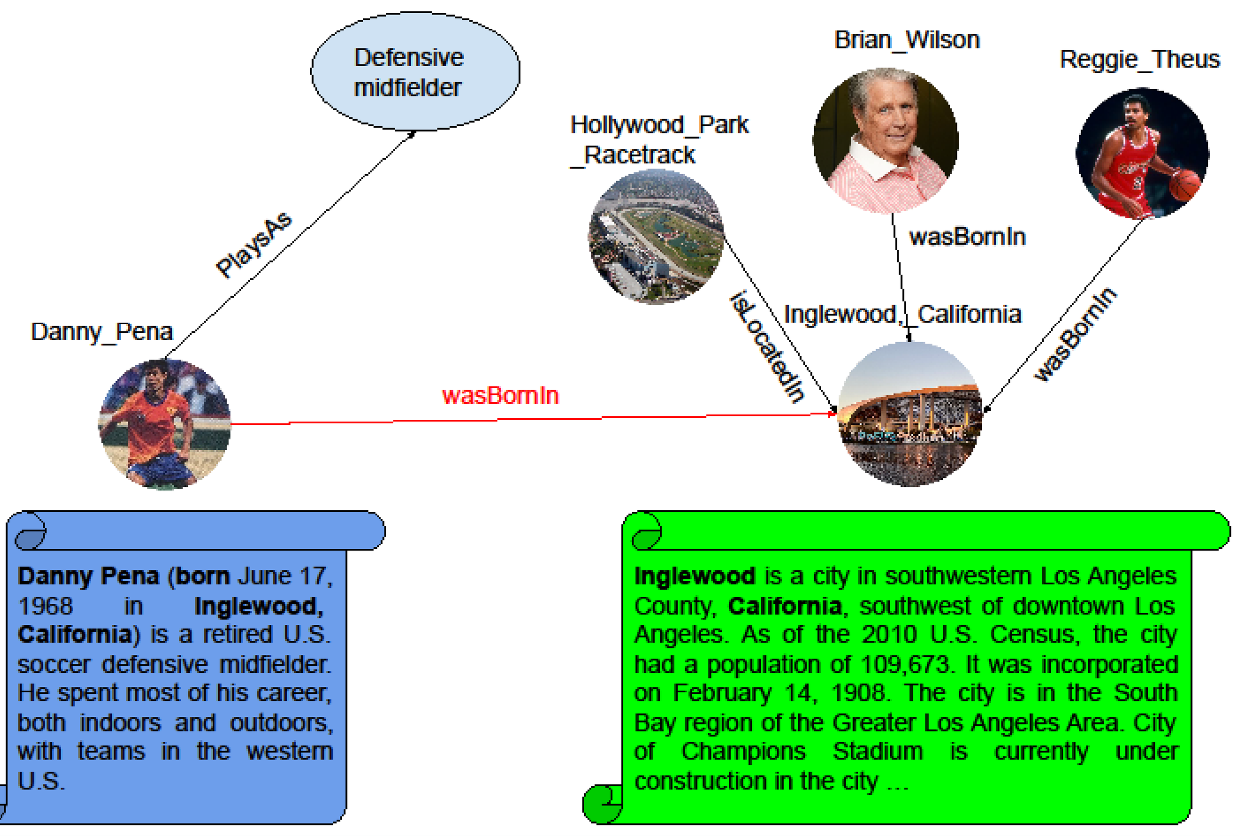
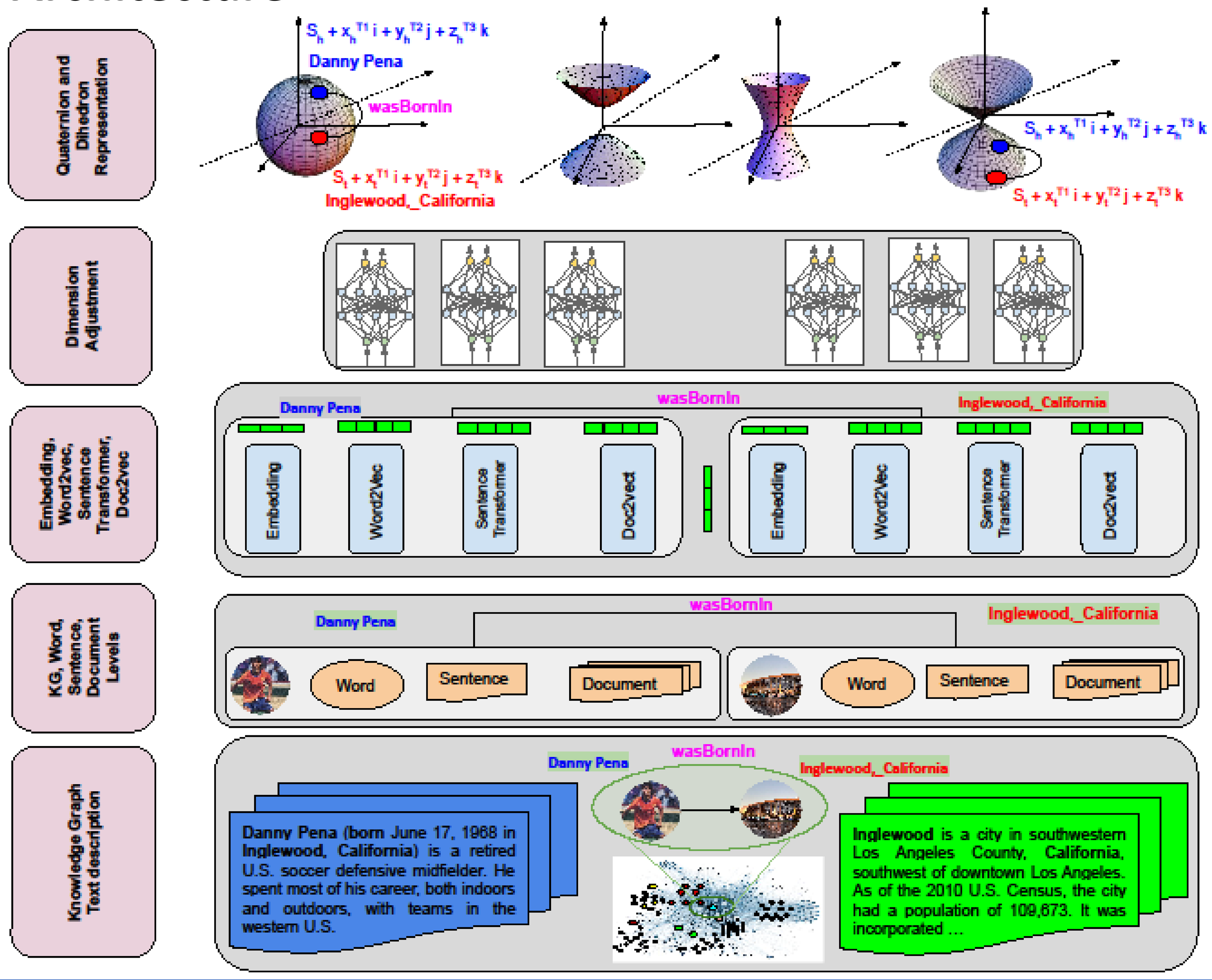
Text-enhanced Knowledge Graph Embedding in Hypercomplex Space

Motivation

- Knowledge is represented in the form of multi-relational directed labeled graphs. Labeled nodes represent entities and labeled edges represent relation.
- Knowledge graph (KG) is highly incomplete. The fact (Danny_Pena, wasBornIn, Ingelwood,_California) exists in the real world but misses in KG.
- There is no direct connection between two entities Danny_Pena and Ingelwood,_California in the KG, but they could be relevant when providing extra knowledge.
- We consider additional textual description of entities, where two entities can be bridged by extracting semantical information from text.



Architecture



5. Represent entities in a natural way that satisfies:
   - Four different information can be incorporated in a unified manner.
   - Consider pairwise interaction between any two information.

   Therefore, entities and relations are represented in the hypercomplex space, where entities are points (red/blue) and relations are rotations (black curve) in the space.

4. Project different representations into the same vector space.

3. Obtain representations from multiple pre-trained language models for all kinds of information from description.

2. Consider four different information to better characterize each entity: structural information in KG, word/sentence/document level information from description.

1. Collect textual descriptions for each entity in the KG: we retrieval the first paragraph in the passage of each entity on the Wikipedia.

Model and Learning

[图片]

- **Entity and relation representation**

$$\mathrm{e} = \boxed{\mathrm{s}_e} + \boxed{\mathrm{x}_e^{\mathcal{T}_1}}\boldsymbol{i} + \boxed{\mathrm{y}_e^{\mathcal{T}_2}}\boldsymbol{j} + \boxed{\mathrm{z}_e^{\mathcal{T}_3}}\boldsymbol{k}$$

Randomly initialized entity embedding

Pre-trained language model embedding

$\mathcal{T}_i \in \{Word2Vec, Fasttext, SentenceTransformer, Doc2Vec\}$

Randomly initialized relation embedding

$$\mathrm{r} = \frac{\boxed{\mathrm{s}_r} + \boxed{\mathrm{x}_r^{\mathcal{T}_1}}\boldsymbol{i} + \boxed{\mathrm{y}_r^{\mathcal{T}_2}}\boldsymbol{j} + \boxed{\mathrm{z}_r^{\mathcal{T}_3}}\boldsymbol{k}}{\sqrt{\mathrm{s}_r^2 + \mathrm{x}_r^{\mathcal{T}_1^2} + \mathrm{y}_r^{\mathcal{T}_2^2} + \mathrm{z}_r^{\mathcal{T}_3^2}}}$$

- **Rotation in the hypercomplex space**

Let $u = s_u + x_u\boldsymbol{i} + y_u\boldsymbol{j} + z_u\boldsymbol{k}, v = s_v + x_v\boldsymbol{i} + y_v\boldsymbol{j} + z_v\boldsymbol{k}$ be two Quaternion or Dihedron numbers in the hypercomplex space, their rotations are defined as follows:

Quaternion product $\otimes_Q$

$$\begin{aligned} u \otimes_Q v := &(s_u s_v - x_u x_v - y_u y_v - z_u z_v) \\ &+ (s_u x_v + x_u s_v + y_u z_v - z_u y_v)\,\mathbf{i} \\ &+ (s_u y_v - x_u z_v + y_u s_v + z_u x_v)\,\mathbf{j} \\ &+ (s_u z_v + x_u y_v - y_u x_v + z_u s_v)\,\mathbf{k} \end{aligned}$$

Dihedron product $\otimes_D$

$$\begin{aligned} u \otimes_D v := &(s_u s_v - x_u x_v + y_u y_v + z_u z_v) \\ &+ (s_u x_v + x_u s_v - y_u z_v + z_u y_v)\,\mathbf{i} \\ &+ (s_u y_v - x_u z_v + y_u s_v + z_u x_v)\,\mathbf{j} \\ &+ (s_u z_v + x_u y_v - y_u x_v + z_u s_v)\,\mathbf{k} \end{aligned}$$

- **Learning**

For a triple in the KG, we optimize its entity and relation embeddings by maximize the following plausibility during training:

$$f(e_h, r, e_t) = -d(e_h \otimes r, e_t) + b_h + b_t$$

In the plausibility above, *d(,)* is a distance function, $\otimes$ could be either the Quaternion or Dihedron product, and b_h, b_t are entity-wise biases.

The model is optimized with the Adagrad optimizational algorithm, where early stopping is employed to prevent overfitting.

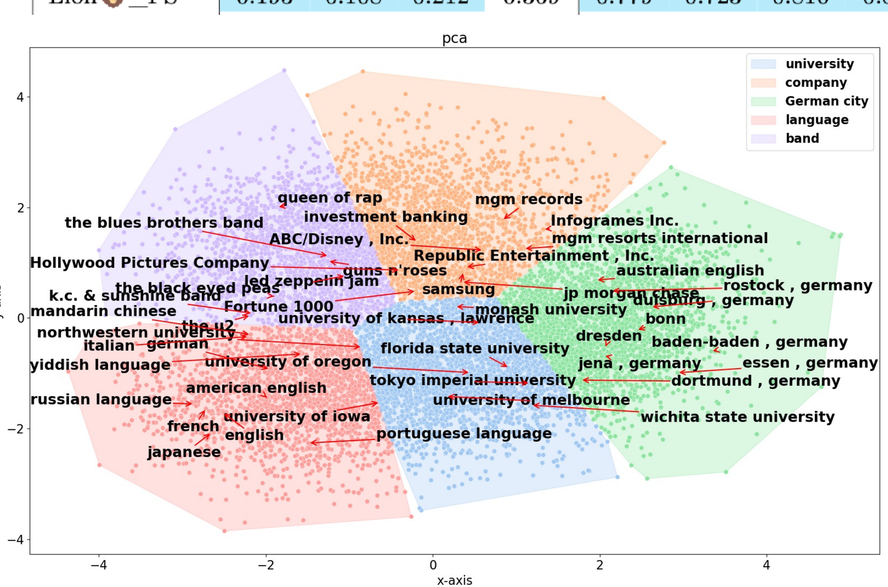
Experiments and Demo

Model	Diabetes				FB-ILT				YAGO-10			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
TransE	0.179	0.098	0.197	0.342	0.739	0.675	0.777	0.852	0.421	0.351	0.461	0.556
AttE	0.176	0.097	0.190	0.341	0.674	0.604	0.711	0.805	0.356	0.294	0.389	0.471
AttH	0.115	0.054	0.118	0.250	0.610	0.521	0.658	0.779	0.313	0.256	0.336	0.431
DKRL	0.162	0.083	0.176	0.328	0.718	0.631	0.779	0.868	0.333	0.239	0.371	0.520
ConMask	0.177	0.094	0.194	0.349	0.698	0.620	0.747	0.842	0.381	0.306	0.421	0.519
Pretrain-KGE	0.159	0.082	0.172	0.323	0.739	0.673	0.780	0.857	0.320	0.231	0.353	0.495
Lion🦁_SD	0.192	0.107	0.208	0.370	0.777	0.722	0.815	0.874	0.433	0.363	0.471	0.562
Lion🦁_FS	0.193	0.108	0.212	0.369	0.779	0.723	0.816	0.878	0.440	0.367	0.478	0.577

- **Link prediction evaluation**

Lion is the best-performed variant of our model. The postfix SD means the model uses Doc2Vec and the postfix FS mean the model uses FastText to represent the document information.

Our model outperforms all baselines on three datasets.



- **Clustering of trained entity embedding**

We cluster trained entity embeddings from our model and reduce their dimension with PCA, so the results can be visualized on the 2D figure.

Each cluster has a topic (see the legend on the top right corner), and each point represents an entity belonging to the topic.

We can see that semantically relevant entities are close to each other.

Triple	Sentence rank	Sentence source	Sentence
Mars_Callahan, created, Zigs_(film)	1	tail	$\boxed{\text{Zigs}}$ is a 2001 English language drama starring Jason Priestley Peter Dobson and Richard Portnow and $\boxed{\text{directed by Mars Callahan.}}$
	2	tail	The film received an r rating by the MPAA.
	3	head	At the age of eleven Callahan toured with a children's musical group through thirty-seven states.

- **The extraction of sentence-level information by our model**

We analyze the importance sentence from the description of head/tail entity with the Shapley value, where the sentence wth higher rank is more important in the extraction of sentence-level information.

Future Work

- About the structural information from KG, in the next work we will consider aggregating information from the neighbors of triple to learn a better representation for entities and relations.
- About the textual information from the description, in the next work we will incorporate multi-lingual descriptions. Because knowledge insufficiency in a language can be alleviated in other languages.

Authors and Affiliations

Mojtaba Nayyeri* mojtaba.nayyeri@ipvs.uni-stuttgart.de University of Stuttgart
Zihao Wang* zihao.wang@ipvs.uni-stuttgart.de University of Stuttgart
Mst. Mahfuja Akter* mahfuja.ruby85@gmail.com University of Bonn
Mirza Mohtashim Alam turzo.mohtasim@gmail.com InfAI-The Institute for Applied Informatics association
Md Rashad Al Hasan Rony rashad.rony@iais.fraunhofer.de Fraunhofer IAIS
Jens Lehmann jens.lehmann@cs.uni-bonn.de University of Bonn
Steffen Staab steffen.staab@ipvs.uni-stuttgart.de University of Stuttgart

*these authors contributed equally

ellis unit | STUTTGART   University of Stuttgart Germany