Towards Generalizable Neuro-Symbolic Reasoners

DaSe Lab

Kansas State University

Monireh Ebrahimi
Senior Cognitive Software Developer @ IBM Watson
monireh.ebrahimi@ibm.com

@monireheb

Project Funded by:
NSF-OIA-2033521
FA9550-18-1-0386

Neuro-Symbolic Reasoning: Outline

For a <u>deep deductive reasoner</u>, we would ideally like to have it on an <u>expressive logic</u>, with <u>transfer</u>, <u>generative</u>, <u>at massive scale</u>, and with high performance.

Method	Logic	Transfer	Generative	Scale	Performance
Memory Networks [5-7]	RDF	Yes	No	Moderate	High
Pointer Networks [8]	RDF, EL+	Yes	Yes	Moderate	High

On the Generalization Capability of Memory Networks for RDF Reasoning

RDF Deductive Reasoning?

- Resource Description Framework Schema (RDFS): Widely used standard for expressing knowledge graphs
  - One of the simplest useful knowledge representation languages beyond propositional logic

- node-edge-node triples such as:

  BarackObama rdf:type President

  President rdfs:subClassOf Human

- Fixed/small set of inference rules, such as:

  rdf:type(x,y) AND rdfs:subClassOf(y,z)THEN rdf:type(x,z)

- Logical consequence:

  BarackObama rdf:type Human

RDF Deductive Reasoning?

$$(x, \text{rdfs:subClassOf}, y), (y, \text{rdfs:subClassOf}, z) \models (x, \text{rdfs:subClassOf}, z) \tag{1}$$

$$(x, \text{rdfs:subPropertyOf}, y), (y, \text{rdfs:subPropertyOf}, z) \models (x, \text{rdfs:subPropertyOf}, z) \tag{2}$$

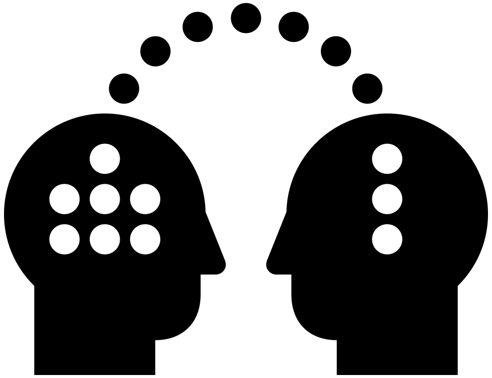
$$(x, \text{rdfs:subClassOf}, y), (z, \text{rdf:type}, x) \models (z, \text{rdf:type}, y) \tag{3}$$

$$(a, \text{rdfs:domain}, x), (y, a, z) \models (y, \text{rdf:type}, x) \tag{4}$$

$$(a, \text{rdfs:range}, x), (y, a, z) \models (z, \text{rdf:type}, x) \tag{5}$$

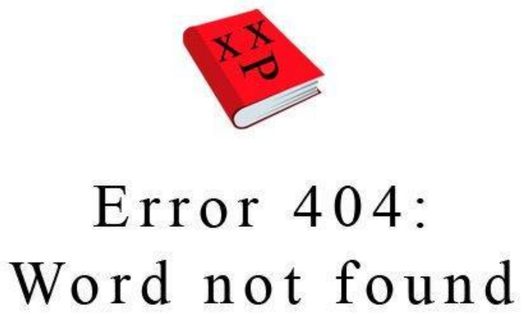
Problem with Previous Works

- What we'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 entity names in the graph up front. The only overlap is in the rdf/s namespace.



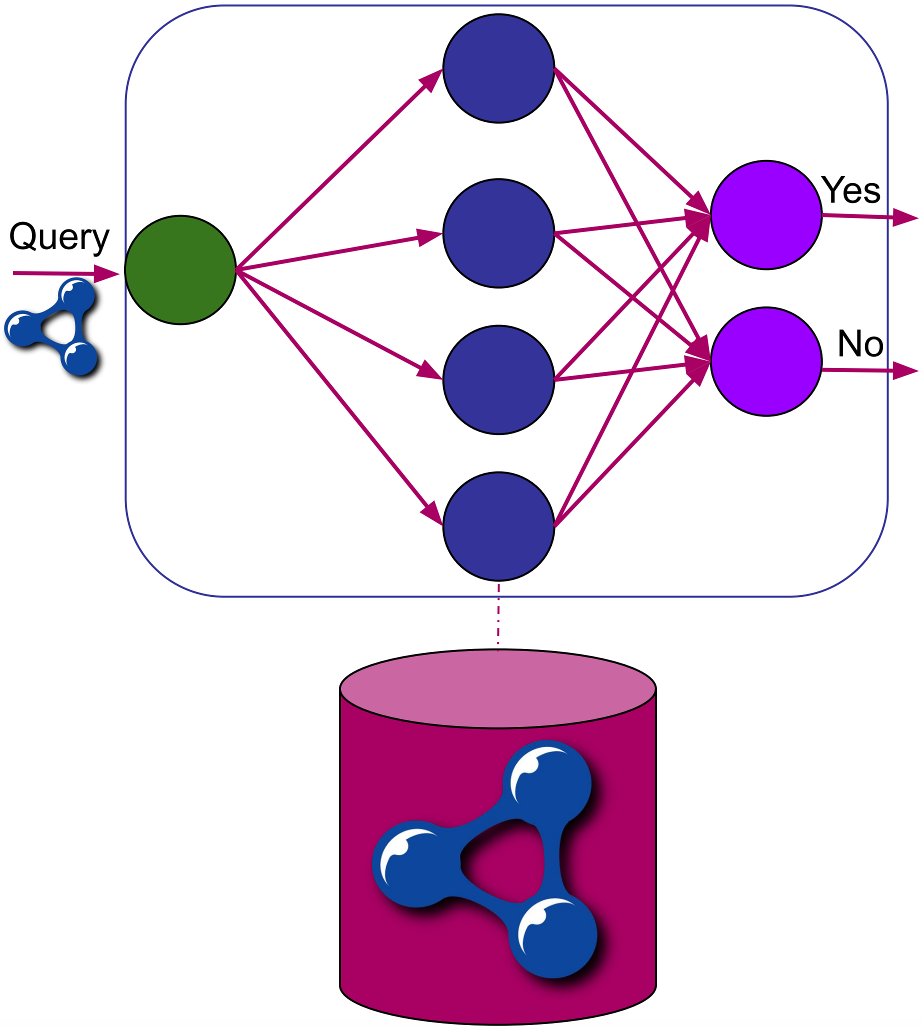
Challenges?

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



Idea Overview

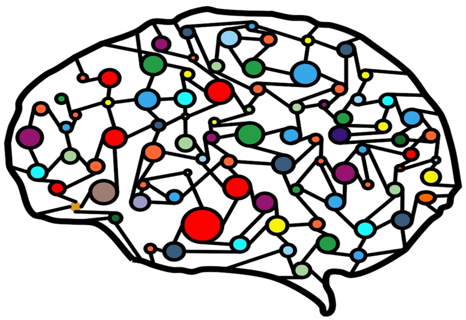
Classification model:



Approach: Why Memory Networks?

- Temporal Structure → **RNN**
- Spatial Structure → **CNN**

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

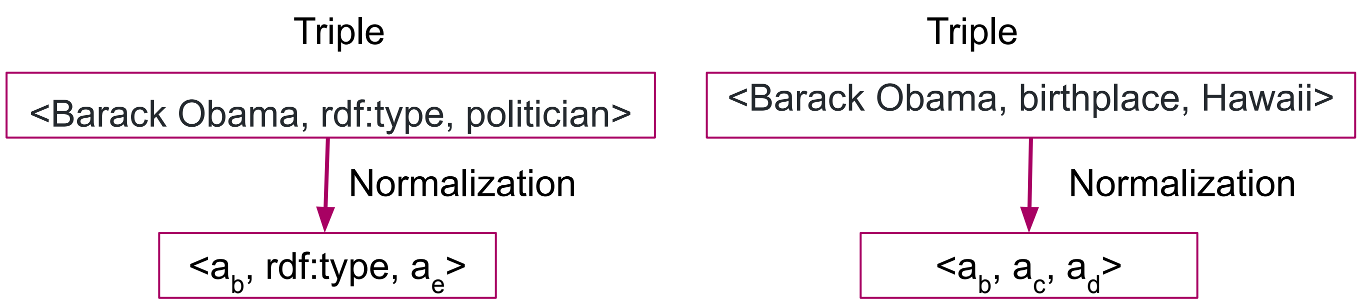


What is Memory Network?

- 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: Normalized Triples

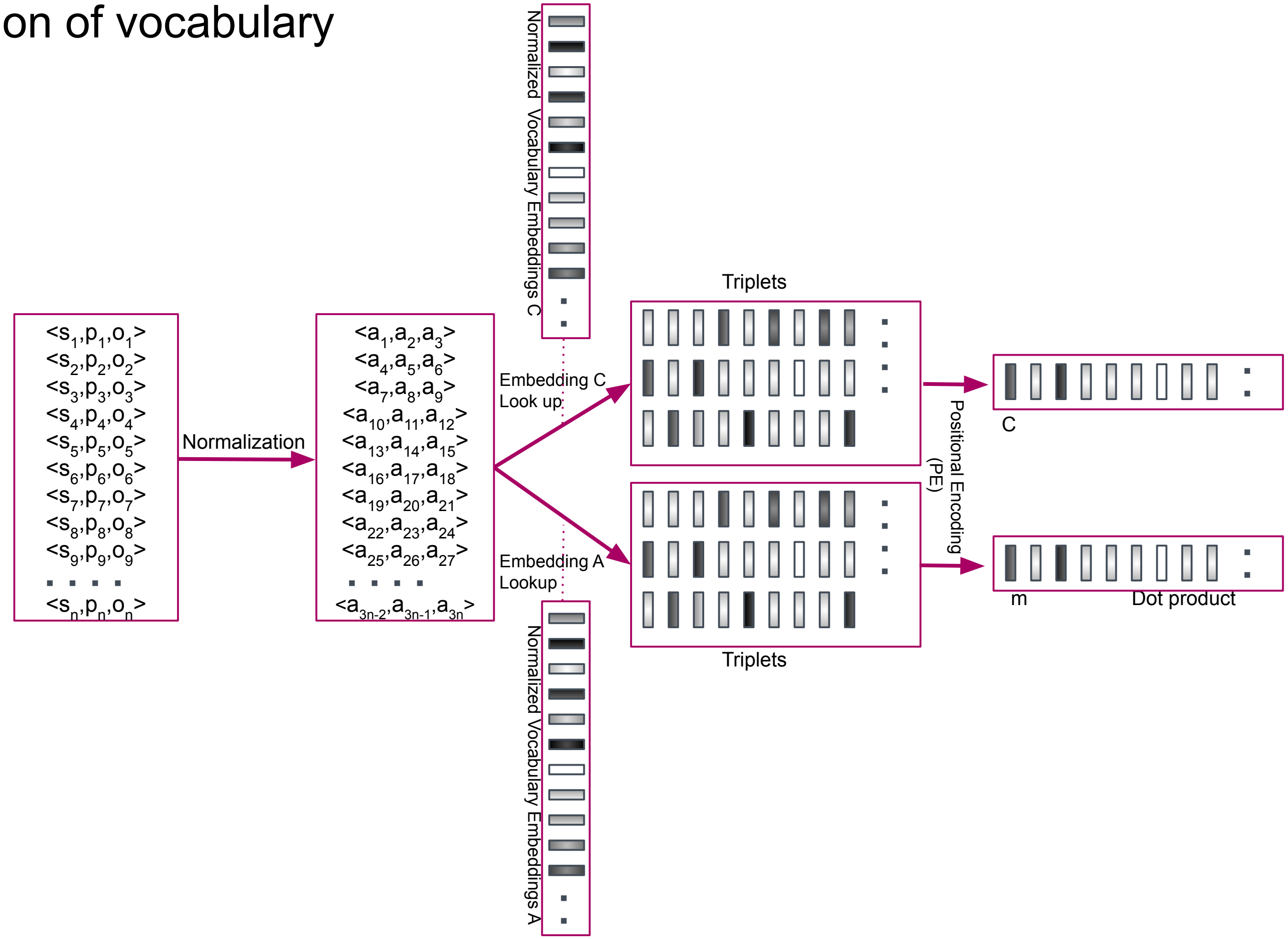
- Syntactic Normalization: Renaming of primitives from the logical language to a set of predefined entity names which will be used across different theories.
  - "**<u>forgetting</u>**" irrelevant label names: Network's learning will be based on the structural information within the theories
  - **<u>Transfer learning</u>** from one knowledge graph to the other



Representation

Knowledge Graph Representation:

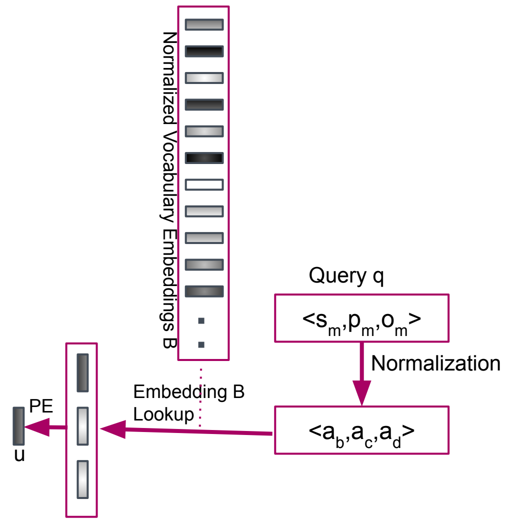
Normalization of vocabulary



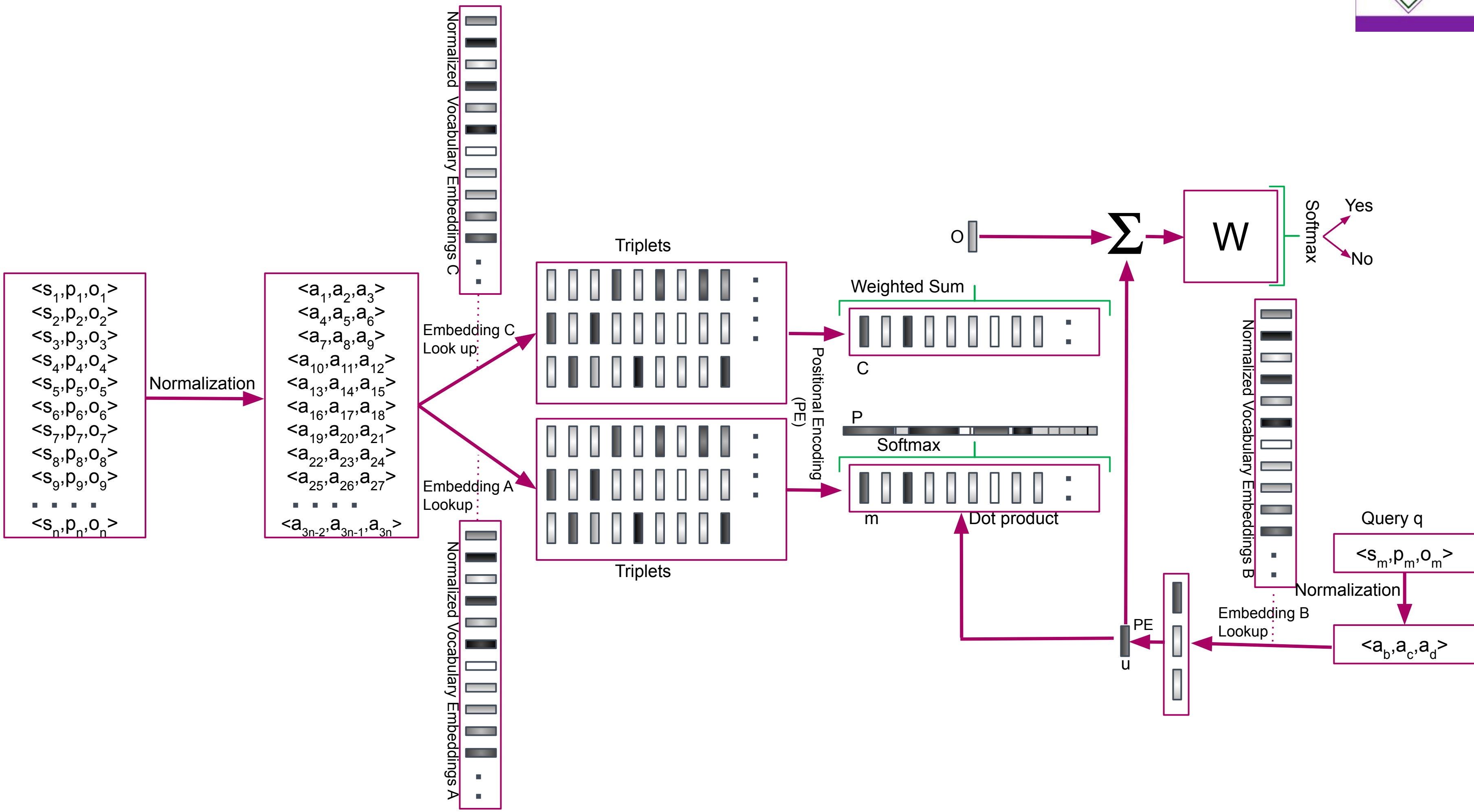
Representation

Query Representation:

Same normalization

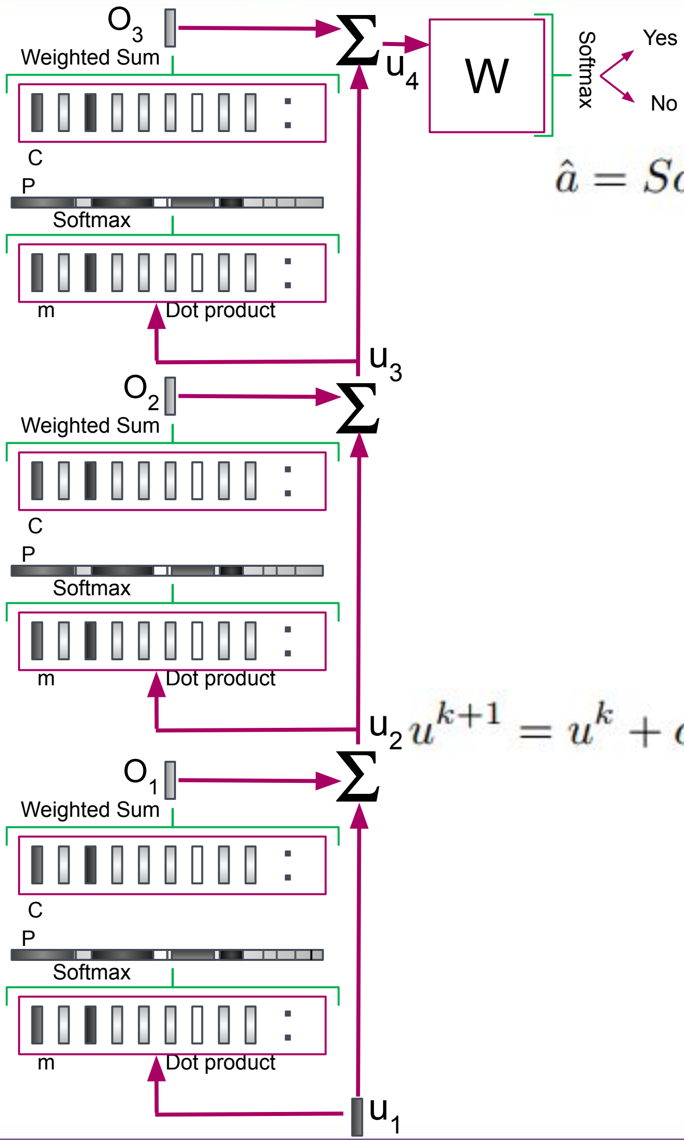


Single layer Model



Mechanics

3-hop Attention Model



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

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

$$o = \sum_i p_i c_i$$

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

DaSe Lab

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

DaSe Lab

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

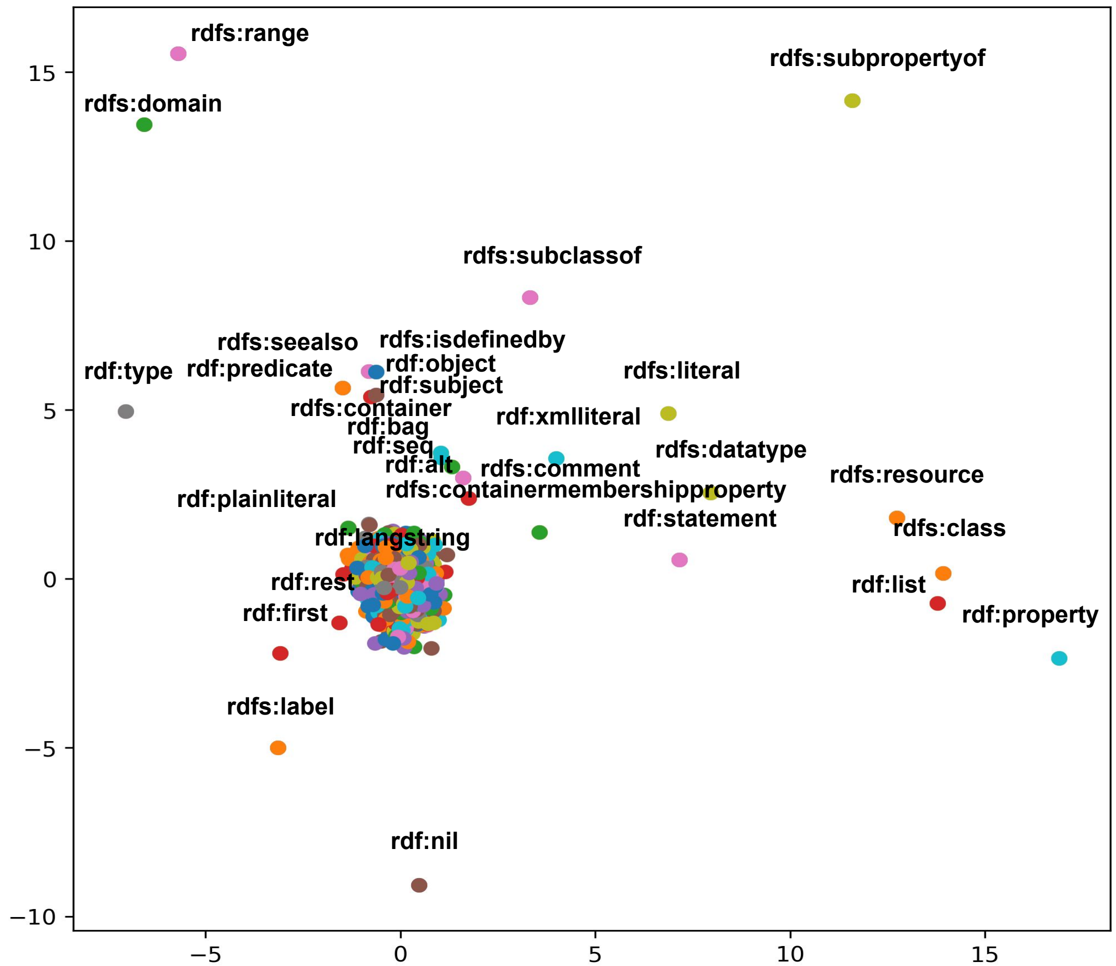
Table 4: Experimental results over each reasoning hop

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



Summary of Contributions

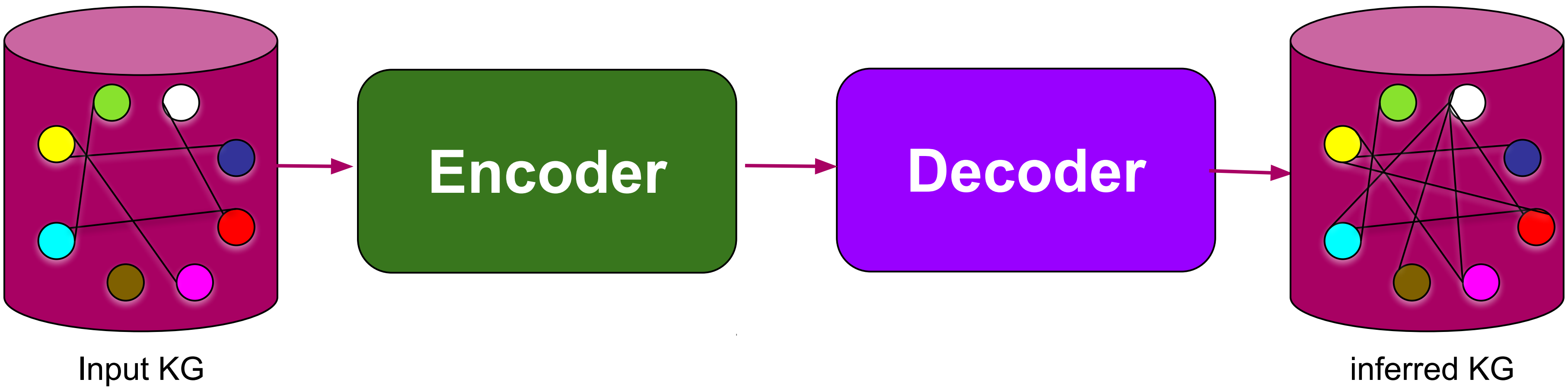
1. Construction of memory networks for emulating the symbolic deductive reasoning.
2. Using normalization approach to enhance their transfer capability.

We have shown the efficacy of our model for cross-domain and cross-knowledge graph deductive reasoning.

On the Generalization Capability of Pointer Networks for RDF & EL+ Reasoning

Idea Overview

Translation model:

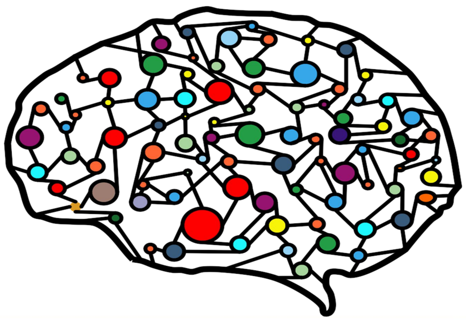


Approach: Why Pointer Networks?

DaSe Lab

- To mimic human reasoning behaviour where one can learn to choose a set of symbols in different locations and copy these symbols to suitable locations to generate new logical consequences based on a set of predefined logical entailment rules

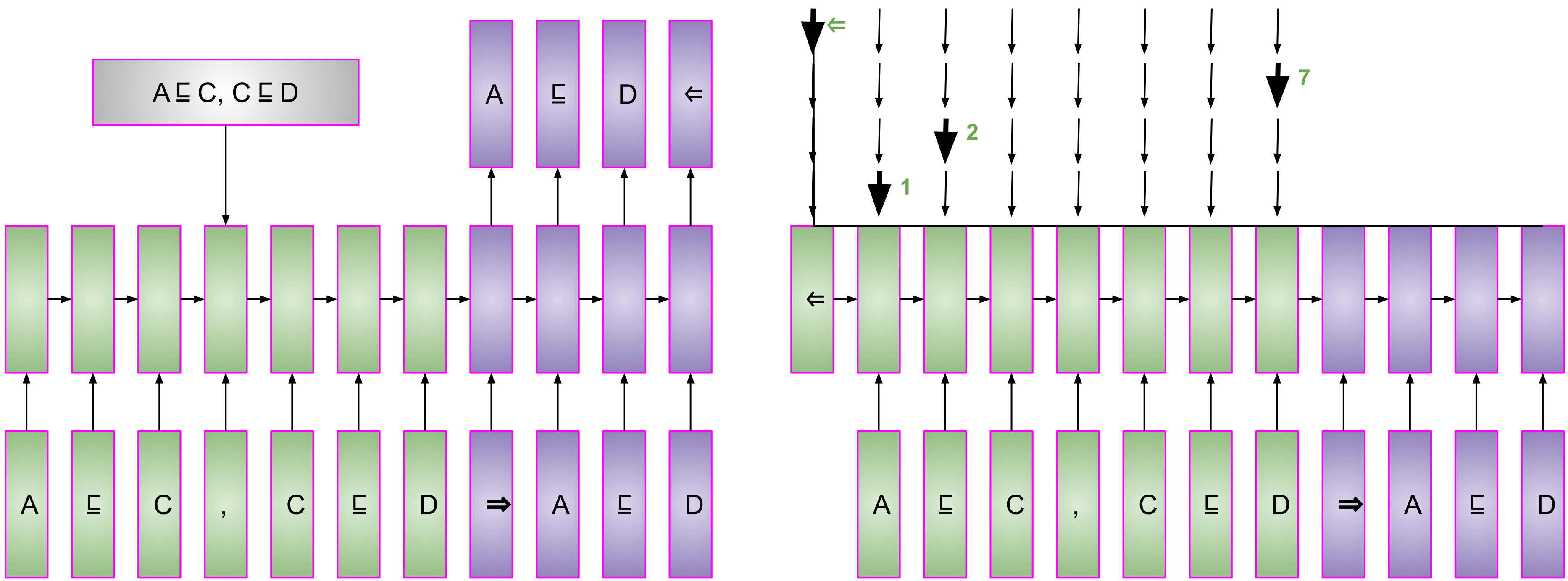
- Successfully applied to:
  - NP-hard Travelling Salesman Problem (TSP)
  - Delaunay Triangulation
  - Convex Hull
  - Text Summarization
  - Code completion
  - Dependency Parsing



What are Pointer Networks?

- Pointer Networks ‘Point’ Input Elements!
- Ptr-Net approach specifically targets problems whose outputs are <u>discrete</u> and <u>correspond</u> to positions in the input.
- At each time step, the distribution of the Attention is the Answer!

Pointer Networks Mechanism



(a) Sequence-to-Sequence (b) Ptr-Net

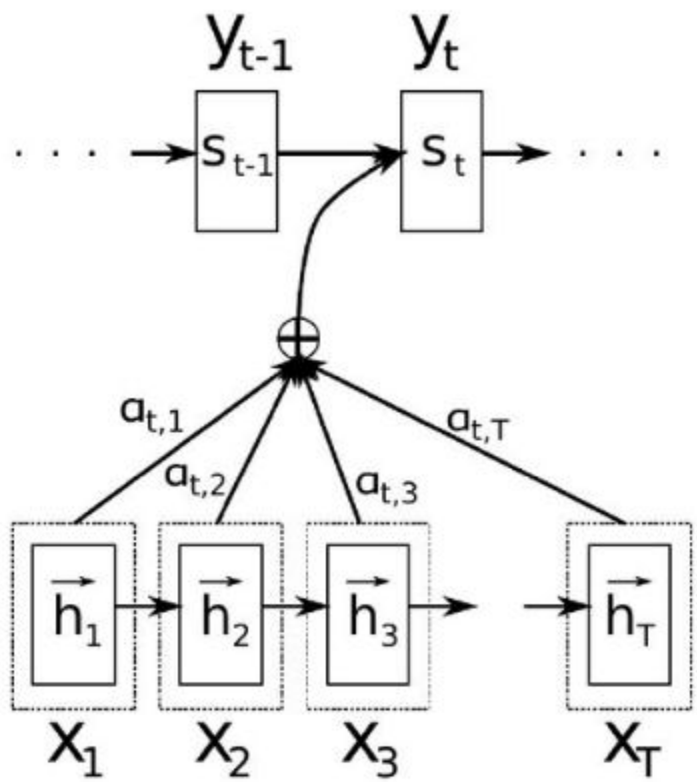
Pointer vs Bahdanau Additive Attention

DaSe Lab

Pointer networks (cont.)

- Based on Attention
- Softmax over a dictionary of **inputs**
- Output models a conditional distribution of the next output token

$$A_j^i = v^T \tanh(W_h h_j + W_s s_i)$$
$$p(C_i | C_1, \ldots, C_{i-1}, \mathcal{P}) = \mathrm{softmax}(A^i)$$



Credits: Vinyals, O., Fortunato, M. & Jaitly, N. Pointer Networks. (2015).
Bahdanau, D., Cho, K. & Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. (2014).

Research Questions

- Can Pointer Networks perform logical deductive reasoning using pointer attention?
- Can other attention-based sequence-to-sequence models like self-attention based popular Transformer architectures successfully perform the same task?
- How well do pointernetwork reasoners perform on completely new knowledge graphs?

Dataset

- **EL+ Dataset: To provide sufficient training input to our network we followed the same synthetic generation procedure as proposed in Eberhart et al. [2]**
  - **Size: 40, 50, and 120 statements with a moderate difficulty**
- **RDF Dataset: We are using the same two datasets used in Makni and Hendler [1] namely a synthetic dataset from Lehigh University Benchmark (LUBM) and a real-world Scientist dataset from DBpedia.**

Baselines

- Transformers
  - Self-attention
  - Achieving the state-of-the-art in tasks such as language modeling and machine translation
- LSTM Decoder
  - Ablation Study
- Graph Words Translation [1]
  - Previous State of the art in RDF Reasoning
- Piece-Wise LSTM [2]
  - Previous State of the art in EL+ Reasoning

Results - Correctness

Logic	KG Size	Pointer Networks		Transformer			LSTM
		SubWordText	Tokenizer	Normalized	Not-Normalized		
					SubWordText	Tokenizer	
RDF	3 - 735	87%	**99%**	5%	25%	4%	0.17%
ER	40	73%	**73%**	8%	8%	0.4 %	0%
	50	68%	**68%**	11%	11%	0.3%	0%
	120	49%	**49%**	15%	NA	NA	0%

- Pointer networks perform remarkably **well** across multiple reasoning tasks while outperforming the previously reported state of the art by a significant margin.
- Our Pointer network model has performed very well (99% accuracy) in conducting the RDF reasoning outperforming the state-of-the-art results obtained by Graph Words Translation (98% accuracy).
- For EL+ reasoning task, our proposed model performs extraordinarily well across all the dataset, and achieves much better results achieving 73% accuracy as opposed to 0.16% accuracy reported in Piecewise LSTM.

Results - Generalizability: Zero-Shot Reasoning

Train \ Test	LUBM	Awards	University
LUBM	*	75%	78%
Awards	79%	*	77%
University	81%	82%	*

Exact Match Accuracy Results for Transfer Learning/Representation: SubWordText Tokenization Encoding

Train \ Test	LUBM	Awards	University
LUBM	*	61%	47%
Awards	96%	*	84%
University	82%	88%	*

Exact Match Accuracy Results for Transfer Learning/ Representation: Whitespace Tokenization Encoding

- We observe that the Pointer Networks preserve their performance even when challenged with knowledge graphs of the domain/vocabulary it <u>has never encountered before</u>.

Summary of Contributions

- A novel paradigm for viewing a symbolic reasoning problem as a pointing problem.
- Pointer Networks are used to neurally resolve symbolic reasoning for the first time.
- The proposed approach is able to transfer its reasoning ability to new domain/vocabulary knowledge graph of same logic.
- We report the state-of-the-art performance of the EL+ and RDF reasoning.

References

[1] Bassem Makni and James A. Hendler. Deep learning for noise-tolerant RDFS reasoning. Semantic Web, 10(5):823–862, 2019.

[2] Aaron Eberhart, **Monireh Ebrahimi**, Lu Zhou, Cogan Shimizu, and Pascal Hitzler. Completion reasoning emulation for the description logic EL+. In Andreas Martin, Knut Hinkelmann, Hans-Georg Fill, Aurona Gerber, Doug Lenat, Reinhard Stolle, and Frank van Harmelen, editors, Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice, AAAI-MAKE 2020, Palo Alto, CA, USA, March 23-25, 2020, Volume I, volume 2600 of CEUR Workshop Proceedings . CEUR-WS.org, 2020.

[3] Patrick Hohenecker and Thomas Lukasiewicz. Ontology reasoning with deep neural networks. J. Artif. Intell. Res., 68:503–540, 2020.

[4] Federico Bianchi and Pascal Hitzler. On the capabilities of logic tensor networks for deductive reasoning. In Andreas Martin, Knut Hinkelmann, Aurona Gerber, Doug Lenat, Frank van Harmelen, and Peter Clark, editors, Proceedings of the AAAI 2019 Spring Symposium on Combining Machine Learning with Knowledge Engineering (AAAI-MAKE 2019) Stanford University, Palo Alto, California, USA, March 25-27, 2019., Stanford University, Palo Alto, California, USA, March 25-27, 2019, volume 2350 of CEUR Workshop Proceedings. CEUR-WS.org, 2019.

[5] **Monireh Ebrahimi**, Aaron Eberhart, Federico Bianchi, and Pascal Hitzler. Towards bridging the neuro-symbolic gap: Deep deductive reasoners.Applied Intelligence, pages1–23, 2021.

[6] **Monireh Ebrahimi**, Md. Kamruzzaman Sarker, Federico Bianchi, Ning Xie, AaronEberhart, Derek Doran, HyeongSik Kim, and Pascal Hitzler. Neuro-symbolic deductivereasoning for cross-knowledge graph entailment. In Andreas Martin, Knut Hinkelmann,Hans-Georg Fill, Aurona Gerber, Doug Lenat, Reinhard Stolle, and Frank van Harme-len, editors,Proceedings of the AAAI 2021 Spring Symposium on Combining MachineLearning and Knowledge Engineering (AAAI-MAKE 2021), Stanford University, PaloAlto, California, USA, March 22-24, 2021, volume 2846 ofCEUR Workshop Proceed-ings. CEUR-WS.org, 2021

[7] **Monireh Ebrahimi**, Md Kamruzzaman Sarker, Federico Bianchi, Ning Xie, AaronEberhart, Derek Doran, and Pascal Hitzler. On the generalization capability of mem-ory networks for reasoning.ICML 2019 Workshop on Understanding and ImprovingGeneralization in Deep Learning, 2019

[8] **Monireh Ebrahimi,** Aaron Eberhart, and Pascal Hitzler. On the capabilities ofpointer networks for deep deductive reasoning.arXiv preprint arXiv:2106.09225, 2021.

Thank You