Completion Reasoning Emulation for the Description Logic EL+

Aaron Eberhart, Monireh Ebrahimi, Lu Zhou, Cogan Shimizu, Pascal Hitzler

Motivation

- Neural networks are great* at learning correct answers
- Many neural-symbolic systems are closer to deep learning than logic
- Can neural-symbolic networks be trained to learn “semantics” instead?
- If a network learns “semantics” then it *should* usually answer correctly...
- It’s not obvious if this is even possible, let alone a good idea

* most of the time, if you do it right, and the training data is good, etc

Research Questions

Research Questions

1. Can a neural network **emulate** reasoner **behavior**?
2. Can we still achieve success ***without*** **embeddings**?
3. Can learning happen with **arbitrary** numerical **names**?
4. Will **intermediate answers** help our evaluation?
5. Is **F1-score sufficient** to evaluate an integrated logic and neural network system? If not, what would be?

EL+

Description	Expression	Semantics
Individual	a	$a \in \Delta^{\mathcal{I}}$
Top	$\top$	$\Delta^{\mathcal{I}}$
Bottom	$\bot$	$\emptyset$
Concept	C	$C^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
Role	R	$R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
Conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
Existential Restriction	$\exists R.C$	$\{ a \mid$ there is $b \in \Delta^{\mathcal{I}}$ such that $(a, b) \in R^{\mathcal{I}}$ and $b \in C^{\mathcal{I}} \}$
Concept Subsumption	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
Role Subsumption	$R \sqsubseteq S$	$R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$
Role Chain	$R_1 \circ \cdots \circ R_n \sqsubseteq R$	$R_1^{\mathcal{I}} \circ \cdots \circ R_n^{\mathcal{I}} \subseteq R^{\mathcal{I}}$

with $\circ$ signifying standard binary composition

DaSe Lab

Completion Rules

$$
\begin{array}{llllll}
(1) & A \sqsubseteq C & C \sqsubseteq D & & \models A \sqsubseteq D \\
(2) & A \sqsubseteq C_1 & A \sqsubseteq C_2 & C_1 \sqcap C_2 \sqsubseteq D & \models A \sqsubseteq D \\
(3) & A \sqsubseteq C & C \sqsubseteq \exists R.D & & \models A \sqsubseteq \exists R.D \\
(4) & A \sqsubseteq \exists R.B & B \sqsubseteq C & \exists R.C \sqsubseteq D & \models A \sqsubseteq D \\
(5) & A \sqsubseteq \exists S.D & S \sqsubseteq R & & \models A \sqsubseteq \exists R.D \\
(6) & A \sqsubseteq \exists R_1.C & C \sqsubseteq \exists R_2.D & R_1 \circ R_2 \sqsubseteq R & \models A \sqsubseteq \exists R.D
\end{array}
$$

EL+

- Each step can be traced back to it's support in the KB
- Initially everything traces directly to the KB with the completion rules

Step 1:

New Fact:	(rule)	support
$C1 \sqsubseteq C3$:	(1)	$C1 \sqsubseteq C2, C2 \sqsubseteq C3$
$C1 \sqsubseteq C4$:	(4)	$C1 \sqsubseteq C2, C1 \sqsubseteq \exists R1.C1, \exists R1.C2 \sqsubseteq C4$
$C1 \sqsubseteq \exists R1.C3$:	(3)	$C1 \sqsubseteq C2, C2 \sqsubseteq \exists R1.C3$
$C1 \sqsubseteq \exists R2.C1$:	(5)	$C1 \sqsubseteq \exists R1.C1, R1 \sqsubseteq R2$
$C1 \sqsubseteq \exists R4.C4$:	(6)	$C1 \sqsubseteq \exists R1.C1, R1 \circ R3 \sqsubseteq R4, C1 \sqsubseteq \exists R3.C4$

- Later the KB support grows as the new facts support other new facts

Step 2:

New Fact:	(rule)	support
$C1 \sqsubseteq C5$:	(2)	$C3 \sqcap C4 \sqsubseteq C5, C1 \sqsubseteq C2, C2 \sqsubseteq C3, C1 \sqsubseteq C2, C1 \sqsubseteq \exists R1.C1, \exists R1.C2 \sqsubseteq C4$

- These supports can guide the training from input KBs to completion

Encoding

KB statement	Vectorization
$CX \sqsubseteq CY$	$\rightarrow [\,0.0,\ \frac{X}{c}, \frac{Y}{c},\ 0.0\,]$
$CX \sqcap CY \sqsubseteq CZ$	$\rightarrow [\ \frac{X}{c},\ \frac{Y}{c},\ \frac{Z}{c},\ 0.0\,]$
$CX \sqsubseteq \exists RY.CZ$	$\rightarrow [0.0,\ \frac{X}{c},\ \frac{-Y}{r},\ \frac{Z}{c}]$
$\exists RX.CY \sqsubseteq CZ$	$\rightarrow [\frac{-X}{r},\ \frac{Y}{c},\ \frac{Z}{c},\ 0.0]$
$RX \sqsubseteq RY$	$\rightarrow [0.0,\ \frac{-X}{r}, \frac{-Y}{r}, 0.0]$
$RX \circ RY \sqsubseteq RZ$	$\rightarrow [\frac{-X}{r}, \frac{-Y}{r}, \frac{-Z}{r}, 0.0]$
Concept names = c	Role names = r

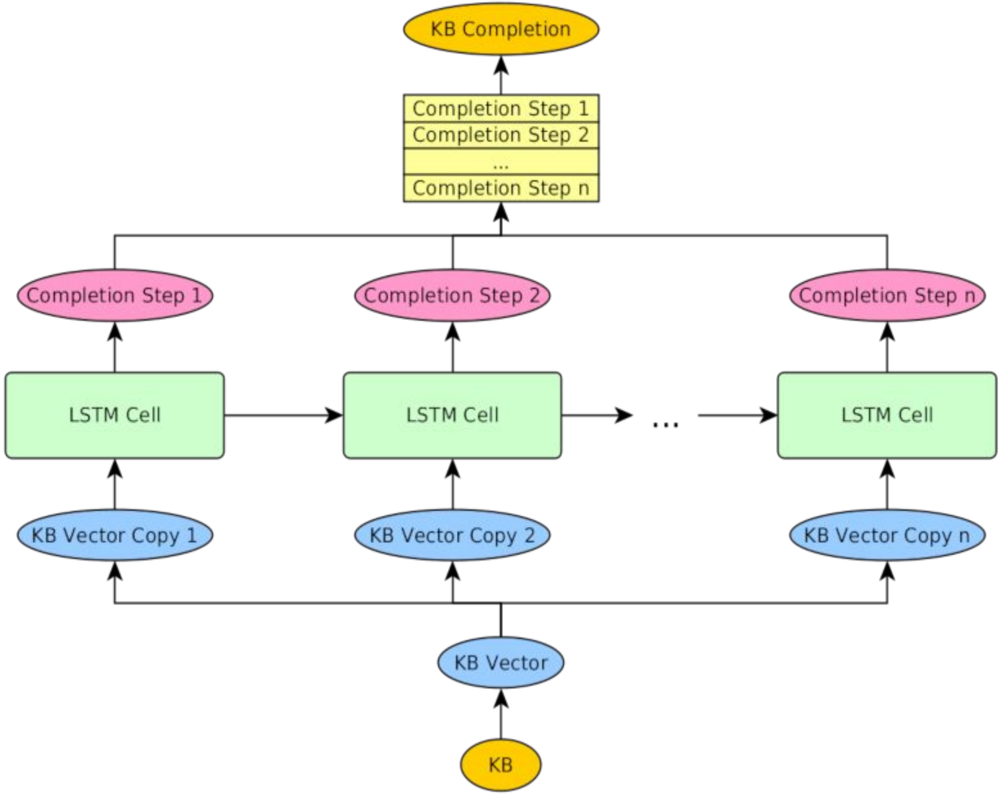
Data Sources

- Sequential-Random Generator
  - Sequence guarantees use of all completion rules
  - Randomized partition with noise
- SNOMED 2012 Sampling
  - Imbalanced use of completion rules
  - EL+ sampled to manageable size

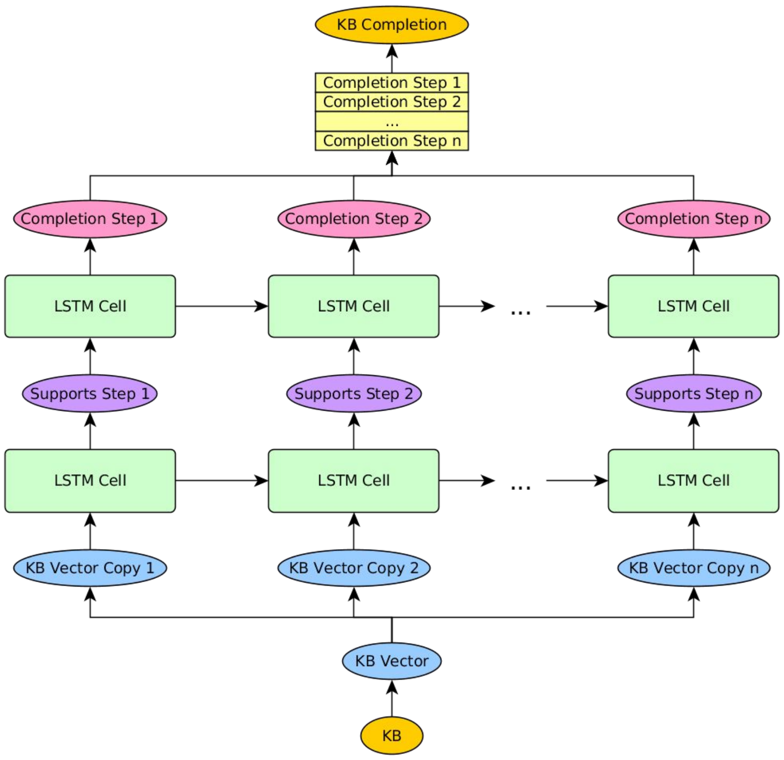
LSTM

- KBs with reasoner steps fed in as a batch of sequences
- Because intermediate answers are known, piecewise training is possible
- Piecewise system has same dimensionality as the layered systems
- Attempting completion on continuous values, not classification, so use regression minimizing MSE

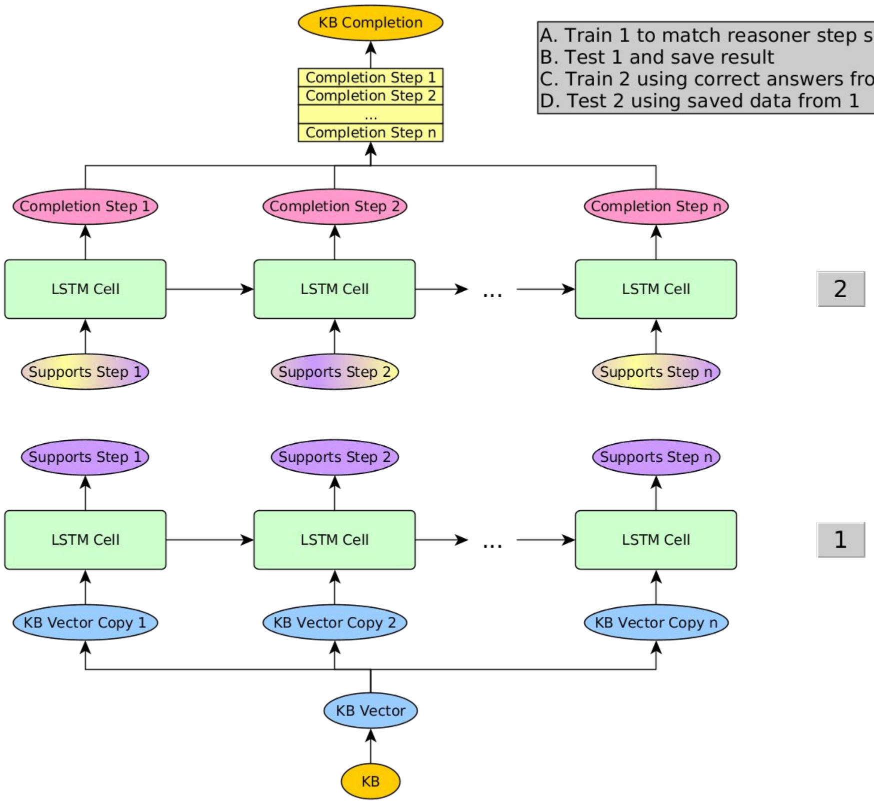
Flat System Architecture



Deep System Architecture



Piecewise System Architecture

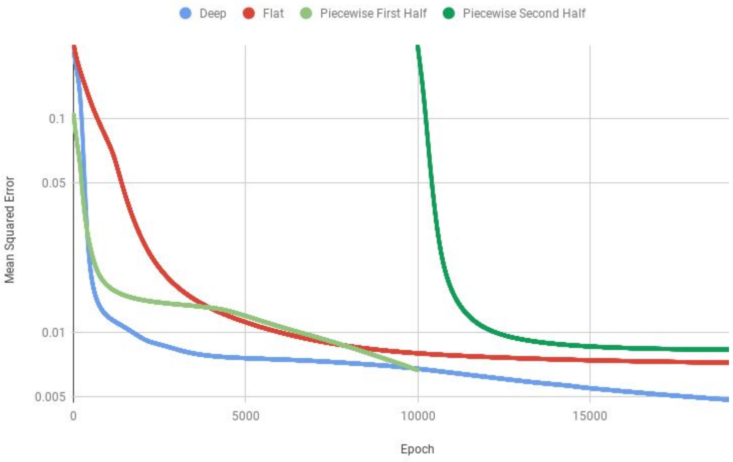


A. Train 1 to match reasoner step support data
B. Test 1 and save result
C. Train 2 using correct answers from reasoner
D. Test 2 using saved data from 1

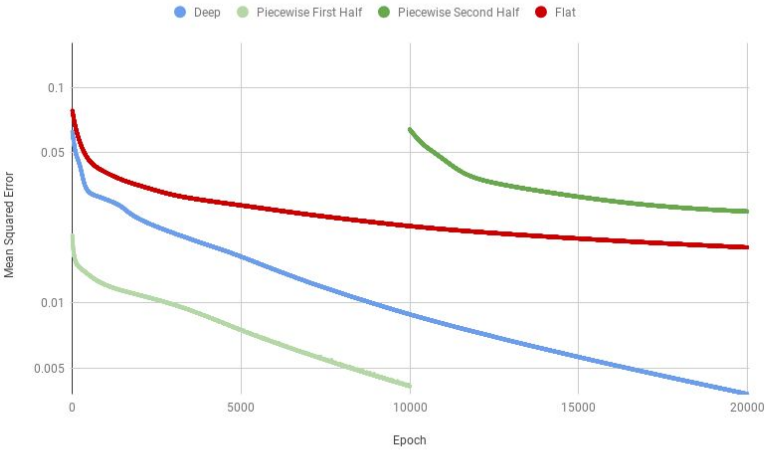
Results - Training

- Piecewise training feasible, presents opportunities for introducing transparency

Synthetic Data



SNOMED Data



Results - Synthetic Example

	Correct Answer	Predicted Answer
Step 0	$C9 \sqsubseteq C11$	$C8 \sqsubseteq C9$
	$C2 \sqsubseteq C10$	$C1 \sqsubseteq C9$
	$C9 \sqsubseteq C12$	$C8 \sqsubseteq C9$
	$C7 \sqsubseteq C6$	$C8 \sqsubseteq \exists R4.C9$
	$C9 \sqsubseteq \exists R4.C11$	$C1 \sqsubseteq \exists R5.C9$
	$C2 \sqsubseteq \exists R4.C9$	$C8 \sqsubseteq \exists R4.C9$
	$C9 \sqsubseteq \exists R5.C9$	$C9 \sqsubseteq \exists R5.C9$
	$C2 \sqsubseteq \exists R5.C11$	
	$C9 \sqsubseteq \exists R7.C12$	
	$C2 \sqsubseteq \exists R6.C12$	
Step 1	$C9 \sqsubseteq C13$	$C8 \sqsubseteq C12$
	$C2 \sqsubseteq C11$	$C2 \sqsubseteq C10$
	$C2 \sqsubseteq C12$	$C1 \sqsubseteq C11$
	$C2 \sqsubseteq \exists R4.C11$	$C1 \sqsubseteq \exists R3.C12$
	$C2 \sqsubseteq \exists R5.C9$	$C1 \sqsubseteq \exists R4.C8$
	$C2 \sqsubseteq \exists R7.C12$	
Step 2	$C2 \sqsubseteq C13$	$C1 \sqsubseteq C12$

Results - SNOMED Example

Good Example

Trial 46

Correct Answer:
C9 ⊑ ∃R3.C2 if something is a anterior hypothalamic region then there is a hypothalamus that it is PartOf
Prediction:
C9 ⊑ ∃R3.C1 if something is a anterior hypothalamic region then there is a side that it is PartOf

Not as Good Example

Trial 3

Correct Answer:
C1 ⊑ ∃R2.C4 if something is a dorsal surface of great toe then there is a surface of body that it is rPartOf
Prediction:
C5 ⊑ ∃R2.C1 if something is a branch of phrenic nerve then there is a dorsal surface of great toe that it is rPartOf

Results - Evaluation

- Levenshtein Distance between every string axiom output and best match in the correct answers
- Character Levenshtein distance uniformly substitutes all numbers with unique characters.
- **Edit distances fail to account for near misses**
- Custom “predicate distance” measures how far away each atom is from its correct value, adding big penalties for R/C misses and small for incorrectly guessed numbers
- Random axioms are generated to compare with results
- An F score is obtained for all of these by counting the number of times a perfect match is found

Results - Precision, Recall, F1

	Atomic Levenshtein Distance			Character Levenshtein Distance			Predicate Distance		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
	Synthetic Data								
Piecewise Prediction	0.138663	0.142208	0.140412	0.138663	0.142208	0.140412	0.138646	0.141923	0.140264
Deep Prediction	**0.154398**	**0.156056**	**0.155222**	**0.154398**	**0.156056**	**0.155222**	**0.154258**	**0.155736**	**0.154993**
Flat Prediction	0.140410	0.142976	0.141681	0.140410	0.142976	0.141681	0.140375	0.142687	0.141521
Random Prediction	0.010951	0.0200518	0.014166	0.006833	0.012401	0.008811	0.004352	0.007908	0.007908
	SNOMED Data								
Piecewise Prediction	0.010530	0.013554	0.011845	0.010530	0.013554	0.011845	0.010521	0.013554	0.011839
Deep Prediction	**0.015983**	0.0172811	**0.016595**	**0.015983**	0.017281	**0.016595**	**0.015614**	0.017281	**0.016396**
Flat Prediction	0.014414	**0.018300**	0.016112	0.0144140	**0.018300**	0.016112	0.013495	**0.018300**	0.015525
Random Prediction	0.002807	0.006803	0.003975	0.001433	0.003444	0.002023	0.001769	0.004281	0.002504

Results - Distances

	Atomic Levenshtein Distance			Character Levenshtein Distance			Predicate Distance		
	From	To	Average	From	To	Average	From	To	Average
	Synthetic Data								
Piecewise Prediction	1.336599	1.687640	1.512119	1.533115	1.812006	1.672560	2.633427	4.587382	3.610404
Deep Prediction	**1.256940**	**1.507150**	**1.382045**	**1.454787**	**1.559751**	**1.507269**	**2.504496**	**3.552074**	**3.028285**
Flat Prediction	1.344946	1.584674	1.464810	1.586281	1.660409	1.623345	2.517655	3.739770	3.128713
Random Prediction	1.598016	1.906369	1.752192	1.970604	1.289533	1.630068	5.467918	10.57324	8.020583
	SNOMED Data								
Piecewise Prediction	1.704931	**2.686562**	**2.195746**	2.016249	2.862737	2.439493	6.556592	**5.857769**	6.207181
Deep Prediction	1.759633	3.052080	2.405857	2.027190	3.328850	2.678020	**4.577427**	6.179389	**5.378408**
Flat Prediction	**1.691738**	2.769542	2.230640	**1.948757**	2.991328	2.470042	5.548226	6.665659	6.106942
Random Prediction	1.814656	3.599629	2.707143	2.094682	**1.621700**	**1.858191**	5.169093	12.392325	8.780709

Contact

Email
aaroneberhart@ksu.edu

GitHub
https://github.com/aaronEberhart/ERCompletionReasoningLSTM