DeepLogic: Towards End-to-End Differentiable Logical Reasoning

using Neural Memory Networks

Nuri C.

March 2019

Imperial College London

Deep learning [1] learns human level tasks.

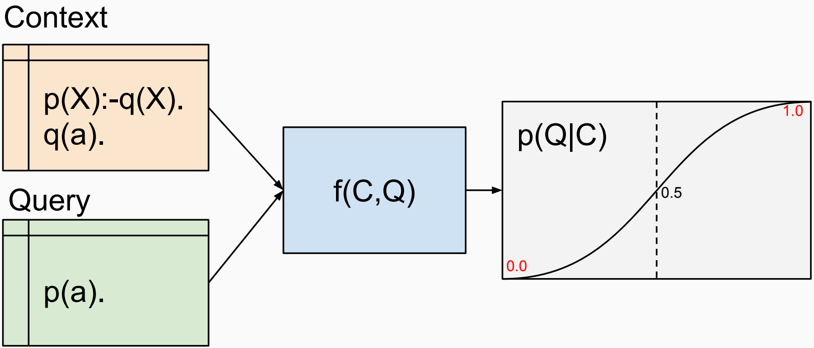
Logic [2] formalises reasoning.

Deep learn to reason logically ?

DeepLogic

1. If and how neural networks learn to represent symbolic constructs from logic?
2. If and how iterative neural networks use those representations to perform reasoning over logic programs?

Setup



Dataset

- 12 tasks each with 20k training and 10k validation, 10k test.
- Normal logic programs with negated atoms and positive atoms as literals, synthetically generated.
- No function applications or recursion allowed.
- Upper case English alphabet for variables, lower case for predicates *and* constants, ex. `p(X),p(q),q(p)`.
- Ground atoms as queries, ex. `p(a)`.
- Maximum arity 2.

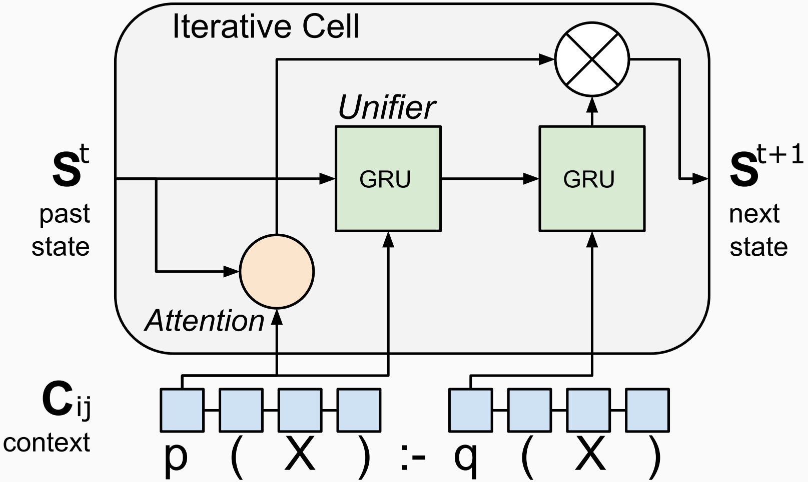
Table 1: Sample programs from tasks 9 to 12.

9: 1 Step NBF	10: 2 Step NBF	11: AND NBF	12: OR NBF
s(X,J) :- -p(J,X).	r(C) :- -o(C).	b(G,B) :- -i(G) , u(B).	y(Z) :- -e(Z).
p(e,x).	o(P) :- l(P).	i(w).	y(Z) :- b(Z).
v(V,Q) :- u(V,Q).	l(o).	g(a).	y(r).
o(N) :- -q(N).	g(u).	u(a).	e(d).
t(x,e).	p(U,L) :- e(U,L).	f(t).	s(a).
m(y,c).	p(X,X).	l(W) :- a(W) , d(W).	b(m).
? s(x,e). 0	? r(u). 1	? b(a,a). 1	? y(a). 1
? s(e,x). 1	? r(o). 0	? b(w,a). 0	? y(d). 0

Dataset

- Training predicates and constants up to 2 characters, longer for test sets. 4, 8 and 12 characters and irrelevant rules for easy, medium and hard sets respectively.
- Training up to 3 hops, longer for multi-hop analysis test (up to 32 hops).
- Publicly available online with generation script, `https://github.com/nuric/deeplogic`

Iterative Memory Attention



Experiments

Training	Multi-task						Curriculum				
Model	LSTM	MAC	DMN	IMA			MAC	DMN	IMA		
Embedding	-	rule	rule	literal		lit+rule	rule	rule	literal		lit+rule
Attention	-	*sm*	σ	σ	*sm*	*sm*	*sm*	σ	σ	*sm*	*sm*
Facts	0.61	0.84	1.00	1.00	1.00	0.98	0.89	1.00	1.00	0.99	0.94
Unification	0.53	0.86	0.87	0.90	0.87	0.85	0.83	0.85	0.88	0.88	0.86
1 Step	0.57	0.90	0.74	0.98	0.94	0.95	0.77	0.62	0.96	0.93	0.92
2 Steps	0.56	0.81	0.67	0.95	0.95	0.94	0.70	0.58	0.95	0.91	0.89
3 Steps	0.57	0.78	0.77	0.94	0.94	0.94	0.64	0.64	0.93	0.86	0.87
AND	0.65	0.84	0.80	0.95	0.94	0.85	0.81	0.70	0.80	0.78	0.83
OR	0.62	0.85	0.87	0.97	0.96	0.93	0.75	0.75	0.96	0.93	0.90
Transitivity	0.50	0.50	0.50	0.50	0.52	0.52	0.50	0.50	0.50	0.50	0.50
1 Step NBF	0.58	0.92	0.79	0.98	0.94	0.95	0.65	0.58	0.96	0.91	0.92
2 Steps NBF	0.57	0.83	0.85	0.96	0.93	0.95	0.57	0.73	0.95	0.90	0.90
AND NBF	0.55	0.82	0.84	0.92	0.93	0.85	0.61	0.61	0.71	0.77	0.83
OR NBF	0.53	0.74	0.75	0.86	0.86	0.86	0.59	0.63	0.86	0.83	0.84
Easy Mean	0.57	0.81	0.79	**0.91**	0.90	0.88	0.69	0.68	0.87	0.85	0.85
Medium Mean	0.52	0.70	0.70	**0.86**	0.81	0.79	0.60	0.61	0.81	0.76	0.74
Hard Mean	0.51	0.63	0.66	**0.83**	0.75	0.72	0.55	0.58	0.76	0.70	0.68

Attention Visualisation

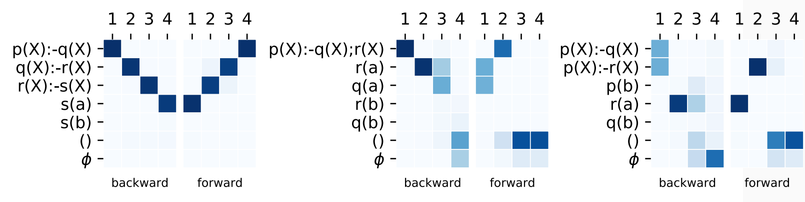
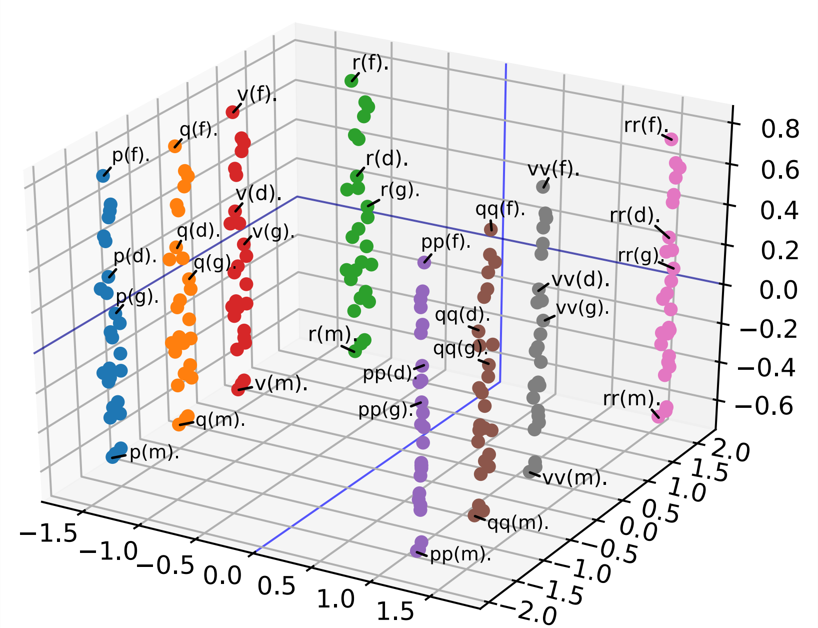
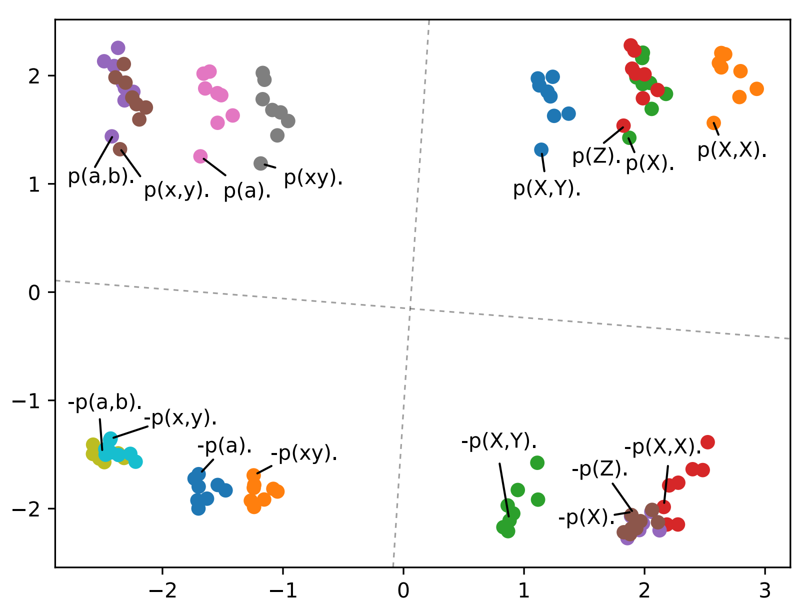


Figure 1: Attention maps produced for query $p(a)$ for IMA with softmax attention performing *backward* chaining in the left column and IMA with literal + rule embedding *forward* chaining in the right column on tasks 5 to 7.

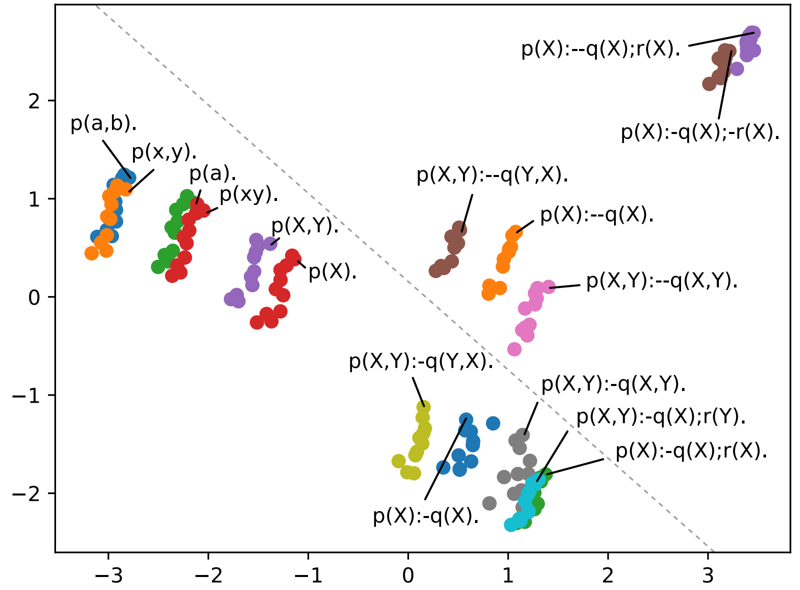
Literal Embedding Visualisation



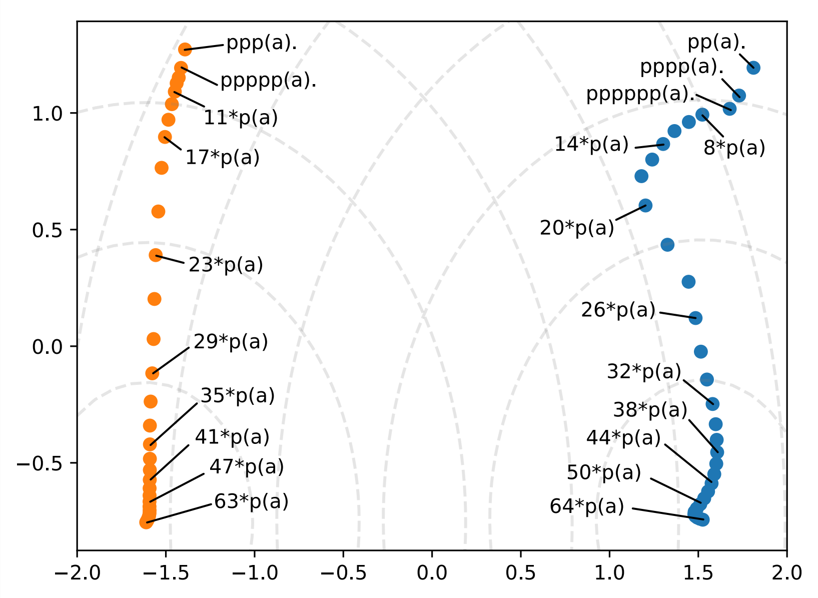
Structurally Different Literals



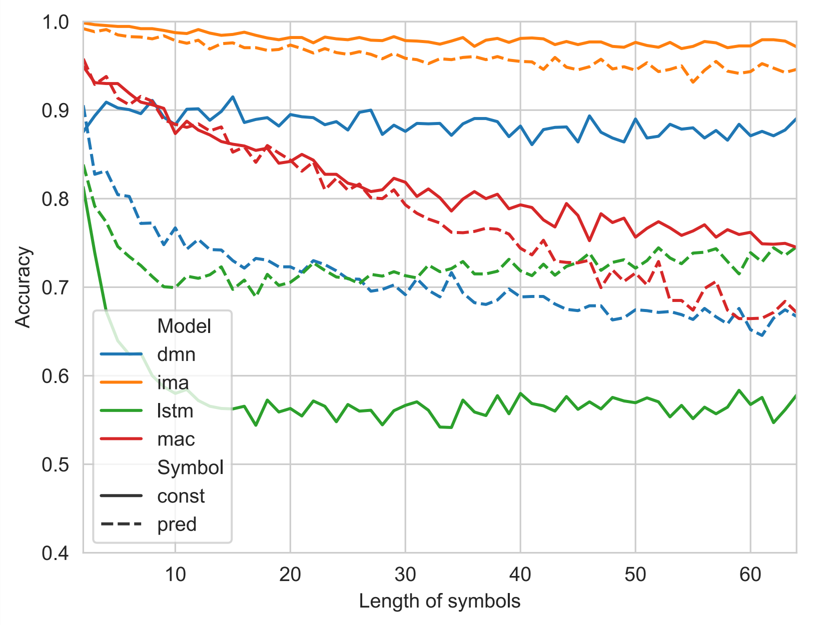
Structurally Different Rules



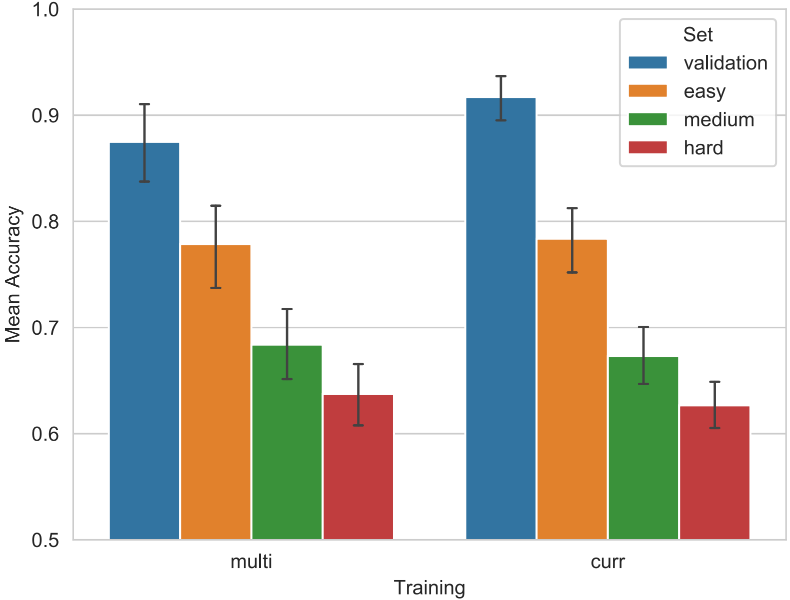
Literal Embedding Saturation



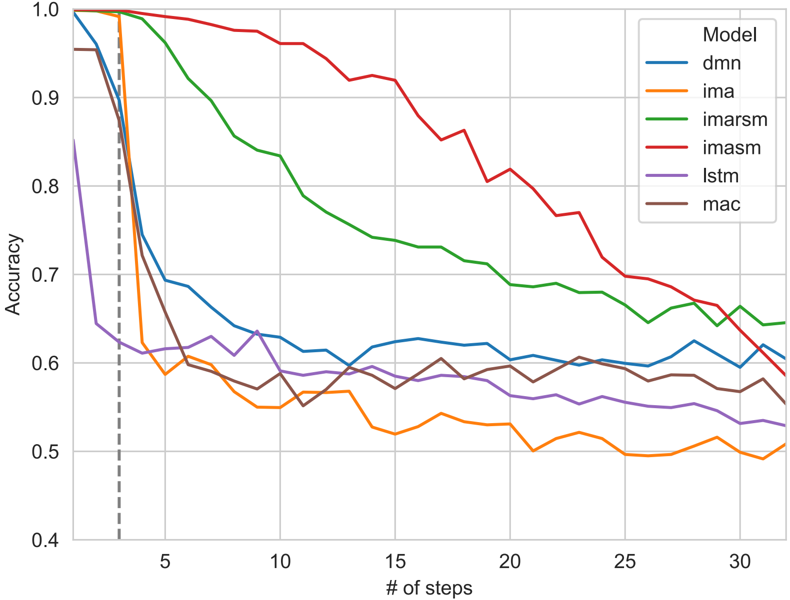
Increasing Symbol Lengths



Training Regime



Multi-hop Reasoning



Concluding Remarks

- Embedding space attempts to uniquely represent symbolic entities, clustered by structural similarity.
- Negation is learnt as separate representation if forced into same embedding space.
- No backtracking, instead attend to multiple rules at the same iteration step, potentially solving in fewer steps.
- Incrementally training on tasks have no advantage despite logically sharing constructs.
- Iterative applications, re-usage and updating of state vector seem to degrade performance over multiple iterations.

Questions

References i

I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*.
The MIT Press, Jan. 2017.

S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach (3rd Edition)*.
Pearson, 2016.

Iteration

$$ca_i^t = [s^t \; ; \; q \; ; \; r_i \; ; \; (s^t - r_i)^2 \; ; \; s^t \odot r_i] \tag{1}$$

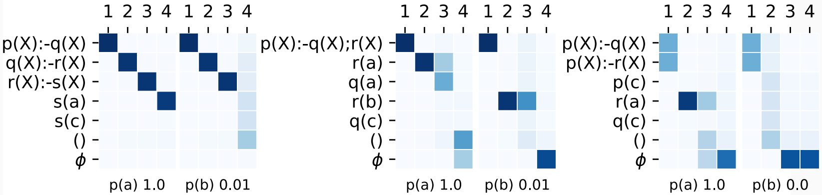
$$\alpha_i^t = \sigma(W^{1 \times \frac{d}{2}}(U^{\frac{d}{2} \times d} ca_i^t + b^{\frac{d}{2}}) + b^1) \tag{2}$$

$$u_{ij}^t = GRU(\mathsf{c}_{ij}, u_{i(j-1)}^t) \tag{3}$$

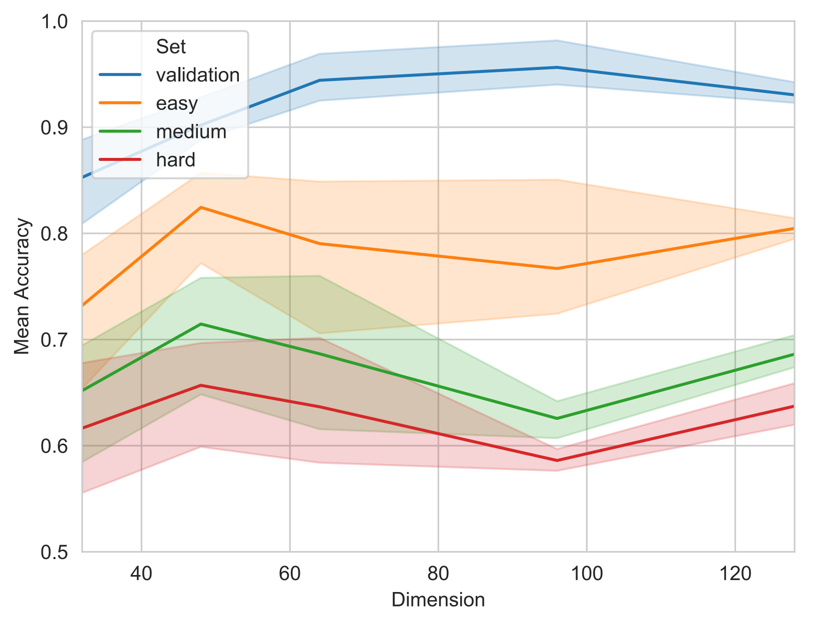
$$s^{t+1} = \sum_i^R \alpha_i^t u_{iL}^t \tag{4}$$

We kick start the process by setting the initial state to the query $s^0 = q$. At every iteration compute features between the current state and the rules ca_i^t to capture which rules are relevant. With initial hidden state $u_{i0}^t = s^t$, compute interaction between state and rule u_{ij}^t, perform weighted sum to get next state s^{t+1}.

Attention of Negative Cases



Hidden Size



Set
validation
easy
medium
hard
1.0
0.9
0.8
0.7
0.6
0.5
Mean Accuracy
40
60
80
100
120
Dimension