1st May 2017

MEng Computer Science

a.p.breeze@durham.ac.uk

Principles and Applications of Differential Neural Computers

Alastair Breeze - Supervised by Stephen McGough, Noura Al Moubayed

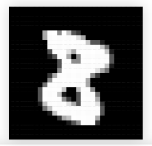
Durham University

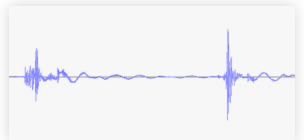
Sequence Problems

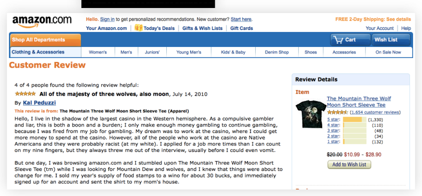
The universe is a giant model of interleaving patterns that can be represented as sequences.

Understanding sequences is key to artificial intelligence, enabling human-like inference models that draw on the past to influence subsequent decisions. Application areas include:

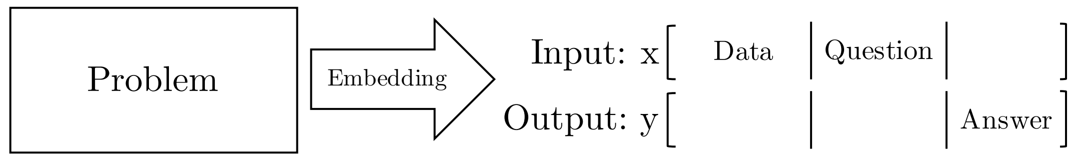
- Natural Language Processing
- Computer Vision
- Healthcare
- Strategic Reasoning







Problems are embedded for feeding as follows:



Optimisations & Visualisations

- Dropout and L1/L2 were investigated.
- Scaling over a distributed GPGPU cluster facilitated complex model training (Fig. 2).
- Lookup by value is core to memory access. We innovated a masked lookup to form key value lookup (Fig. 4).

Fig. 2: Distributed learning architecture.

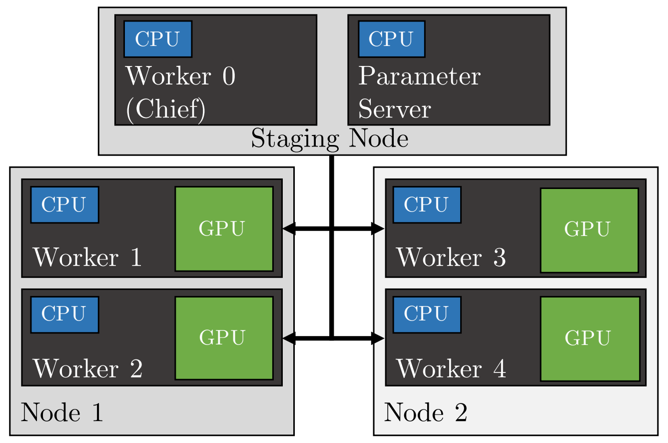


Fig. 3: DNC memory table writes and reads on a copy task.

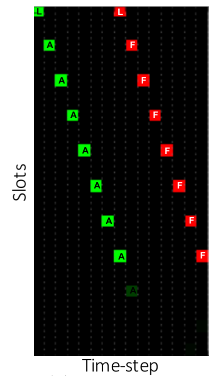
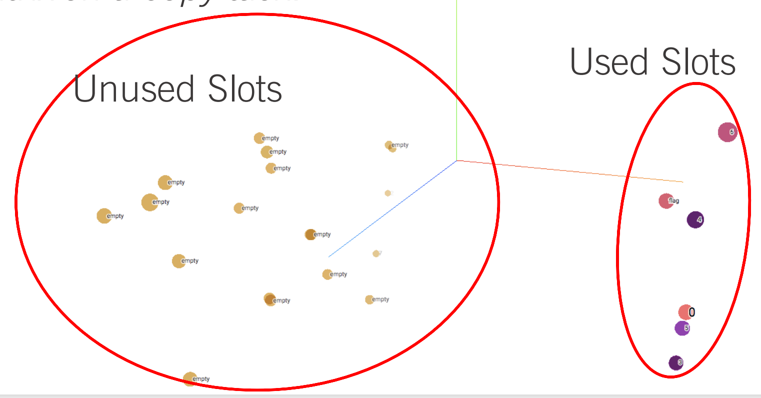


Fig. 4: t-SNE 3D visualisation of the DNC memory matrix on a copy task.



Results

- The DNC was optimised for data structure and real world problems demonstrating versatility.
- In some cases the DNC outperformed an LSTM [3] reaching 100% accuracy and visualisations validated the inner mechanisms (Fig. 3).
- The penny drop curve was coined describing sudden learning underpinning algorithmic learning (Fig. 5).
- Profiling on a high performance cluster achieves near linear speed up (Fig. 6).

Fig. 5: Learning a copy task over various models. Penny drop circled.

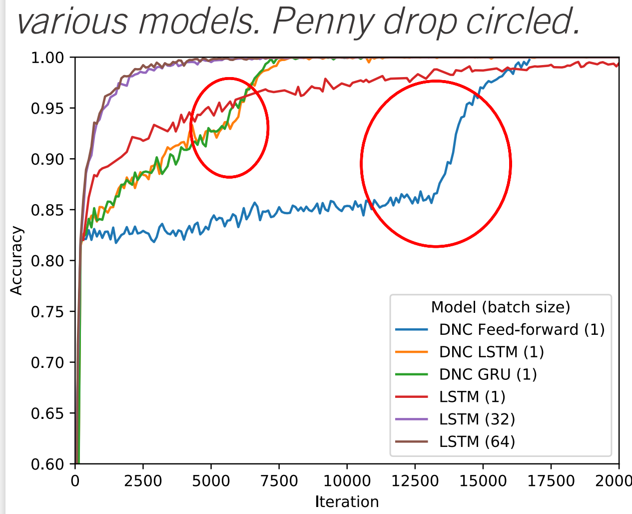
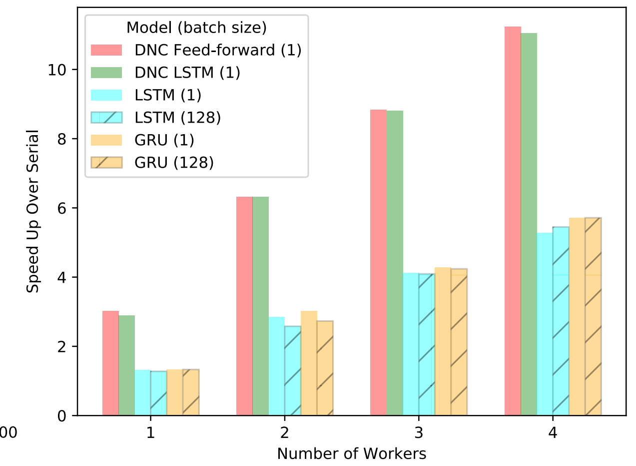


Fig. 6: Speed up over a distributed architecture.



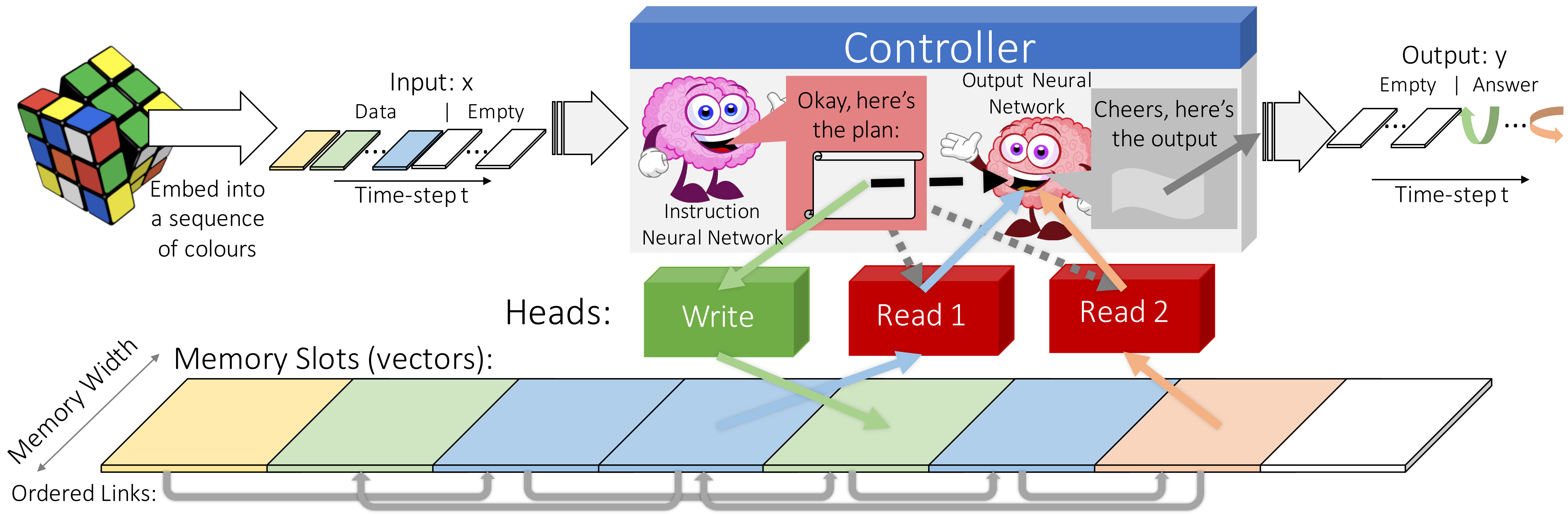
Model

DeepMind pioneered the Neural Turing Machines (NTM) [1] proceeded by the Differential Neural Computer (DNC) [2] that bridged Turing Machines [4] using deep neural networks to train an algorithm.

Fig. 1 displays the end-to-end pipeline of a DNC. It begins by embedding the problem sequentially for feeding iteratively to a controller that executes operations on internal state before regressing output.

The controller can be feed-forward or recurrent neural network with optimisation being an open problem we investigated.

Fig. 1: DNC Architecture end-to-end pipeline for the Rubik's Cube task.



Conclusion

- **Validated** the DNC against state of the art, highlighting merits of each.
- **Optimised** as a scalable system, identifying and exploring bottlenecks.
- **Visualised** the inner workings of the models.
- **Explored** innovative ideas to improve the models forming a new state of the art.

References

1. *Graves, A., Wayne, G. & Danihelka, I. (2014), 'Neural Turing Machines', CoRR abs/1410.5401.*
2. *Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I. et al. (2016), 'Hybrid Computing Using a Neural Network with Dynamic External Memory', Nature.*
3. *Hochreiter, S. & Schmidhuber, J. (1997), 'Long Short-Term Memory', Neural Comput. 9(8), 1735–1780.*
4. *Turing, A. M. (1937), 'On Computable Numbers, with an Application to the Entscheidungsproblem', In Proc. London Mathematical Society 2(1), 230–265.*