Quantum Embedding of Knowledge for Reasoning

Dinesh Garg[1], Shajith Ikbal[1], Santosh K. Srivastava[1], Harit Vishwakarma[2], Hima Karanam[1], L Venkata Subramaniam[1]

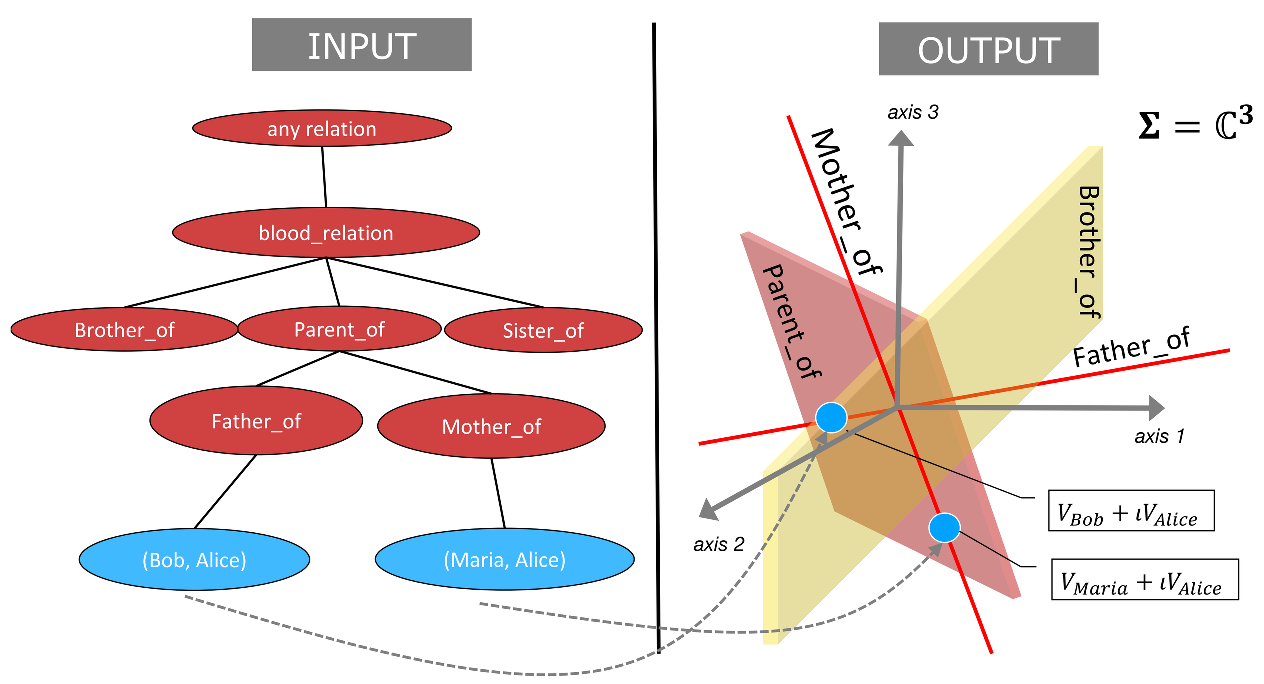
[1]IBM Research AI, India. [2]Dept. of Computer Sciences, UW-Madison, USA

garg.dinesh, shajmoha, sasriva5@in.ibm.com, hvishwakarma@cs.wisc.edu, hkaranam, lvsubram@in.ibm.com

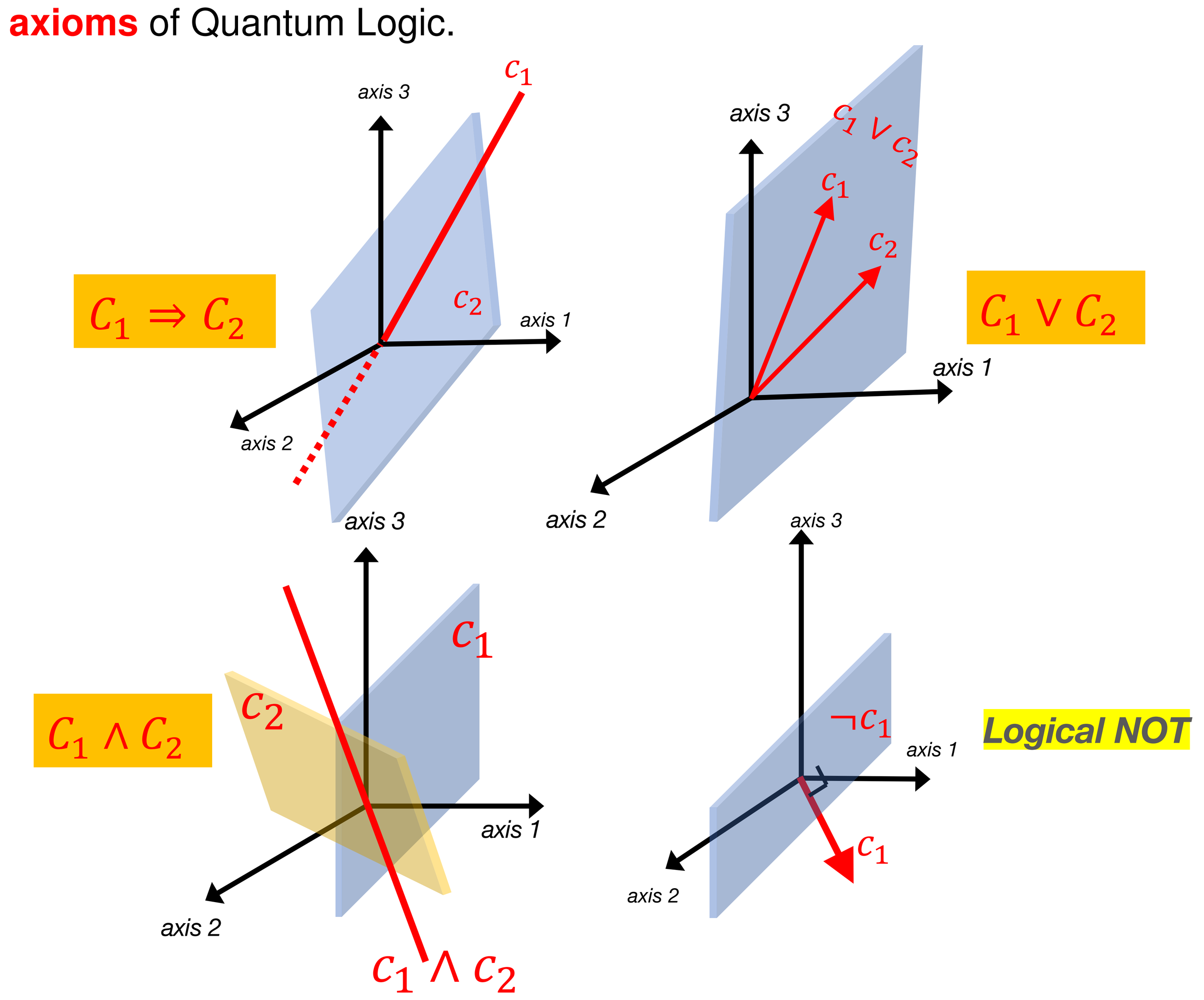
What and Why?

What are we doing? (1) Technique to embed Symbolic KB into a vector space while preserving logical structure. (2) Ability to perform logical operations on such embeddings in a manner similar to the Boolean Logical operations on a symbolic KB.
Why are we doing? Such embeddings can be leveraged by sub-symbolic (e.g. neural) methods to accomplish complex reasoning tasks, including (a) Knowledge Completion (Inductive Reasoning), and (b) Complex Membership Queries (Deductive Reasoning).

Idea Behind Quantum Embedding

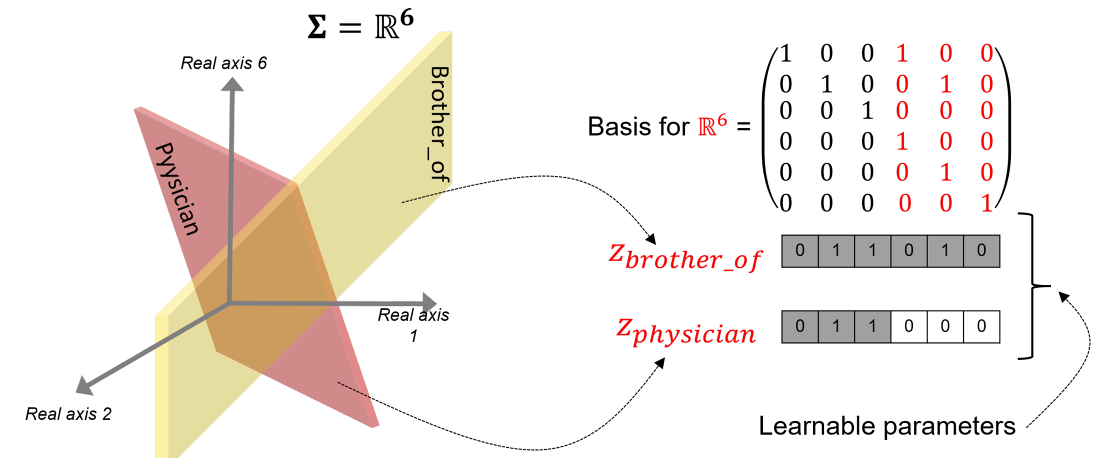


- Each **predicate (unary or binary)** should be represented by a **linear subspace** of a (real or complex) vector space Σ, where $\Sigma = \mathbb{R}^d$ (or $\mathbb{C}^d$) for some integer d.
- All the **entities (or entity pairs)** should be denoted by **(complex) vectors** in a way that they lie in each of the predicate subspaces to which they belong.
- The **axes** of Σ represent **latent semantic attributes** of the entities and entity pairs.
- In general, Σ could be any **finite/infinite dimensional Hilbert space.**
- The idea is inspired from the theory of **Quantum Logic** [1] and hence, embedding is constrained in a way that **geometry** of the predicate subspaces and entity vectors respect the **axioms** of Quantum Logic.

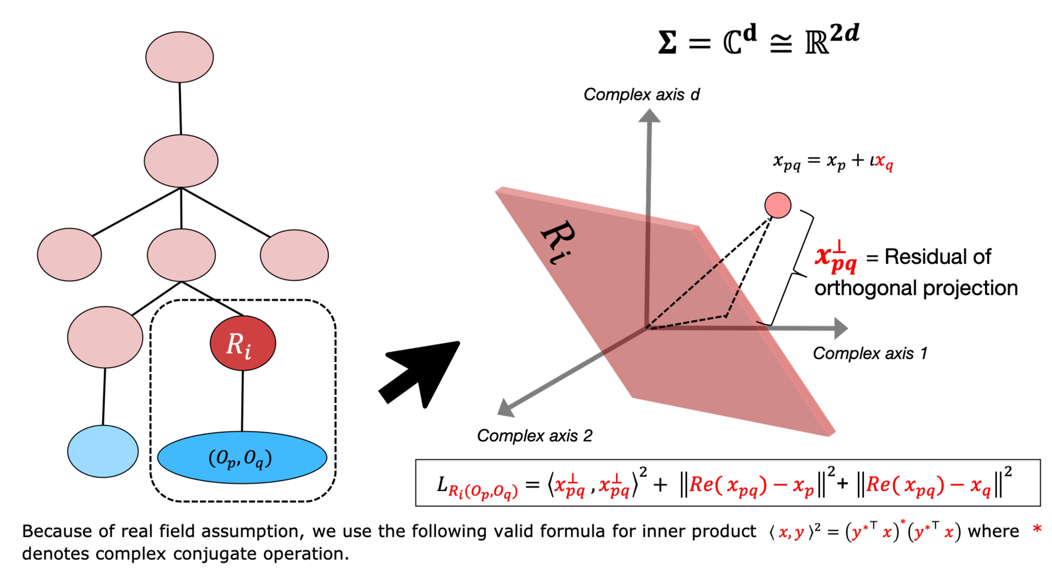


Learning Quantum Embedding

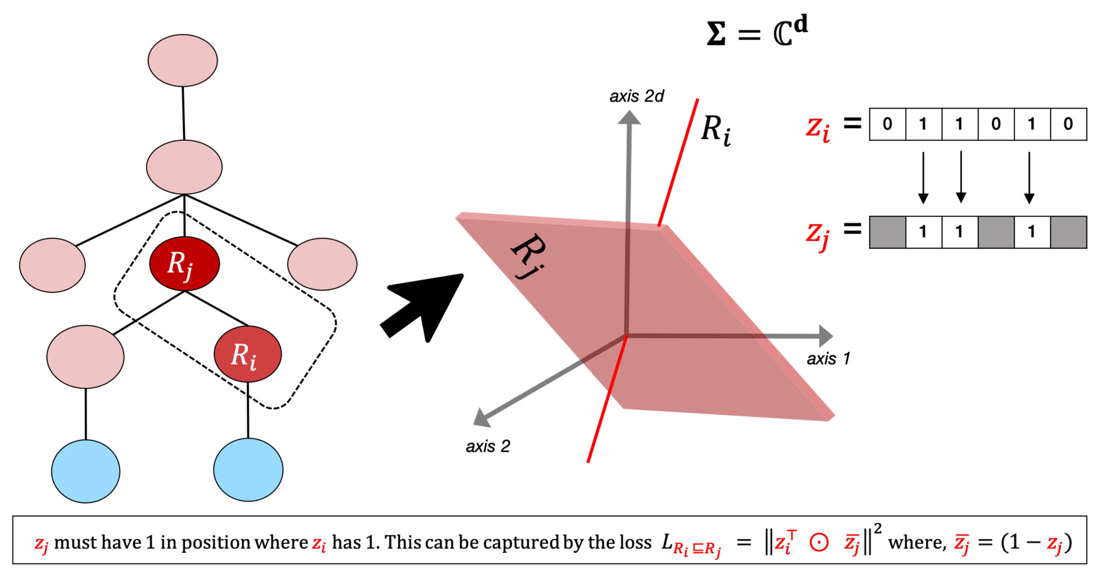
Axis-Parallel Quantum Embedding: For computational reasons, we restrict to the axis-parallel subspaces of $\mathbb{C}^d$. We learn axis-parallel subspaces of $\mathbb{C}^d$ indirectly by learning indicator vectors z for standard basis of $\mathbb{R}^{2d}$ (because it is isomorphic to $\mathbb{C}^d$ under real field). We also proved that distributive law holds for axis-parallel subspaces which otherwise does not hold true for QL in general.



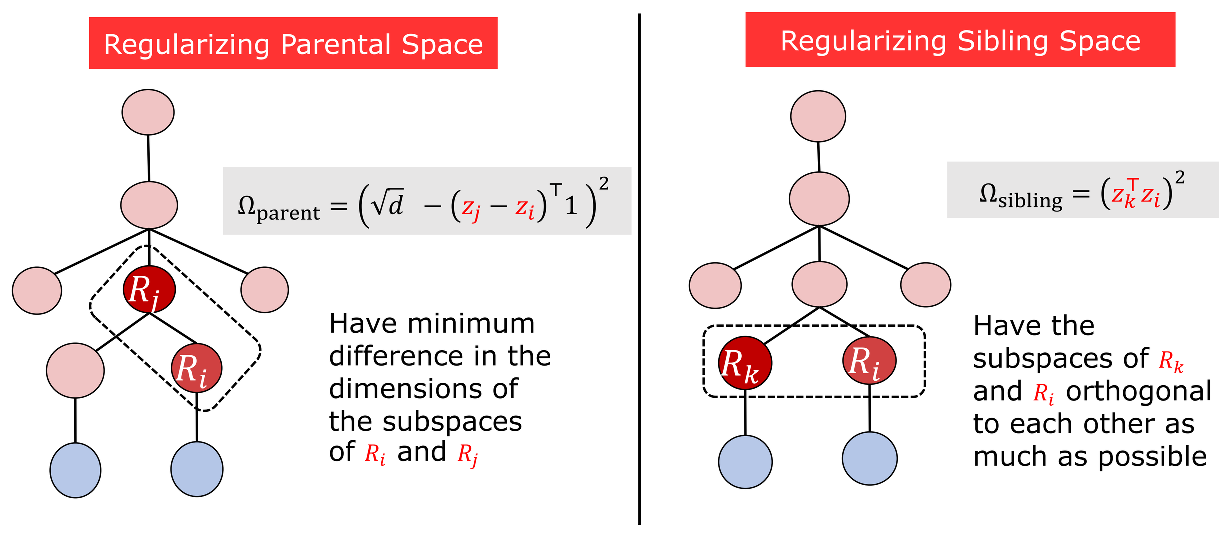
Loss for Membership Constraint:



Loss for Implication Constraint:



Parent-Sibling Regularization Loss:



Overall Learning Problem - Embed2Reason (E2R)

$$\min_{z_i, V_e} \left(L_{R_i(e_1,e_2)} + L_{R_i \sqsubseteq R_j} + \Omega_{parent} + \Omega_{sibling} \right)$$

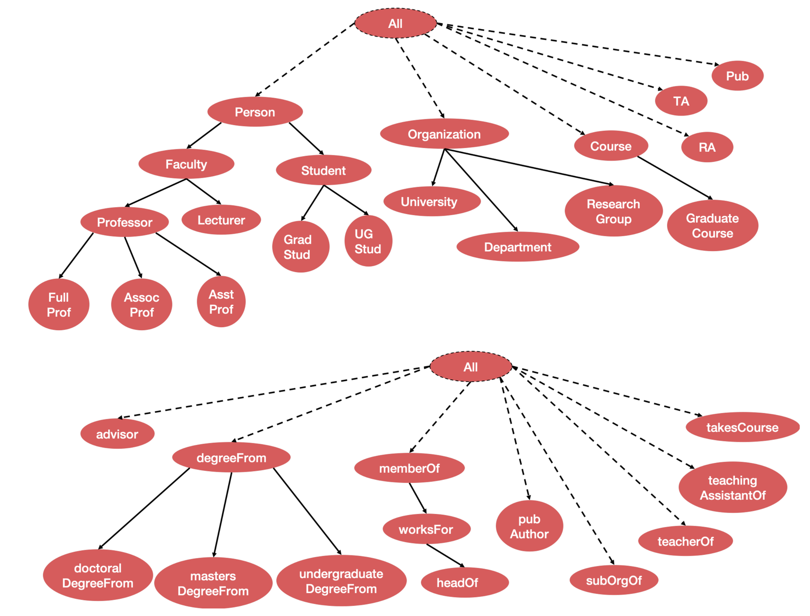
$s.t.$ z_i are binary valued vectors

By approximating integer constraints, we convert E2R program into an unconstrained optimization problem.

Experiments and Results

- Used SGD to get (approximate) local minima of E2R.
- Resulting embeddings are an approximation to the quantum embedding.
- Evaluated on two different tasks – (i) link prediction, and (ii) reasoning.

Task	Link Prediction	Finding Members of a Complex Concept
Type of Reasoning	Predictive	Deductive
Dataset	FB15K, WN18	LUBM1U
T-Box	Absent	Present
A-Box	Present	Present
Predicate Types	[WN18] Only binary Predicates [FB15K] Only binary Predicates	Both Unary and Binary
Input KB Size	[A-Box for WN18] 141442 triples [A-Box for FB15K] 483142 triples	[A-Box] 69628 triples [T-Box] 18 triples
Test Set Size	[WN18] 5000 [FB15K] 59071	8 membership queries for non-leaf concepts
Performance Metrics	Mean Rank, MRR, Hits@1, Hits@10	Mean Rank, MRR, Hits@1, Hits@10
Baslines	TransE[2], ComplEx [3]	TransE [2], ComplEx [3]



8 test queries for LUBM: members of *Professor, Faculty, Person, Student, Course, Organization, MemberOf, WorksFor*. Used $d = 100$, and TransE [2], ComplEx [3] as baselines.

Data	Mean Rank			MRR			Hits@1 (%)			Hits@10 (%)		
	E2R	TE	CE	E2R	TE	CE	E2R	TE	CE	E2R	TE	CE
FB15K	72.0	68.4	114.0	**0.96**	0.49	0.61	**96.4**	34.8	49.8	**96.4**	76.7	81.2
WN18	5780.2	409.9	468.1	0.71	0.63	0.90	71.1	41.0	87.4	71.1	93.2	95.25
LUBM1U	**220.1**	1292.6	5742.9	**0.46**	0.26	0.12	**45.4**	18.97	12.5	45.4	49.1	12.59

Insights

- E2R often ranks a ground truth entity either at Rank-1 or at a quite low rank.
- For WN18 dataset, binary relation satisfy transitivity property (e.g. hypernym) and have inverse relations (e.g. hypernym/hyponym).
- Baselines approaches are primarily distanced based and hence capture transitivity/inversion properties better than E2R.

References

[1] Garrett Birkhoff and John Von Neumann. The logic of quantum mechanics. *The Annals of Mathematics*, 37(4):823–843, 1936.

[2] Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *NIPS*, pages 2787–2795, 2013.

[3] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Eric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *ICML*, pages 2071–2080, 2016.

NeurIPS, December 8 - 14, 2019, Vancouver, Canada