

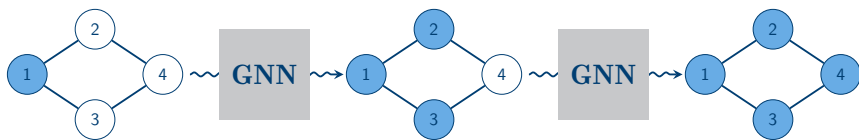
Deep Equilibrium Algorithmic Reasoning

Dobrik Georgiev¹, JJ Wilson, Davide Buffelli², Pietro Liò¹

¹Department of Computer Science and Technology, University of Cambridge, UK

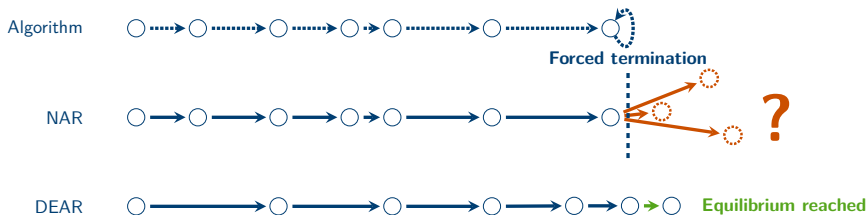
²MediaTek Research

A quick primer on neural algorithmic reasoning (NAR)



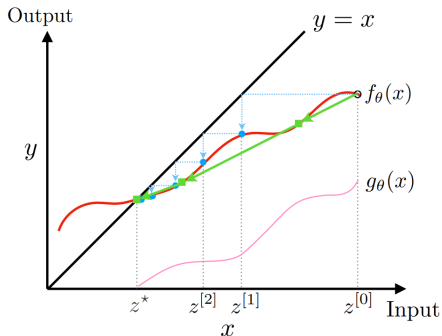
- ▶ A Graph Neural Network (GNN) neurally executing BFS
- ▶ Robust performance is achieved through **alignment**

Equilibrium is a missed alignment



- ▶ Termination was “forced” (also assumes test-time information)
- ▶ Denotations of `for` loops require finding a fixed point

How? DEQs [Bai et al., 2019]



In order to find \mathbf{x} , s.t.

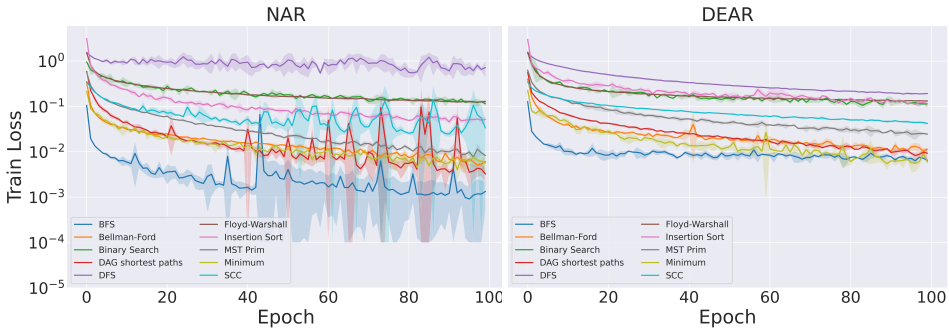
$$f_{\theta}(\mathbf{x}) = \mathbf{x}$$

find the roots of

$$g_{\theta}(\mathbf{x}) = f_{\theta}(\mathbf{x}) - \mathbf{x}$$

- ▶ Extra inductive bias
- ▶ Does not require termination signal
- ▶ Any choice of root finding

Can you really train them this way?



And are they as accurate?

Algorithm	NAR \blacklozenge	NAR \blacklozenge (Triplet-MPNN)	NAR \diamond (LT)	DEAR (ours)
Weighted graphs				
Bellman-F.	97.06% \pm 0.40	97.23% \pm 0.15	95.39% \pm 1.42	96.78% \pm 0.43
Floyd-W.	52.53% \pm 0.98	61.86% \pm 1.57	49.30% \pm 0.53	55.75% \pm 2.20
DSP	94.21% \pm 1.77	93.32% \pm 1.60	88.30% \pm 1.04	89.81% \pm 0.14
MST Prim	93.56% \pm 0.77	92.01% \pm 1.50	87.69% \pm 1.17	88.67% \pm 0.74
Unweighted graphs				
BFS	99.85% \pm 0.09	99.69% \pm 0.29	99.51% \pm 0.06	98.73% \pm 0.37
DFS	16.89% \pm 5.73	31.20% \pm 4.02	29.07% \pm 2.32	40.62% \pm 0.44
SCC	40.70% \pm 1.39	46.84% \pm 1.70	39.33% \pm 1.52	43.63% \pm 1.19
Arrays				
Search (Binary)	94.67% \pm 2.31	93.33% \pm 2.31	84.33% \pm 8.33	59.00% \pm 12.3
Minimum	97.67% \pm 5.73	96.67% \pm 2.31	94.00% \pm 2.00	97.22% \pm 3.82
Sort (Ins.)	27.07% \pm 10.3	63.67% \pm 39.97	33.8% \pm 12.04	86.93% \pm 3.87

\blacklozenge – termination given at train+test time \diamond – termination given at train time only

And are they as accurate?

Algorithm	NAR [◆]	NAR [◆] (Triplet-MPNN)	NAR [◇] (LT)	DEAR (ours)
Overall	71.42%	77.58%	70.07%	<u>75.42%</u>

◆ – termination given at train+test time

◇ – termination given at train time only

They cannot be fast as well

Algorithm		NAR [♦]	DEAR
Bellman-F.	↑	0.0118	0.0215
Floyd-W.	↑	0.0916	0.1102
DSP	↓	0.1334	0.0345
MST Prim	↓	0.0708	0.0297
BFS	↑	0.0094	0.0137
DFS	↓↓	0.2440	0.0478
SCC	↓↓	0.4017	0.0253
Search (Binary)	≈	0.0125	0.0131
Minimum	↓	0.0684	0.0174
Sort (Ins.)	↓↓	0.5680	0.0260

See you at the poster session!

Deep Equilibrium Algorithmic Reasoning

Dobrik Georgiev, JJ Wilson, Davide Buffelli, Pietro Liò



Background

Neural Algorithmic Reasoners learn to simulate algorithms. NAR iterations match algorithm iterations. Termination signal is provided at train and, often, at test time. The equilibrium property - not established.

We present Deep Equilibrium Algorithmic Reasoning (DEAR), a novel approach for executing algorithms, which **requires no termination supervision**, achieves **competitive performance to more computationally expensive models** and has **substantially faster inference times**.

Oh, DEAR!?! But...

Why? Denotational semantics assigns (strictly) compositional meanings to programs.

$w = \text{let } x \text{ be } b \text{ do } e$

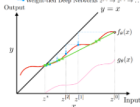
$w = \text{if } b \text{ then } (e_1) \text{ else } (e_2)$

$[\![w]\!] = \text{if } b \text{ then } (e_1) \text{ else } (e_2) = \text{let } x \text{ be } \text{State}_b \text{ in } \text{if } (b) \text{ then } (e_1) \text{ else } (e_2)$

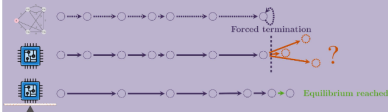
Achieving this for loops requires finding the equilibrium (a.k.a. fixed point).

How? Deep Equilibrium Models (Bai et al.)

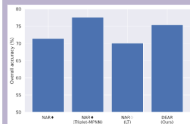
- Deep Equilibrium Models
- Weight-tied Deep Networks $z^{(t)} \rightarrow z^{(t+1)} \rightarrow \dots$



Algorithms possess equilibriums and so should neural algorithmic reasoners!



DEAR is accurate... and efficient →



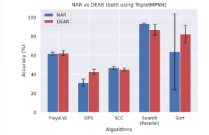
Equilibrium is a useful inductive bias

Algorithm	NAR*	NAR* (2x)	NAR (2x)	DEAR	DEAR (2x)
Unweighted graphs					
Robust-E	07.6875 ± 0.46	07.2250 ± 0.10	06.8875 ± 1.40	06.0275% (E)	06.2975 ± 0.42
Find-W	12.0250 ± 0.99	08.4875 ± 0.22	08.9875 ± 0.02	-	08.2975 ± 0.26
DFP	04.2175 ± 1.17	02.3275 ± 0.00	04.4875 ± 1.10	-	04.4475 ± 0.15
MMF Pvis	03.0275 ± 0.57	00.2175 ± 0.00	04.0275 ± 1.17	02.0175% (E)	04.0475 ± 0.24
Avg	08.7975 ± 0.69	04.9875 ± 0.20	08.5575 ± 0.26	-	08.2975 ± 0.44
Avg E	04.7975 ± 1.10	04.9475 ± 0.10	08.3975 ± 1.10	-	04.0275 ± 0.10
Unweighted graphs					
DFP	04.9875 ± 0.69	04.9875 ± 0.20	04.5575 ± 0.26	04.47%	04.7975 ± 0.47
DFP	04.9875 ± 0.12	02.2075 ± 0.00	00.7075 ± 0.00	2.00%	04.0275 ± 0.44
Avg E	04.7975 ± 1.10	04.9475 ± 0.10	08.3975 ± 1.10	-	04.0275 ± 0.10
Common fully connected graphs					
Search (Binary)	04.4275 ± 0.10	00.3075 ± 0.00	04.4875 ± 0.30	-	00.9875 ± 0.12
Minimum	07.4275 ± 0.12	04.4775 ± 0.10	04.9875 ± 0.10	-	07.2275 ± 0.42
Sort (Ins.)	04.4275 ± 0.10	04.4775 ± 0.10	05.4875 ± 0.10	-	04.9875 ± 0.10
Overall	11.4275	09.3875	10.8875	-	10.6875

DEARs are highly parallel

Algorithm	NAR*	DEAR
Beltrano-F.	↑ 0.0118	0.0215
Find-W.	↑ 0.0016	0.1102
DFP	↓ 0.1234	0.0163
MMF Pvis	↓ 0.0738	0.0297
DFP	↑ 0.0884	0.0127
DFP	↓ 0.2440	0.0475
SCC	↓ 0.0017	0.0252
Search (Binary)	→ 0.0125	0.0124
Minimum	↓ 0.0004	0.0174
Sort (Ins.)	↓ 0.0260	0.0260

DEARs are foundational



Paper

Code



Bai, S., Kolter, J. Z., and Koltun, V. (2019). Deep equilibrium models. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, Advances in Neural Information Processing Systems, volume 32. Curran Associates, Inc.