Towards Fusion of Semantic Knowledge into Deep Learning Models

I.	**Neural-Symbolic Systems: Brief Overview and Ongoing Work**	**M. Ebrahimi (WSU)**
II.	**What is Multimodal Machine Learning?**	**J. Francis (CMU & Bosch)**
III.	**Multimodal Sense-making: a Natural Ground for Neural Symbolic Systems**	**A. Oltramari (Bosch)**
IV.	**Discussion**	

- **A. Questions/Comments**
- **B. "I am working on it!" Share your experience in 60 seconds**

Neural-Symbolic Systems: Representation and Reasoning Approaches

Monireh Ebrahimi

Data Semantics Laboratory (DaSe Lab)
Wright State University

Advisor:
Dr. Pascal Hitzler

Why Neuro-Symbolic? Neural Approaches

- **Connectionist machine learning systems:**
  - ✓ Very effective for some machine learning tasks
  - ✓ Robustness to noise
  - ✗ Not explainable and understandable
  - ✗ Poor in working with symbolic data
  - ✗ Not created for injection of domain knowledge
  - ✗ Requires large amounts of data for training
  - ✗ Really poor on handling new data/vocabulary/domain

Why Neuro-Symbolic?

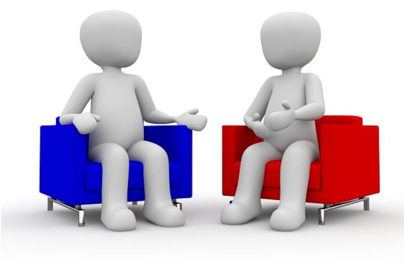
- **Symbolic systems:**
  - ✘ Poor at machine learning problems
  - ✘ Sensitive to noise
  - ✓ Explainable and and understandable
  - ✓ Really good at working with symbolic data
  - ✓ Easy background knowledge injection
  - ✓ No training is required
  - ✓ Transferable to new data/vocabulary/domain in the same logic

Why Neuro-Symbolic?

- **Neuro-Symbolic systems:**
  - ✓ Very powerful for machine learning tasks
  - ✓ Robust to data noise
  - ✓ More explainable and understandable
  - ✓ Easy domain knowledge injection
  - ✓ Good at working with symbols

Neuro-Symbolic Gap?

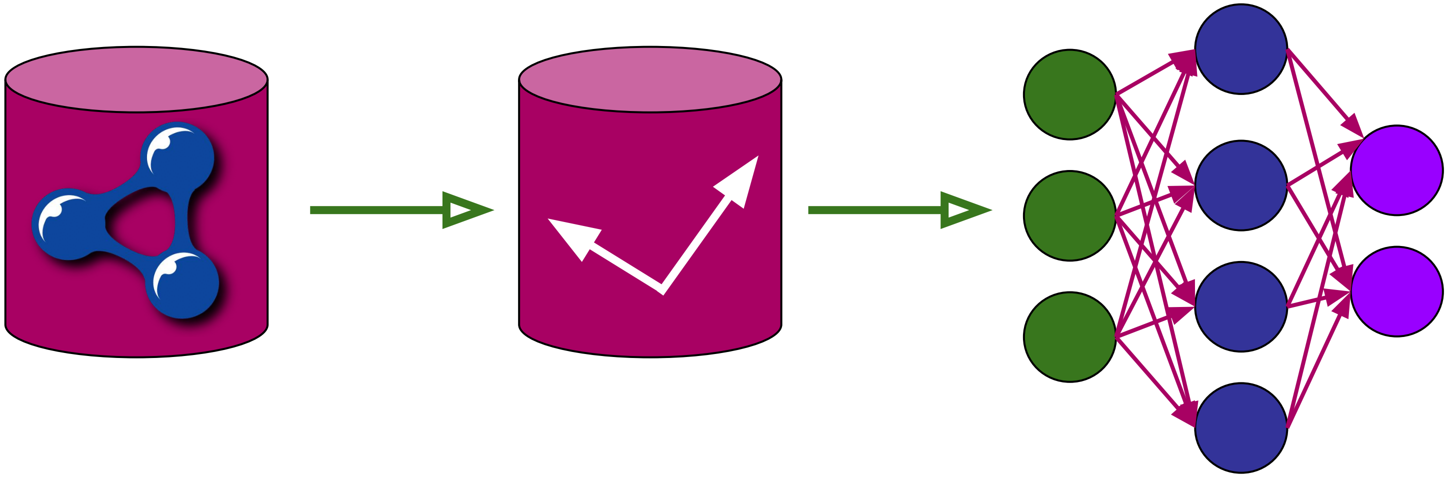
- **Neural-Symbolic gap:**
  - Neural learners use numerical low level representations
  - Symbolic reasoners use atomic symbolic representations



- **How to do that?**
  - Representation?
  - Reasoning Approach?

Neuro-Symbolic RDF Reasoning: Current Approaches

- How can we represent the RFDS logic symbols in a way that it can fed to neural learners?
  - Graph embedding approaches for link prediction problem
    - knowledge base completion is the most relevant effort for bridging this gap



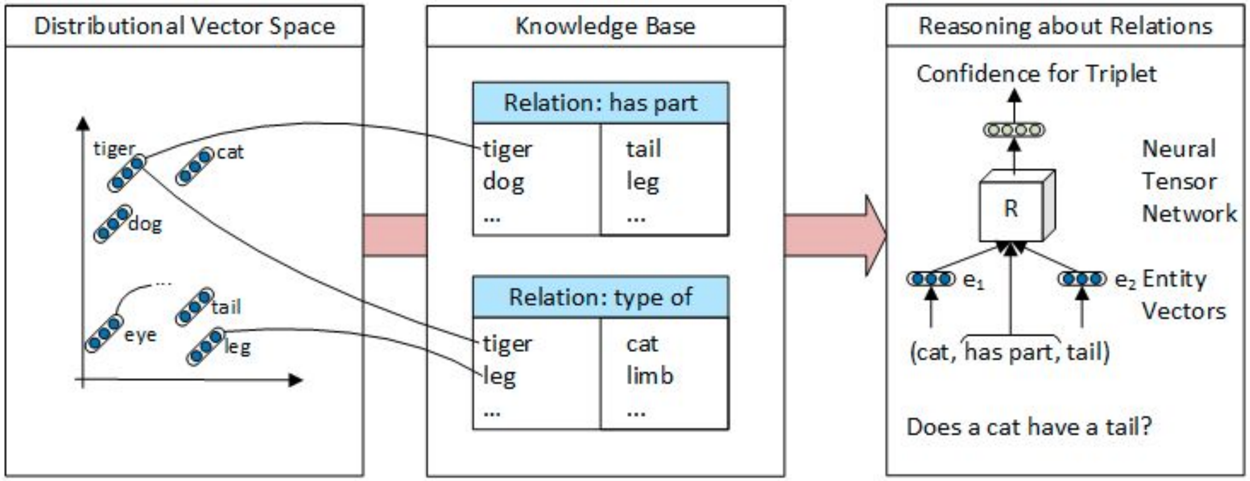
Knowledge Graph Reasoning: Current Approaches (Representation)

- Embedding of knowledge graphs:
  - Translational distance models:
    - TransE: $h + r \approx t$
    - Maximizing $f_r(h,t)$ for the KG facts
    - Does not support n-ary relationships
    - TransH, TransR, TransD
  - Semantic Matching models:
    - Maximizing $f_r(h,t) = h^t M_r t$ for the KG facts
    - M_r is relation adjacency matrix
    - RESCAL, DistMult, ComplEx, HOPE
    - High-Order Proximity preserved Embedding (HOPE)

Knowledge Graph Reasoning: Current (Approaches)

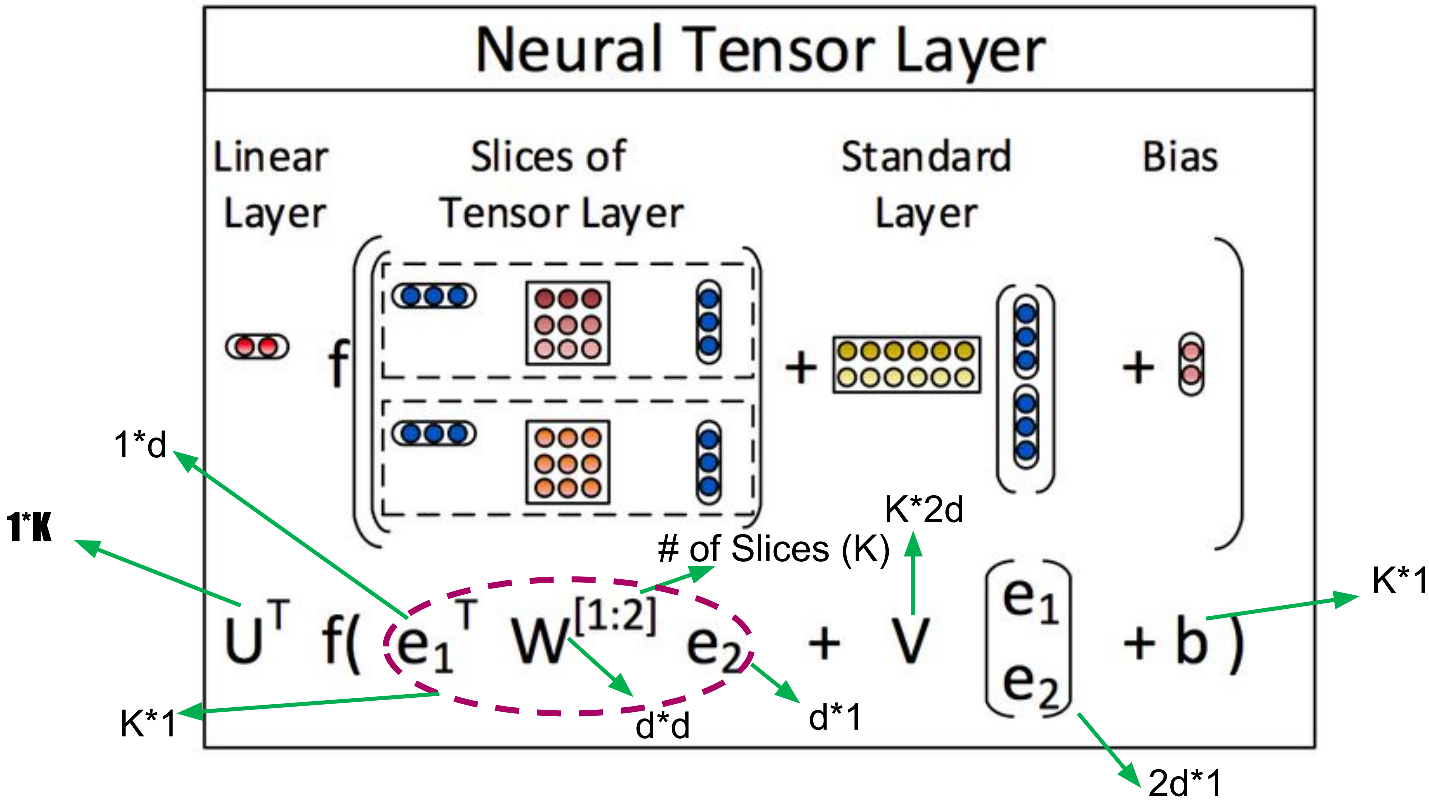
Neural Tensor Networks

- The Neural Tensor Network(NTN), replaces a standard linear neural network layer with a tensor layer
- The model computes the score of how likely it is that two entities are in a certain relationship by the following NTN based function: $g(e_1, R, e_2)$



Knowledge Graph Reasoning: Current (Approaches)

Neural Tensor Networks



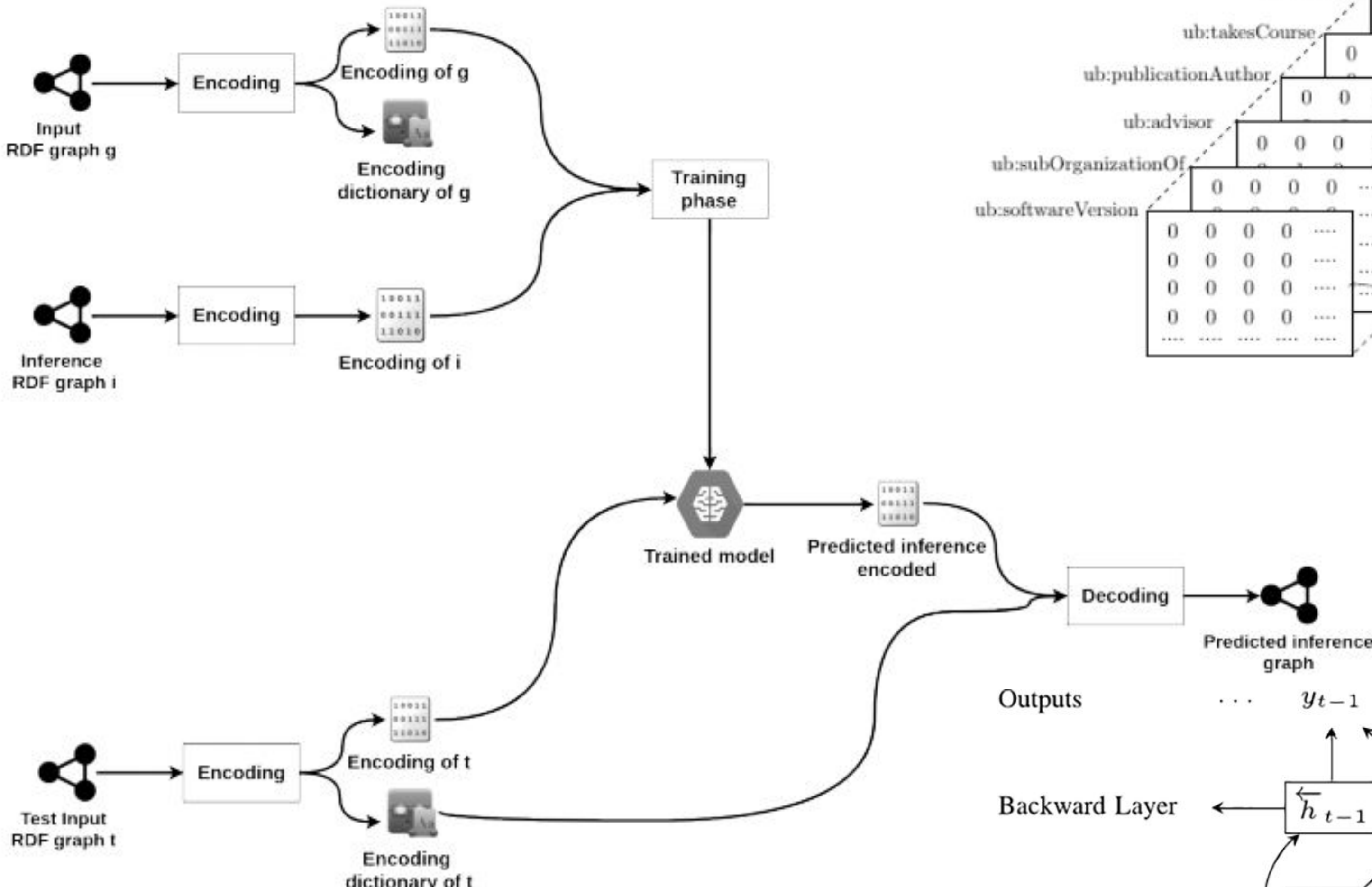
Neuro-Symbolic RDF Reasoning: Current Approaches

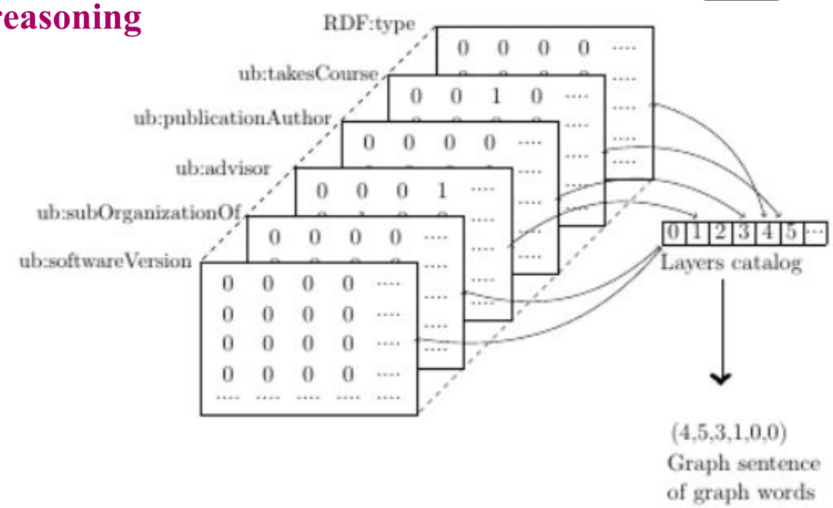
- RDF embedding Techniques:
  - Graph Kernels for RDF:
    - intersection graph kernel
    - intersection tree kernel
    - Weisfeiler-lehman graph kernels
    - h-hop neighbourhood-based graph kernel for LOD
- 2vec RDF embedding:
  - RDF2vec

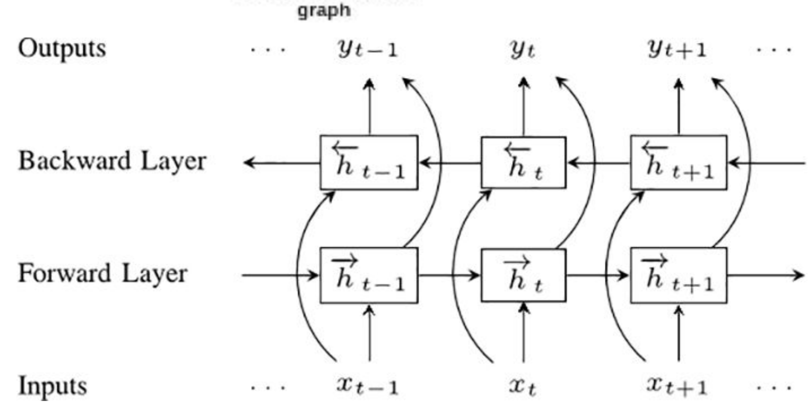
Neuro-Symbolic Reasoning: Current Approaches

DaSe Lab

B. Makni and J. Hendler / Deep learning for noise-tolerant RDFS reasoning





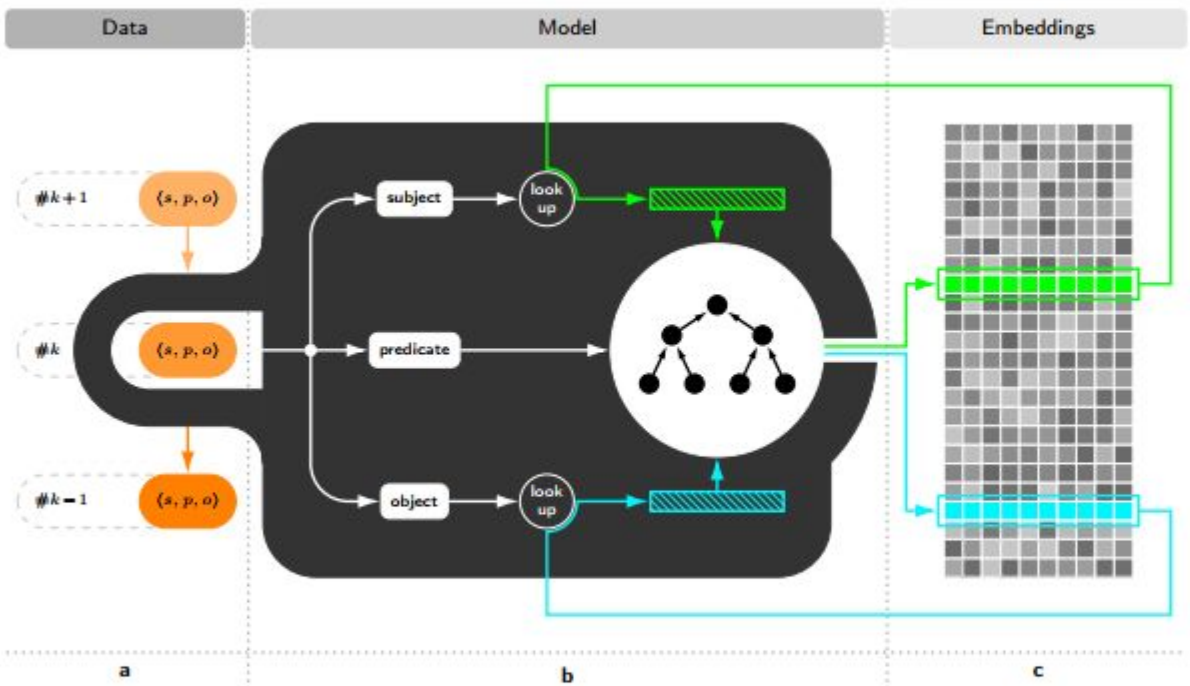


Neuro-Symbolic Reasoning: Current Approaches

DaSe Lab

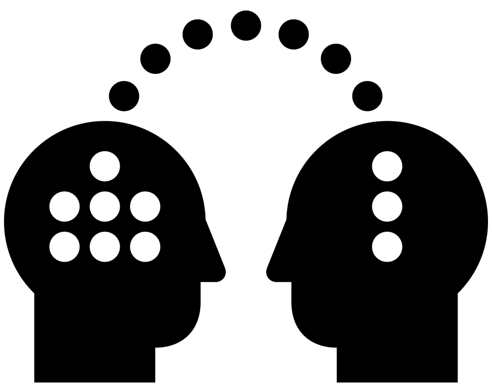
Recursive Reasoning Network (RRN):

P. Hohenecker and T. Lukasiewicz/Ontology Reasoning with Deep Neural Networks



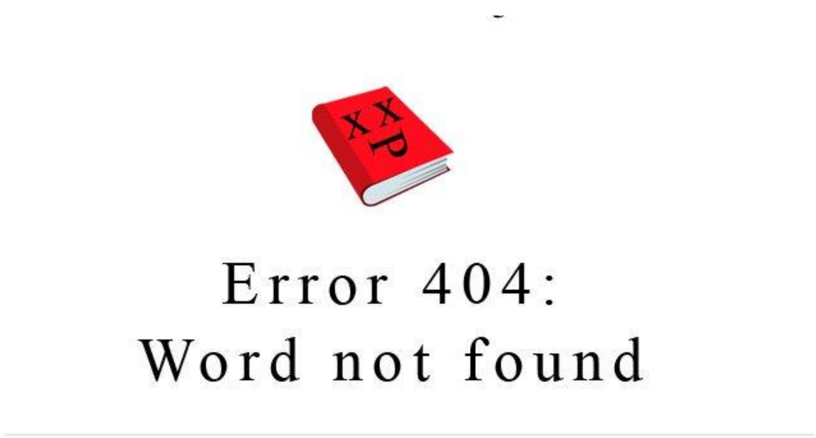
Problem with Previous Works

- What you’d really like to do is:
  - Train a deep learning system so that you can present a new, unseen graph to it, and it can correctly derive the deductively inferred triples.
- Note:
  - You don’t know the IRIs in the graph up front. The only overlap is the IRIs in the rdf/s namespace.



Issues?

- Out-of-Vocabulary Problem
- Reasoning based on the **similarity/relatedness** and geometric-based proximity of real-valued vectors as opposed to deductive reasoning



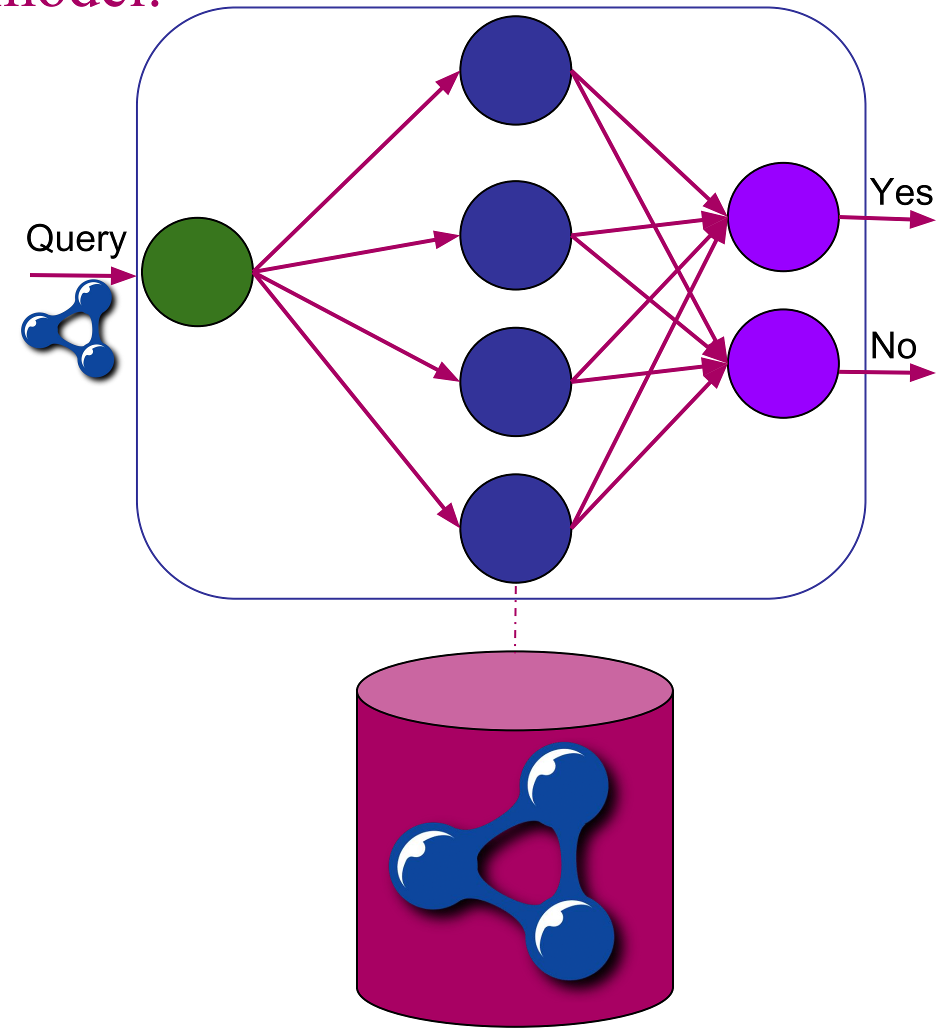
RDFS Deductive Reasoning via Deep Memory Networks

**Monireh Ebrahimi, Md Kamruzzaman Sarker,
Federico Bianchi, Ning Xie, Aaron Eberhart,
Derek Doran, Pascal Hitzler**

Under Review

Idea overview

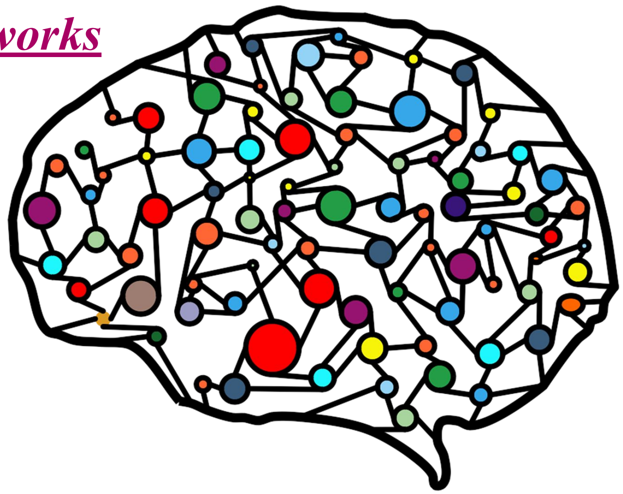
Classification model:



Approach:
Why Memory Networks?

- Temporal Structure → RNN
- Spatial Structure → CNN

- Out of order access
- Long-term dependency
- Unordered set
- E.g., question answering on story → ***Memory Networks***



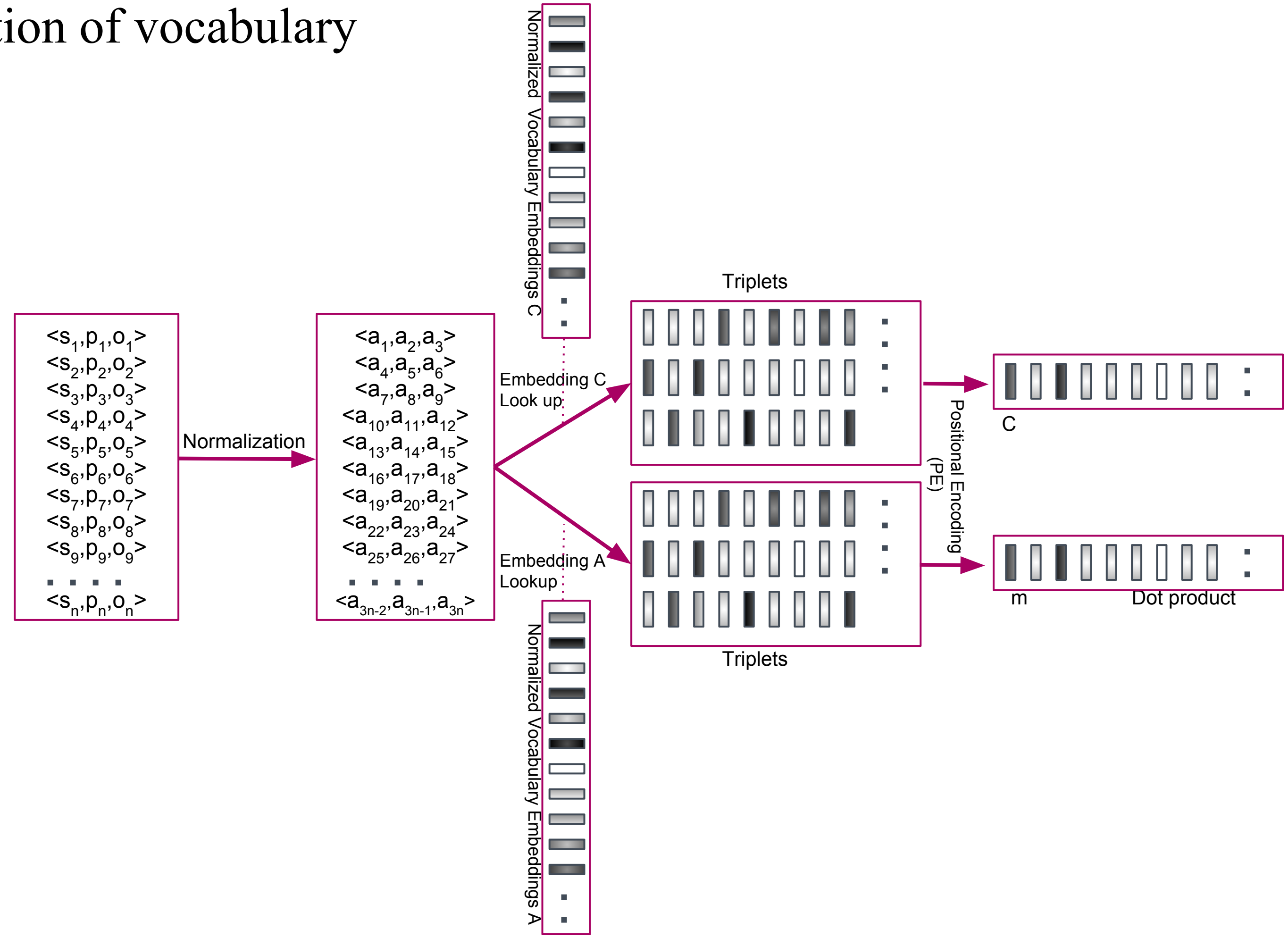
What is Memory Network?

- End to end memory networks (Memn2n)
  - Neural network model with external memory
  - Reads from memory with soft attention (as opposed to hard attention in memory networks)
    - Multiple hops/lookups over memory
  - End to end training using back-propagation

Representation

Knowledge Graph Representation:

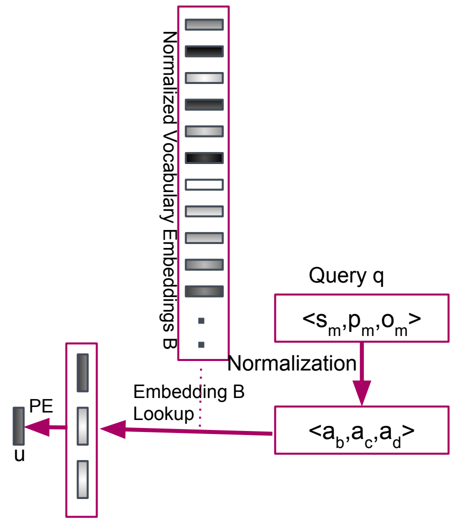
Normalization of vocabulary



Representation

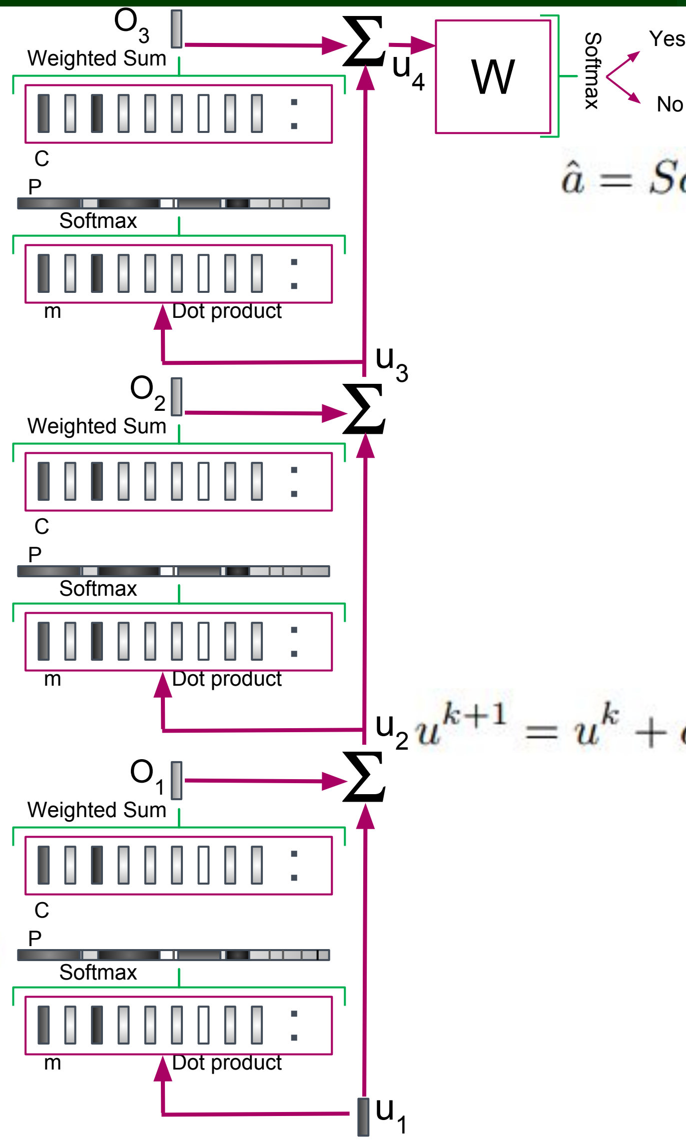
Query Representation:

Same normalization



Mechanics

3-hop attention model



$$\hat{a} = Softmax(W(u^{K+1}))$$

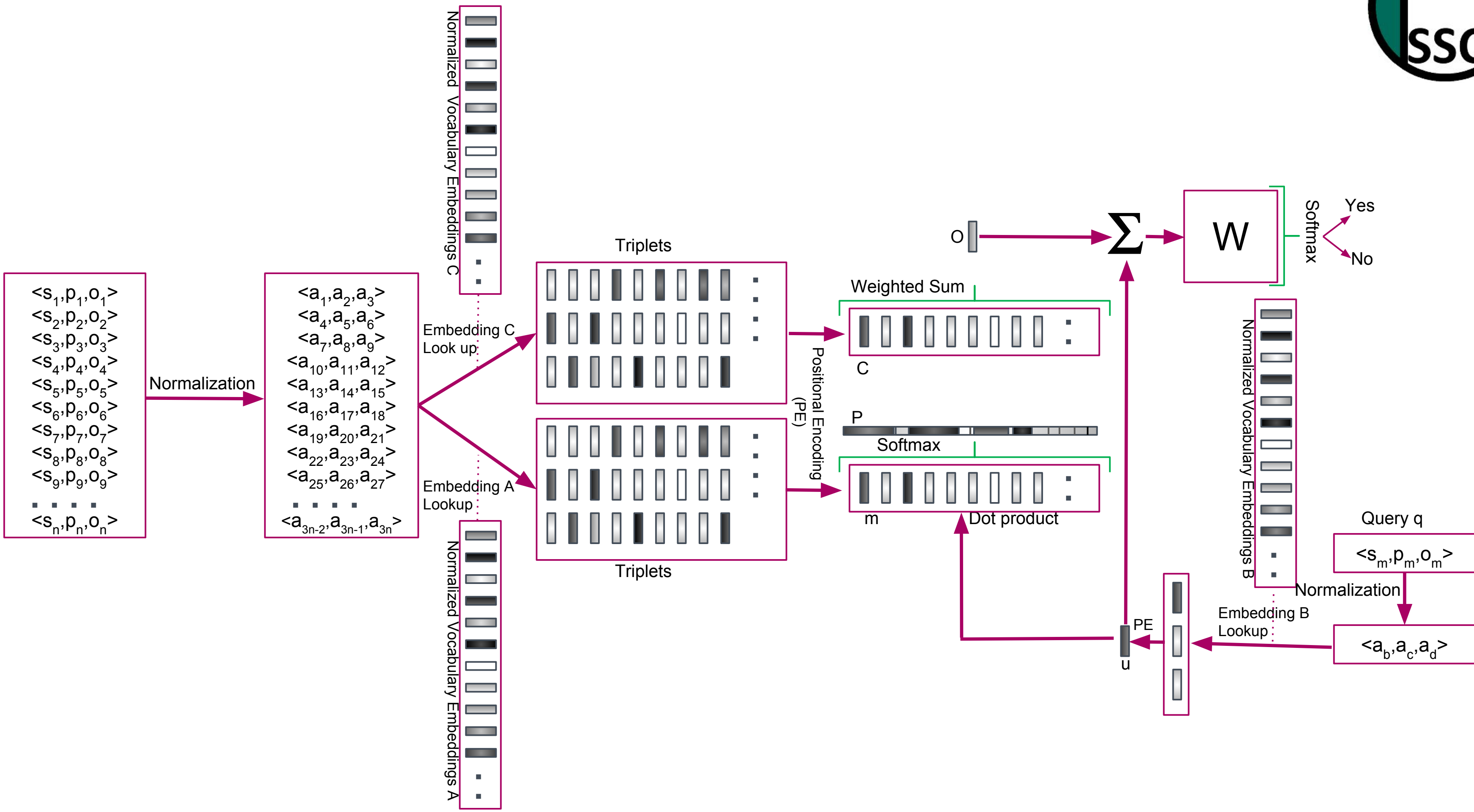
$$u^{k+1} = u^k + o^k$$

$$o = \sum_i p_i c_i$$

$$p_i = Softmax(u^T(m_i))$$

Model

Single layer Model



Dataset

Dataset	Ontologies
OWL-Centric	Amino Acid Ontology schema, Biological Pathway Exchange (BioPAX) schema, COmmon Semantic MOdel (COSMO), dbpedia-schema, Descriptions and situation, Disease, Dolce, Dublin_core schema, Gene, General formal ontology (GFO), Human Phenotype, Institutional Ontology, Metadata for Ontology Description and publication, Ontology for Biomedical Investigations, Phenotypic quality, Schema.org, University of Lehigh benchmark, Xenopus anatomy and development, Yet Another More Advanced Top-level Ontology (YAMATO).
Linked Data	AGROVOC Linked Dataset, Amsterdam Museum Linked Open Data, The Apertium Bilingual Dictio- naries on the Web of Data, A Curated and Evolving Linguistic Linked Dataset (Asit), EARTh: an Environmental Application Reference Thesaurus in the Linked Open Data Cloud data, lemonUby - a large, interlinked, syntactically-rich lexical resource for ontologies, Linked European Television Heritage data, Linked Web APIs Dataset: Web APIs meet Linked Data.
OWL-Centric Test Set	Animal Health Surveillance Ontology, Cryptographic ontology of Semantic interoperability for rapid integration and deployment, Drug Abuse Ontology, Drug target ontology, General Ontology for Linguistic Description (GOLD), Identification ontology, Inline Hockey League pattern ontology, Knowledge processing ontology for Robots, Minimal category of list ontology, Provenance and Plans ontology , SAREF: the Smart Appliances REFerence ontology, Tatian Corpus of Deviating Examples (T-CODEX) ontology

Table 1: List of ontologies used to create our datasets

Test Dataset	#KG	Base						Inferred						Invalid
		#Facts	#Ent.	%Class	%Indv	%R.	%Axiom.	#Facts	#Ent.	%Class	%Indv	%R.	%Axiom.	#Facts
OWL-Centric	2464	996	832	14	19	3	0	494	832	14	0.01	1	20	462
Linked Data	20527	999	787	3	22	5	0	124	787	3	0.006	1	85	124
OWL-Centric Test Set	21	622	400	36	41	3	0	837	400	36	3	1	12	476
Synthetic Data	2	752	506	52	0	1	0	126356	506	52	0	1	0.07	700

Table 2: Statistics of various datasets used in experiments

Results

Training Dataset	Test Dataset	Valid Triples Class			Invalid Triples Class			Accuracy
		Precision	Recall /Sensitivity	F-measure	Precision	Recall /Specificity	F-measure	
OWL-Centric Dataset	Linked Data	93	98	96	98	93	95	**96**
OWL-Centric Dataset (90%)	OWL-Centric Dataset (10%)	88	91	89	90	88	89	**90**
OWL-Centric Dataset	OWL-Centric Test Set [b]	79	62	68	70	84	76	**69**
OWL-Centric Dataset	Synthetic Data	65	49	40	52	54	42	**52**
OWL-Centric Dataset	Linked Data [a]	54	98	70	91	16	27	86
OWL-Centric Dataset [a]	Linked Data [a]	62	72	67	67	56	61	91
OWL-Centric Dataset(90%) [a]	OWL-Centric Dataset(10%) [a]	79	72	75	74	81	77	80
OWL-Centric Dataset	OWL-Centric Test Set [a b]	58	68	62	62	50	54	58
OWL-Centric Dataset [a]	OWL-Centric Test Set [a b]	77	57	65	66	82	73	73
OWL-Centric Dataset	Synthetic Data [a]	70	51	40	47	52	38	51
OWL-Centric Dataset [a]	Synthetic Data [a]	67	23	25	52	80	62	50
Baseline								
OWL-Centric Dataset	Linked Data	73	98	83	94	46	61	43
OWL-Centric Dataset (90%)	OWL-Centric Dataset (10%)	84	83	84	84	84	84	82
OWL-Centric Dataset	OWL-Centric Test Set [b]	62	84	70	80	40	48	61
OWL-Centric Dataset	Synthetic Data	35	41	32	48	55	45	48

[a] More Tricky Nos & Balanced Dataset
[b] Completely Different Domain.

Table 3: Experimental results of proposed model

Results

Test Dataset	Hop 0			Hop 1			Hop 2			Hop 3			Hop 4			Hop 5			Hop 6			Hop 7			Hop 8			Hop 9			Hop 10		
	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F	P	R	F
Linked Data[a]	0	0	0	80	99	88	89	97	93	77	98	86	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Linked Data[b]	2	0	0	82	91	86	89	98	93	79	100	88	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OWL-Centric [c]	19	5	9	31	75	42	78	80	78	48	47	44	4	34	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Synthetic	32	46	33	31	87	38	66	55	44	25	45	32	29	46	33	26	46	33	25	46	33	25	46	33	24	43	31	25	43	31	22	36	28

[a] LemonUby Ontology
[b] Agrovoc Ontology
[c] Completely Different Domain

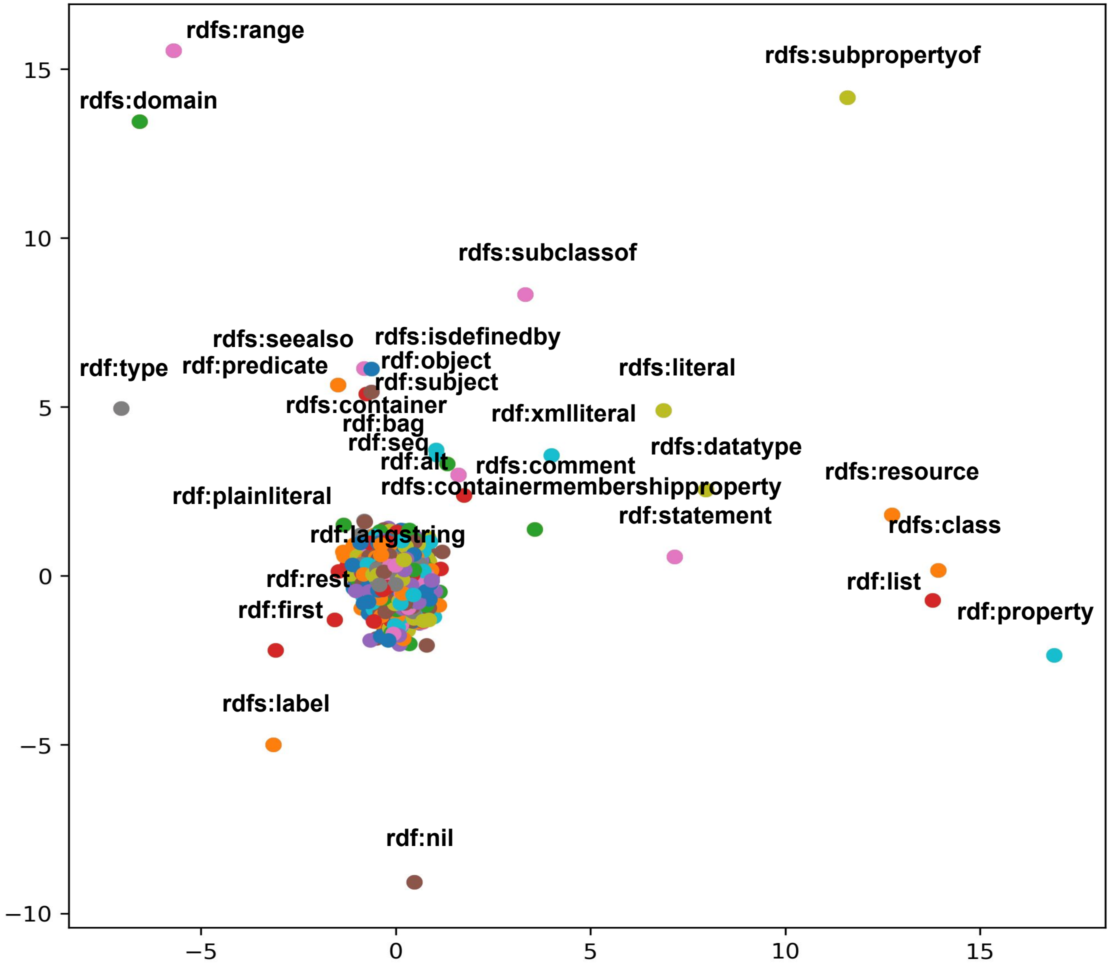
Table 4: Experimental results over each reasoning hop

Dataset	Hop 1	Hop 2	Hop 3	Hop 4	Hop 5	Hop 6	Hop 7	Hop 8	Hop 9	Hop 10
OWL-Centric [a]	8%	67%	24%	0.01%	0%	0%	0%	0%	0%	0%
Linked Data[b]	31%	50%	19%	0%	0%	0%	0%	0%	0%	0%
Linked Data[c]	34%	46%	20%	0%	0%	0%	0%	0%	0%	0%
OWL-Centric[d]	5%	64%	30%	1%	0%	0%	0%	0%	0%	0%
Synthetic Data	0.03%	1.42%	1%	1.56%	3.09%	6.03%	11.46%	20.48%	31.25%	23.65%

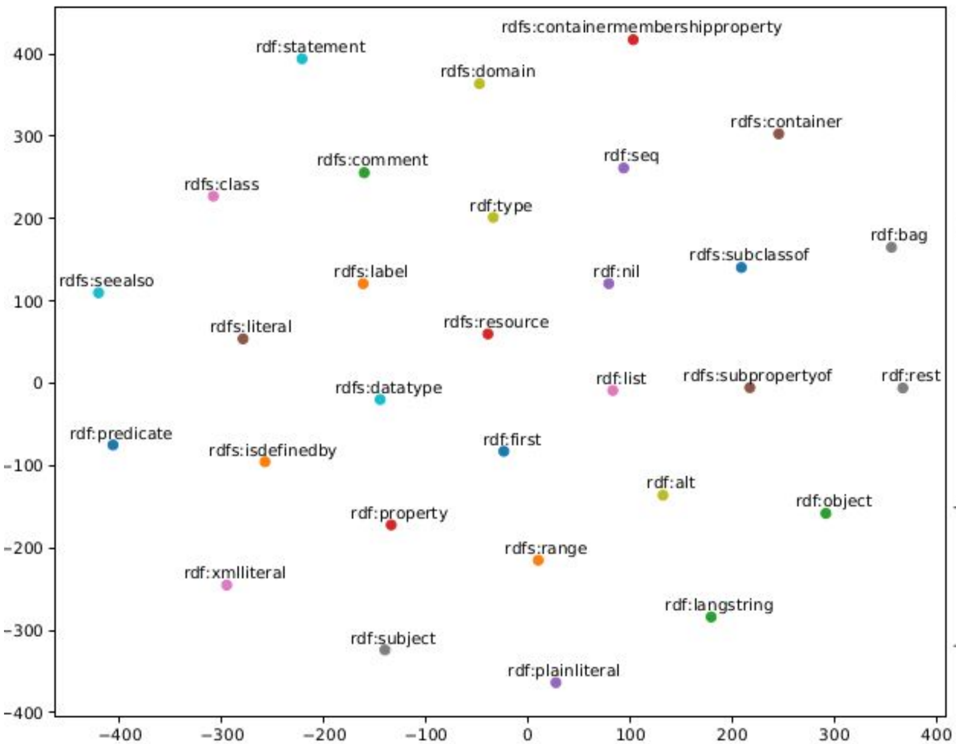
[a] *Training Set*
[b] LemonUby Ontology
[c] Agrovoc Ontology
[d] Completely Different Domain

Table 5: Data distribution per knowledge graph over each reasoning hop

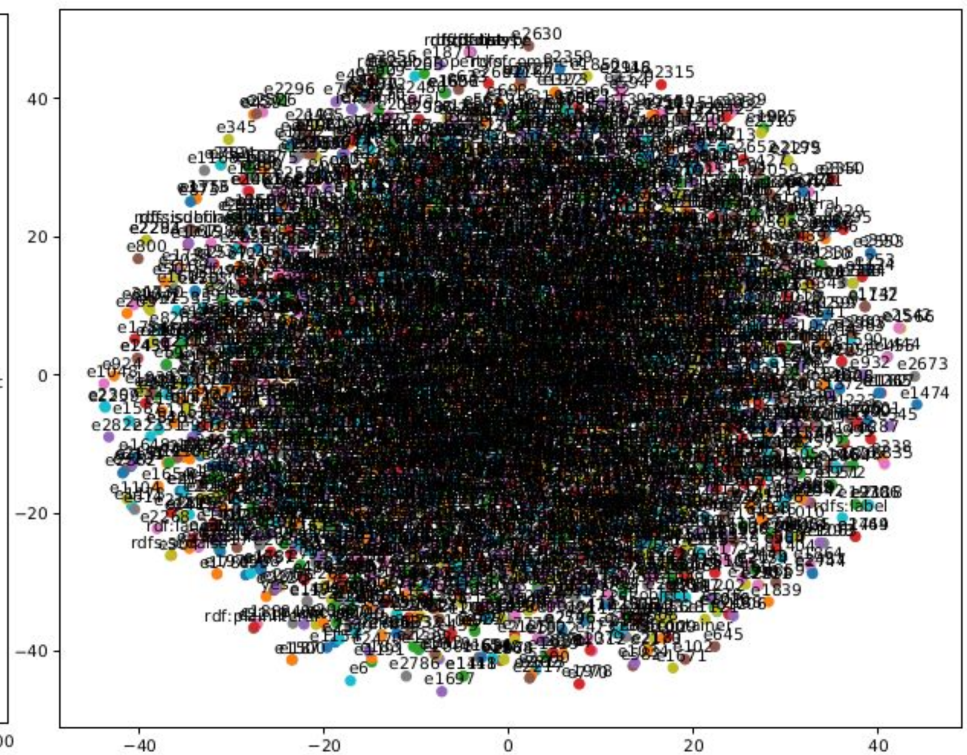
PCA projection of embeddings for the whole general vocabulary



tsne projection of embeddings for the whole general vocabulary



(a) Embeddings for RDF(S) namespace



(b) Embeddings for the whole general vocabulary

Summary of Contributions

1. Construction of memory networks for emulating the symbolic deductive reasoning.
2. Using normalization approach to enhance their transfer capability. We examine the efficacy of our model for cross-domain and cross-knowledge graph deductive reasoning.

Current & Future Work

DaSe Lab

SSC

ER Logic:

$C \sqsubseteq D$ --> P1 C D

$C1 \sqcap C2 \sqsubseteq D$ --> P2 C1 C2 D

$C \sqsubseteq \exists R.D$ --> P3 C R D

$\exists R.C \sqsubseteq D$ --> P4 R C D

$R1 \sqsubseteq R$ --> P5 R1 R

$R1 \circ R2 \sqsubseteq R$ --> P6 R1 R2 R

Completion Rules:

(1) $C \sqsubseteq D, A \sqsubseteq C \models A \sqsubseteq D$

(2) $C1 \sqcap C2 \sqsubseteq D, A \sqsubseteq C1, A \sqsubseteq C2 \models A \sqsubseteq D$

(3) $C \sqsubseteq \exists R.D, A \sqsubseteq C \models A \sqsubseteq \exists R.D$

(4) $\exists R.C \sqsubseteq D, A \sqsubseteq \exists R.B, B \sqsubseteq C \models A \sqsubseteq D$

(5) $R \sqsubseteq S, A \sqsubseteq \exists R.B \models A \sqsubseteq \exists S.B$

(6) $R1 \circ R2 \sqsubseteq R, A \sqsubseteq \exists R1.B, B \sqsubseteq \exists R2.C \models A \sqsubseteq \exists R.C$

Thank You