Using AOC* to Evolve LSTM RNN** for Time Series Prediction

A Ph.D. Candidacy Proposal

by

A. ElSaid

*Ant Colony Optimization

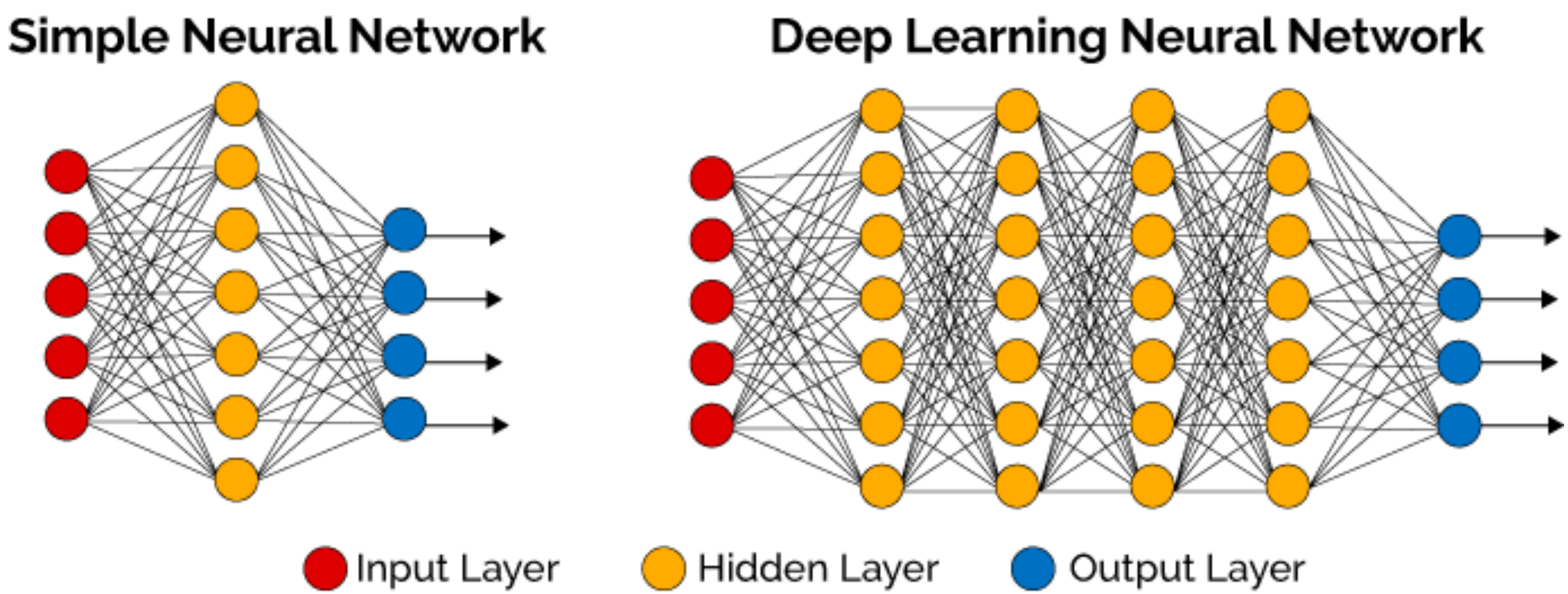
**Long Short Term Memory Recurrent Neural Networks

Main Points

- Machine deep-learning and what it offers.
- Why RNN?
- Why LSTM RNN?
- Why to optimize deep neural networks and Why ACO?
- Work done so far.
- Preliminary results.
- Proposed work.
- Timeline.

Machine deep-learning

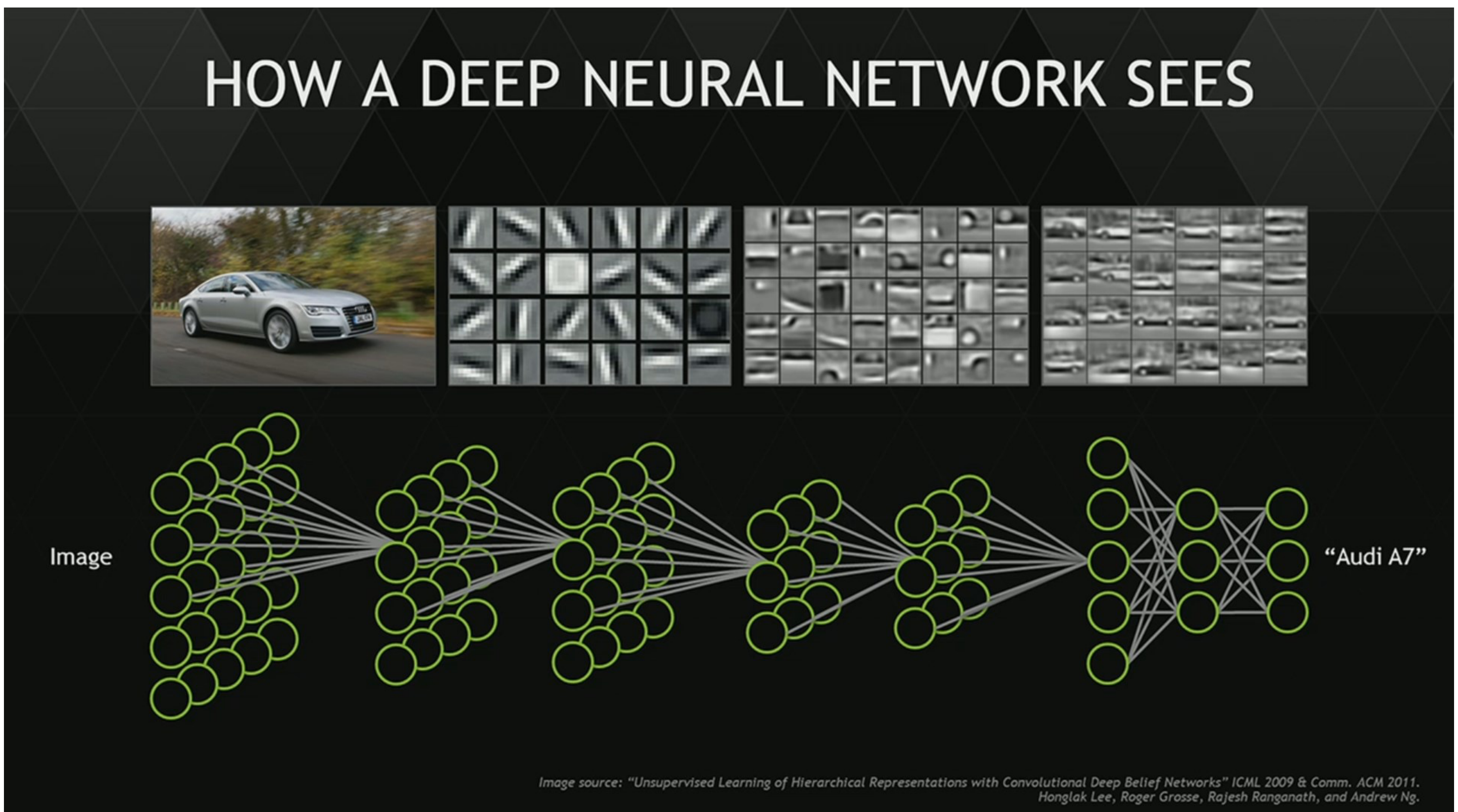
- Machine deep learning is the current track of advancing simple neural network structures.



https://www.quora.com/What-is-the-difference-between-Neural-Networks-and-Deep-Learning

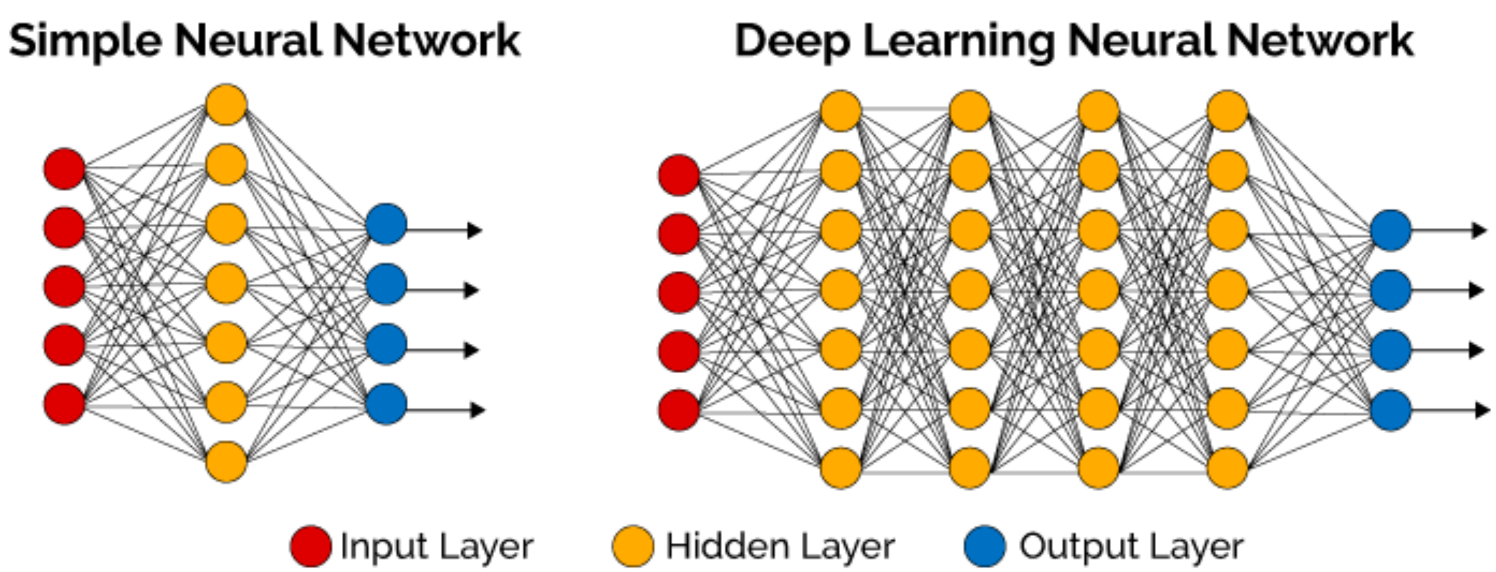
Machine deep-learning

- Adds more abstraction levels to information for better analysis and decision making



Machine deep-learning

- This is made possible party because of the optimization and gradient generation approximation methods used in obtaining the gradients in the stochastic gradient decent:
  - Dropout
  - Batch regularization
  - Regularization
  - Momentum/velocity

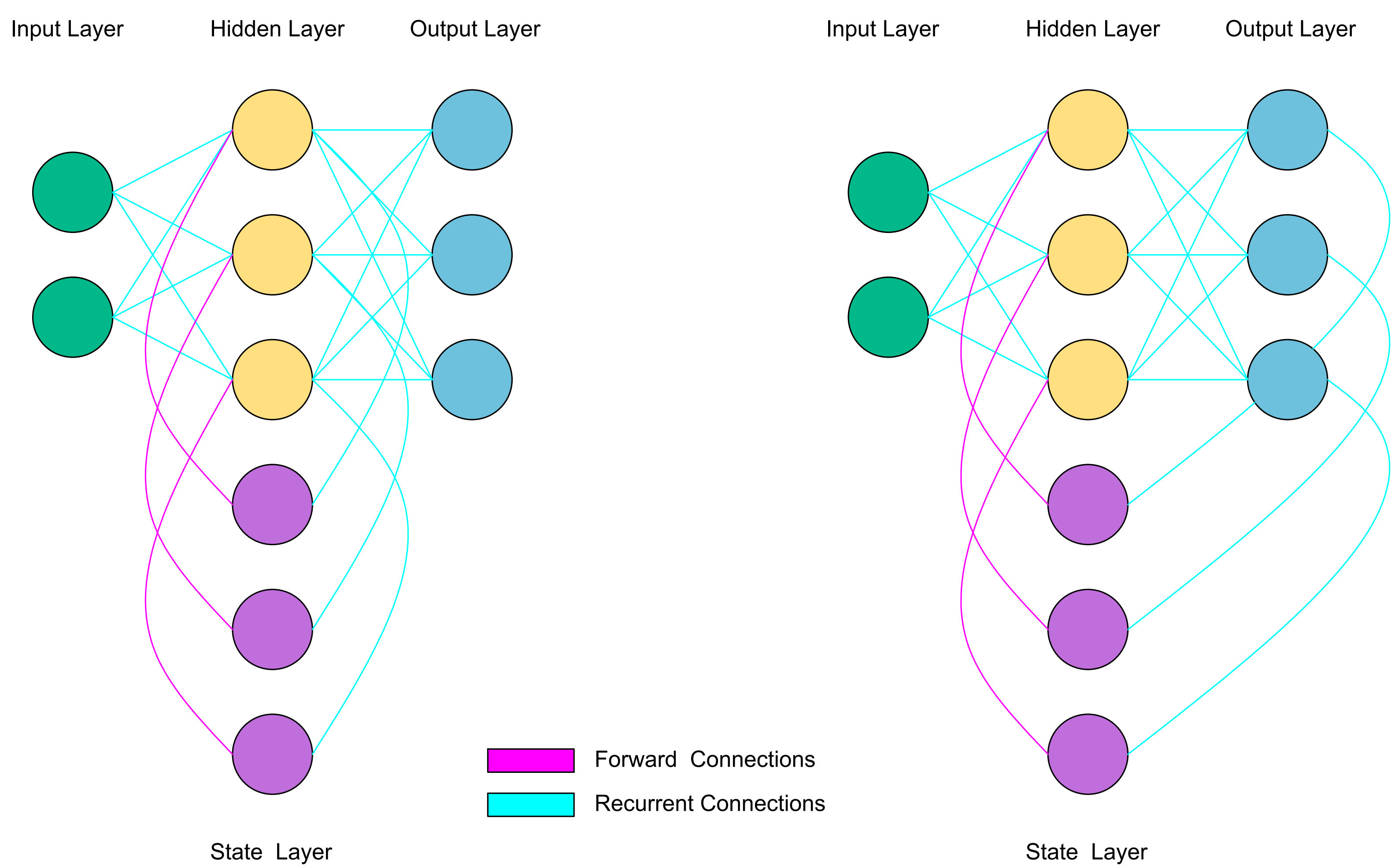


https://www.quora.com/What-is-the-difference-between-Neural-Networks-and-Deep-Learning

Recurrent Neural Networks

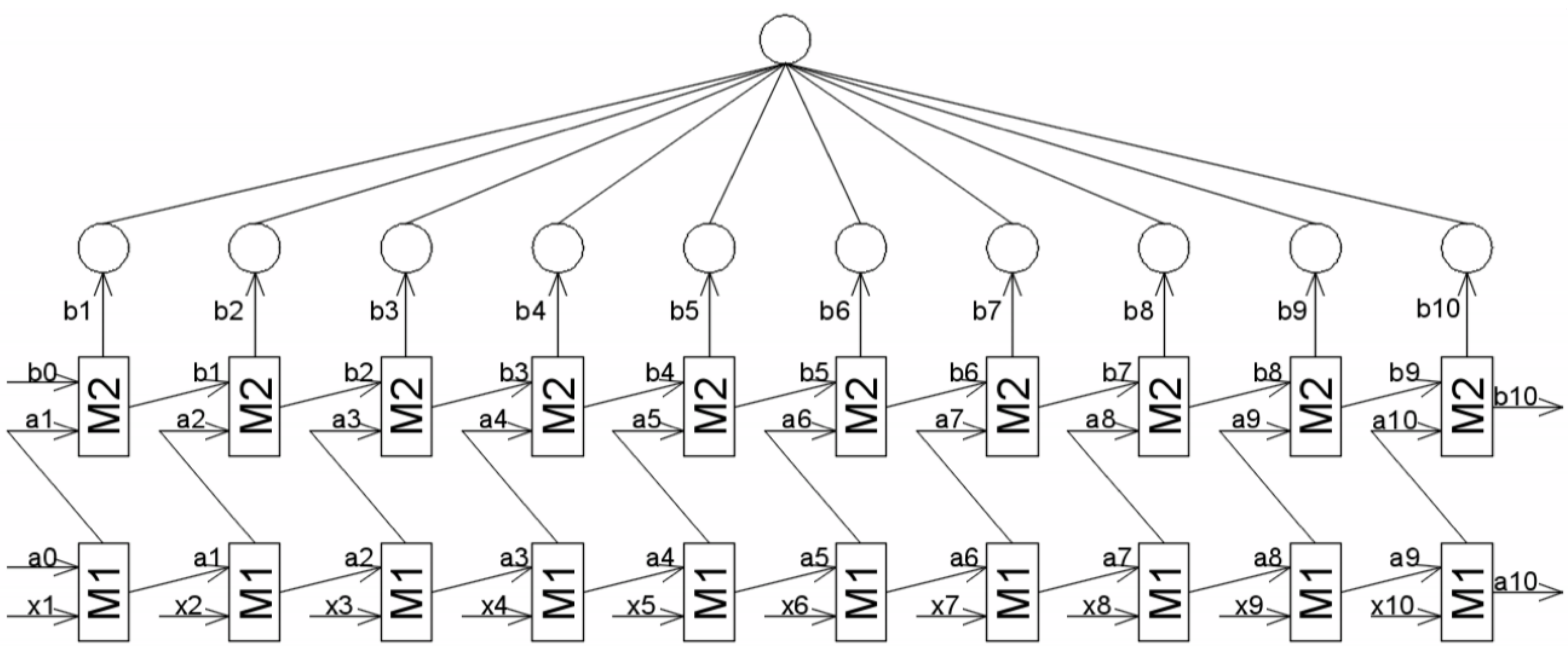
- Passes information from past time series to the contemporary data analysis… *The past can influence current decision making*.
- Below are two popular implementations of the RNN.

Elman network   Jordan network



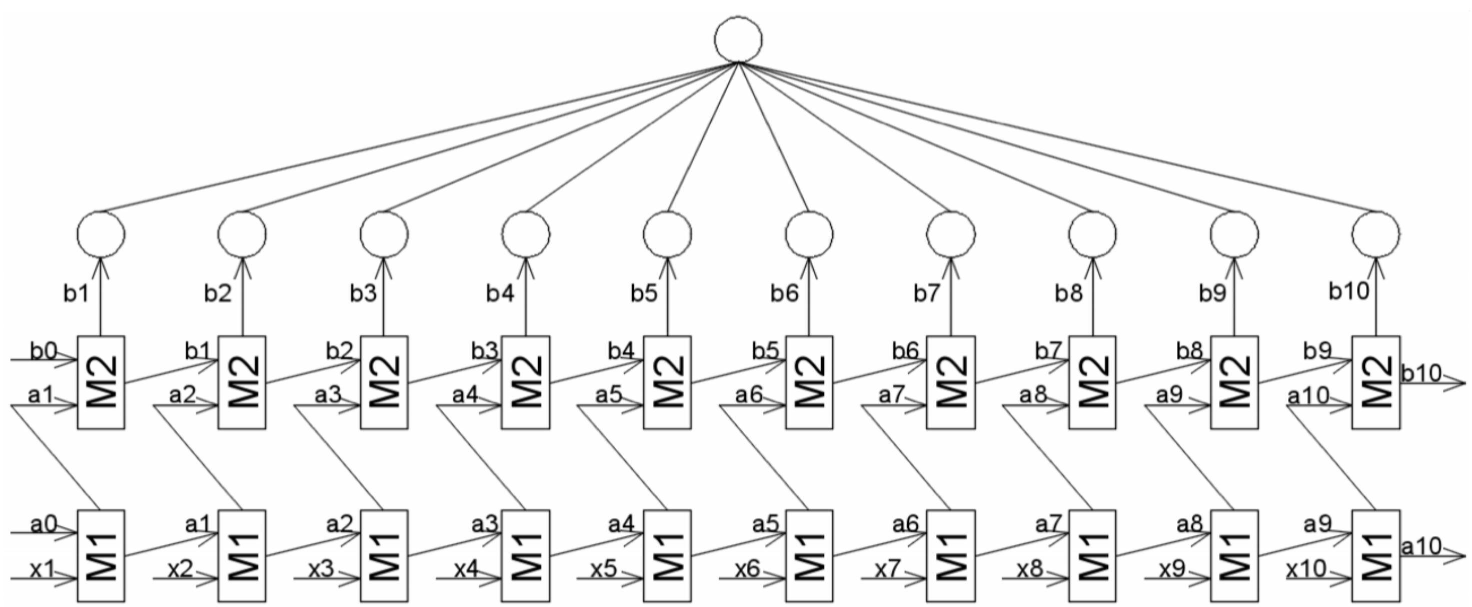
LSTM RNN

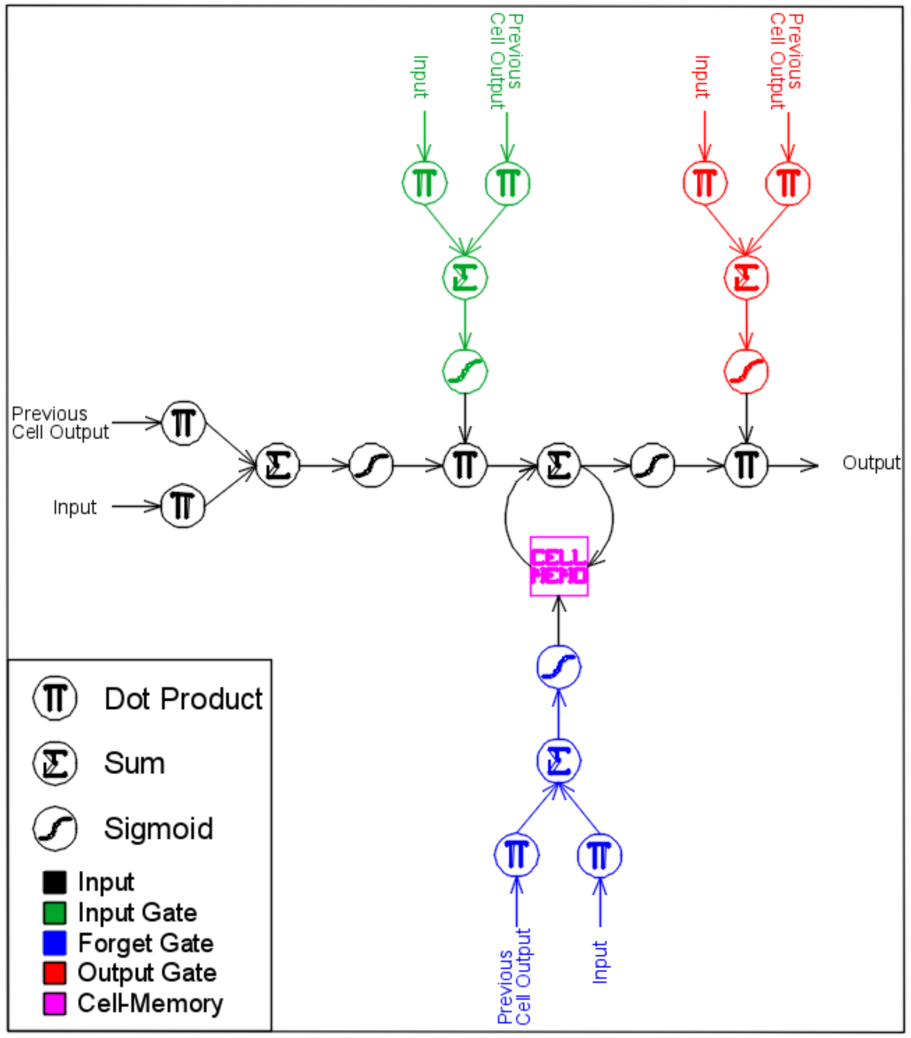
1. A further step to pass information from previous time series data to current contemporary analysis



LSTM RNN

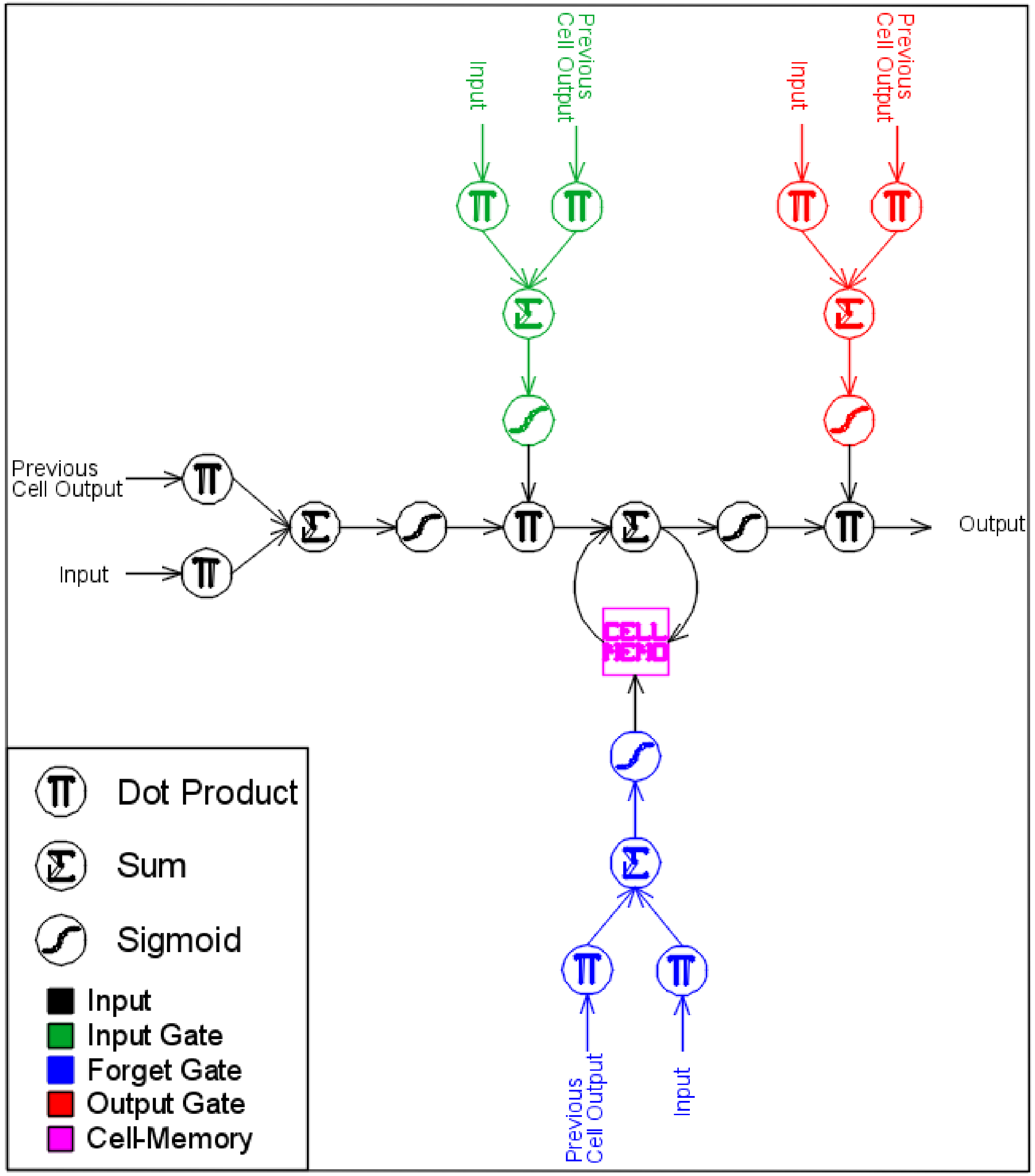
2. Recurrent connections passes information, and …





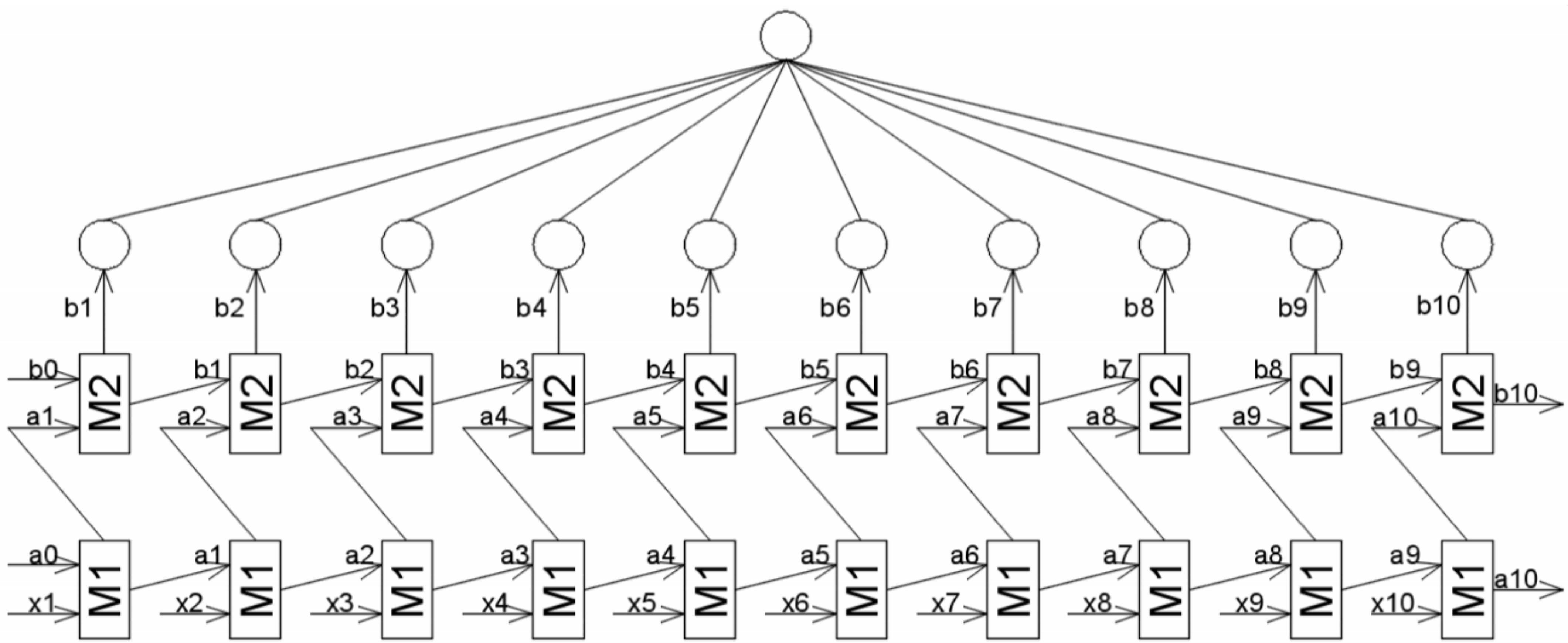
LSTM RNN

3. There are gates that control the flow of information from all the sources (inputs and recurrent) to adjust for sound output… furthermore,

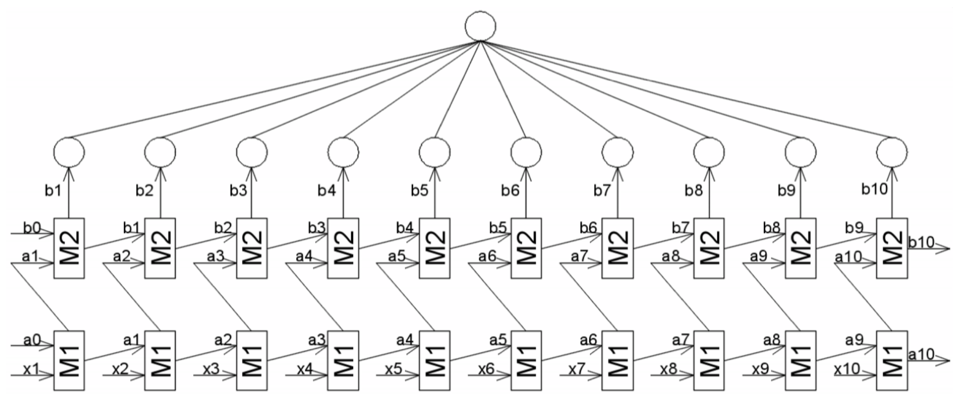


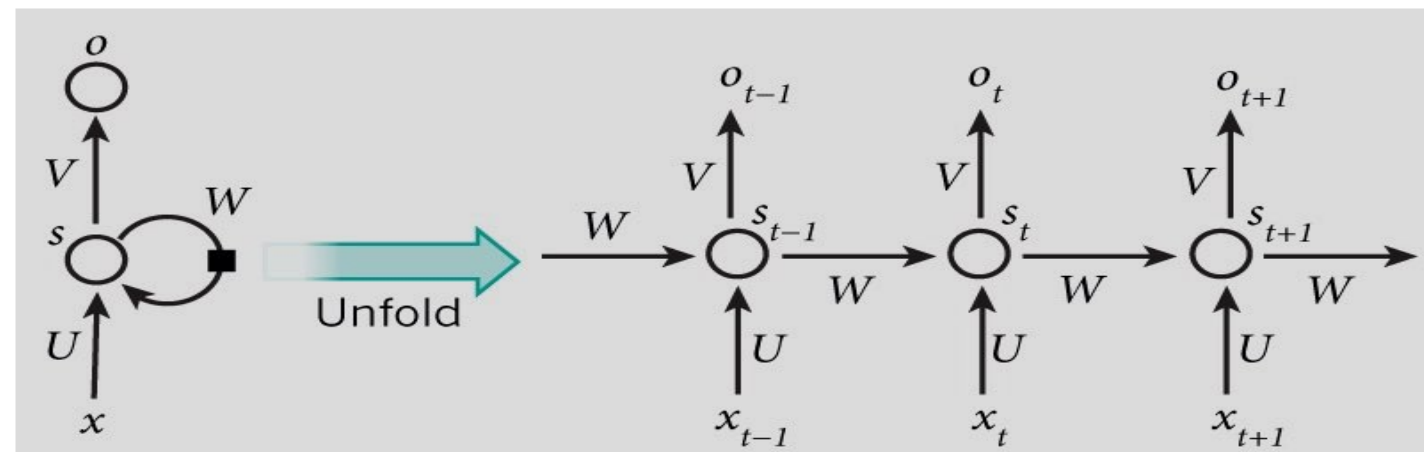
LSTM RNN

4. LSTM RNN with its gated and memory cells offers a solution for an epidemic which deep neural networks suffer from: *vanishing an exploding gradients*.



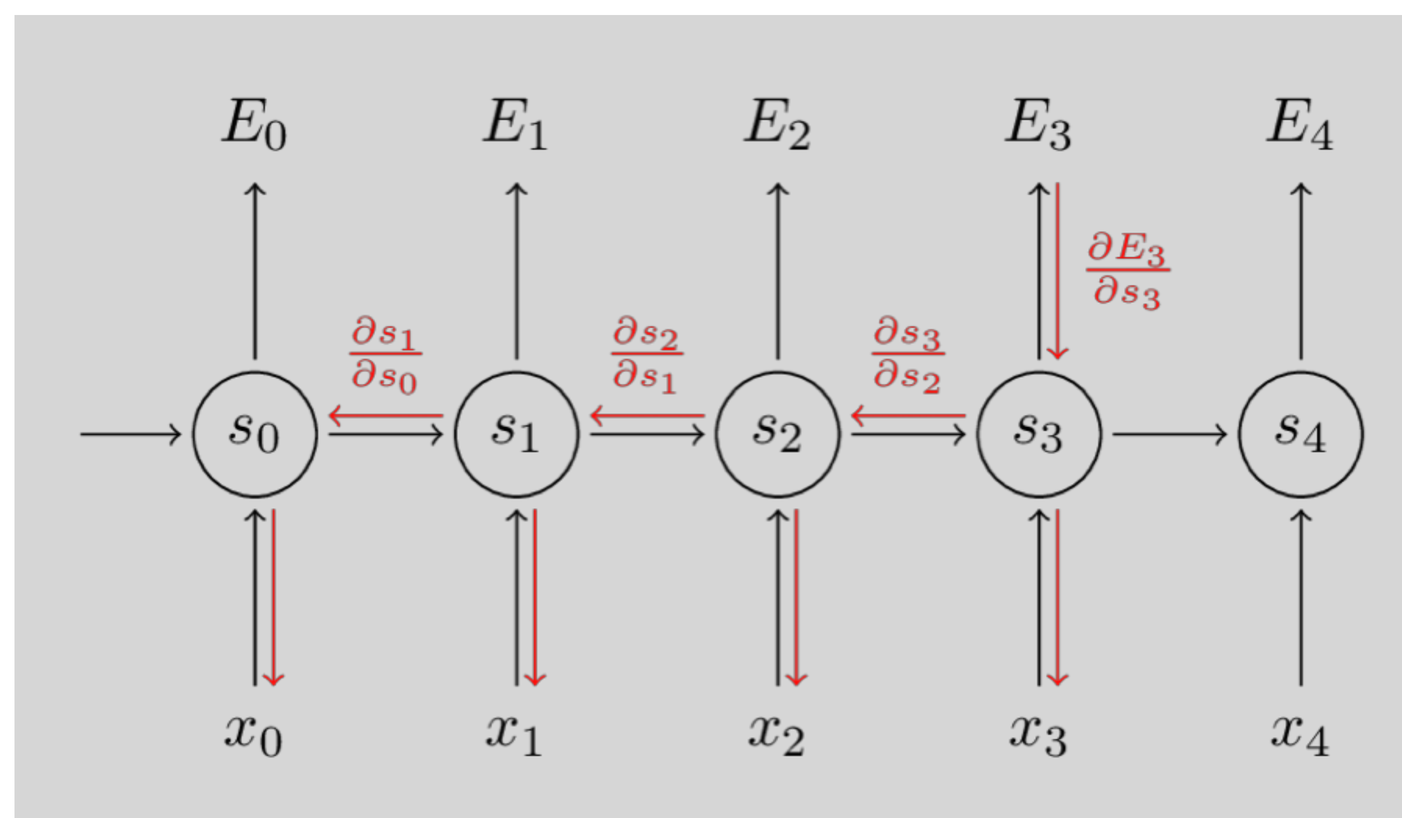
LSTM RNN and BPTT



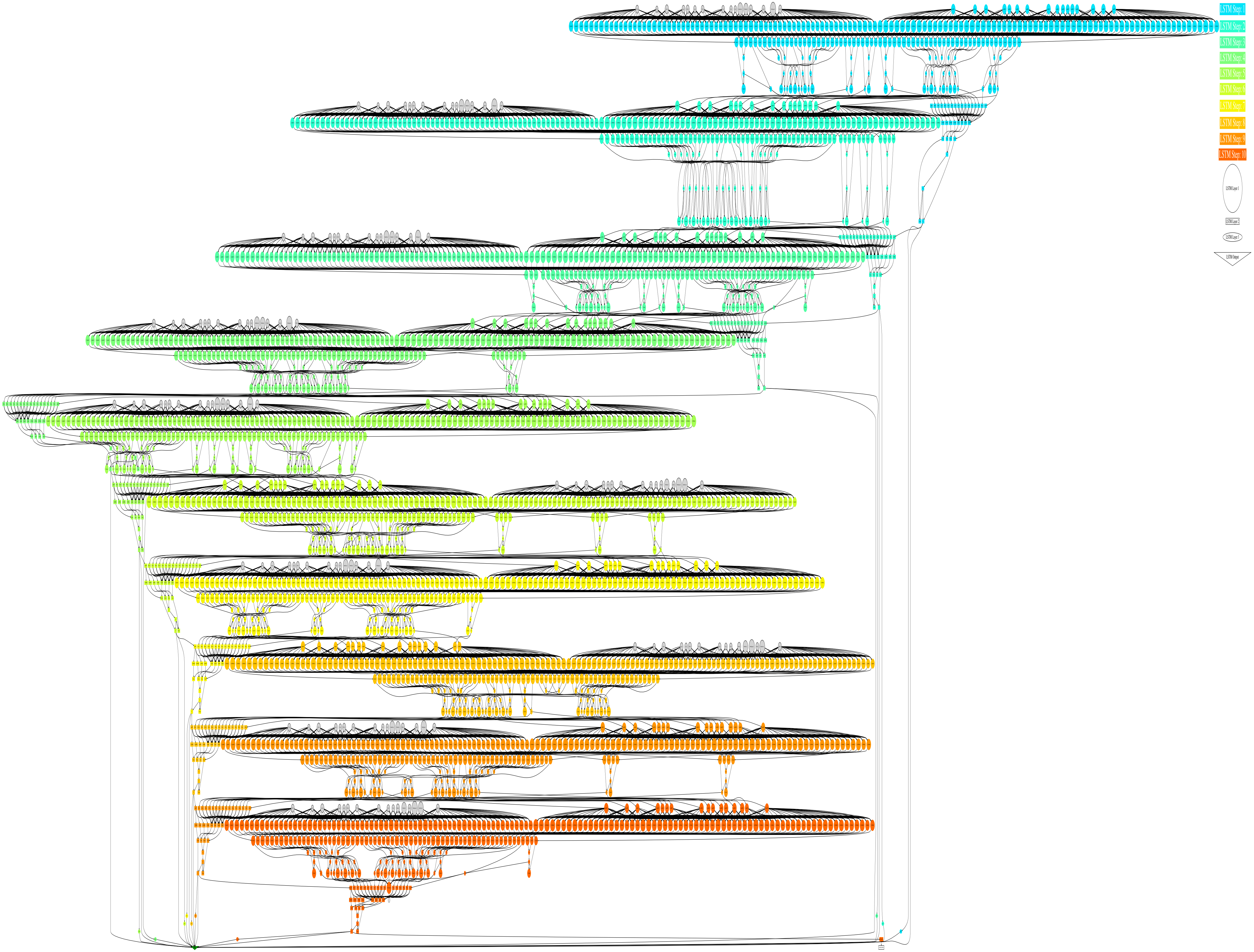


$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^{3} \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial s_3} \left(\prod_{j=k+1}^{3} \frac{\partial s_j}{\partial s_{j-1}} \right) \frac{\partial s_k}{\partial W}$$

Vanishing/Exploding Gradients



http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-2-implementing-a-language-model-rnn-with-python-numpy-and-theano/



LSTM Step: 1
LSTM Step: 2
LSTM Step: 3
LSTM Step: 4
LSTM Step: 5
LSTM Step: 6
LSTM Step: 7
LSTM Step: 8
LSTM Step: 9
LSTM Step: 10
LSTM Layer 1
LSTM Layer 2
LSTM Layer 3
LSTM Output

LSTM RNN Optimization

- Naturally, a fully connected network is not the optimum since some of the connection might induce noise rather than contributing to a sound output.
- No matter how good is the structure of the network, how deep it is, how well it is trained, the best model it will reach to will have imperfections because of these noises.
- So, the network has to be either trimmed, groomed, or evolve to its optimum structure.

LSTM RNN Optimization

- There several methods used to optimize an LSTM RNN, like, cells dropout, non-recurrent connections regularization, and NEAT*.

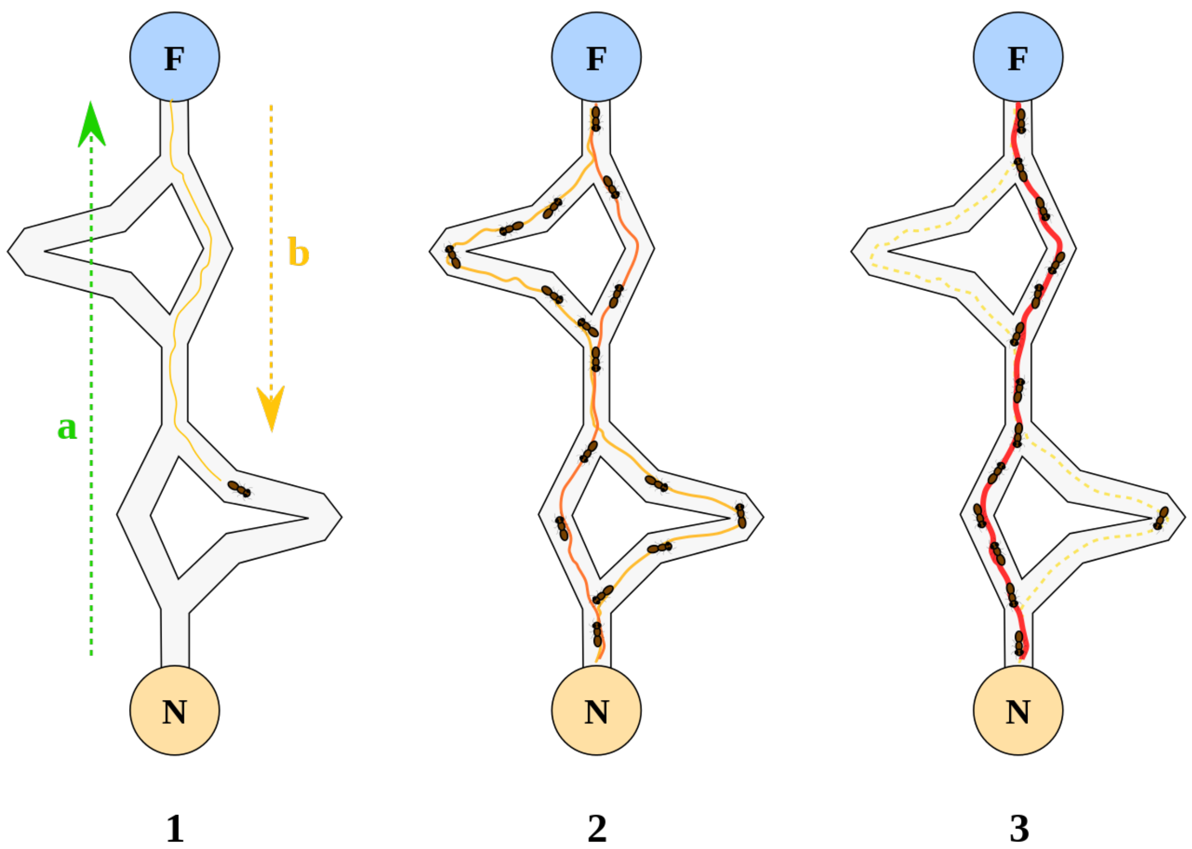
- There are significant number of studies performed on the above mentioned methods.

- To the best of my knowledge, ACO is not used in LSTM RNN topology optimization, yet.

*NeuroEvolution of Augmenting Topologies

Why ACO?

- The main advantage of ACO is that it is not a pure random evolution algorithm, but rather a bio-inspired metaheuristics method.
- It is a distributed approach using agents called artificial ants.
- These artificial ants resemble biological ants
- Each ant is independent and communicates with other members of the colony through a chemical called pheromone.
- Ants randomly explore areas, however they tend to follow paths with pheromones left on them by other ants, and upon finding food they mark their return path with more pheromone.
- Pheromones decay over time, and paths with the most pheromone represent the most promising paths to food.



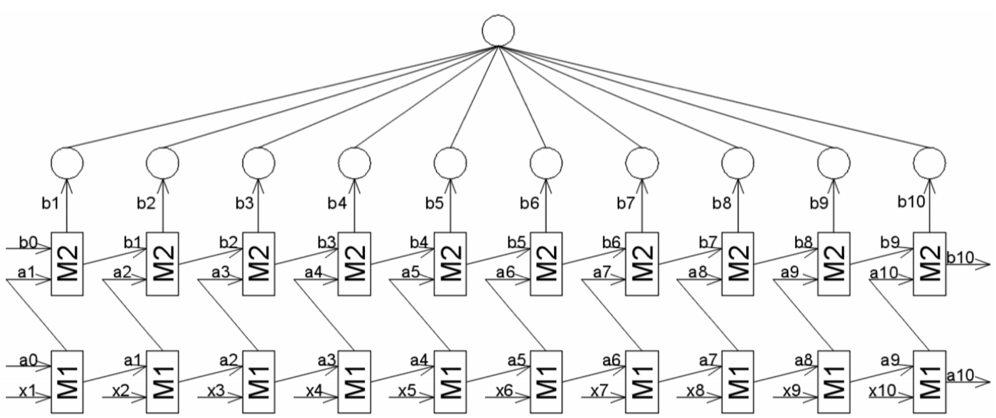
https://upload.wikimedia.org/wikipedia/commons/thumb/a/af/Aco_branches.svg/1280px-Aco_branches.svg.png

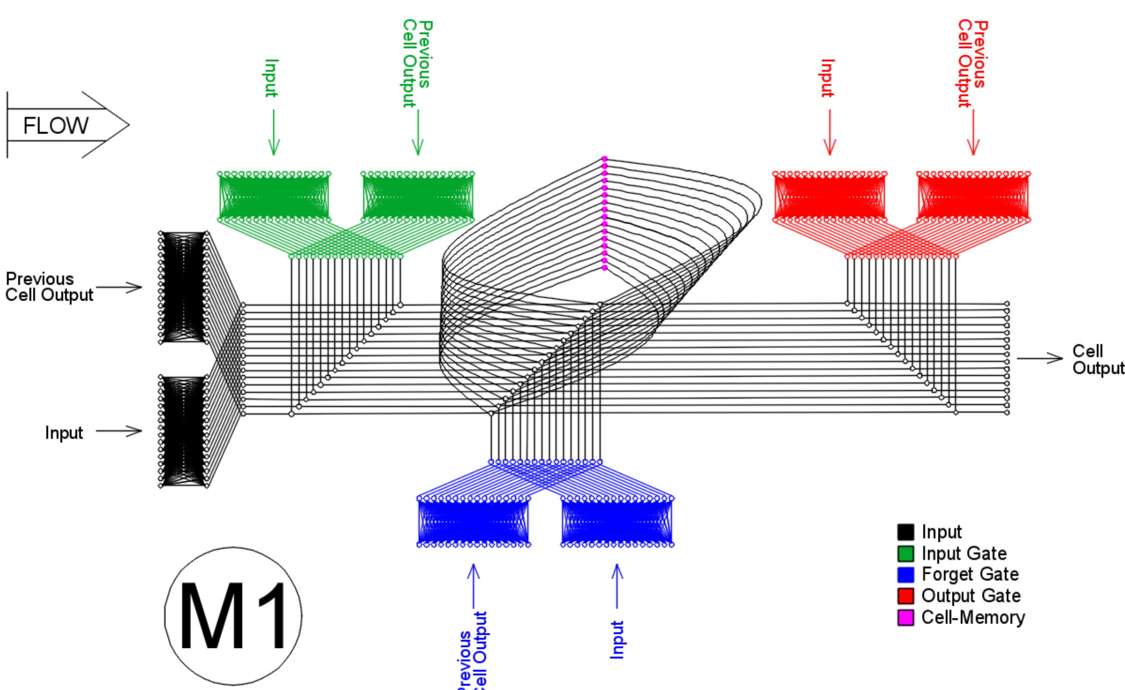
Why ACO?

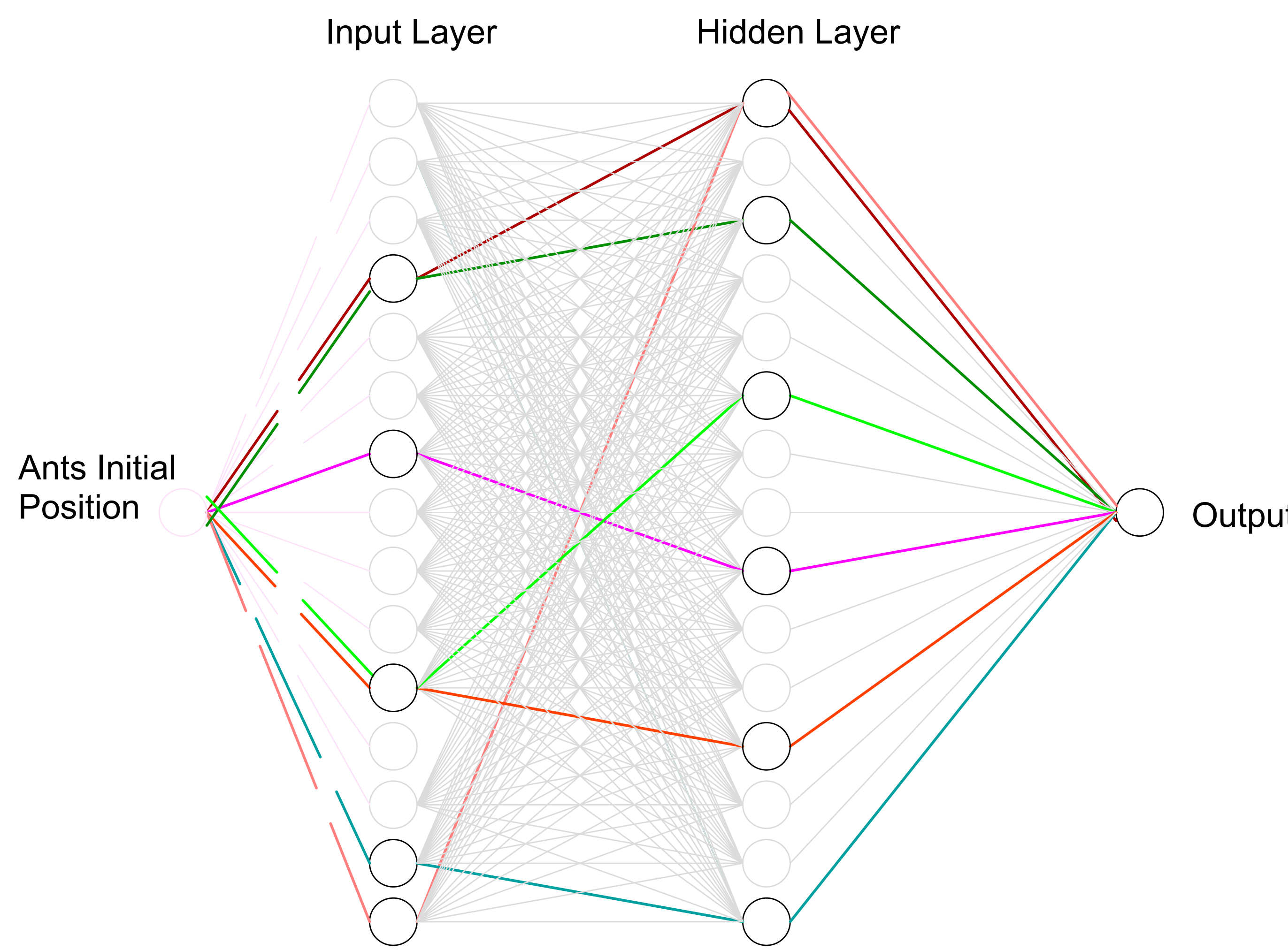
https://6legs2many.files.wordpress.com/2013/06/neivamyrmex_army_ants_raiding_trail.jpg

What is Done So Far

- The preliminary implementation done so far is as fellows:







Preliminary Results

Table 7: K-Fold Cross Validation Results

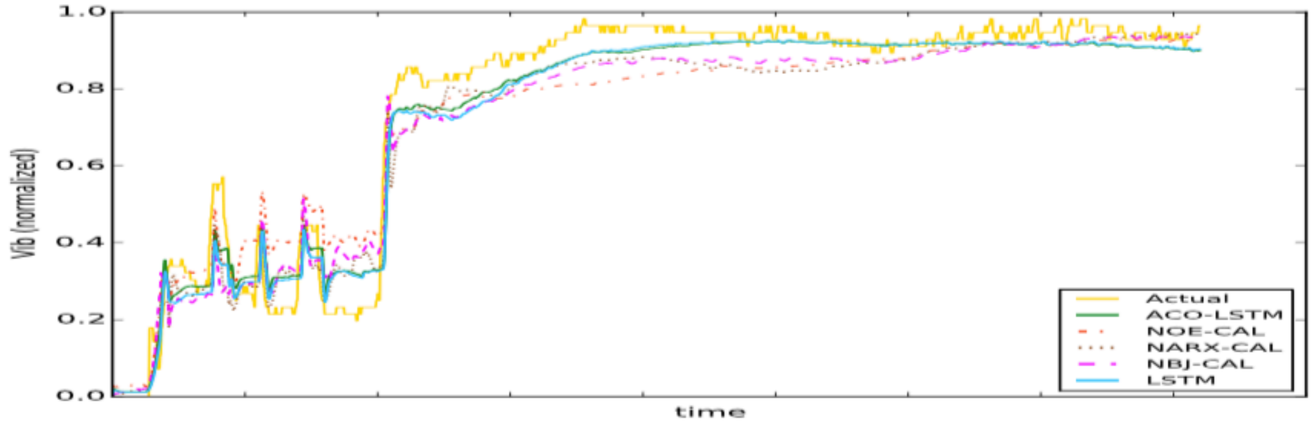
	Prediction Errors (MAE)				
	LSTM	**NOE**	**NARX**	**NBJ**	**ACO**
Subsample 1	8.34%	10.6%	8.13%	8.40%	7.80%
Subsample 2	4.05%	6.96%	6.08%	7.34%	3.70%
Subsample 3	6.76%	16.8%	11.2%	13.6%	3.49%
Mean	**0.0638**	**0.1145**	**0.0847**	**0.0977**	**0.0501**
Std. Dev.	**0.0217**	**0.0497**	**0.0258**	**0.0333**	**0.0245**

NOE : Nonlinear Output Error model

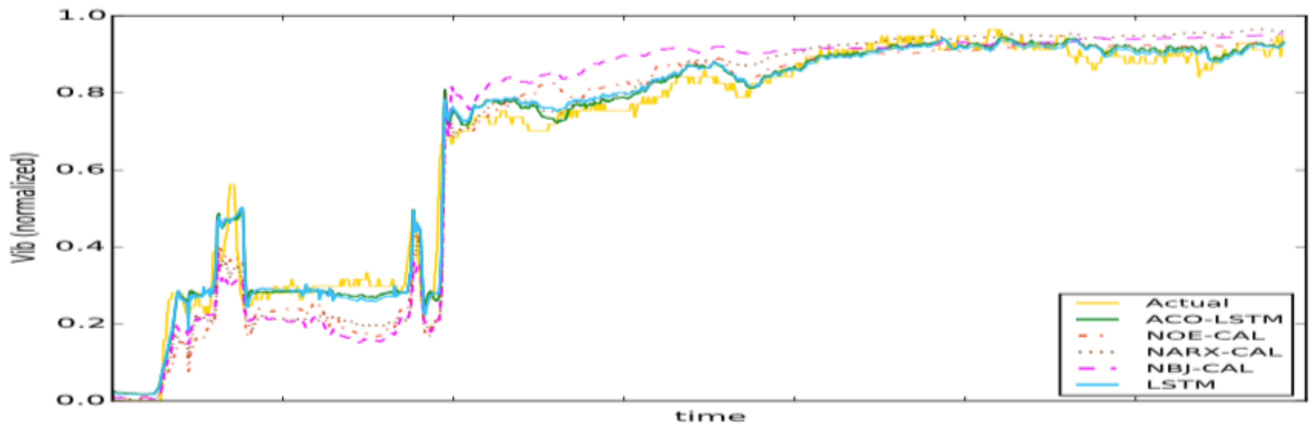
NARX : Nonlinear AutoRegressive eXogenous model

NBJ : Nonlinear Box-Jenkins model

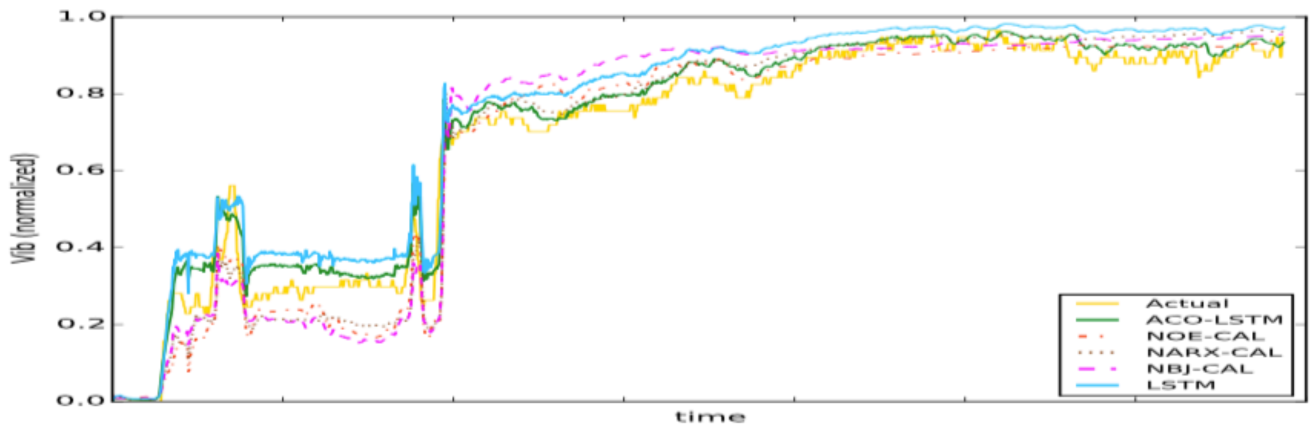
Preliminary Results



(a) Subsample 1

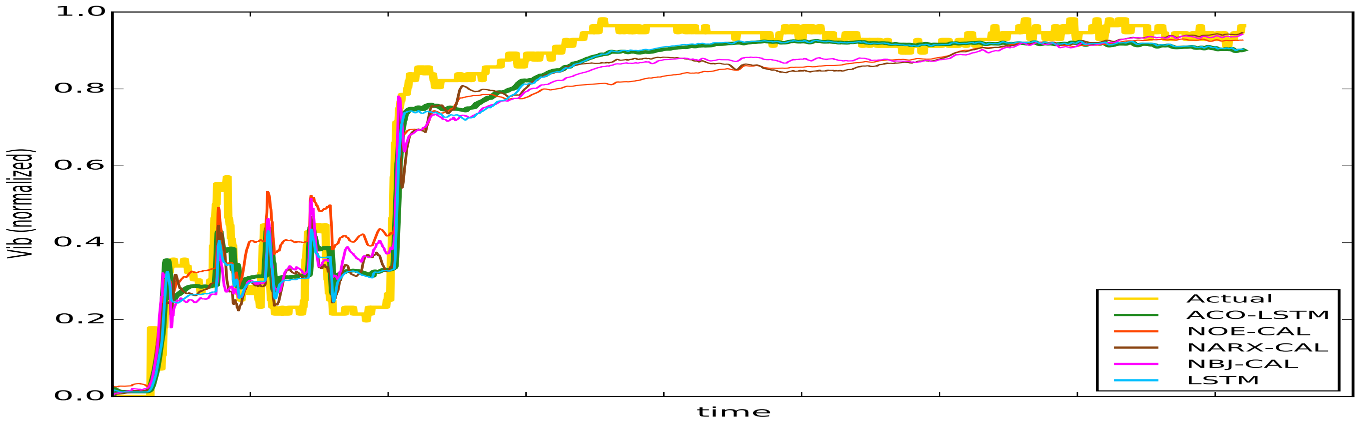


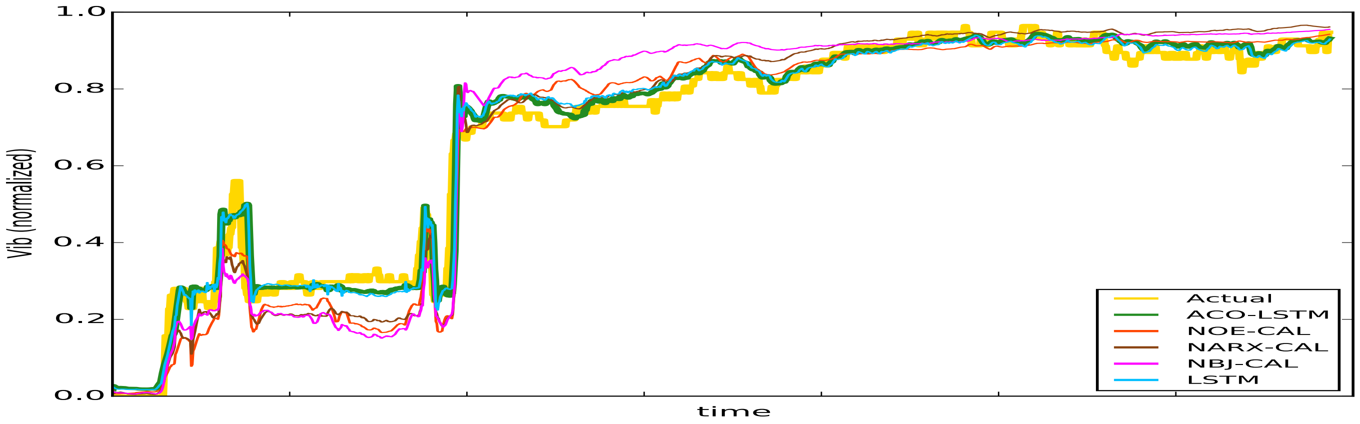
(b) Subsample 2

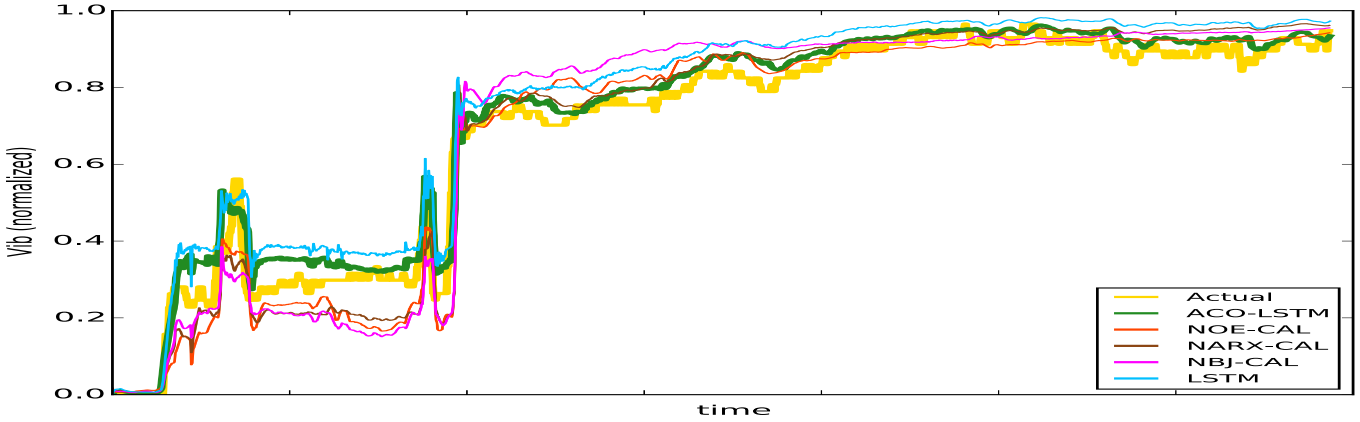


(c) Subsample 3

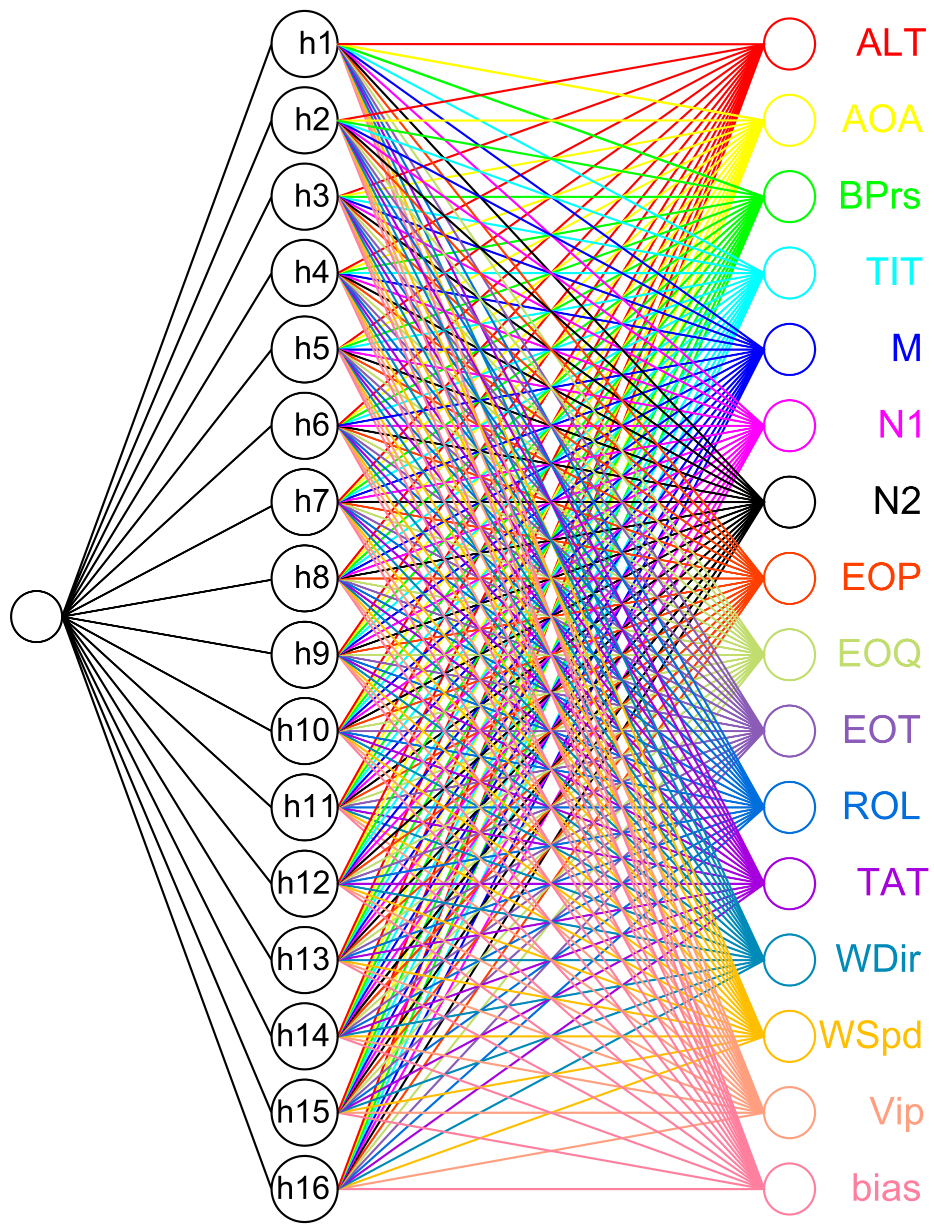
Preliminary Results

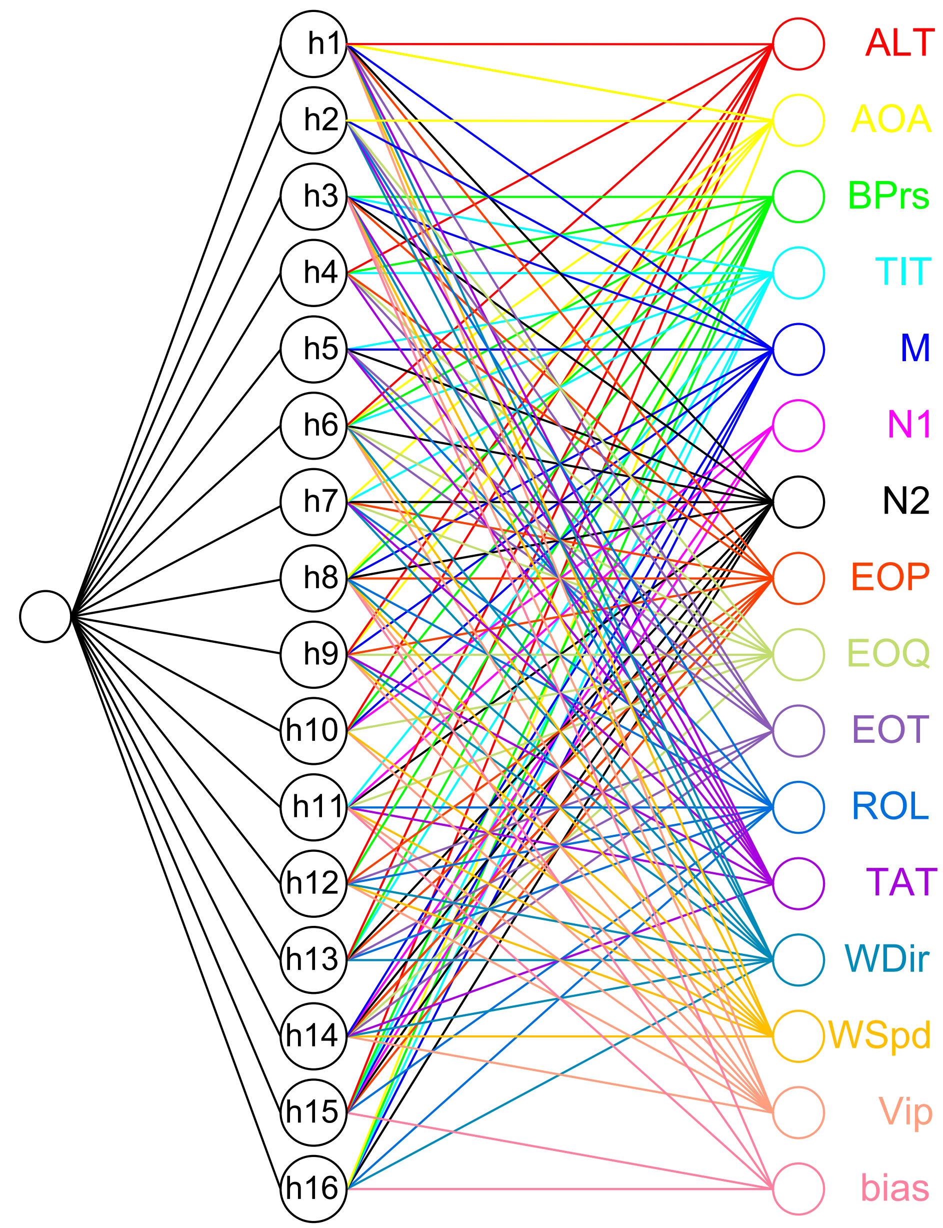






Preliminary Results





Preliminary Results

$$
mesh\_1 = \begin{array}{r|cccccccccccccccc|}
 & h1 & h2 & h3 & h4 & h5 & h6 & h7 & h8 & h9 & h10 & h11 & h12 & h13 & h14 & h15 & h16 \\
i1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\
i2 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
i3 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
i4 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\
i5 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
i6 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
i7 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\
i8 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\
i9 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
i10 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
i11 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\
i12 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\
i13 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\
i14 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\
i15 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
bias & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1
\end{array}
$$

$$
mesh\_2 = \left|\begin{array}{cccccccccccccccc}
h1 & h2 & h3 & h4 & h5 & h6 & h7 & h8 & h9 & h10 & h11 & h12 & h13 & h14 & h15 & h16 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{array}\right|
$$

Proposed Work

<table>
<tr><th colspan="6">Research Timeline</th></tr>
<tr><td>Mar. - Apr.</td><td>June - July.</td><td colspan="3">Aug. - Oct.</td><td>Nov.</td></tr>
<tr><td>ACO LSTM Cells Implementation</td><td></td><td colspan="3"></td><td></td></tr>
<tr><td></td><td>ACO LSTM Whole Structure Implementation</td><td colspan="3"></td><td></td></tr>
<tr><td rowspan="2"></td><td rowspan="2"></td><td colspan="3">Optimization Evaluation</td><td rowspan="2"></td></tr>
<tr><td>Network Dropout</td><td>Network Regularization</td><td>Hyper-parameters Tuning (tentative)</td></tr>
<tr><td></td><td></td><td colspan="3"></td><td>Documentation</td></tr>
</table>

- Apply ACO to the whole LSTM topology in the same time.
- Examine other network topology optimization technologies to compare its performance to ACO: *Dropout and Regularization*.
- Further examine the effectiveness of using ACO to conclude the actual input contributors to the output.
  - This will be done by introducing randomly generated inputs which will induce noise, and investigate if the optimization process eliminates them.
  - This step's results might also offer an opportunity to understand the group effect of the input parameters.
- Use a time series data that belongs to a different industry than the data used previously, to investigate the generality of the LSTM neural network topology.

Thank you!

Questions!?